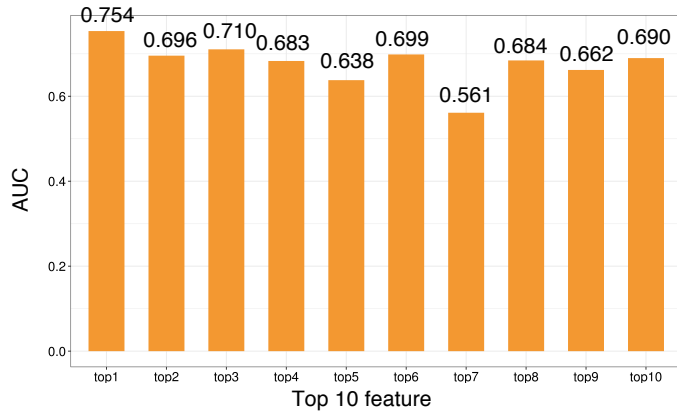


Figure S1 Details of recruited patients from 14 clinical sites. The total number of enrollment patients and the malignancy prevalence from different provinces of China was illustrated.

A.



B.

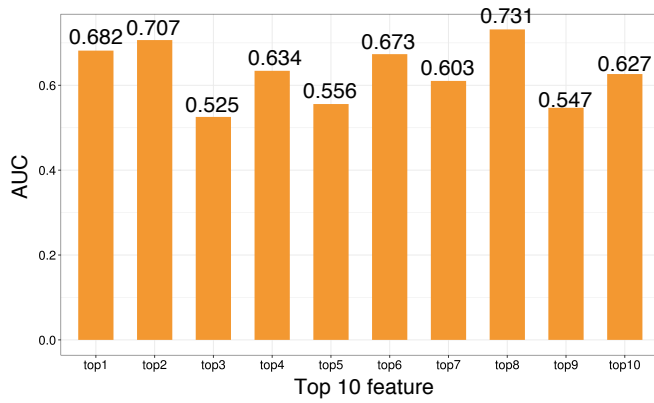


Figure S2 AUC performance of top 10 model markers. A. AUC performance of top 10 individual model feature in the training set B. AUC performance of top 10 individual model feature in the test set.

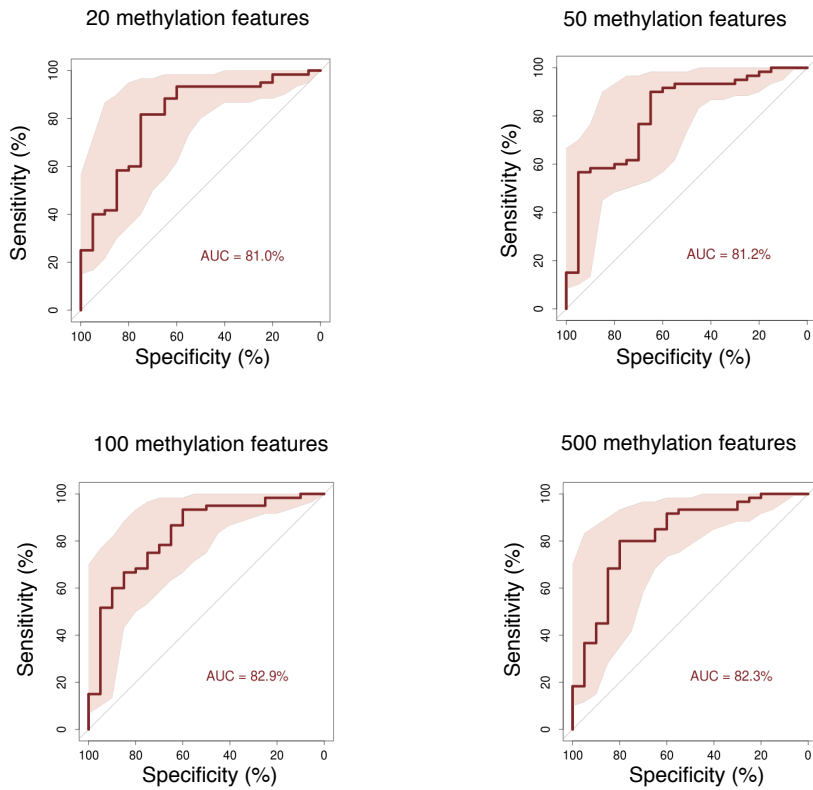
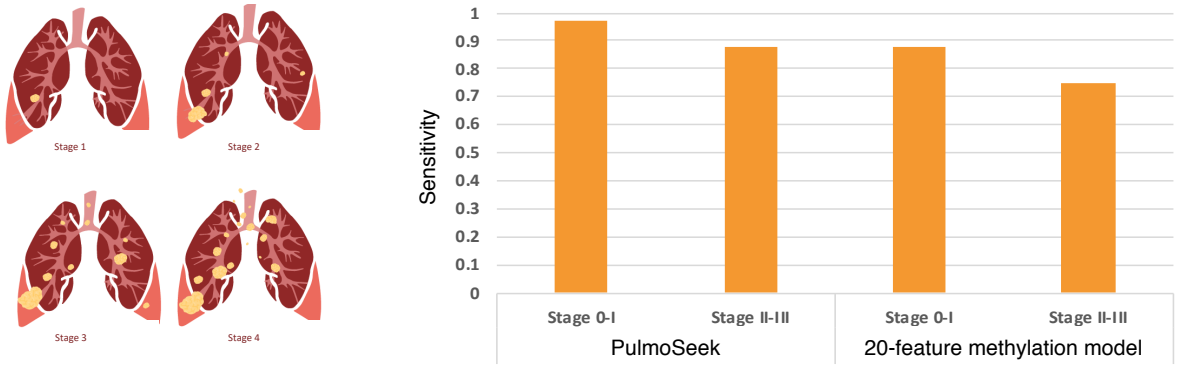
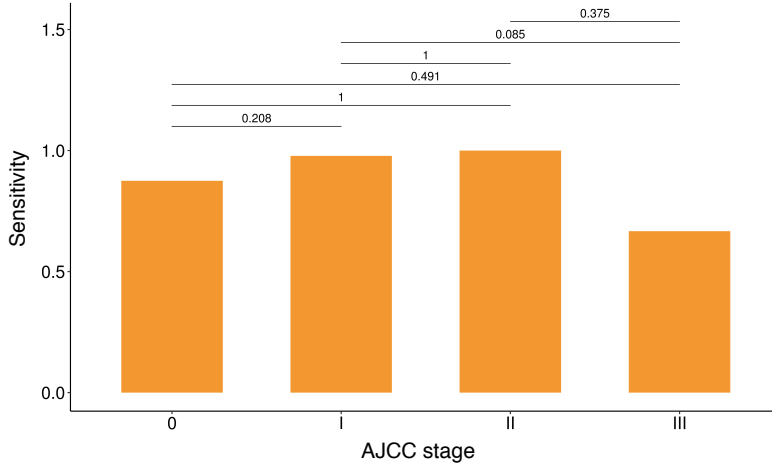


Figure S3 Performance of lung nodule model using different features in the test set. Representative receiver operating curves (ROCs) displays the classification performance of the methylation classifier using different numbers of features in the test set.

A.



B.



C.

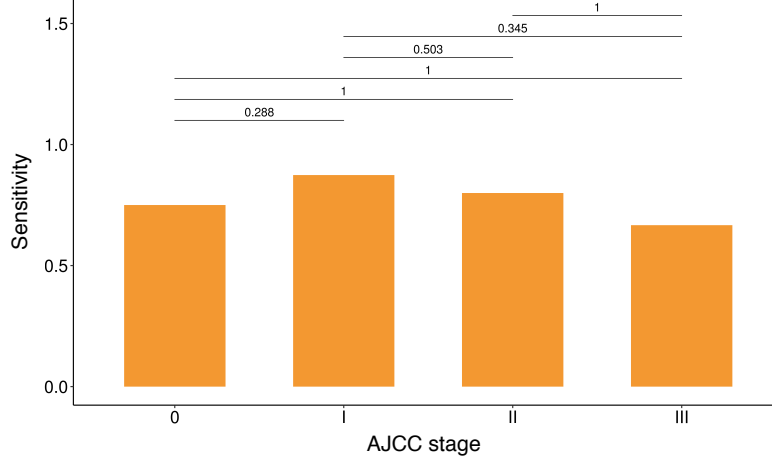


Figure S4 PulmoSeek performance in different stages (combined test and validation set). A. Top 100 (PulmoSeek) and Top 20 methylation models' performance in different stages in the combined test and validation set. Sensitivity of PulmoSeek in stage 0-I 0.971 [0.942-0.993], in stage II-III 0.875 [0.625-1.000]. Sensitivity of 20-feature methylation model in stage 0-I 0.877 [0.819-0.935], in stage II-III 0.750 [0.500-1.000] B. Comparison of PulmoSeek performances within different groups of stages in combined test and validation set C. Comparison of 20-feature model performances within different groups of stages in combined test and validation set. Stage 0 (n=8), stage I (n=135), stage II (n=5) and stage III (n=3). Comparisons of sensitivities for different stages were calculated using Fisher exact test.

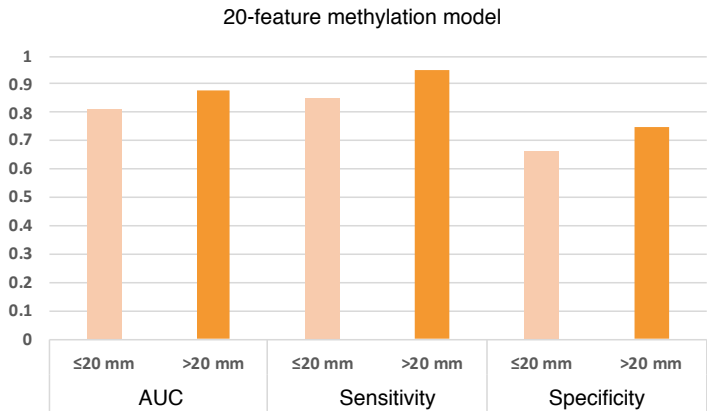
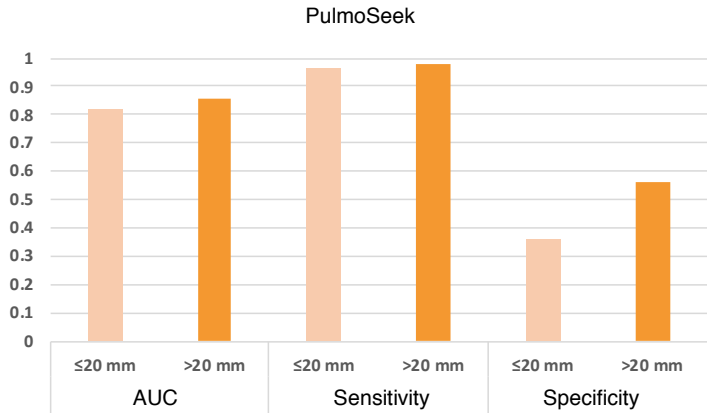
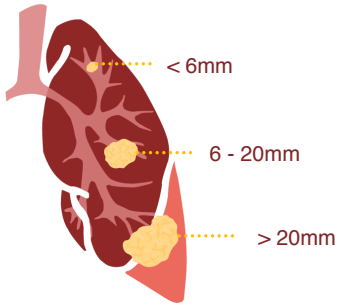


Figure S5 PulmoSeek performance in different nodule sizes. Top 100 (PulmoSeek) and Top 20 methylation models' performances in different nodule sizes in the combined test and validation set. *PulmoSeek*: AUC in ≤20 mm group 0.820 [0.742-0.891], in >20 mm group 0.860 [0.740-0.964]; Sensitivity in ≤20 mm group 0.966 [0.573-0.991], in >20 mm group 0.977 [0.628-1.000]; Specificity in ≤20 mm group 0.364 [0.273-0.932], in >20 mm group 0.562 [0.375-0.938]; *20-feature methylation model*: AUC in ≤20 mm group 0.811 [0.730-0.880], in >20 mm group 0.879 [0.775-0.967]; Sensitivity in ≤20 mm group 0.846 [0.650-0.966], in >20 mm group 0.963 [0.860-1.000]; Specificity in ≤20 mm group 0.659 [0.364-0.841], in >20 mm group 0.750 [0.438-0.938]. ≤20 mm group (n=160), >20 mm group (n=59).

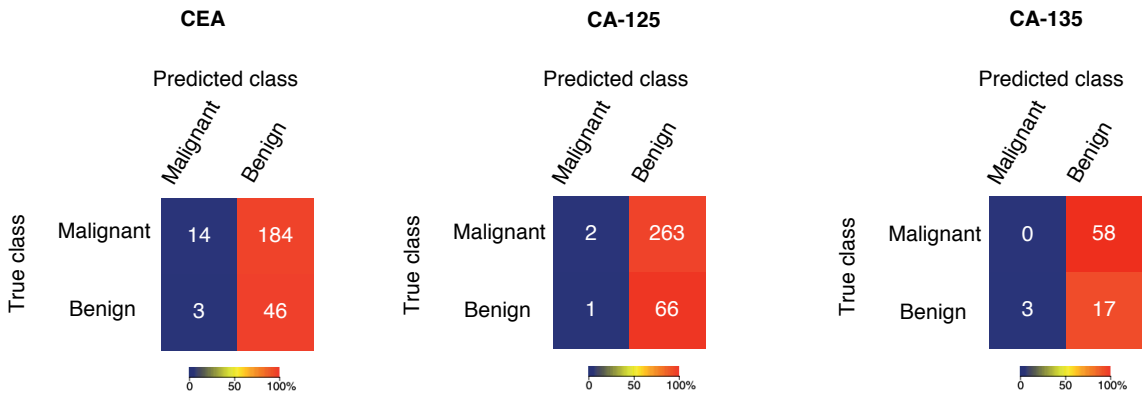
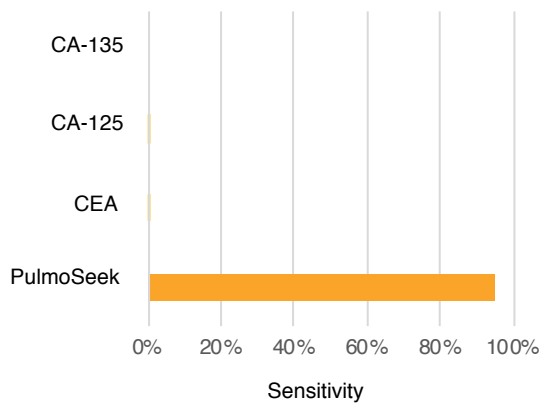


Figure S6 Cancer biomarker (CEA, CA-125, CA-135) performance. Classical cancer biomarkers (CEA, CA-125, CA-135) performance illustrated using confusion matrices.

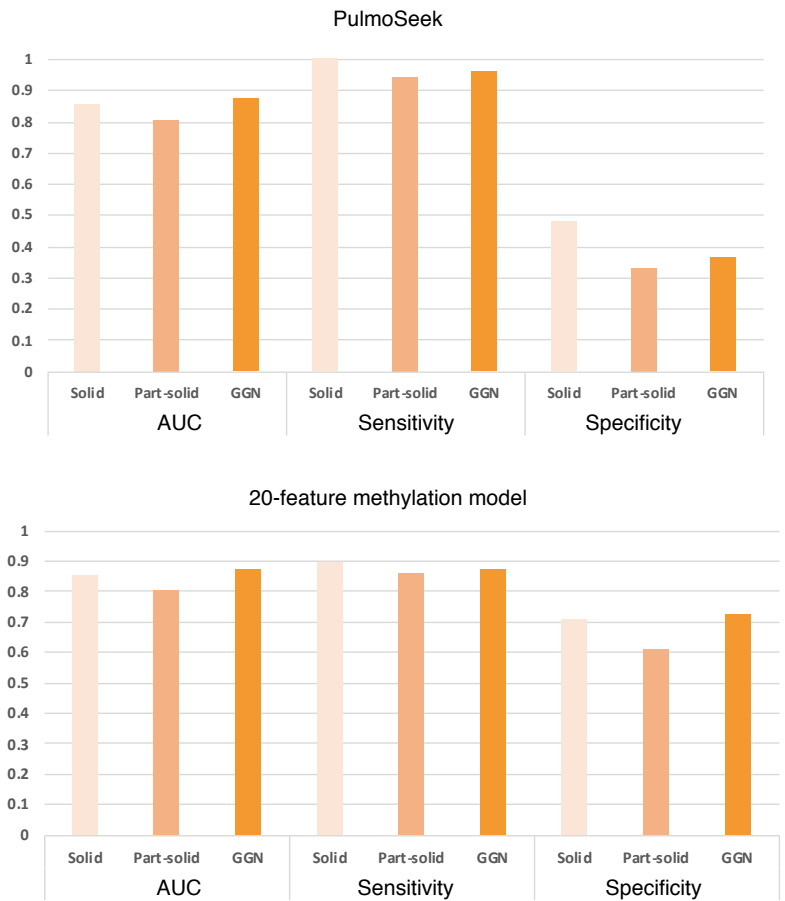
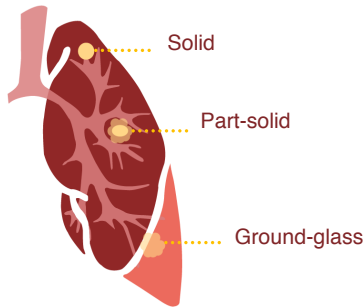


Figure S7 PulmoSeek performance in different nodule subtypes. Top 100 (PulmoSeek) and Top 20 methylation models' performance in different nodule subtypes (Solid, Part-solid and GGN) in the combined test and validation set. *PulmoSeek*: AUC in solid 0.858 [0.761-0.938], in part-solid 0.808 [0.695-0.906], in GGN 0.878 [0.766-0.963]; Sensitivity in solid 1.000 [0.702-1.000], in part-solid 0.947 [0.509-1.000], in GGN 0.964 [0.518-1.000]; Specificity in solid 0.484 [0.387-0.903], in part-solid 0.333 [0.222-0.944], in GGN 0.364 [0.182-1.000]; *20-feature methylation model*: AUC in solid 0.854 [0.759-0.934], in part-solid 0.803 [0.686-0.909], in GGN 0.872 [0.752-0.963]; Sensitivity in solid 0.894 [0.809-1.000], in part-solid 0.860 [0.702-0.965], in GGN 0.875 [0.696-0.982]; Specificity in solid 0.710 [0.484-0.871], in part-solid 0.611 [0.222-0.833], in GGN 0.727 [0.182-1.000]. Solid group (n=78), part-solid group (n=75), GGN group (n=67).

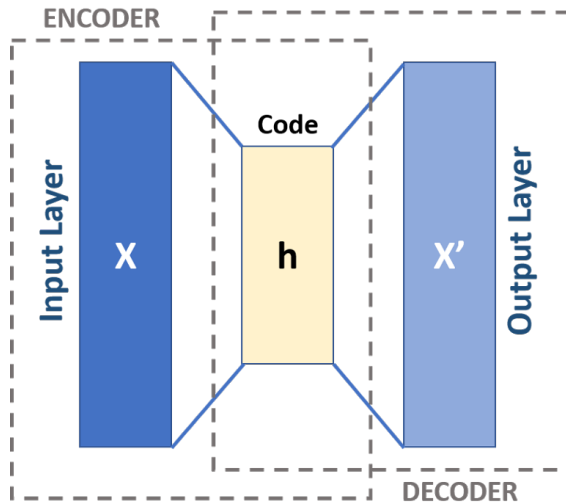


Figure S8 The general architecture of an autoencoder (AE) neural network. The AE model was composed of two modules, i.e., an encoder and a decoder, which were both implemented by neural networks. The encoder encoded the input matrix X into a hidden vector h . The decoder reconstructed a matrix X' from h , which should be as close to X as possible. The loss function of this model was formulated as: $L = \|X - X'\|_2 / (r \times c)$, where r and c were the number of rows and columns of X , respectively. After the model was trained, the encoder was used again to retrieve the hidden vector for each input, which was used as the representation of each input in the downstream pipeline.

Table S1 Characteristics of patients in training, test and validation sets stratified by cancer status

Category	Training Set			Test Set			Validation Set		
	Malignant Nodules	Benign Nodules	P value	Malignant Nodules	Benign Nodules	P value	Malignant Nodules	Benign Nodules	P value
Total Number	253	56		60	20		100	40	
Median Age - yr	58 (28-85)	50 (28-75)	***	59 (36-76)	50.5 (30-75)		55 (23-84)	54 (35-72)	
Gender – No.									
Male	103	32		31	12		50	21	
Female	150	24		29	8		50	19	
Smoking history									
Status – No.									
Never	194	41		39	14		71	28	
Former	29	5		12	4		12	5	
Current	29	10		8	2		17	7	
No data	1	0		1	0		0	0	
Pack-year	30 (3–90)	30 (5–120)		35 (2–100)	30 (3.5–40)		30 (4–200)	17.5 (1–40)	
Lesion size									
< 5 mm	0	0		0	0		0	0	
5 – 20 mm	180	43		43	14		73	30	
20 – 30 mm	68	12		16	6		27	10	
>30 mm	0	0		0	0		0	0	
No data	5	1		1	0		0	0	
Location – no.									
Upper lobe	98	27		26	10		26	20	
Lower lobe	155	29		34	10		74	20	

Age and nodule size between malignant and benign groups were compared using Student's T test. Gender and nodule locations were compared using Fisher' exact test. *P<0.05, **P<0.01, ***P<0.001

Table S2 Characteristics of patients in training, test and validation sets stratified by nodule types

Categories	Training Set		Test Set		Validation Set		Total
	Malignant	Benign	Malignant	Benign	Malignant	Benign	
GGN	85	7	24	2	32	9	159
Part-solid	99	17	15	6	42	12	191
Solid	69	32	21	12	26	19	179
Total	253	56	60	20	100	40	

Table S3 Characteristics of patients in training, test and validation sets stratified by cancer stages

<i>Cancer Stage</i>	Training Set	Test Set	Validation Set	Total
0	17	6	2	25
I	207	44	86	337
II	13	6	0	19
III	9	2	0	11
IV	0	0	0	0
Unknown	7	2	12	21
Total	253	60	100	413

Table S6 Pulmonary nodule model (20-, 50- and 500- feature) performance metrics

All Nodule Sizes			
	20-feature model		
	Test Set (N=80)	Validation Set (N=140)	Test and Validation Set (N=220)
Accuracy (95%CI)	0.800 (0.675-0.912)	0.836 (0.757-0.886)	0.823 (0.736-0.877)
Sensitivity (95%CI)	0.817 (0.617-0.983)	0.910 (0.770-0.990)	0.875 (0.719-0.975)
Specificity (95%CI)	0.750 (0.500-0.950)	0.650 (0.300-0.800)	0.683 (0.417-0.833)
Positive predictive value (95%CI)*	0.907 (0.848-0.979)	0.867 (0.778-0.919)	0.881 (0.812-0.925)
Negative predictive value (95%CI)*	0.577 (0.429-0.900)	0.743 (0.568-0.952)	0.672 (0.512-0.895)
Positive predictive value (95%CI) ^{&}	0.266 (0.172-0.635)	0.224 (0.135-0.335)	0.235 (0.153-0.339)
Negative predictive value (95%CI) ^{&}	0.974 (0.953-0.996)	0.985 (0.967-0.998)	0.980 (0.962-0.995)
	50-feature model		
	Test Set (N=80)	Validation Set (N=140)	Test and Validation Set (N=220)
Accuracy (95%CI)	0.800 (0.713-0.912)	0.793 (0.729-0.857)	0.795 (0.741-0.859)
Sensitivity (95%CI)	0.800 (0.667-0.967)	0.850 (0.760-0.980)	0.831 (0.744-0.969)
Specificity (95%CI)	0.800 (0.550-0.950)	0.650 (0.375-0.800)	0.700 (0.467-0.817)
Positive predictive value (95%CI)*	0.923 (0.859-0.981)	0.859 (0.790-0.915)	0.881 (0.823-0.925)
Negative predictive value (95%CI)*	0.571 (0.454-0.867)	0.634 (0.521-0.905)	0.609 (0.519-0.857)
Positive predictive value (95%CI) ^{&}	0.308 (0.185-0.658)	0.212 (0.143-0.323)	0.235 (0.162-0.339)
Negative predictive value (95%CI) ^{&}	0.973 (0.957-0.994)	0.975 (0.961-0.995)	0.974 (0.963-0.993)
	500-feature model		
	Test Set (N=80)	Validation Set (N=140)	Test and Validation Set (N=220)
Accuracy (95%CI)	0.838 (0.625-0.912)	0.821 (0.636-0.871)	0.827 (0.645-0.873)
Sensitivity (95%CI)	0.900 (0.517-0.967)	0.950 (0.580-0.990)	0.931 (0.581-0.975)
Specificity (95%CI)	0.650 (0.550-1.000)	0.500 (0.350-0.875)	0.550 (0.450-0.917)
Positive predictive value (95%CI)*	0.885 (0.853-1.000)	0.826 (0.788-0.933)	0.847 (0.821-0.949)
Negative predictive value (95%CI)*	0.684 (0.392-0.895)	0.800 (0.421-0.957)	0.750 (0.423-0.889)
Positive predictive value (95%CI) ^{&}	0.222 (0.177-1.000)	0.174 (0.142-0.384)	0.187 (0.160-0.438)
Negative predictive value (95%CI) ^{&}	0.983 (0.946-0.996)	0.989 (0.942-0.998)	0.986 (0.946-0.995)

* Cancer prevalence = 78% in the current cohort
 & Cancer prevalence = 10%, reported in an intended-use population

Table S7 PulmoSeek performance in sub-validation cohorts with different prevalence**Sub-validation Cohorts 1 – Malignancy Prevalence 79%**

PulmoSeek	23 B + 90 M
Accuracy (95%CI)	0.858 (0.681-0.920)
Sensitivity (95%CI)	1.000 (0.633-1.000)
Specificity (95%CI)	0.304 (0.174-0.957)
Positive predictive value (95%CI)	0.849 (0.826-0.985)
Negative predictive value (95%CI)	0.720 (0.370-1.000)

Sub-validation Cohorts 2 – Malignancy Prevalence 50%

PulmoSeek	7 B + 7 M
Accuracy (95%CI)	0.500 (0.357-0.857)
Sensitivity (95%CI)	0.857 (0.143-1.000)
Specificity (95%CI)	0.500 (0.350-0.875)
Positive predictive value (95%CI)	0.500 (0.364-1.000)
Negative predictive value (95%CI)	0.900 (0.200-1.000)

Sub-validation Cohorts 3 – Malignancy Prevalence 23%

PulmoSeek	12 B + 3 M
Accuracy (95%CI)	0.667 (0.444-1.000)
Sensitivity (95%CI)	1.000 (1.000-1.000)
Specificity (95%CI)	0.571 (0.286-1.000)
Positive predictive value (95%CI)	0.400 (0.286-1.000)
Negative predictive value (95%CI)	1.000 (1.000-1.000)

Table S8 Validation of the PulmoSeek performance with different histology types in the combined test and validation set

	Histology Type	Negative	Positive	Total	Sensitivity	Specificity
Malignant	AIS	1	7	8	87.5%	
	MIA	1	20	21	95.2%	
	IA	2	108	110	98.2%	
	SC	1	9	10	90.0%	
	SCLC	0	1	1	100.0%	
	Others	0	10	10	100.0%	
	Sum of Malignant	5	155	160	96.9%	
Benign	GRAN	1	5	6		
	HAM	2	4	6		
	INF	12	15	27		
	Others	10	11	21		
	Sum of Benign	25	35	60		47.6%

AIS: adenocarcinoma *in situ*; MIA: minimally invasive adenocarcinoma; IA: invasive adenocarcinoma; SC: squamous cell; SCLC: small cell
 GRAN: granuloma; HAM: hamartoma; INF: inflammation

Table S9 Characteristics of 6-20mm patients in training, test and validation sets stratified by cancer status

Category	Test Set			Validation Set		
	Malignant Nodules	Benign Nodules	P value	Malignant Nodules	Benign Nodules	P value
Total Number	42	14		71	29	
Age - yr	56 (36-76)	52.5 (30-75)		54 (23-75)	53 (35-72)	
Gender – no.						
Male	19	8		31	15	
Female	23	6		40	14	
Category						
GGN	20	2		27	8	
Part-solid	11	3		26	9	
Solid	11	9		18	12	
Cancer stage						
0 - I	37			63		
IIA	1			0		
IIB	1			0		
III	1			0		
IV	0			0		
Unknown	2			8		

Age between malignant and benign groups was compared using Student's T test. Gender and nodule subtype were compared using Fisher's exact test. *P<0.05, **P<0.01, ***P<0.001

Table S10 PulmoSeek performance metrics in 6-10 mm, 11-15 mm and 16-20 mm lesion sizes

PulmoSeek	6-10 mm (N=35)	11-15 mm (N=38)	16-20 mm (N=24)
AUC (95%CI)	0.670 (0.479-0.837)	0.929 (0.831-0.997)	1.000 (1.000-1.000)
Accuracy (95%CI)	0.750 (0.444-0.806)	0.769 (0.641-0.949)	0.920 (0.800-1.000)
Sensitivity (95%CI)	1.000 (0.417-1.000)	1.000 (0.500-1.000)	1.000 (0.762-1.000)
Specificity (95%CI)	0.250 (0.083-0.833)	0.308 (0.077-1.000)	0.500 (0.250-1.000)
Positive predictive value (95%CI)*	0.727 (0.600-0.875)	0.743 (0.684-1.000)	0.913 (0.875-1.000)
Negative predictive value (95%CI)*	1.000 (0.250-1.000)	1.000 (0.477-1.000)	1.000 (0.444-1.000)
Positive predictive value (95%CI) &	0.129 (0.077-0.280)	0.138 (0.107-1.000)	0.182 (0.129-1.000)
Negative predictive value (95%CI) &	1.000 (0.857-1.000)	1.000 (0.942-1.000)	1.000 (0.974-1.000)

* Cancer prevalence = 78%, in the current cohort

& Cancer prevalence = 10%, reported in an intended-use population

Table S11 PulmoSeek performance compared to integration of PET-CT after PulmoSeek in the combined test and validation set

AGE 40-70, with PET-CT record

Solid nodules only (N=14)		
	PulmoSeek	PulmoSeek + PET-CT *
False Positive Ratio	14.2% (2 out of 14)	16.6% (2 out of 12)
False Negative Ratio	0% (0 out of 0)	100% (2 out of 2)
Combined solid and part-solid nodules (N=27)		
	PulmoSeek	PulmoSeek + PET-CT *
False Positive Ratio	14.8% (4 out of 27)	11.7% (2 out of 17)
False Negative Ratio	0% (0 out of 0)	80.0% (8 out of 10)

AGE 56-70, with PET-CT record

Solid only (N = 9)		
	PulmoSeek	PulmoSeek + PET-CT *
False Positive Ratio	22.2% (2 out of 9)	28.5% (2 out of 7)
False Negative Ratio	0% (0 out of 0)	100% (2 out of 2)
Combined solid and part-solid nodules (N = 27)		
	PulmoSeek	PulmoSeek + PET-CT *
False Positive Ratio	12.5% (2 out of 16)	18.1% (2 out of 11)
False Negative Ratio	0% (0 out of 0)	100% (5 out of 5)

* PET-CT testing on the positive group using PulmoSeek