JCI The Journal of Clinical Investigation

Accurate diagnosis of pulmonary nodules using a non-invasive DNA methylation test

Wenhua Liang, … , Jian-Bing Fan, Jianxing He

J Clin Invest. 2021. https://doi.org/10.1172/JCI145973.

Clinical Medicine In-Press Preview Genetics Oncology

BACKGROUND. Current clinical management of patients with pulmonary nodules involves either repeated LDCT/CT scans or invasive procedures yet causes significant patient misclassification. An accurate non-invasive test is needed to identify malignant nodules and reduce unnecessary invasive tests.

METHOD. We developed a diagnostic model based on targeted DNA methylation sequencing of 389 pulmonary nodule patients' plasma samples, and then validated in 140 plasma samples independently. We tested the model in different stages and subtypes of pulmonary nodules.

RESULTS. A 100-feature model was developed and validated for pulmonary nodule diagnosis: the model achieved a ROC-AUC of 0.843 on 140 independent validation samples with an accuracy of 0.800. The performance was well maintained in, 1) 6-20 mm size subgroup (N=100), with a sensitivity of 1.000 and adjusted NPV of 1.000 at 10% prevalence; 2) stage I malignancy (N=90), with a sensitivity of 0.971; 3) different nodule types - solid nodules (N=78) with a sensitivity of 1.000 and adjusted NPV of 1.000, part-solid nodules (N=75) with a sensitivity of 0.947 and adjusted NPV of 0.983, and ground-glass nodules (N=67) with a sensitivity of 0.964 and adjusted NPV of 0.989 at 10% prevalence. This methylation […]

Accurate Diagnosis of Pulmonary Nodules Using a Non-invasive DNA Methylation Test

Wenhua Liang[1,*], Zhiwei Chen[2,3,*], Caichen Li[1,*], Jun Liu[1,*], Jinsheng Tao[2,*], Xin Liu[3], Dezhi Zhao[2], Weiqiang Yin[1], Hanzhang Chen[1], Chao Cheng[4], Fenglei Yu[5], Chunfang Zhang[6], Lunxu Liu[7], Hui Tian[8], Kaican Cai[9], Xiang Liu[10], Zheng Wang[11], Ning Xu[12], Qing Dong[13], Liang Chen[14], Yue Yang[15], Xiuyi Zhi[16], Hui Li[2], Xixiang Tu[2], Xiangrui Cai[17], Zeyu Jiang[2], Hua Ji[17,18], Lili Mo[1], Jiaxuan Wang[1], Jian-Bing Fan[19,2,§], Jianxing He[1,9,§]

1, Department of Thoracic Surgery and Oncology, The First Affiliated Hospital of Guangzhou Medical University, China National Center for Respiratory Medicine, China State Key Laboratory of Respiratory Disease & National Clinical Research Center for Respiratory Disease, Guangzhou, China 510120

2, AnchorDx Medical Co., Ltd., Unit 502, No. 8, 3rd Luoxuan Road, International Bio-Island, Guangzhou, China 510300

3, AnchorDx, Inc., 46305 Landing Pkwy, Fremont, CA, United States 94538

4, Department of Thoracic Surgery, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China 510080

5, Department of Thoracic Surgery, The Second Xiangya Hospital of Central South University, Changsha, China 410011

6, Department of Thoracic Surgery, Xiangya Hospital, Central South University, Changsha, Chnia 410008

7, Department of Thoracic Surgery, West China Hospital of Sichuan University, Chengdu, China 610041

8, Department of Thoracic Surgery, Qilu Hospital of Shandong University, Jinan, China 250012

9, Department of Thoracic Surgery, Nanfang Hospital of Southern Medical University, Guangzhou, China 510515

10, Department of Thoracic Surgery, The Second Hospital, University of South China, Hengyang, China 421000

11, Department of Thoracic Surgery, Shenzhen People's Hospital, Shenzhen, China 518020

12, Department of Thoracic Surgery, Anhui Chest Hospital, Hefei, China 230022

13, Department of Thoracic Surgery, The Forth Affiliated Hospital of Harbin Medical University, Harbin, China 150001

14, Department of Thoracic Surgery, Jiangsu Province Hospital, Nanjing, China 210029

15, Department of Thoracic Surgery, Beijing Cancer Hospital, Beijing, China 100142

16, Department of Thoracic Surgery, Xuanwu Hospital, Capital Medical University, Beijing, China 100053

17, College of Computer Science, Nankai University, Tianjin, China 300353

18, Laboratory for Foundations of Computer Science, School of Informatics, University of Edinburgh, United Kingdom

19, Department of Pathology, School of Basic Medical Science, Southern Medical University, Guangzhou, China 510515

*, these authors contributed equally to this work.

§, Corresponding authors: Jian-Bing Fan (Phone: 86.18701841892; Email: jianbingfan1115@smu.edu.cn), Department of Pathology, Southern Medical University, 1838 ShaTai Road, Guangzhou, 510515, China; Wenhua Liang (Phone: 86.13710249454; Email: liangwh@gird.cn) or Jianxing He (Phone: 86.13802777270; Email: drjianxing.he@gmail.com), Department of Thoracic Surgery and Oncology, The First Affiliated Hospital of Guangzhou Medical University, 151 Yanjiang Road, Guangzhou, 510120, China.

Abstract

Background. Current clinical management of patients with pulmonary nodules involves either repeated LDCT/CT scans or invasive procedures yet causes significant patient misclassification. An accurate non-invasive test is needed to identify malignant nodules and reduce unnecessary invasive tests.

Method. We developed a diagnostic model based on targeted DNA methylation sequencing of 389 pulmonary nodule patients' plasma samples, and then validated in 140 plasma samples independently. We tested the model in different stages and subtypes of pulmonary nodules.

Results. A 100-feature model was developed and validated for pulmonary nodule diagnosis: the model achieved a ROC-AUC of 0.843 on 140 independent validation samples with an accuracy of 0.800. The performance was well maintained in, 1) 6-

20 mm size subgroup (N=100), with a sensitivity of 1.000 and adjusted NPV of 1.000 at 10% prevalence; 2) stage I malignancy (N=90), with a sensitivity of 0.971; 3) different nodule types - solid nodules (N=78) with a sensitivity of 1.000 and adjusted NPV of 1.000, part-solid nodules (N=75) with a sensitivity of 0.947 and adjusted NPV of 0.983, and ground-glass nodules (N=67) with a sensitivity of 0.964 and adjusted NPV of 0.989 at 10% prevalence. This methylation test, called PulmoSeek, outperformed PET-CT and two clinical prediction models (Mayo and Veterans Affairs) in discriminating malignant pulmonary nodules from benign ones.

Conclusion. This study suggests that the blood-based DNA methylation model may provide a better test for classifying pulmonary nodules, which could help facilitate the accurate diagnosis of early-stage lung cancer from pulmonary nodule patients and guide clinical decisions.

Funding The National Key Research and Development Program of China; Science and Technology Planning Project of Guangdong Province; The National Natural Science Foundation of China National.

Introduction

Lung cancer is the leading cause of cancer-related mortality globally (1). It has been shown that the prognosis of lung cancer is highly correlated with the stage of the disease at diagnosis, with a 5-year overall survival rate decreasing dramatically from 85% for stage IA to 6% for stage IV disease (2). This makes lung cancer screening a highly favorable strategy to save lives and related-medical costs.

The National Lung Screening Trial (NLST) has demonstrated that lung cancer screening by low-dose CT (LDCT) reduces 20% mortality among current and former smokers of high lung cancer risk (>55 years old, >30 packs per year) which has led to a quick adoption of LDCT screening worldwide (3). Although LDCT does identify small nodules more effectively than conventional X-rays, this advantage comes with the challenge of distinguishing the small percentage of malignant nodules (~10-20%) from the majority of the detected nodules that are deemed benign (4). Clinical nodule assessment tools, such as Mayo Clinic and Veterans Affairs models, based on imaging parameters as well as other risk factors are widely used (5). However, the sensitivity of these tools are largely affected by nodule size and location. Suspected lung cancer lesions identified by LDCT can be further diagnosed via invasive approaches (e.g. bronchoscopy, transthoracic needle aspiration, and surgery); however, complications may emerge, including hemorrhage, infection, pneumothorax, and even death. To avoid high false-positive rate, the new Lung-

RADS classification and guidelines set the detection of nodules of 6 mm as the threshold for positivity. Nevertheless, positive CT scans can still be indecisive clinically, particularly for the class of intermediate-risk nodules (usually range from 6 to 20 mm in size, with a 5-65% probability of malignance as calculated by the clinical assessment tools) (6).

Liquid biopsy has been considered as an easier and safer, more cost effective, and less invasive method for cancer diagnosis and monitoring. Most noninvasive early detection approaches depend on identification of tumor-derived nucleic acids or proteins present in blood. For example, a blood test of proteomic biomarkers - PANOPTIC has been developed (7). ctDNA is exquisitely specific for an individual's tumor, therefore it can bypass the issue of false positivity encountered with other circulating biomarkers. Advancement in digital PCR and next-generation sequencing (NGS)-based technologies have drastically improved accuracy and sensitivity of ctDNA analysis in the detection of early-stage cancers (8). This fast-developing field has drawn attention from international societies such as IASLC which advocats using liquid biopsy in the management of non-small cell lung cancer (NSCLC).

In this study, we conducted ctDNA methylation profiling, instead of somatic mutation detection, to develop and validate a blood-based pulmonary nodule diagnosis test. When combined with standard care, it provides a more accurate clinical measurement for pulmonary nodule management.

Results

Clinical cohort. A total of 585 LDCT positive patients were enrolled from thoracic surgery departments of 14 clinical sites across 8 different provinces in China. The percentage of malignancy based on pathological diagnosis from each province ranged from 75% to 88% (Supplemental Figure S1). Of which, 56 samples were excluded from analysis due to failed experimental QC, e.g. inadequate cfDNA amount extracted from plasmas. The remaining 529 patients' plasma samples (116 benign and 413 malignant) were used for DNA methylation profiling, model development and validation. An overview of the study design is shown in Figure 1 and the demographic characteristics for the 529 patients are shown in Table 1.

The 529 plasma samples were first split into a model development set and an independent validation set at 3:1 ratio. Furthermore, the model development set was divided into a training set (56 benign + 253 malignant) and a test set (20 benign + 60 malignant) in a way that the distribution of malignancy, age and gender of the

test set matched to the training set, as shown in Figure 1C. The percentage of malignancy was 82% and 75% in the training and test set, respectively. The samples used for model development were primarily early-stage non–small cell lung cancer. Specifically, stage I + II cancers comprised of 94% or 98% from the total cancer patients in the training set and the test set, respectively. Benign and malignant samples were matched with respect to gender and smoking status ($p>0.05$). The average size of the nodules in the benign group is 15.8 mm [9.6-22.0 mm], which is statistically smaller ($p<0.05$) than that of the malignant group which is 16.4 mm [9.9-22.9 mm]. A summary of nodule types and AJCC stage information is shown in Supplemental Table S2 and S3.

Development and validation of the diagnosis model "PulmoSeek" for pulmonary nodule diagnosis. Methylation profiles of 309 plasma samples (Supplemental Table S1, training set) were analyzed using AnchorDx's proprietary targeted methylation sequencing platform with a panel of 12,899 pre-selected lung cancer-specific methylation regions, corresponding to 105,844 CpG sites (9). A specific methylation signature was selected based on its performace of differentiating malignant from benign nodules.

The derived classification model, comprised of 500 methylation target regions (features) achieved an ROC-AUC of 0.823 [0.771-0.884] in the test set. Compared to the 500-feature model, the top 10 features within the model showed AUC values between 0.561-0.754 in the training set and 0.525-0.720 in the test set, demonstrating the necessity for building a multiple feature-based model (Supplemental Figure S2). For further downstream analysis, we annotated the selected 500 CpG features and performed a gene enrichment analysis. A total of 89 Gene Ontology (GO) categories were significantly enriched (Table S4). The enriched categories include tissue proliferation and differentiation such as "embryonic morphogenesis" (q-value=$10^{-9.3}$), cell fate commitment (q-value=$10^{-4.7}$), stem cell proliferation (q-value=$10^{-4.6}$), and "epithelial tube morphogenesis" (q-value=$10^{-3.0}$). In addition, transcriptional factor activity, such as "RNA polymerase II-specific DNA-binding transcription activator activity", were also significantly enriched (q-value=$10^{-7.5}$). This result suggested that specific epigenetic signaling responsible for cell differentiation/reprogramming might be essential for pulmonary nodule development.

The performance of the model remained stable during a recursive feature elimination process: the smallest number of features that maintained an AUC within 1% of the 0.829 was 20, with an AUC of 0.810 [0.783-0.850] in the test set (Supplemental

Figure S3). This indicates that a robust signature is maintained across different numbers of features selected.

We then subsequent analyzed on a 100-feature model, named PulmoSeek, which had a high overall AUC of 0.829 [0.719-0.942], that yielded the highest sensitivity of 0.933 [0.533-0.983] at a specificity of 0.600 [0.500-1.000] in the test set, corresponding to an accuracy of 0.850 [0.625-0.912] (Figure 2A and 2C, Table 2). The detailed information for each methylation feature of PulmoSeek was listed in Supplemental Table S5. Given excessive false-positives and over-diagnosis in LDCT screening, it is most desireable to prevent patients with benign nodules from unnecessary invasive procedures that can be achieved under the premise of a high true-positive rate. We assessed PulmoSeek's performance with regard to its negative predictive value (NPV) and positive predictive value (PPV). In the current study cohort of 78% prevalence, the NPV was 0.750 [0.396-0.929] and the PPV was 0.875 [0.852-1.000] in the test set (Table 2). The sensitivities of the top 20-, 50- and 500-feature models were 0.800 [0.675-0.912], 0.800 [0.713-0.912] and 0.900 [0.517-0.967] respectively, as shown in Supplemental Table S6.

We then used an independent cohort of 140 patient plasma samples (40 benign and 100 malignant, Supplemental Table S2, validation set) to further evaluate the

203 performance of PulmoSeek. PulmoSeek achieved an AUC of 0.843 [0.769–0.918]
204 (Figure 2A and 2D) with sensitivity of 0.990 [0.610-1.000] at specificity of 0.325
205 [0.200-0.875], and an overall accuracy of 0.800 [0.657-0.871]. The NPV was 0.929
206 [0.444-1.000] and the PPV was 0.786 [0.758-0.938]. In an intended-use population
207 with a prevalence of malignanat nodule at 10% (10), the NPV was calculated as
208 0.997 [0.947-1.000] (Table 2). We further splitted the validation cohort into three
209 sub-cohorts from high to low prevalences. We found that the NPV increased from
210 0.790 [0.370-1.000] to 1.000 [1.000-1.000] when the sub-cohort prevalence
211 decreased from 79% to 23% (Supplemental Table S7).
212
213 The performance of PulmoSeek in patients with nodules of different histological
214 types was further explored. Robust sensitivity for different subtypes, including
215 minimally invasive adenocarcinoma (MIA, 95.2%), invasive adenocarcinoma (IA,
216 98.2%) and squamous cell carcinoma (SCC, 90.0%) were observed (Supplemental
217 Table S8).
218
219 We also compared the performance of PulmoSeek to two clinical assessment models
220 - the Mayo Clinic and Veterans Affairs models that are based on clinical information
221 and radiological characteristics, including nodule size and location among others. In
222 the validation set, PulmoSeek outperformed both of the clinical models with an AUC

of 0.843 [0.769-0.918] vs. AUC of 0.591 [0.482-0.688] for the Mayo Clinic model and 0.544 [0.442-0.640] for the VA model (Figure 2B)

Classification accuracy of the model in very early-stage lung cancers. Very early-stage cancer (TNM Stage I) poses the greatest challenge for cancer diagnosis using a liquid biopsy (11). We tested PulmoSeek in different Stage I sub-stages in the validation cohort: it achieved sensitivities of 0.941 and 1.00 in Stage IA (N=85) and Stage IB (N=5), more specifically, 0.864, 0.950 and 1.000 in Stage IA1 (N=22), IA2 (N=40) and IA3 (N=23), respectively (Figure 3A and B). In the combined test and validation set, PulmoSeek detected malignancies with sensitivity of 0.971 [0.942-0.993] for Stage 0-I and 0.875 [0.625-1.000] for later stage cancers (Supplemental Figure S4A). The decreased sensitivity in late stage cancers could be due to the limited number of late-stage samples (N=8) which is not statistically significant (P=0.248). Besides, the differences in performance for PumoSeek in different groups were also calculated and we observed no significant difference between each group (Supplemental Figure S4B and S4C). Taken together, these results validated the accuracy of PulmoSeek, especially in detecting very early-stage lung cancers.

PulmoSeek outperformed clinical prediction models and conventional cancer biomarkers tests in indeterminate nodules. Diagnosis of indeterminate pulmonary nodules (IPN, nodules ranging between 6-20 mm in size) is challenging for clinicians due to the lack of well-specified optimal action strategies (12). The 6-20 mm size nodules comprised about 70% of the test set (56 of 80) and the independent validation set (100 of 140) in this study (Supplemental Table S9). PulmoSeek achieved an AUC of 0.762 [0.610-0.913], sensitivity of 0.905 [0.429-0.976] and specificity of 0.500 [0.286-1.000] in the test set (Figure 4A and 4B, Table 2). In the independent validation set, PulmoSeek achieved an AUC of 0.844 [0.759-0.932], sensitivity of 1.000 [0.577-1.000], specificity of 0.300 [0.172-0.931] (Figure 4A and 4D, Table 2 and Supplemental Table S10). For nodules above 20 mm (N=59), PulmoSeek had an AUC of 0.860 [0.740-0.964] with sensitivity of 0.977 [0.628-1.000] and specificity of 0.562 [0.375-0.938] (Supplemental Figures S5).

When compared to the Mayo Clinic and Veterans Affairs models, PulmoSeek outperformed both clinical models in the validation set in which an AUC of 0.602 [0.482-0.719] was obtained with the Mayo Clinic model and an AUC of 0.512 [0.402-0.633] was obtained with the Veterans Affairs model (Figure 4C).

Consistent with previous studies, conventional cancer biomarkers like carcinoembryonic antigen (CEA), cancer antigen 125 (CA-125) and cancer antigen 135 (CA-135) alone failed to effectively identify malignant nodules in our cohort (13). The corresponding sensitivity of CEA, CA-125 and CA-135 was only 0.010, 0.030 and 0.030, respectively, as compared to sensitivity of 0.950 by using PulmoSeek (Supplemental Figure S6).

PulmoSeek outperformed PET-CT in different nodule types, including ground-glass nodule (GGN). Positron emission tomography/computed tomography (PET-CT) is known to be more accurate than CT alone for characterizing solid-type pulmonary nodules, resulting in fewer equivocal findings (14). Thus, low to intermediate risk nodules are usually recommended to be further evaluated by PET-CT. However, PET-CT performance drops significantly for sub-solid nodules (part-solid and GGN). We assessed the performance of PulmoSeek in comparison to PET-CT on the participants with established PET-CT records in our independent validation set. The accuracy of te PulmoSeek was significantly higher than PET-CT: It correctly classified 8 out of 10 patients in the solid nodule subgroup, 9 out of 11 in the part-solid nodule subgroup and 5 out of 5 in the ground-glass nodule subgroup; while PET-CT correctly classified 6 out of 10 patients in the solid nodule subgroup, 7 out of 11 in the part-solid nodule subgroup and 0 out of 5 in the ground-glass

nodule subgroup (Figure 5). This performance was maintained across all nodule types in the combined test and independent validation sets: the model demonstrated a sensitivity of 1.000 [0.702-1.000] in the solid subgroup (N=78), 0.947 [0.509-1.000] in the part-solid subgroup (N=75) and 0.964 [0.518-1.000] in the GGN subgroup (N=67) (Supplemental Figure S7).

A strategy of integrating liquid biopsy-based ctDNA and protein marker analysis followed by PET-CT imaging for cancer screening has been proposed (15). We tried to assess this strategy in our cohort by testing the performance of PET-CT on the malignant nodules identified by our methylation model. In both solid and part-solid nodule groups, integration of PET-CT did not reduce false-positive rate. Rather, it introduced a significant number of false-negatives: in solid nodules, PulmoSeek had a false-positive rate of 14.2% (2 out of 14 misclassified) while integration of PET-CT resulted in a false-positive rate of 16.2% (2 out of 12) and a false-negative rate of 100% (2 out of 2); In all nodules, PulmoSeek had a false-positive rate of 14.8% (4 out of 27 misclassified) while integration of PET-CT had a false-positive rate of 11.7% (2 out of 17) and a false-negative rate of 80% (8 out of 10) (Supplemental Table S11).

Discussion

In this study, we analyzed ctDNA methylation profiles in 529 pulmonary nodule patients from 14 hospitals in China, and developed and validated a model called PulmoSeek for pulmonary nodule diagnosis. Notably, PulmoSeek demonstrated high sensitivity and NPV at a moderate specificity across different lesion locations, nodule types and stages of lung cancer. To the best of our knowledge, this is the largest retrospective study so far to validate a blood-based methylation model for lung nodule diagnosis.

Most recently, a prospective, interventional study of more than 10,000 women using a multi-omics blood test coupled with PET-CT imaging demonstrated its clinical potential for early cancer screening (15). However, this strategy may not apply to lung cancer screening efficiently: 1) using a sequencing-based method for initial screening can be costly and throughput-limited; 2) a ctDNA somatic mutation assay may yield high number of false positives that would need a PET-CT to filter out; however, it is reported that PET-CT performs sub-optimally in characterizing sub-solid nodules.

Our study demonstrated a potential new diagnosis work-up of LDCT followed by a blood-based test as a less invasive and cost-effective strategy to identify early-stage lung cancer. The cost-effectiveness advantage is significant: chest LDCT costs less

than $50 (in China) and the process only takes a few minutes (16). This is clinically critical because in the United States, there is a well-documented high rate of pulmonary nodules (17). Similarly, in China, it is estimated that over 100M people live with lung nodules and the number is growing quickly each year (18). LDCT screening has already been widely adopted as a major tool for lung cancer screening. We believe that a strategy of coupling LDCT with PulmoSeek is more practical and suited for population-based lung cancer screening.

It has been shown that in the National Lung Screening Trial (NLST) study, all enrolled patients had pulmonary nodules ≥4 mm in diameter and the false positive rate was over 96.4% after 3 rounds of LDCT screening (6). This is particularly impactful for nodules between 6-20 mm, i.e. indeterminate nodules (pCA 5-65%), which account for the majority of the nodules identified by LDCT (50-76%) and where the risk of malignancy is hard to determine with current clinical risk assessment models (19). Current guidelines suggest further evaluation with PET-CT scan, endobronchial ultrasound guided transbronchial forceps biopsy (EBUS-TBB) or transthoracic needle aspiration (TTNA). Integrated PET-CT imaging shows good sensitivity (~88%) and specificity (~75%). However, this performance is limited to solid nodules and there are still possibilities of false-positives (e.g. granulomatous disease) and false-negatives (e.g. carcinoid). It has been reported that the sensitivity

of PET-CT dropped to 50% in part-solid nodules and even lower than 20% in GGN nodules (20); the performance of EBUS-TBB on peripheral pulmonary nodules is largely dependent on nodule size: the diagnostic sensitivity is significantly higher for nodules of >20 mm than for those ≤20 mm (~50% sensitivity), with only 35% sensitivity for nodules between 5-10 mm in diameter (21). The EBUS-TBB procedure is also expertise dependent. TTNA has a 1% risk of hemorrhage and less patient compliance (22). Despite of the high medical costs associated with those three approaches, clinicians can still be left uncertain in management decisions, leading to potentially overdiagnosis and/or overtreatment. To fulfill the unmet clinical and economical needs, it calls for an alternative/complementary, noninvasive approach for nodule management: at one end, prompt necessary treatment when the nodule is in the early stages of lung cancer and at another end, minimize testing when the nodule is deemed benign.

PulmoSeek provides a potential solution to meet all of the above needs: this blood-based assay was developed on a group of pathology confirmed nodules mostly at early stages (Stage I+II, 92%) from thoracic departments and with a high prevalence of lung cancer (78%). In the current study, PulmoSeek achieved an AUC of 0.843, high sensitivity of 0.990 and NPV of 0.929 in the independent validation set. It outperformed current clinical assessment models (23). Ultimately, a rule-out test is

likely to be most clinically beneficial in the group of a lower prevalence of lung cancer: doctors would have a reliable test to rule out the "true negatives" and effectively reduce the "uncertain cases" so to avoid over-treatment. When adjusted to an average prevalence of 10%, the model had a very high NPV of 0.997 with specificity over 40%. This suggests that PulmoSeek alone could reduce more than 40% of unnecessary invasive procedures on benign nodules with less than 0.3% false negative rate. The superior NPV and sensitivity of PulmoSeek compare favorably to other published rule-out models to date, of which the NPV range between 85-98% (depending on the prevalence) and sensitivity between 85-97% (24-27). Those models are usually used in combination with clinical parameters like age, smoking status, nodule size/location and classic cancer biomarkers (e.g. CEA).

PulmoSeek demonstrated a robust performance in very early stages of lung cancer (Stage 0 and I). The slightly lower sensitivity observed in the later stage cancers (87.5%) was due to misclassification of 1 out of the 8 late-stage samples. This misclassified sample is a part-solid nodule of smaller nodule size (11mm), the subgroup PulmoSeek showed a relatively lower performance in the current study. We further tested with another independent cohort (N=12) of late stage nodule samples and PulmoSeek correctly identified all of them (Liang et al., unpublished

observations). Nevertheless, a larger number of cohort is required to further validate performane of PulmoSeek in late stage cancers.

In addition, PulmoSeek is accurate in diagnosing 6-20 mm indeterminate pulmonary nodules and sub-solid nodules (part-solid and GGN nodules), unlike other tests that are limited to solid nodules. We are currently combining PulmoSeek with LDCT image AI to further augment the overall diagnostic performance (Liang et al., unpublished observations).

To gauge the potential clinical utility, a trade-off value was calculated as following: specificity/(1 - sensitivity) ≥ (prevalence/1 - prevalence) × harm/benefit, where the harm/benefit ratio is the net harm of a false-negative test to the net benefit of a true-negative test (28)); PulmoSeek produced a harm/benefit value of 292.5 (e.g. 292.5 true-negative results accompanied with one false-negative result) in the intended-use population (10% prevalence). These results suggest that the trade-off is acceptable and warrant a future clinical utility study. As a matter of fact, a large prospective clinical validation study - The Thunder Project – has started in 2018 aiming to enroll more than 10,000 patients across 23 top hospitals in China (NCT03651986) (29). As of January, 2021, over 9,500 patients have already been enrolled (Zhong et al., unpublished observations).

In summary, we have developed and validated a ctDNA methylation assay for diagnosis of malignant and benign pulmonary nodules. It showed superior performance as compared to existing clinical procedures. Coupled with LDCT, it could become a robust tool for pulmonary nodule management and lung cancer screening.

Methods

Study design and participants. We performed a multicenter, retrospective diagnostic study using plasma samples collected from 14 hospitals' thoracic departments in China. 585 patients with malignant and benign pulmonary nodules were enrolled from May 2017 to Feburary 2019. The participating hospitals were the First Affiliated Hospital of Guangzhou Medical University, Xiangya Second Hospital of Central South University, The First Affiliated Hospital of Sun Yat-sen University, Shenzhen People's Hospital, Southern Hospital of Southern Medical University, Jiangsu Provincial People's Hospital, West China Hospital of Sichuan University, Xuanwu Hospital of Capital Medical University, Beijing Cancer Hospital, Qilu Hospital of Shandong University, The Second Affiliated Hospital of Nanhua University, Thoracic hospital of Anhui Province, Xiangya Hospital of Central South University, The Fourth Affiliated Hospital of Harbin Medical University.

Adult patients of 18 years old or older were included with the following criteria: either sex, single pulmonary nodules detected by standard- or low-dose CT screening with nodule size between 5 to 30 mm, nodule types of solid nodules (SN), part-solid nodules (mixed ground-glass nodules, mGGN) and pure ground-glass nodules (pGGN). Exclusion criteria included pregnant or lactating females, patients with 2 or more nodules with lesion size ≥5 mm, patients with metastasis symptoms like pleural effusion or mediastinal lymph node's shorter diameter >10 mm, patients without confirmed pathological diagnosis after surgery, or patients with cancer confirmed pathologically within 2 years prior to enrollment (except for nonmelanoma skin cancer). All patients underwent pathological examination and the detailed deidentified clinical information including demographics, LDCT imaging reports and pathology reports were transferred to the investigators.

Procedures. All blood samples were collected in Streck cell-free DNA BCT tubes (Streck, Inc. Cat# 218962) according to manufacturer's instructions and shipped to AnchorDx's certified molecular diagnosis laboratory. Plasma was separated immediatedly from the whole blood samples upon receiving, using standard protocol described previously (9) and stored at -80 degree until use. Repeated freezing and thawing of plasma was avoided to prevent cfDNA degradation and genomic DNA

contamination from white blood cells (WBC). cfDNA was isolated by the Thermo MagMAX™ Cell-Free DNA Kit (Thermo Fisher Scientific, Cat# A29319) according to the manufacturer's protocol. The concentration of cfDNA was measured by Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, Cat# Q32854) and the quality was examined using the Agilent High Sensitivity DNA Kit (Cat# 5067-4626).

Full details of sample preparation and targeted cfDNA methylation sequencing were described previously (9). In brief, bisulfite conversion was performed using the EZ DNA Methylation-Lightning Kit (Cat# D5031, Zymo Research) according to the manufacturer's protocol. Targeted genome methylation analysis was conducted using a proprietary AnchorIRIS technology on 10 ng input cfDNA. AnchorIRIS™ pre-library construction was carried out using AnchorDx EpiVisio™ Methylation Library Prep Kit (AnchorDx, Cat# A0UX00019) and AnchorDx EpiVisio™ Indexing PCR Kit (AnchorDx, Cat# A2DX00025). The amplified pre-hybridization libraries were subsequently purified using the IPB1 Magnetic Beads and the concentration was determined using the Qubit™ dsDNA HS Assay Kit. Pre-hybridization libraries containing more than 400 ng DNA were considered qualified for target enrichment. Next, target enrichment was performed using AnchorDx EpiVisio™ Target Enrichment Kit (AnchorDx, Cat# A0UX00031). A custom-made

lung cancer methylation panel (see below), which consists of 12,899 pre-selected regions enriched for lung cancer specific methylations was used for this study.

After probe hybridization, specific portions of the DNA libraries bound with biotinylated probes were pulled down using the Dynabeads M270 streptavidin beads (Thermo Fisher Scientific, Cat# 65306). These enriched libraries were further amplified with P5 and P7 primers by the KAPA HiFi HotStart Ready Mix (KAPA Biosystems, Cat# KK2602) and PCR product was then purified with Agencourt AMPure XP Magnetic Beads (Beckman Coulter, Cat# A63882). The resulted libraries were sequenced on the NovaSeq 6000 system (Illumina Inc.).

Lung cancer-specific methylation panel development. Early stage lung cancer methylation profiles were generated by targeted bisulfite sequencing - DNA extracted from a total of 232 tissue samples, including 133 benign pulmonary nodule samples (inflammation, granulomas, tuberculosis, fungal infection, hamartomas and sclerosing hemangioma) and 99 malignant pulmonary nodules (invasive adenocarcinoma, minimal invasisve adenocarcinoma, adenocarcinoma in situ, squamous cell lung cancer), were analyzed by the TruSeq Methyl Capture EPIC Library Kit (Illumina, Cat# FC-151-1002). Differentially methylated CpG sites were discovered using R package DSS version 2.14.0 (30). By using a filtering criteria of

480 $p<0.001$ and delta (i.e. group difference) >0.02, hypermethylated and
481 hypomethylated sites were identified. Together with the lung cancer-specific DNA
482 methylation markers discovered from The Cancer Genome Atlas (TCGA) database,
483 we developed a targeted methylation panel consisting of 12,899 lung cancer
484 associated informative methylation regions, covering 105,844 CpG sites.
485
486 *Sequencing data analysis*. Sequencing data were processed as previously reported
487 [9]. Briefly, the sequencing quality was evaluated by the Illumina Sequencing
488 Analysis Viewer and the FastQC software (Babraham Bioinformatics, Cambridge,
489 UK). Sequencing adapters and 3'-low quality bases were trimmed from raw
490 sequencing reads using a custom algorithm and then aligned to C->T *in silico*
491 converted hg19 reference genome, using Bismark version 0.17.0 (Bowtie2 as the
492 default aligner behind Bismark). Aligned reads were then evaluated by Picard
493 version 2.5.0 for metrics that measured the performance of target-capture based
494 bisulfite sequencing assays (http://broadinstitute.github.io/picard).The biases of
495 specific motifs or GC enriched regions were excluded. After the preliminary analysis,
496 we calculated the average coverage as well as the missing rate for each CpG site.
497 The CpG sites with coverage less than 30X and/or with missing rate >0.20 were
498 filtered out.
499

Differential methylation signature analysis. Differential methylation (DM) analysis was performed on the training cohort of lung cancer patients and controls using R package DSS version 2.14.0 (30). Differentially methylated CpG sites were identified by comparing malignant to benign samples ($p<0.001$, $delta>0.02$), and further assembled into differentially methylated regions (DMRs). Targeted regions of the capture panel covered by DMRs (at least 50% bases of a target region covered) were selected as candidate features to build classification models of malignant/benign state.

Deep learning based benign-malignant prediction modeling. Methylation features were selected by calculating the co-methylated reads (reads having at least 3 methylated CpGs within a sliding window of 5 CpGs or at least 2 methylated CpGs within a sliding window of 3 CpGs) ratios within the DMRs (9). Then, in light of the heuristic nature of various methylation metrics such as co-methylation and epiallele (31), an autoencoder (AE) neural network (32) was applied to further construct the representative methylation features. The AE is a type of unsupervised neural network with wide applications particularly in image processing. A general autoencoder architecture was shown in Supplemental Figure S8. We took advantage of it when analyzing methylation sequencing data to convert intractable high

519 dimensional sequencing reads into lower dimensional numerical representative
520 features (31).
521
522 In our model, the input matrix X represented each DMR and the hidden vector h was
523 the low dimensional representative feature of the DMR methylation status after
524 training. The encoder was implemented by ResNet model based convolutional
525 neural network model (33). For the decoder, to reconstruct the region from h,
526 deconvolutional layers, composing of the reverse operations of the convolutional
527 layers in the encoder, were implemented (34). The whole model was further
528 optimized by the Adam algorithm (35).
529
530 We then built a gradient boosted trees-based classifier with Scikit-Learn LightGBM
531 using the AE-based methylation features. During the training process, we tuned the
532 number of trees, maximum tree depth and the number of leaves used by the
533 lightGBM model, as these were major parameters to overcome the overfitting
534 problem. The learning rate and other parameters were kept at its default values. The
535 number of leaves was set up between 3-20, depth of trees between 3-10 and up to
536 1,000 for each model.
537

Comparison to Mayo Clinic and Veterans Affairs models. The Mayo Clinic model for malignancy in pulmonary nodules calculated the malignancy probability as a function of 3 clinical and 3 radiographic variables (36):

Probability of malignancy = $e^x/(1+e^x)$

$x = -6.8272 + (0.0391 \times \text{age}) + (0.7917 \times \text{smoking}) + (1.3388 \times \text{cancer}) + (0.1274 \times \text{Nodule diameter}) + (1.0407 \times \text{spiculation}) + (0.7838 \times \text{upper lobe})$

The Veterans Affairs model for malignancy in pulmonary nodules calculated the malignancy probability as a function of 3 clinical and 1 radiographic variable (37):

Probability of malignancy = 100 * [e(Logx) / (1 + e(Logx))]

x = −8.404 +2.061 * Smoke + 0.779 * Age / 10 + 0.112 * Diameter + 0.567 * Yearsquit/10, where smoke is 1 if a current or former smoker (otherwise 0), Age/10 is age in years divided by 10, Diameter is the largest diameter of the nodule in millimeters, and yearsquit/10 is the number of years since quitting smoking divided by 10.

Gene set enrichment analysis. We performed gene set enrichment analysis using the R-package Metascape (38).

Statistics. Statistical analysis was performed as described in each corresponding figure legend, and sample sizes are given in each corresponding figure legend. Categorical variables including gender, nodule subtypes etc., were compared using Fisher's exact test. Sensitivities of different AJCC stages for malignant nodules were also compared with Fisher's exact test. Continuous variables like age were compared

using Student's T test. 95% confidence intervals were calculated based on 2,000 bootstrap resamplings of the classification results. The sensitivity, specificity, accuracy, PPV, and NPV of PulmoSeek and other models in detecting malignant nodules were obtained by comparison to pathological outcomes. The receiver operating characteristic curves (ROCs) were obtained using the pROC R package (version 1.15.3). The positive and negative classifications of PulmoSeek were determined by the cutoff value (0.960) using the Youden's Index, while positive and negative values for CEA, CA-125 and CA-135 were determined by the clinical report. Unless otherwise specified, all statistical tests were two-sided. False discovery rate (FDR, Benjamini-Hochberg method) correction was used for multiple test correction. All statistical analysis was performed with R software, version 3.32.

Study approval. This study was approved by the institutional review boards (IRB) at the hospitals involved. Written consent has been obtained from each participant.

Authors Contributions

J.H. and J.B.F. conceived the study. W.L., Z.C., J.B.F. and J.H. designed the experiments. H.L., D.Z. and Z.C. performed the experiments. H.J., Z.Y., X.C., J.T. and X.T. performed the data modelling. All authors contributed to data acquisition. J.T., Z.C. and W.L. contributed to data visualization. Xin L., W.L., Z.C., J.B.F. and

J.H. contributed substantially to the development of this manuscript. All authors reviewed and approved the manuscript.

Acknowledgements

We thank all the participants for their generosity. This study was supported by the China National Science Foundation (NO. 82022048 & 81871893), the Key Project of Guangzhou Scientific Research Project (NO. 201804020030), National Key Research and Development Project (NO. 2017YFC0907903, 2017YFC1309002, & 2017YFC0112704), the Scheme of Guangzhou Economic and Technological Development District for Leading Talents in Innovation and Entrepreneurship (NO. 2017-L152), Scheme of Guangzhou for Leading Talents in Innovation and Entrepreneurship (NO. 2016007), Scheme of Guangzhou for Leading Team in Innovation (NO. 201909010010), Science and Technology Planning Project of Guangdong Province, China (NO. 2017B020226005), High-level University Construction Project of Guangzhou Medical University (NO. 20182737, 201721007, 201715907, 2017160107), Guangdong High Level Hospital Construction "Reaching Peak" Plan.

Competing Interests

The authors J.B.F., Z.C., J.T., Xin L., D.Z., H.L., X.T., Z.J. are current employees of AnchorDx Medical Co., Ltd, or AnchorDx, Inc. All other authors declare no competing financial interest.

References

1. Didkowska J, et al. Lung cancer epidemiology: contemporary and future challenges worldwide. *Annals of translational medicine.* 2016;4(8).
2. Torre LA, et al. *Lung cancer and personalized medicine*. Springer; 2016:1-19.
3. Kramer BS, et al. SAGE Publications Sage UK: London, England; 2011.
4. McWilliams A, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *New England Journal of Medicine.* 2013;369(10):910-9.
5. Nair VS, et al. Accuracy of models to identify lung nodule cancer risk in the National Lung Screening Trial. *American journal of respiratory and critical care medicine.* 2018;197(9):1220-3.
6. Pinsky PF, et al. Performance of Lung-RADS in the National Lung Screening Trial: a retrospective assessment. *Annals of internal medicine.* 2015;162(7):485-91.
7. Silvestri GA, et al. Assessment of plasma proteomics biomarker's ability to distinguish benign from malignant lung nodules: results of the PANOPTIC (Pulmonary Nodule plasma proteomic classifier) trial. *Chest.* 2018;154(3):491-500.
8. Postel M, et al. Droplet-based digital PCR and next generation sequencing for monitoring circulating tumor DNA: a cancer diagnostic perspective. *Expert review of molecular diagnostics.* 2018;18(1):7-17.
9. 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.* 2019;9(7):2056.
10. MacMahon H, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017. *Radiology.* 2017;284(1):228-43.
11. Castro-Giner F, et al. Cancer diagnosis using a liquid biopsy: challenges and expectations. *Diagnostics.* 2018;8(2):31.
12. Team NLSTR. Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine.* 2011;365(5):395-409.
13. Hassanein M, et al. The state of molecular biomarkers for the early detection of lung cancer. *Cancer prevention research.* 2012;5(8):992-1006.
14. Groheux D, et al. FDG PET-CT for solitary pulmonary nodule and lung cancer: literature review. *Diagnostic and interventional imaging.* 2016;97(10):1003-17.
15. Lennon AM, et al. Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention. *Science.* 2020.

16. Jaine R, et al. Cost-effectiveness of a low-dose computed tomography screening programme for lung cancer in New Zealand. *Lung Cancer.* 2018;124:233-40.
17. Gould MK, et al. Recent trends in the identification of incidental pulmonary nodules. 2015;192(10):1208-14.
18. Yaguang F, et al. China national guideline of classification, diagnosis and treatment for lung nodules (2016 version). *Zhongguo Fei Ai Za Zhi.* 2016;19(12).
19. Peikert T, et al. Radiomics-based Management of Indeterminate Lung Nodules? Are We There Yet? *American Journal of Respiratory and Critical Care Medicine.* 2020(ja).
20. Garcia-Velloso MJ, et al. Assessment of indeterminate pulmonary nodules detected in lung cancer screening: Diagnostic accuracy of FDG PET/CT. *Lung Cancer.* 2016;97:81-6.
21. Schuhmann M, et al. Endobronchial ultrasound for peripheral lesions: a review. *Endoscopic ultrasound.* 2013;2(1):3.
22. Chockalingam A, and Hong K. Transthoracic needle aspiration: the past, present and future. *Journal of thoracic disease.* 2015;7(Suppl 4):S292.
23. Choi HK, et al. Models to estimate the probability of malignancy in patients with pulmonary nodules. 2018;15(10):1117-26.
24. Li X-j, et al. A blood-based proteomic classifier for the molecular characterization of pulmonary nodules. *Science translational medicine.* 2013;5(207):207ra142-207ra142.
25. Trivedi NN, et al. Analytical validation of a novel multi-analyte plasma test for lung nodule characterization. *Biomed Res Rev.* 2018;2:1-10.
26. Ren S, et al. Early detection of lung cancer by using an autoantibody panel in Chinese population. *Oncoimmunology.* 2018;7(2):e1384108.
27. Chen C, et al. Ultrasensitive DNA hypermethylation detection using plasma for early detection of NSCLC: a study in Chinese patients with very small nodules. 2020;12(1):1-11.
28. Mazzone PJ, et al. Evaluating molecular biomarkers for the early detection of lung cancer: when is a biomarker ready for clinical use? An official American Thoracic Society policy statement. *American journal of respiratory and critical care medicine.* 2017;196(7):e15.
29. Liang W, et al. Evaluating the diagnostic accuracy of a ctDNA methylation classifier for incidental lung nodules: protocol for a prospective, observational, and multicenter clinical trial of 10,560 cases. 2020;9(5):2016.
30. Feng H, et al. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. 2014;42(8):e69-e.
31. Barrett JE, et al. Quantification of tumour evolution and heterogeneity via Bayesian epiallele detection. *BMC bioinformatics.* 2017;18(1):1-10.
32. De R, et al. Learning internal representations by error propagation. *Parallel distributed processing: Explorations in the microstructure of cognition, 1: Foundations.* 1986:318-62.
33. He K, et al. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016:770-8.
34. Zeiler MD, et al. *2010 IEEE Computer Society Conference on computer vision and pattern recognition*. IEEE; 2010:2528-35.
35. Kingma DP, and Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:14126980.* 2014.
36. Swensen SJ, et al. The probability of malignancy in solitary pulmonary nodules: application to small radiologically indeterminate nodules. *Arch Intern Med.* 1997;157(8):849-55.

37. Gould MK, et al. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. *Chest.* 2007;131(2):383-8.
38. Zhou Y, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun.* 2019;10(1):1-10.

Figure Legends

Figure 1 Study flow of participants in the study. Total 585 enrolled; 30 excluded due to limited cfDNA extracted (<5ng) and 26 excluded due to failing sequencing QC; The model was developed, tested on 389 samples and validated independently on 140 samples; The model was further validated on 100 indeterminate nodules (6-20 mm sizes) in the validation set.

Figure 2 PulmoSeek performance compared to Mayo Clinic/VA model in all nodule sizes. A representative receiver operating curve (ROC) displays the classification performance of PulmoSeek (A) In the test set, the area under the curve (AUC) was 0.83 [0.72-0.94]; In the validation set, the AUC was 0.84 [0.77–0.92] (B) In the validation set, the AUC of Mayo Clinic classifier was 0.59 [0.48-0.69] and the AUC of Veteran's Affairs(VA) classifier was 0.54 [0.44-0.64]. (C) Confusion matrices for PulmoSeek comparing the true class with the predicted class for benign (n=20) and malignant (n=60) nodule samples, and distribution of PulmoSeek scores (range, 0 to 1) in the test set. (D) Confusion matrices for

PulmoSeek comparing the true class with the predicted class for benign (n=40) and malignant (n=100) nodule samples, and distribution of PulmoSeek scores (range, 0 to 1) in the validation set.

Figure 3 PulmoSeek performance in early-stage lung cancer. In the independent validation set (A) PulmoSeek performance in early-stage cancer: sensitivity is 100% in Stage 0 (n=2), 94.1% in Stage IA (n=85) and 100% in Stage 1B (n=5) (B) PulmoSeek performance in Stage IA sub-stages: sensitivity is 86.4% in Stage IA1 (n=22), 95.0% in Stage IA2 (n=40) and 100% in Stage 1A3 (n=23).

Figure 4 PulmoSeek performance compared to Mayo Clinic/VA model in 6-20mm nodule sizes. A representative receiver operating curve (ROC) displays the classification performance of PulmoSeek (A) In the test set, the AUC was 0.76 [0.61–0.91]; In the validation set, the AUC was 0.84 [0.76-0.93] (B) In the validation set, the AUC of Mayo Clinic classifier was 0.60 [0.48-0.72] and the AUC of VA classifier was 0.51 [0.40-0.63] (C) Confusion matrices for PulmoSeek comparing the true class with the predicted class for benign (n=14) and malignant (n=43) nodule samples, and distribution of PulmoSeek scores (range, 0 to 1) in the test set. (D) Confusion matrices for PulmoSeek comparing the true class with the predicted class

721 for benign (n=30) and malignant (n=73) nodule samples, and distribution of
722 PulmoSeek scores (range, 0 to 1) in the validation set.
723
724 **Figure 5 PulmoSeek performance in different nodule types and comparison to**
725 **PET-CT.** In the independent validation set samples with PET-CT records, the
726 diagnosis result for each patient using PulmoSeek (box shape) and PET-CT
727 (diamond shape) was shown. The green color indicated the sample was diagnosed
728 correctly and the red color incorrectly. The PulmoSeek correctly identified 8 out 10
729 patients in the solid nodule subgroup, 9 out of 11 in the part-solid nodule subgroup
730 and 5 out of 5 in the ground-glass nodule subgroup. The PET-CT correctly identified
731 6 out 10 patients in the solid nodule subgroup, 7 out of 11 in the part-solid nodule
732 subgroup and 0 out of 5 in the ground-glass nodule subgroup.
733

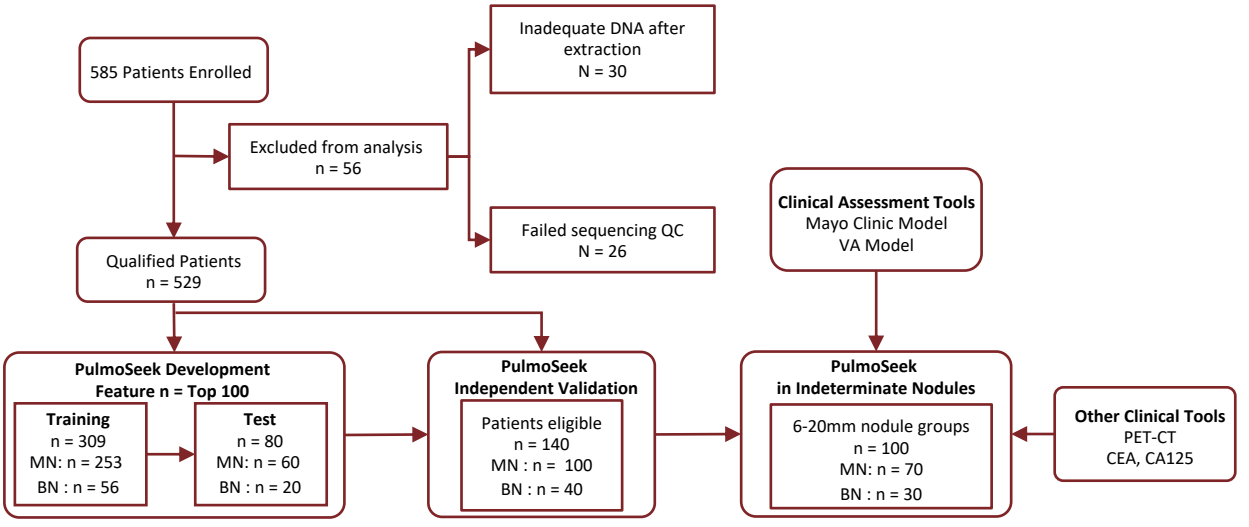


Figure 1 Study flow of participants in the study. Total 585 enrolled; 30 excluded due to limited cfDNA extracted (<5ng) and 26 excluded due to failing sequencing QC; The model was developed, tested on 389 samples and validated independently on 140 samples; The model was further validated on 100 indeterminate nodules (6-20 mm sizes) in the validation set.

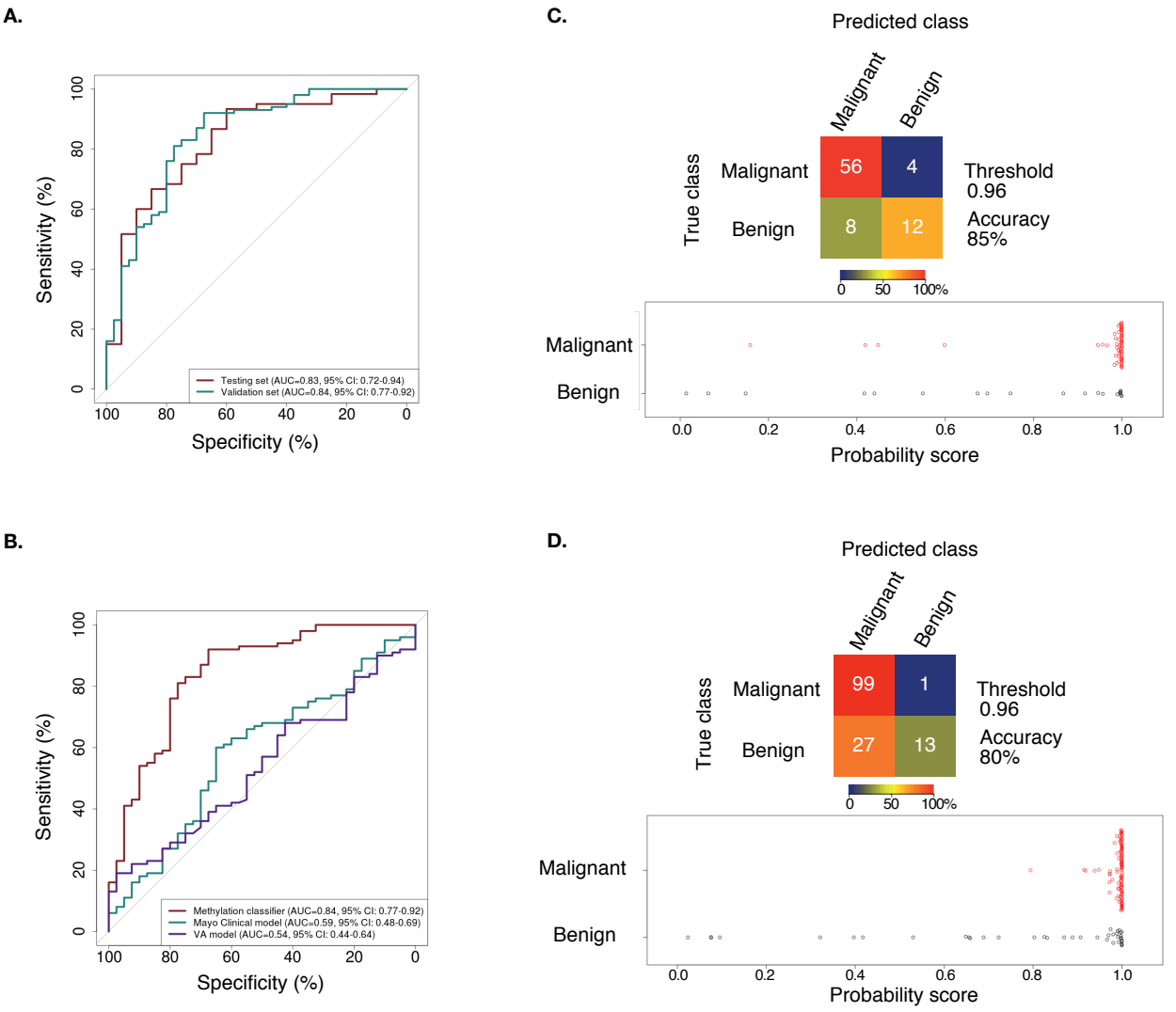
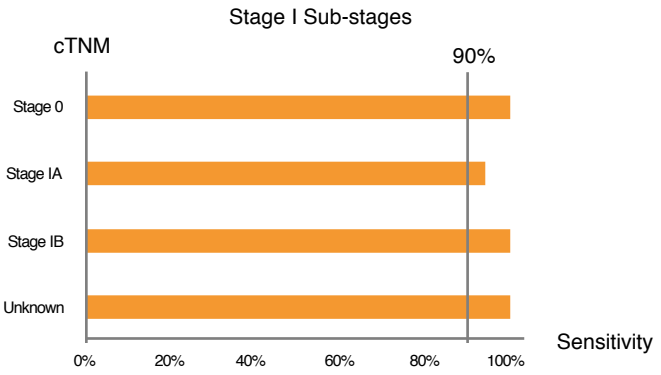


Figure 2 PulmoSeek performance compared to Mayo clinic/VA model in all nodule sizes. A representative receiver operating curve (ROC) displays the classification performance of PulmoSeek (**A**) In the test set, the area under the curve (AUC) was 0.83 [0.72-0.94]; In the validation set, the AUC was 0.84 [0.77–0.92] (**B**) In the validation set, the AUC of Mayo Clinic classifier was 0.59 [0.48-0.69] and the AUC of Veteran's Affairs(VA) classifier was 0.54 [0.44-0.64]. (**C**) Confusion matrices for PulmoSeek comparing the true class with the predicted class for benign (n=20) and malignant (n=60) nodule samples, and distribution of PulmoSeek scores (range, 0 to 1) in the test set. (**D**) Confusion matrices for PulmoSeek comparing the true class with the predicted class for benign (n=40) and malignant (n=100) nodule samples, and distribution of PulmoSeek scores (range, 0 to 1) in the validation set.

A.



B.

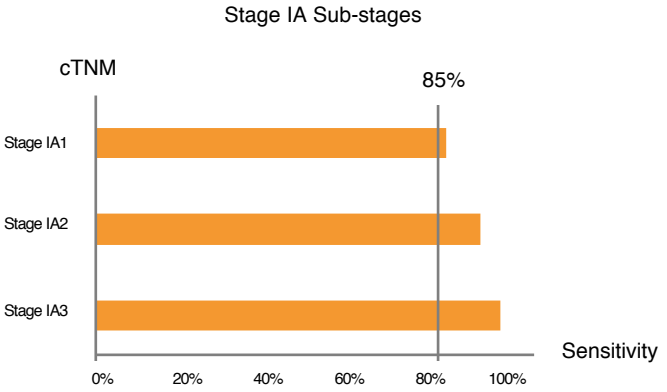


Figure 3 PulmoSeek performance in early-stage lung cancer. In the independent validation set (**A**) PulmoSeek performance in early-stage cancer: sensitivity is 100% in Stage 0 (n=2), 94.1% in Stage IA (n=85) and 100% in Stage 1B (n=5) (**B**) PulmoSeek performance in Stage IA sub-stages: sensitivity is 86.4% in Stage IA1 (n=22), 95.0% in Stage IA2 (n=40) and 100% in Stage 1A3 (n=23).

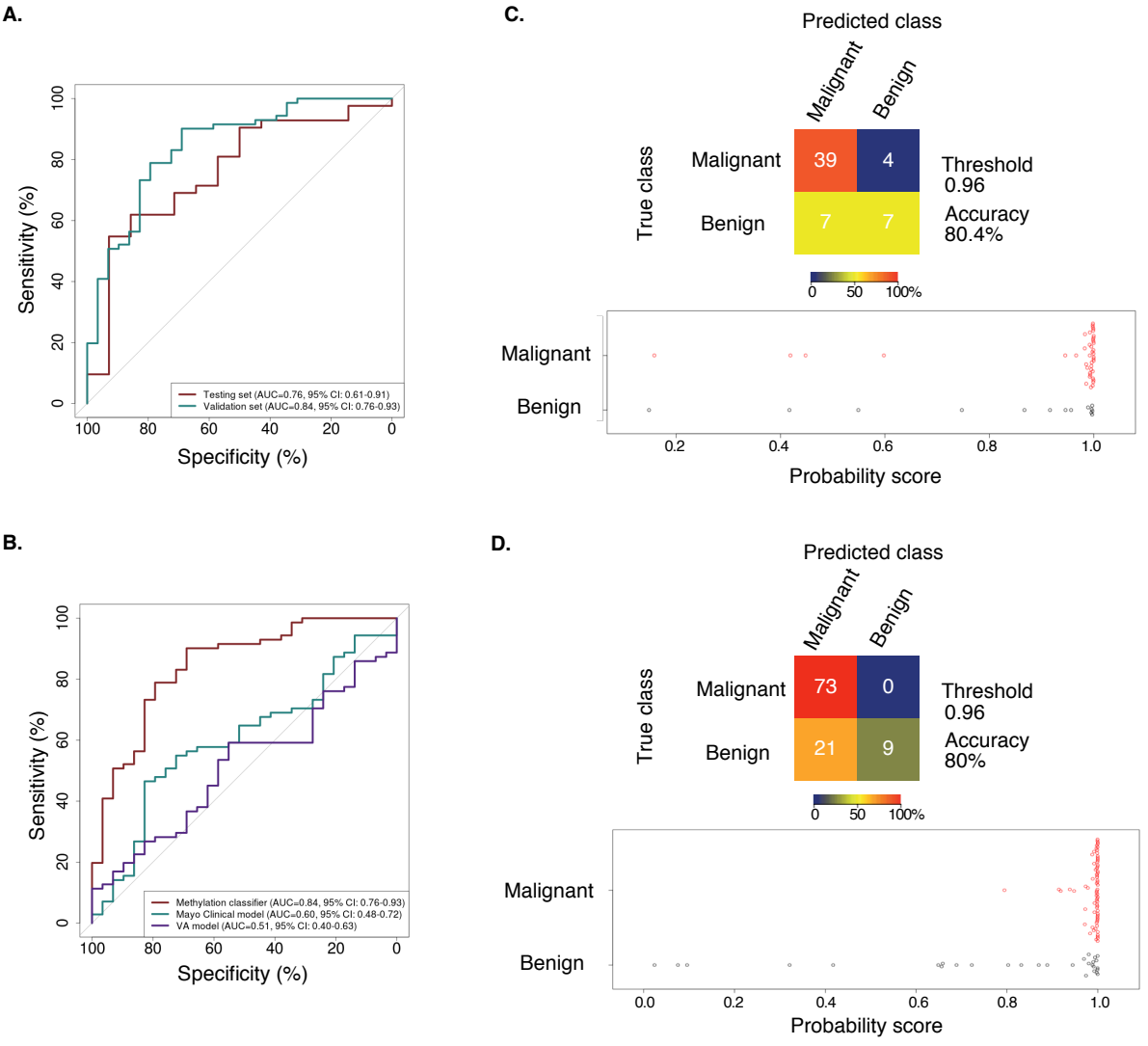


Figure 4 PulmoSeek performance compared to Mayo clinic/VA model in 6-20mm nodule sizes. A representative receiver operating curve (ROC) displays the classification performance of PulmoSeek (**A**) In the test set, the AUC was 0.76 [0.61–0.91]; In the validation set, the AUC was 0.84 [0.76-0.93] (**B**) In the validation set, the AUC of Mayo Clinic classifier was 0.60 [0.48-0.72] and the AUC of VA classifier was 0.51 [0.40-0.63] (**C**) Confusion matrices for PulmoSeek comparing the true class with the predicted class for benign (n=14) and malignant (n=43) nodule samples, and distribution of PulmoSeek scores (range, 0 to 1) in the test set. (**D**) Confusion matrices for PulmoSeek comparing the true class with the predicted class for benign (n=30) and malignant (n=73) nodule samples, and distribution of PulmoSeek scores (range, 0 to 1) in the validation set.

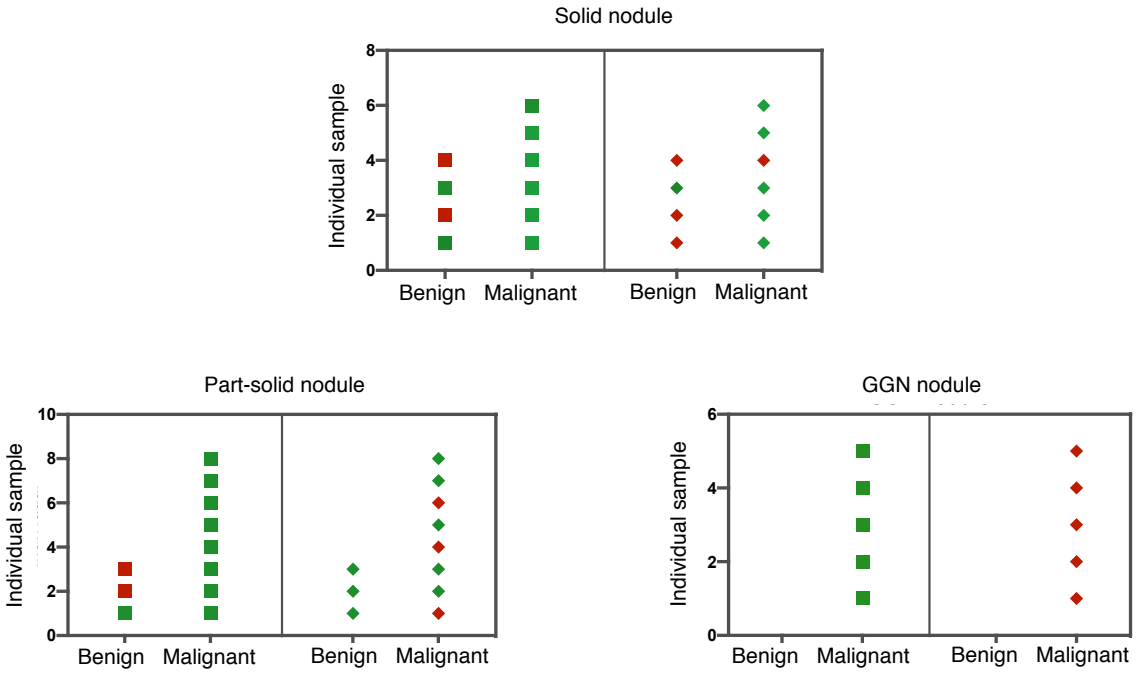


Figure 5 PulmoSeek performance in different nodule types and comparison to PET-CT. In the independent validation set samples with PET-CT records, the diagnosis result for each patient using PulmoSeek (box shape) and PET-CT (diamond shape) was shown. The green color indicated the sample was diagnosed correctly and the red color incorrectly. The PulmoSeek correctly identified 8 out 10 patients in the solid nodule subgroup, 9 out of 11 in the part-solid nodule subgroup and 5 out of 5 in the ground-glass nodule subgroup. The PET-CT correctly identified 6 out 10 patients in the solid nodule subgroup, 7 out of 11 in the part-solid nodule subgroup and 0 out of 5 in the ground-glass nodule subgroup.

Table 1 Demographic and clinical characteristics of study participants

Characteristics	
Subjects	
Age - yr	55.8 (23 - 85)
Gender – no.	
Male	249
Female	280
Smoking history Status – no.	
Never	387
Former	67
Current	75
No data	0
Pack-year	7.69 (0–200)
Nodules	
Size – no.	
< 5 mm	0
5 – 20 mm	365
20 – 30 mm	157
>30 mm	0
No data	7
Location – no.	
Upper lobe	322
Lower lobe	207
Histopathology	
Benign diagnosis – no.	
Granuloma	56
Hamartoma	12
Other	48
Cancer diagnosis – no.	
Small-cell	1
Non-small cell	412
Adenocarcinoma	381
Squamous cell	17
Large cell	0
Others	14
Unknown	0

Table 2 PulmoSeek performance metrics

	All Nodule Sizes			
		PulmoSeek		Mayo Clinical Model
	Test Set (N=80)	Validation Set (N=140)	Test and Validation Set (N=220)	Validation Set (N=140)
Accuracy (95%CI)	0.850 (0.625-0.912)	0.800 (0.657-0.871)	0.818 (0.664-0.868)	0.357 (0.307-0.407)
Sensitivity (95%CI)	0.933 (0.533-0.983)	0.990 (0.610-1.000)	0.969 (0.594-0.994)	0.130 (0.070-0.200)
Specificity (95%CI)	0.600 (0.500-1.000)	0.325 (0.200-0.875)	0.417 (0.350-0.900)	0.925 (0.850-1.000)
Positive predictive value (95%CI)*	0.875 (0.852-1.000)	0.786 (0.758-0.938)	0.816 (0.799-0.947)	0.812 (0.600-1.000)
Negative predictive value (95%CI)*	0.750 (0.396-0.929)	0.929 (0.444-1.000)	0.833 (0.437-0.944)	0.298 (0.273-0.322)
Positive predictive value (95%CI)&	0.206 (0.176-1.000)	0.140 (0.122-0.403)	0.156 (0.142-0.425)	0.161 (0.063-1.000)
Negative predictive value (95%CI)&	0.988 (0.946-0.997)	0.997 (0.947-1.000)	0.992 (0.949-0.998)	0.905 (0.894-0.915)
	6-20mm Nodule Sizes			
		PulmoSeek		Mayo Clinical Model
	Test Set (N=56)	Validation Set (N=100)	Test and Validation Set (N=156)	Validation Set (N=100)
Accuracy (95%CI)	0.804 (0.536-0.875)	0.800 (0.640-0.870)	0.801 (0.628-0.840)	0.300 (0.290-0.320)
Sensitivity (95%CI)	0.905 (0.429-0.976)	1.000 (0.577-1.000)	0.965 (0.540-0.991)	0.014 (0.000-0.042)
Specificity (95%CI)	0.500 (0.286-1.000)	0.300 (0.172-0.931)	0.372 (0.256-0.930)	1.000 (1.000-1.000)
Positive predictive value (95%CI)*	0.844 (0.783-1.000)	0.780 (0.747-0.954)	0.801 (0.773-0.955)	1.000 (1.000-1.000)
Negative predictive value (95%CI)*	0.636 (0.318-0.857)	1.000 (0.436-1.000)	0.800 (0.411-0.923)	0.293 (0.290-0.299)
Positive predictive value (95%CI)&	0.167 (0.118-1.000)	0.139 (0.118-0.484)	0.146 (0.126-0.474)	1.000 (1.000-1.000)
Negative predictive value (95%CI)&	0.979 (0.926-0.994)	1.000 (0.945-1.000)	0.990 (0.943-0.996)	0.901 (0.900-0.904)

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