Accurate diagnosis of high-risk pulmonary nodules using a non-invasive epigenetic biomarker test

Running head: Epigenetic Test for High-Risk Lung Nodule

Pei-Hsing Chen[1,2], M.D.; Tung-Ming Tsai M.D[3]; Tzu-Pin Lu, Ph.D[2,4]; Hsiao-Hung Lu[2], M.D.; Dorian F. Pamart MSc[5]; Aristotelis Kotronoulas, Ph.D.[5]; Marielle Herzog, Ph.D.[5]; Jacob V. Micallef, Ph.D.[6]; Hsao-Hsun Hsu[2,3], M.D., Ph.D.; Jin-Shing Chen[2], M.D., Ph.D.

[1]Institute of Biomedical Engineering, College of Medicine and College of Engineering, National Taiwan University, Taipei City, Taiwan

[2]Department of Surgery, National Taiwan University Hospital and National Taiwan University College of Medicine, No. 1, Sec. 1, Ren'ai Rd., Zhongzheng Dist., Taipei City 100, Taiwan

[3]Department of Surgical Oncology, National Taiwan University Cancer Center and National Taiwan University College of Medicine No. 57, Ln. 155, Sec. 3, Keelung Rd., Da'an Dist., Taipei City 106, Taiwan

[4]Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, No. 17, Xuzhou Rd., Zhongzheng Dist., Taipei City 100, Taiwan

[5]Belgian Volition SRL, 22 Rue Phocas Lejeune, Parc Scientifique Crealys, 5032 Isnes, Belgium

[6]Volition Diagnostics UK Limited, 93-95 Gloucester Place, London, W1U 6JQ, UK

Author to whom correspondence and proofs should be sent:

Jin-Shing Chen, M.D., Ph.D.

Department Chief, Department of Surgery, National Taiwan University Hospital and National Taiwan University College of Medicine, No. 1, Sec. 1, Ren'ai Rd., Zhongzheng Dist., Taipei City 100, Taiwan

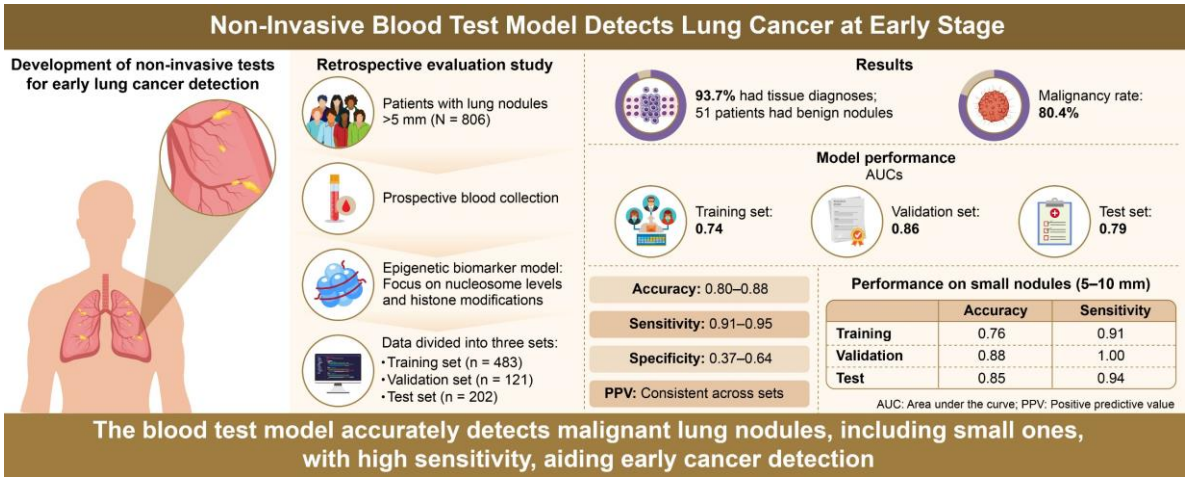
Tel.: +886-2-2322-0322; Fax: +886-2-3322-9601; E-mail: chenjs@ntu.edu.tw

Conflicts of interest: Dorian Pamart, Aristotelis Kotronoulas, and Marielle Herzog are employees of Belgian Volition, Jacob Micallef is an employee of Volition Diagnostics UK. Aristotelis Kotronoulas, Marielle Herzog and Jacob Micallef are shareholders of VolitionRX Limited.

Presentations: ESMO 2023-minioral-133MO Differentiation of malignant and benign lung nodules using epigenetically modified nucleosomes in plasma. *Annals of Oncology* 34 (2023): S234.

Synopsis

An epigenetic biomarker (EB) model demonstrated high sensitivity and accuracy for early lung cancer detection in 806 patients with pulmonary nodules. It performed well across imaging types and stages, including small nodules, reducing false negatives in minimally invasive surgery.



ABSTRACT

Background: Accurate non-invasive tests to improve the early detection and diagnosis of lung cancer are urgently needed, given that large tumors may metastasize or be resistant to treatment. However, no regulatory-approved blood tests are available for the early detection of lung cancer. We aimed to improve the classification of pulmonary nodules to identify malignant ones in a high-prevalence patient group.

Methods: A study involving 806 participants with undiagnosed nodules larger than 5 mm, identified via CT imaging, focused on assessing nucleosome levels and histone modifications in circulating blood. Nodules were classified as malignant or benign. A logistic regression analysis was performed. For model development, the data were randomly divided into training (n = 483) and validation (n = 121) datasets. The model's performance was then evaluated using a separate testing dataset (n = 202).

Results: Among patients, 755 (93.7%) had a tissue diagnosis. The overall malignancy rate in the cohort was 80.4%. For all datasets, the AUCs were as follows: training, 0.74; validation, 0.86; and test, 0.79 (accuracy range: 0.80–0.88). Sensitivity showed consistent results across all datasets (0.91, 0.95, and 0.93, respectively), whereas specificity ranged from 0.37 to 0.64. For smaller nodules (5–10 mm), the model recorded accuracy values of 0.76, 0.88, and 0.85. Sensitivity values of 0.91, 1.00, and 0.94 further highlight the robust diagnostic capability of the model. The performance of the model across the RADS categories was evaluated, and it demonstrated consistent accuracy.

Conclusion: Our EB panel detected non-small cell lung cancer early in a high-risk patient group with high sensitivity and accuracy. The EB model was particularly effective in identifying high-risk lung nodules, including small, part-solid, and non-solid nodules, and provided further evidence for external validation.

Keywords:

Lung cancer, epigenetic biomarkers, histone PTMs, circulating nucleosomes, pulmonary nodule, liquid biopsy

INTRODUCTION

Lung cancer remains the leading cause of cancer-related mortality worldwide.[1] Early detection using low-dose computed tomography (LDCT) reduces death rates, as evidenced in pivotal trials such as the National Lung Screening Trial (NLST)[2] and the Dutch-Belgian Lung Cancer Screening Trial(NELSON).[3] NLST reported a 20% reduction in lung cancer mortality with the implementation of LDCT screening.[2] Despite these outcomes, indeterminate pulmonary nodules, which comprise 50–76% of the nodules identified in the LDCT, present a significant challenge.[4, 5] The likelihood of malignancy increases with nodule size, with those measuring 7–29 mm exhibiting a malignancy risk ranging from 1.7% to 22%.[6] Lung nodules detected using LDCT screening are often <20 mm, complicating the biopsy process.[6-8] Consequently, while close monitoring is the primary strategy, larger tumors may develop resistance or metastasize during observation.

Efforts have been made to develop robust, sensitive, and non-invasive tests to diagnose pulmonary nodules.[9] Despite advancements, there is currently no regulatory body-approved and widely adopted blood test for early detection of lung cancer.[4] Tumor cells release various biomolecules such as cell-free DNA (cfDNA), circulating tumor DNA (ctDNA), exosomes, micro-RNA, circular RNA, circulating tumor cells (CTCs), and DNA-methylated fragments. However, while these biomarkers are effective diagnostic biomarkers, some experts remain skeptical of liquid biopsies. The detection of molecular changes in evolving tumor cells requires highly sensitive and specific assays for ctDNA mutations.[10, 11] The diagnostic sensitivity of liquid biopsy tests is hampered by very low levels of somatic molecular alterations in patients with early-stage cancer, who constitute the majority of the population after LDCT screening.[12, 13]

Changes in DNA methylation in specific regions, such as promoter CpG islands, may signify early molecular events in tumor initiation.[14] Patients with cancer have distinct

histone post-translational modifications (PTMs) in circulating nucleosomes, indicating their potential as cancer biomarkers.[15, 16]

We aimed to develop an epigenetic biomarker (EB) model based on circulating nucleosomes, including histone variant and histone methylation, to evaluate the risk of malignancy in pulmonary nodules and to achieve a more accurate classification of pulmonary nodules, particularly through focusing on identifying malignant nodules in thoracic surgery scenarios.

METHODS

Study design and patients

This prospective blood specimen collection and retrospective evaluation study was approved by the Institutional Review Boards (201905009RIFC) of the participating hospitals and researchregistry-10711. In guidelines for Asia, it is recommended that nodules with a diameter of ≥5 mm undergo clinical management.[17] We recruited 806 participants with undiagnosed nodules larger than 5 mm, identified on computed tomography (CT) scans and classified as high risk by the attending physician. Adult patients of either sex aged ≥18 years were eligible for inclusion if they met the following criteria: pulmonary nodules >5 mm detected using standard- or low-dose CT (LDCT) screening, and with nodules categorized as solid nodules, part-solid nodules (mixed ground-glass nodules), or pure non-solid nodules. Participants were recruited from the outpatient clinics of the National Taiwan University Hospital and the National Taiwan University Cancer Center, both teaching hospitals. The study was conducted from August 2019 to July 2021. Exclusion criteria included patients exhibiting metastatic symptoms such as pleural effusion, patients unwilling to undergo blood sampling, patients without a confirmed pathological diagnosis post-surgery, or patients with cancer confirmed pathologically within two years prior to enrollment. Written informed consent was obtained from the patient for blood sampling.

Blood sampling

All chest blood samples were prospectively collected before surgery, either during the initial nodule check or during the admission period. Blood (10 mL) was collected in K2-EDTA blood tubes (Sarstedt, Nümbrecht, Germany) within two weeks prior to the initiation of surgery. Blood collection and CT-based response evaluations were conducted for patients undergoing observation at 1–2 weeks intervals. Blood samples were centrifuged at 3000 g for 10 min at 15-30°C. The plasma was then stored at -80 °C until the nucleosome analysis was conducted.

Quantification of circulating nucleosomes using immunoassays

All samples were tested using Nu.Q® assays (Belgian Volition SRL, Isnes, Belgium). two nucleosome structures were measured using Nu.Q® H3.1 and Nu.Q® H3K27Me3 immunoassays, according to the manufacturer's instructions. These sandwich immunoassays, based on chemiluminescence technology, were performed using the IDS-i10 automated analyzer system (Immunodiagnostic System Ltd., Boldon, UK). Briefly, 50 µL of K2-EDTA plasma was incubated with acridinium ester labeled anti-nucleosome detection antibody. Magnetic particle beads coated with the corresponding monoclonal anti-histone variant H3.1, or anti-histone modification H3K27Me3 capture antibody, were added. After washing, trigger solutions were added, and the light emitted by the acridinium ester was measured using a luminometer. The results were expressed in relative light units, and concentration (expressed in ng/mL) was extrapolated using four-parameter logistic regression of a reference standard curve. All samples were analyzed in duplicate. If the sample concentration was higher than the lowest concentration and the %CV of the determined concentration was >20%, the analysis was repeated.

Chest CT imaging and radiological analysis

For each participant who underwent a chest CT scan, the lung and mediastinal image series were reconstructed with a slice thickness of 1.000–1.250 mm if the scans were performed at the National Taiwan University Hospital and National Taiwan University Cancer Center. The slice thickness was at least 5.000 mm. The chest CT images were initially evaluated before drawing blood to confirm inclusion. Diagnosis of the nodules was based on the pathological outcomes. If a nodule disappeared on subsequent imaging and a radiologist confirmed this, it was classified as benign.

Imaging analysis was overseen by Jin-Shing Chen, a senior thoracic surgeon with extensive experience exceeding three decades. Additional team members, Pei-Hsing Chen, and Tung-Ming Tsai, with 10 and 17 years of experience, respectively, interpreted the scans and delineated the regions of interest. For subgroup analysis, Lung-imaging, reporting, and data system (Lung-RADS) version 1.1 guidelines were used.[18] In the external validation of the model, Hsiao-Hung Lu performed a blinded assessment, categorizing the nodules as malignant or benign based on their spiculation characteristics.

Tumor size was determined preoperatively based on thin-section CT findings. All tumors were subsequently evaluated to estimate the extent of ground-glass opacity (GGO) using a thin-section CT scan with a 5.000 mm collimation. The solid component, part-solid component, and GGO were defined as areas of increased opacification that completely obscured the underlying vascular markings, as described in previous studies.[19, 20] GGO was defined as an area of slight, homogeneous increase in density that did not obscure the underlying vascular markings.

Operation policy and pathology evaluation

The operation policy followed the American College of Chest Physicians or American Association for Thoracic Surgery recommendations[21] to evaluate and treat nodules >5 mm.[17] Malignant and benign tumors were defined based on the 2021 WHO Classification of Lung Tumors.[22] The subcategories of adenocarcinoma included in the malignant group were atypical adenomatous hyperplasia (AAH), adenocarcinoma in situ, minimally invasive adenocarcinoma, and invasive adenocarcinoma.[23]

Theory/calculation

EB model development for benign-malignant predictions

To advance the prediction of benign and malignant states in lung cancer, we developed an EB model specifically focusing on nucleosome levels and histone modifications in circulating blood. For model building, we employed a logistic regression approach to predict benign and malignant states. For model development, we randomly allocated 25% of the data to the test dataset (n = 202). The remaining 75% was further divided, with 80% (n = 483) used for training and 20% (n = 121) for validation. (Figure 1.)

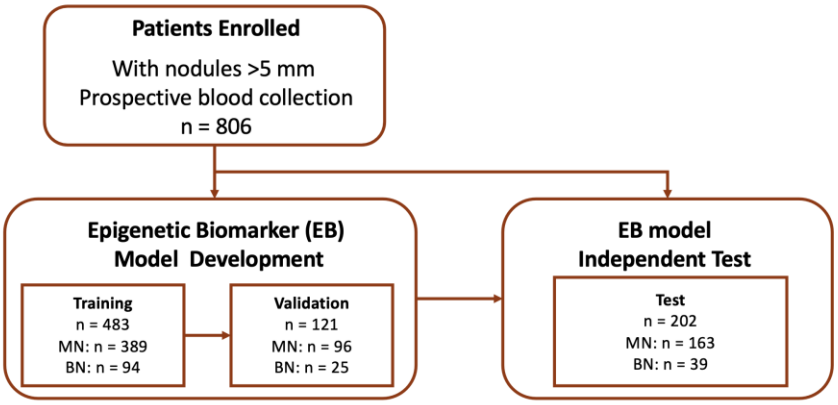


Figure 1. **Flowchart of participant enrollment and model development.** A total of 806 patients with nodules >5 mm were enrolled. The Epigenetic Biomarker (EB) model was trained on 483 samples, validated on 121 samples, and independently tested on 202 samples. The model was built based on the training dataset, with the cutoff determined using the validation dataset, and its performance evaluated on the test dataset.

The training dataset was used for model development and the calculation of its coefficients. Key biomarkers included histone isoform nucleosome levels (Nu.Q® H3.1) and methylated lysine 27 of histone H3 (Nu.Q H3K27Me3). A validation dataset was used to fine-tune the model and ensure its predictive capability.

Given the benefits of video-assisted thoracoscopic surgery, such as reduced invasiveness, accurate localization, and quick recovery, [21,24, 25] the main challenge for thoracic surgeons is to accurately identify malignant lung nodules and minimize the risk of delayed diagnosis in high-risk populations identified through CT/LDCT screening. This necessitates decreasing the false-negative rate while maintaining an adequate PPV. Achieving this requires a sensitivity of >80% while maintaining an adequately high PPV to confidently identify lung cancer in high-prevalence populations. To this end, the probability cutoff for cancer diagnosis was determined by maintaining the model's sensitivity at >0.80 and then selecting the optimal threshold using the

Youden index. Finally, this cutoff was applied to the test dataset, and the diagnostic performance was calculated.

Comparison with the Mayo Clinic and Veteran Affairs (VA) models

The Mayo Clinic model for malignancy in pulmonary nodules calculates the probability of malignancy using three clinical and three radiographic variables. The formula is as follows:

$$\text{probability of malignancy} = e^x/(1 + e^x),$$

where $x = -6.8272 + (0.0391 \times \text{age}) + (0.7917 \times \text{smoking}) + (1.3388 \times \text{cancer}) + (0.1274 \times \text{nodule diameter}) + (1.0407 \times \text{spiculation}) + (0.7838 \times \text{upper lobe})$, e is Euler's number, a mathematical constant approximately equal to 2.71828.[26]

The VA model for malignancy in pulmonary nodules calculates the probability of malignancy using three clinical variables and one radiographic variable. The formula is:

$$\text{probability of malignancy} = 100 \times (e^x/[1 + e^x]), \text{ where } x = -8.404 + 2.061 \times \text{smoke} + 0.779 \times \text{age}/10 + 0.112 \times \text{diameter} + 0.567 \times \text{yearsquit}/10,$$

where smoking is 1 if the patient is a current or former smoker (otherwise 0); age divided by 10 is the age in years divided by 10; diameter is the largest diameter of the nodule in millimeters; yearsquit/10 is the number of years since quitting smoking divided by 10; and e is the Euler's number.[27]

Statistics

All statistical analyses were conducted using R software (version 4.4.1). Statistical analyses were conducted as outlined in each figure legend with sample sizes provided accordingly. Categorical variables such as sex and nodule sub-type were compared using Fisher's exact test. The sensitivities of the different Lung-RADS for malignant nodules were compared using Fisher's exact test. Continuous variables such as age were compared using Student's *t*-test, and 95% confidence intervals (CIs) were calculated based on Wald confidence intervals for proportions.

The sensitivity, specificity, accuracy, PPV, and negative PV (NPV) of the model and other models for differentiating malignant nodules were assessed by comparing the pathological outcomes and imaging studies (for those with vanished nodules only). Receiver operating characteristic (ROC) and area under the curve (AUC) were calculated using pROC R package (version 1.15.3) software.

RESULTS

Patient demographics and clinicopathologic features

In total, 806 patients who were CT-positive/LDCT-positive were recruited from the thoracic surgery departments of the National Taiwan University Hospital and the National Taiwan University Cancer Center. Of these, 755 (93.7%) were tissue-diagnosed. The remaining 51 (6.3%) patients had nodules that disappeared on subsequent imaging, confirmed by a radiologist, and defined as benign. The malignancy rate in the entire cohort was 80.4% (158 benign, 648 malignant). An overview of the study design is provided in Figure 1, and the demographic characteristics of the 806 patients are detailed in Table 1. The diagnoses in the cohort were predominantly early-stage lung cancer (AAH, stage 0, stage I, or stage II), comprising 84.3% (546/648) of

all cancer patients in the cohort. The mean nodule size in the entire cohort was 25 mm. Additionally, 78.6% (630/806) of the patients had never smoked. Adenocarcinoma and its subcategories constituted the majority of malignant cases (92.6%, 600/648). No statistically significant differences in the distribution of malignancy, age, sex, tumor components, tumor size, Lung-RADS scores, or smoking history ($p > 0{\cdot}05$) were observed among the three datasets. The demographic and clinical characteristics of the participants are presented in Table 1.

The EB model and lung cancer diagnostic accuracy

The cohort data were divided into an independent test dataset (n = 202, 25%) and the EB model development subset (75%). The training dataset (n = 483) comprised 80% of the EB model development subset and was used for the model development. The validation dataset (n = 121) comprised 20% of the model development subset and was used for the optimal threshold selection.

We developed the EB model by screening multiple combination models from five quantitative epigenetic features (Supplementary Table 1) derived from blood tests during the pre-training model tuning. For feature selection, we analyzed the relationship between the AUC values and the number of primary features. The AUC plateaued when the number of primary features reached two, indicating that adding more features did not significantly improve the AUC. Therefore, we identified two independent predictors of malignancy using multivariate logistic regression analysis. This was then applied to the training dataset using Nu.Q® H3.1 and Nu.Q® H3K27Me3. Other potential predictors not associated with malignancy were excluded from the final model. The prediction model was calculated as follows:

$$\text{Probability of malignant SPN} = e^x / 1 + e^x$$

$$X = 1.78668 + 0.07821 \times H3.1 - 0.15885 \times H3K27Me3,$$

where H3.1 is the level of the histone variant H3.1, and H3K27Me3 is the histone modification H3K27Me3. Analysis of the relationship between the AUC values, detailed in Figure 2A and 2B, provided AUC values for all the datasets. Positive and negative classifications for the model were determined using a cutoff value (0.755). We validated the performance of the EB model using an independent test set that showed consistently good performance.

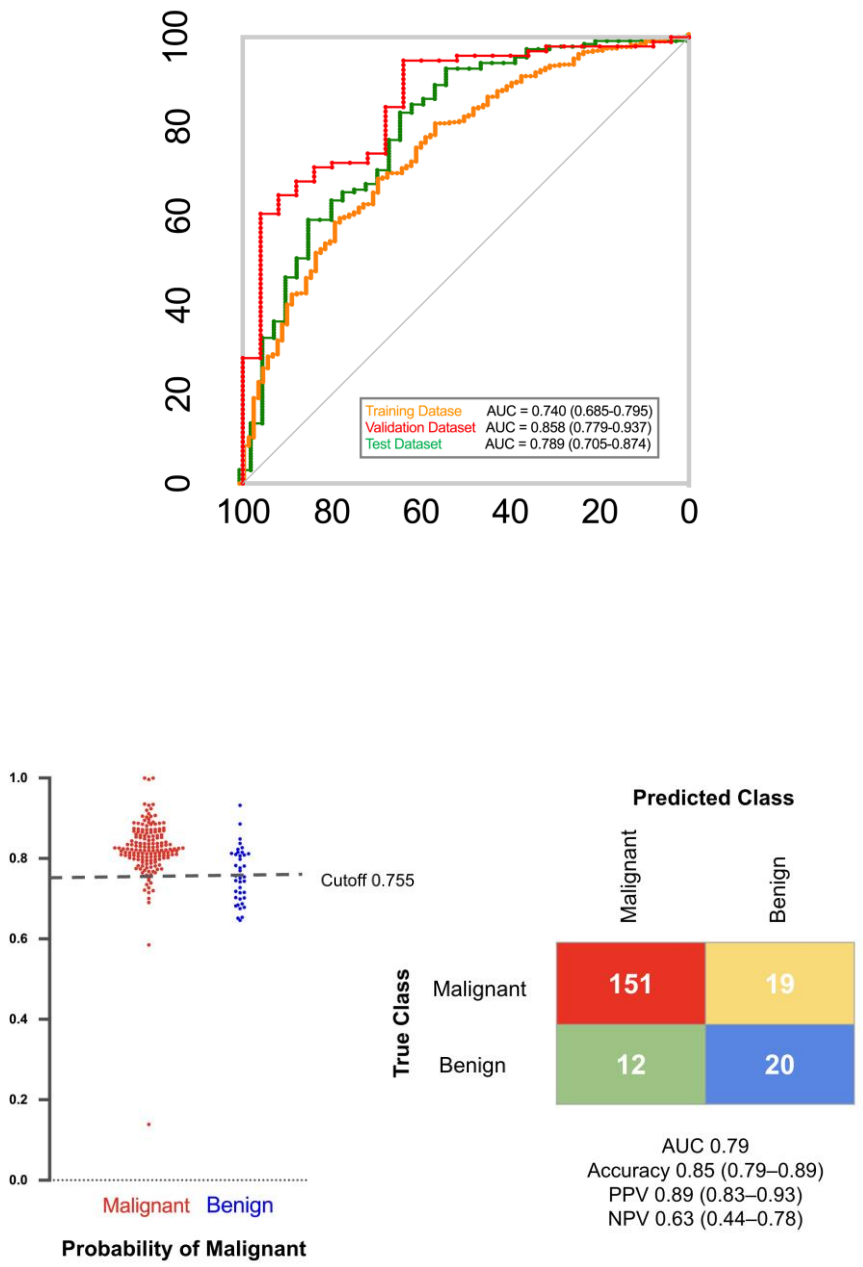


Figure 2. (EB) model. (A) A representative ROC curve illustrates the classification performance of the Epigenetic Biomarker across the training, validation, and test datasets.

(B) The left panel shows predicted probabilities for malignant (red) and benign (blue) nodules. The right panel presents the confusion matrix of test dataset at a cutoff value

When comparing the model performance across all cohorts, specifically for small nodules (5–10 mm), the results indicated that the model remained robust even for smaller nodules. In the overall dataset, the AUCs for the EB model were 0.74, 0.86,

and 0.79 for the training, validation, and test datasets, respectively, with accuracies ranging from 0.80 to 0.88. Sensitivity was high across all datasets, with values of 0.91, 0.95, and 0.93, respectively, while specificity ranged from 0.37 to 0.64. The PPV and NPV were consistent, indicating the reliability of the model for predicting true positives and negatives.

The model maintained strong performance for small nodules (5–10 mm). The AUCs of the training, validation, and test datasets were 0.70, 0.89, and 0.80, respectively, with accuracies of 0.76, 0.88, and 0.85, respectively, indicating that the diagnostic accuracy of the model remained high for smaller nodules. The sensitivities were 0.91, 1.00, and 0.94, respectively, indicating that the model correctly identified the majority of malignant cases. Although lower than sensitivity, specificity was sufficient in most of the subgroups to complement high sensitivity, with values of 0.27, 0.62, and 0.54, respectively.

Lung-RADS score analysis in the test dataset

The model's performance across different RADS categories was evaluated using a test dataset. In RADS 2 (n = 73), the model achieved an AUC of 0.84, an accuracy of 0.84, a sensitivity of 0.90, and a specificity of 0.57. In RADS 3 (n = 19), the AUC was 0.84, with an accuracy of 0.84, sensitivity of 1.00, and specificity of 0.40. In RADS 4A+4B (n = 81), the AUC was 0.82, with an accuracy of 0.82, sensitivity of 0.92, and specificity of 0.44. In RADS 4X (n = 29), the model performed best, with an AUC of 0.97, accuracy of 0.97, sensitivity of 0.96, and specificity of 1.00. These results show high diagnostic accuracy, especially in the higher RADS categories, indicating the clinical utility of the model for assessing pulmonary nodules.

EB model performance in different nodule types

The model demonstrated the detection of lung cancer with accuracy independent of the tumor components; 0.84 for solid and part-solid nodules and 0.86 for GGO nodules in the test dataset. Both GGO and part-solid nodules showed higher PPV when maintaining a similar threshold. The PPVs were 0.86 (95% CI 0.76–0.92) for solid nodules, 0.914 (95% CI 0.76–0.98) for part-solid nodules, and 0.91 (95% CI 0.80–0.97) for GGO nodules. These results highlight the high diagnostic accuracy of the model across different tumor components, indicating its potential utility in the assessment of pulmonary nodules.

Conventional cancer diagnostic model comparison

For the external validation (Figure 3), the EB model achieved an AUC of 0.858 (95% CI 0.779–0.937), significantly outperforming the Mayo Clinic model (AUC, 0.570 [95% CI 0.446–0.694]) and the VA model (AUC, 0.503 [95% CI 0.386–0.621]). The accuracy, sensitivity, and specificity of the EB model are detailed in Table 2. Its superior AUC indicates a higher overall performance compared with the Mayo Clinic and VA models. This finding indicates the potential effectiveness of the EB model in accurately predicting outcomes compared with established clinical models.

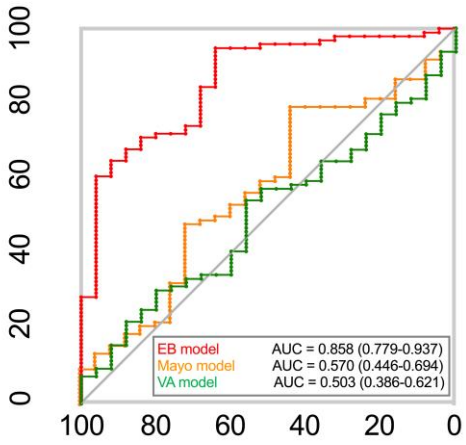


Figure 3: Confusion matrices for Epigenetic Biomarker (EB) model comparing the true class with the predicted class for benign and malignant nodule samples with the Mayo Clinic and Veteran Affairs (VA) models.

DISCUSSION

Early cancer detection is an effective method for reducing cancer-specific mortality. We analyzed the EB profiles of 806 patients with pulmonary nodules and developed an EB model for pulmonary nodule diagnosis, it showed high sensitivity and accuracy with good PPVs at moderate specificity across various imaging characteristics, nodule types, and stages of lung cancer. The EB model also maintained adequate performance, even for small nodules ranging from 5 to 10 mm, which would help decrease the false-negative rate concerning minimally invasive surgery. In addition, sandwich nimmuo-based assays are simple and relatively inexpensive. This is the first retrospective study to validate a blood-based EB model for diagnosing lung nodules.

Previous studies have attempted to enhance lung cancer risk assessment using blood-based biomarkers.[28-30] Various biosources from liquid biopsy, including cfDNA, ctDNA, CTCs, exosomes, and tumor-educated platelets, have been extensively investigated for their role in lung cancer diagnosis. However, none of these tests has been implemented clinically because their sensitivities and specificities are typically insufficient for

clinical decision-making.[12, 13, 30] Alterations in the epigenome, such as DNA methylation and histone modification, play pivotal roles in carcinogenesis. DNA methylation levels and global histone modification patterns may predict cancer recurrence and prognosis across a wide variety of cancer types.[31, 32] In lung cancer, these changes affect significant signaling pathways, including the ERK family, NF-kB, and Hedgehog pathways. Additionally, epigenetic markers are potential biomarkers for early screening, monitoring, and therapeutic strategies in non-small cell lung cancer (NSCLC).[16]

These PTMs can work together or independently to promote the activation or suppression of chromatin-mediated gene expression. These include regulation of inflammatory cytokines, cell cycle arrest, senescence, apoptosis, growth factors, antioxidants, and tumor suppressor genes associated with lung cancer.[16] We focused on the histone variant H3.1 levels and the histone modification H3K27me3. Regarding the prognostic effect of H3K27me3 in various human cancers, H3K27me3 overexpression is linked to a more malignant behavior and worse prognosis in patients with prostate,[33] esophageal,[34] nasopharyngeal,[35] and hepatocellular[36] carcinoma. Conversely, in breast, ovarian, and pancreatic cancers,[31] and in renal cell carcinoma,[37] reduced H3K27me3 expression is associated with a worse prognosis. In lung cancer patients, a lower level of H3K27me3 in tissues has been associated with carcinogenesis [38], whereas a high level of circulating H3K27Me3-nucleosomes in blood has been associated with lung cancer at diagnosis and during treatment[39]. Our model focused on differentiating between benign and malignant nodules. The precise role of H3K27me3 in distinguishing between normal and malignant populations still requires further investigation.

Beyond LDCT screening, liquid biopsies can identify various biomolecular markers, offering insights into the disease status. Integrating liquid biopsy with model training

shows great promise for early-stage diagnoses.[40] However, many current liquid biopsy methods targeting early cancer detection lack the sensitivity to reliably identify early-stage cancers or small nodules.[41-44] The EB model showed satisfactory accuracy, PPV, and NPV for small nodules, specifically within the 5–10 mm range (Table 2). This simple epigenetic regression model had both strengths and limitations. The AUC and accuracy of the model improved significantly from the training to the validation dataset, indicating better predictive performance with new data. The sensitivity was particularly strong, with the validation dataset achieving a perfect score (1.000), emphasizing its effectiveness in detecting malignant nodules, which is essential for early cancer detection. However, specificity was sub-optimal, particularly in the training set. While specificity improved in the validation and test datasets, the risk of false positives remained an issue. Given the benefits of video-assisted thoracoscopic surgery, the main challenge for thoracic surgeons is to accurately identify malignant lung nodules and minimize the risk of delayed diagnosis in high-risk populations identified through CT/LDCT screening.

We applied the updated Lung-RADS to retrospectively evaluate nodule characteristics.[18] The Lung-RADS categorizes nodules based on their likelihood of being malignant, with classification depending on characteristics such as size, attenuation, growth pattern, and other features that may indicate a higher risk of cancer.[18] Nodules classified under Lung-RADS categories 1 and 2 have an estimated malignancy risk of <1%, whereas those in category 3 have a 1–2% risk. Nodules in category 4A have a 5–15% risk, whereas those in category 4B or 4X have a risk of >15%.[18] Most lung cancers identified through screening were observed in nodules categorized as Lung-RADS 3 or 4.

Table 3 shows that the model effectively maintained good accuracy, PPV, and sensitivity in Lung-RADS categories 2 and 3, which are often associated with a lower malignancy risk. The accuracy of both RADS 2 and RADS 3 was consistently high at 0.84, indicating that the model was reliable for correctly classifying nodules as either malignant or benign within these groups. The PPV remained robust, with RADS 2 at 0.90 and RADS 3 at 0.82, showing that when the model predicted a nodule as malignant, it was generally accurate and that the likelihood that these nodules were malignant was high. Furthermore, the sensitivity findings highlighted the model's reliability in correctly identifying malignant nodules, ensuring that few malignancies were missed. Overall, the model maintained strong performance across these metrics, demonstrating its reliability and effectiveness even in categories with a lower pre-test probability of malignancy, supporting its utility in lung cancer screening programs.

In CT lung cancer screening, detected nodules are often part-solid or non-solid; these types of nodules are more likely to be malignant than solid nodules, even when their size is considered.[45] Therefore, achieving high accuracy and PPV in non-solid and part-solid nodules is crucial; our model performed very well in these categories. As shown in Table 4, the model's accuracy was consistent across all nodule types, with 0.84 for both solid and part-solid nodules, and slightly higher at 0.86 for non-solid GGO nodules. The sensitivity remained high, particularly for solid nodules at 0.96, followed by 0.91 for non-solid nodules, and 0.89 for part-solid nodules. This indicates that the model effectively correctly identified malignant nodules across these types. Moreover, the PPV was consistently strong, with 0.86 for solid nodules, 0.91 for part-solid nodules, and 0.91 for non-solid nodules. This suggests that when the model predicts a nodule as malignant, there is a high probability that it will be malignant, regardless of the nodule type. The model showed excellent performance in identifying malignancies in both part-solid and non-solid nodules, which are more likely to be malignant than solid

nodules. This strong performance in terms of accuracy, sensitivity, and PPV underlines the reliability and effectiveness of the model for lung cancer screening, particularly for nodules that present a higher risk of malignancy.

This study had some limitations. We only enrolled thoracic department participants with definitive pathological diagnoses, which may limit the generalizability of the findings. Additionally, the imaging sources were not standardized across the studies, leading to potential image quality and interpretation variability. There was also a lack of integration between the imaging and clinical characteristics of the model parameters, owing to the simplified methodology. Such integration would have provided a more comprehensive assessment. Moreover, the study enrolled a higher proportion of women than men, more never-smokers than smokers, and more patients with adenocarcinoma than those with squamous carcinoma. The study was conducted without using a central laboratory in Taiwan and involved a shipping process that may have introduced variability. We plan to establish an EB model test at a central laboratory-provided service in Taiwan to ensure consistency and reliability. Furthermore, this study was conducted retrospectively, which raises the possibility of overfitting during the model development. Lack of external validation further limitats in ensuring the findings' robustness. A prospective, multi-institutional study with a larger and more diverse patient cohort is required to confirm these observations and enhance the generalizability of the results.

Our study showed that high sensitivity and accuracy in the early detection of NSCLC can be achieved using a panel of EBs in plasma. Concerning detecting high-risk lung cancer, the EB model performed well in detecting small, part-solid, and non-solid nodules, which are the majority in lung cancer screening. This model may reduce false-negative results and facilitate early diagnosis.

Consent

Written informed consent was obtained from the patient for publication of these cases. A copy of the written consent is available for review by the Editor-in Chief of this journal on request.

Acknowledgements

This work was supported by the surgery department of the National Taiwan University Hospital and by grants from the Ministry of Science and Technology, Taiwan (MOST 113-2314-B-002-038-MY3).

Funding:

This study was supported by grants from the Ministry of Science and Technology, Taiwan (MOST 113-2314-B-002-038-MY3) and VolitionRX Limited. Both Volition and the authors collaboratively contributed to the study's development and interpretation of the results.

Tables

Table 1. Participants' baseline characteristics (n = 806)

Patient characteristics	Whole cohort (n = 806)	Training dataset (n = 483) n (%)	Validation dataset (n =121) n (%)	Test set (n = 202) n (%)	*p*-value
Mean age (years) (range)	59.44 (23–89)	59.04 (26–88)	60.03 (33–89)	60.02 (23–85)	.51
Female	511 (63.40%)	306 (63.35%)	79 (65.29%)	126 (62.38%)	.87
Non-smoker	630 (78.16%)	379 (78.47%)	92 (76.03%)	159 (78.71%)	.88
History of alcohol consumption	87 (10.79%)	53 (10.97%)	13 (10.74%)	21 (10.40%)	.97
Lung cancer family history	280 (34.74%)	164 (33.95%)	41 (33.88%)	75 (37.13%)	.75
Nodule type					.28
Solid	357 (44.29%)	217 (44.93%)	51 (42.15%)	89 (44.06%)	
Part-solid	183 (22.70%)	114 (23.60%)	25 (20.66%)	44 (21.78%)	
GGO	266 (33.00%)	152 (31.47%)	45 (37.19%)	69 (34.16%)	
Lung-RADS					.30

	284	164	47	73	
2	(35.24%)	(33.95%)	(38.84%)	(36.14%)	
	69	40	10	19	
3	(8.56%)	(8.28%)	(8.26%)	(9.41%)	
	107	71	19	17	
4A	(13.28%)	(14.70%)	(15.70%)	(8.42%)	
	346	208	45	93	
4B, 4X	(42.92%)	(43.06%)	(37.19%)	(46.04%)	
Nodule size (cm)					.29
	236	136	38	62	
<1cm	(29.28%)	(28.16%)	(31.40%)	(30.69%)	
	274	173	44	57	
1–2 cm	(34.00%)	(35.82%)	(36.36%)	(28.22%)	
	296	174	39	83	
>2 cm	(36.72%)	(36.02%)	(32.23%)	(41.09%)	
	2.05 ± 1.70	2.00 ± 1.64	1.92 ± 1.63	2.24 ± 1.87	.15
Mean tumor size:	(0.3–10.2)	(0.3–10.1)	(0.5–9.6)	(0.4–10.2)	
cm (range)					
Nodule location					.16
	211	125	35	51	
Right upper lobe	(26.18%)	(25.88%)	(28.93%)	(25.25%)	
Right middle	67	49	9	9	
lobe	(8.31%)	(10.14%)	(7.44%)	(4.45%)	
Right lower	171	98	28	45	
lobe	(21.22%)	(20.29%)	(23.14%)	(22.28%)	

	231	137	31	63	
Left upper lobe	(28.66%)	(28.36%)	(25.62%)	(31.19%)	
	113	68	18	27	
Left lower lobe	(14.12%)	(14.08%)	(14.88%)	(13.37%)	
	13	6	0	7	
Others*	(1.61%)	(1.24%)	(0.00%)	(3.47%)	
	648	389	96	163	.95
Malignancy	(80.40%)	(80.54%)	(79.34%)	(80.69%)	

Nodule size (%) < 1 cm, 1–2 cm, > 2 cm

*Patients with a pleural lesion, a hilum lesion, or an inter-fissure lesion

Abbreviations: COPD, chronic obstructive pulmonary disease; FEV1, forced expiratory volume in 1 s; FVC, forced vital capacity; GGO, ground-glass opacity; IQR, interquartile range

Table 2. Performance metrics

	Epigenetic simple regression model		
	Training dataset	Validation dataset	Test dataset
All nodule sizes	(n = 483)	(n =121)	(n = 202)
AUC	0.74	0.86	0.79
Accuracy	0.80	0.88	0.85
	(0.77–0.84)	(0.81–0.94)	(0.79–0.89)
Sensitivity	0.91	0.95	0.93
	(0.87–0.93)	(0.88–0.98)	(0.87–0.96)
Specificity	0.37	0.64	0.51
	(0.28–0.48)	(0.43–0.81)	(0.35–0.67)
PPV[A]	0.86	0.91	0.89
	(0.82–0.89)	(0.83–0.96)	(0.83–0.93)
NPV[A]	0.49	0.76	0.63
	(0.37–0.61)	(0.53–0.91)	(0.44–0.78)
Nodules sized 5–10 mm	(n = 142)	(n = 43)	(n = 61)
AUC	0.70	0.89	0.80
Accuracy (95% CI)	0.76	0.88	0.85
	(0.68–0.83)	(0.75–0.96)	(0.74–0.93)
Sensitivity (95% CI)	0.91	1.000	0.94
	(0.83–0.95)	(0.86–1.000)	(0.82–0.98)
Specificity (95% CI)	0.27	0.62	0.54
	(0.14–0.46)	(0.32–0.85)	(0.26–0.80)

PPV (95% CI)[a]	0.81	0.86	0.88
	(0.72–0.87)	(0.69–0.95)	(0.75–0.95)
NPV (95% CI)[a]	0.47	1.000	0.70
	(0.25–0.71)	(0.60–1.000)	(0.35–0.92)

[A]Cancer prevalence, 80.4% in the current cohort

[B]Cancer prevalence, 76.0% in the current subgroup cohort

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value

Table 3. Performance metrics in the test dataset according to Lung-RADS

	Lung-RADS			
	2	3	4A + 4B	4X
	(n = 73)	(n = 19)	(n = 81)	(n = 29)
AUC				
Accuracy	0.84	0.84	0.82	0.97
	(0.73–0.91)	(0.60–0.97)	(0.71–0.89)	(0.82–1.00)
Sensitivity	0.90	1.00	0.92	0.96
	(0.79–0.96)	(0.73–1.00)	(0.82–0.97)	(0.79–1.00)
Specificity	0.57	0.40	0.44	1.00
	(0.30–0.81)	(0.07–0.83)	(0.22–0.69)	(1.00–1.00)
PPV	0.90	0.82	0.85	1.00
	(0.79–0.96)	(0.56–0.95)	(0.74–0.92)	(0.84–1.00)
NPV	0.57	1.000	0.62	0.67
	(0.30–0.81)	(0.20–1.00)	(0.32–0.85)	(0.13–0.98)

Abbreviations: AUC, area under the receiver operating characteristic curve; NPV, negative predictive value; PPV, positive predictive value; Lung-RADS, lung imaging reporting and data system

Table 4. Performance metrics in the test dataset according to tumor component

	Component		
	Solid	Part-solid	GGO
All nodule sizes	(n = 89)	(n =44)	(n = 69)
Accuracy	0.84	0.84	0.86
	(0.75–0.91)	(0.70–0.93)	(0.75–0.93)
Sensitivity	0.96	0.89	0.91
	(0.87–0.99)	(0.73–0.96)	(0.80–0.97)
Specificity	0.45	0.63	0.55
	(0.24–0.68)	(0.26–0.90)	(0.25–0.82)
PPV	0.86	0.914	0.91
	(0.76–0.92)	(0.76–0.98)	(0.80–0.97)
NPV	0.75	0.56	0.55
	(0.43–0.93)	(0.23–0.85)	(0.25–0.82)

Abbreviation: GGO**,** ground-glass opacity; NPV, negative predictive value; PPV, positive predictive value

Figure Legends

Figure 1. **Flowchart of participant enrollment and model development.** A total of 806 patients with nodules >5 mm were enrolled. The Epigenetic Biomarker (EB) model was trained on 483 samples, validated on 121 samples, and independently tested on 202 samples. The model was built based on the training dataset, with the cutoff determined using the validation dataset, and its performance evaluated on the test dataset.

Figure 2. (EB) model. (A) A representative ROC curve illustrates the classification performance of the Epigenetic Biomarker across the training, validation, and test datasets.
(B) The left panel shows predicted probabilities for malignant (red) and benign (blue) nodules. The right panel presents the confusion matrix of test dataset at a cutoff value of 0.755, yielding an accuracy of 85%.

Figure 3: Confusion matrices for Epigenetic Biomarker (EB) model comparing the true class with the predicted class for benign and malignant nodule samples with the Mayo Clinic and Veteran Affairs (VA) models.

REFERENCES

1. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. *CA Cancer J Clin*. 2010;60(5):277-300. doi:10.3322/caac.20073

2. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365(5):395-409. doi:10.1056/NEJMoa1102873

3. De Koning HJ, Van Der Aalst CM, De Jong PA, et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial. *N Engl J Med*. 2020;382(6):503-513. doi:10.1056/NEJMoa1911793

4. He J, Wang B, Tao J, et al. Accurate classification of pulmonary nodules by a combined model of clinical, imaging, and cell-free DNA methylation biomarkers: a model development and external validation study. *Lancet Digit Health*. 2023;5(10):e647-e656. doi:10.1016/S2589-7500(23)00125-5

5. Vachani A, Tanner NT, Aggarwal J, et al. Factors that influence physician decision making for indeterminate pulmonary nodules. *Ann Am Thorac Soc*. 2014;11(10):1586-1591. doi:10.1513/AnnalsATS.201405-197BC

6. Gierada DS, Pinsky P, Nath H, Chiles C, Duan F, Aberle DR. Projected outcomes using different nodule sizes to define a positive CT lung cancer screening examination. *J Natl Cancer Inst*. 2014;106(11):dju284-dju. doi:10.1093/jnci/dju284

7. Chang G-C, Chiu C-H, Yu C-J, et al. Low-dose CT screening among never-smokers with or without a family history of lung cancer in Taiwan: a prospective cohort study. *Lancet Respir Med*. 2024;12(2):141-152. doi:10.1016/S2213-2600(23)00338-7

8. Gould MK, Donington J, Lynch WR, et al. Evaluation of individuals with pulmonary nodules: when is it lung cancer? Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest*. 2013;143(5)(suppl):e93S-e120S. doi:10.1378/chest.12-2351

9. Kammer MN, Lakhani DA, Balar AB, et al. Integrated biomarkers for the management of indeterminate pulmonary nodules. *Am J Respir Crit Care Med*. 2021;204(11):1306-1316. doi:10.1164/rccm.202012-4438OC

10. Dawson S-J, Tsui DWY, Murtaza M, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med*. 2013;368(13):1199-1209. doi:10.1056/NEJMoa1213261

11. Abbosh C, Birkbak NJ, Wilson GA, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature*. 2017;545(7655):446-451. doi:10.1038/nature22364

12. Casagrande GMS, Silva MO, Reis RM, Leal LF. Liquid biopsy for lung cancer: up-to-date and perspectives for screening programs. *Int J Mol Sci*. 2023;24(3):2505. doi:10.3390/ijms24032505

13. Malapelle U, Pisapia P, Pepe F, et al. The evolving role of liquid biopsy in lung cancer. *Lung Cancer*. 2022;172:53-64. doi:10.1016/j.lungcan.2022.08.004

14. Kerr KM, Galler JS, Hagen JA, Laird PW, Laird-Offringa IA. The role of DNA methylation in the development and progression of lung adenocarcinoma. *Dis Markers*. 2007;23(1-2):5-30. doi:10.1155/2007/985474

15. Van Den Ackerveken P, Lobbens A, Pamart D, et al. Epigenetic profiles of elevated cell free circulating H3.1 nucleosomes as potential biomarkers for non-Hodgkin lymphoma. *Sci Rep*. 2023;13(1). doi:10.1038/s41598-023-43520-0

16. Bajbouj K, Al-Ali A, Ramakrishnan RK, Saber-Ayad M, Hamid Q. Histone modification in NSCLC: molecular mechanisms and therapeutic targets. *Int J Mol Sci*. 2021;22(21):11701. doi:10.3390/ijms222111701

17. Bai C, Choi C-M, Chu CM, et al. Evaluation of pulmonary nodules: Clinical Practice Consensus Guidelines for Asia. *Chest*. 2016;150(4):877-893. doi:10.1016/j.chest.2016.02.650

18. Kastner J, Hossain R, Jeudy J, Dako F, Mehta V, Dalal S, Dharaiya E, White C. *Lung-RADS*. version 1.0 versus Lung-RADS. version 1.1: Comparison of Categories Using Nodules from the National Lung Screening Trial. *Radiology*; 2021. Vol 300:199-206.

19. Ye T, Deng L, Wang S, et al. Lung adenocarcinomas manifesting as radiological part-solid nodules define a special clinical subtype. *J Thorac Oncol*. 2019;14(4):617-627. doi:10.1016/j.jtho.2018.12.030

20. Hattori A, Matsunaga T, Hayashi T, Takamochi K, Oh S, Suzuki K. Prognostic impact of the findings on thin-section computed tomography in patients with subcentimeter non–small cell lung cancer. *J Thorac Oncol*. 2017;12(6):954-962. doi:10.1016/j.jtho.2017.02.015

21. Jacobson FL, Austin JHM, Field JK, et al. Development of the American Association for Thoracic Surgery guidelines for low-dose computed tomography scans to screen for lung cancer in North America: recommendations of The American Association for Thoracic Surgery Task Force for Lung Cancer Screening and Surveillance. *J Thorac Cardiovasc Surg*. 2012;144(1):25-32. doi:10.1016/j.jtcvs.2012.05.059

22. Nicholson AG, Tsao MS, Beasley MB, et al. The 2021 WHO classification of lung tumors: impact of advances since 2015. *J Thorac Oncol*. 2022;17(3):362-387. doi:10.1016/j.jtho.2021.11.003

23. Travis WD, Brambilla E, Noguchi M, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol*. 2011;6(2):244-285. doi:10.1097/JTO.0b013e318206a221

24. Chen J-S, Cheng Y-J, Hung M-H, Tseng Y-D, Chen K-C, Lee Y-C. Nonintubated thoracoscopic lobectomy for lung cancer. *Ann Surg*. 2011;254(6):1038-1043.

doi:10.1097/SLA.0b013e31822ed19b

25. Chen P-H, Hsu H-H, Yang S-M, et al. Preoperative dye localization for thoracoscopic lung surgery: hybrid versus computed tomography room. *The Annals of Thoracic Surgery*. 2018/09/16;106(6):1661-1667. doi:10.1016/j.athoracsur.2018.07.030

26. Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD, Edell ES. The probability of malignancy in solitary pulmonary nodules: application to small radiologically indeterminate nodules. *Arch Intern Med*. 1997;157(8):849-855

27. Gould MK, Ananth L, Barnett PG. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. *Chest*. 2007;131(2):383-388. doi:10.1378/chest.06-1261

28. Ostrow KL, Hoque MO, Loyo M, et al. Molecular analysis of plasma DNA for the early detection of lung cancer by quantitative methylation-specific PCR. *Clin Cancer Res*. 2010;16(13):3463-3472. doi:10.1158/1078-0432.CCR-09-3304

29. Belinsky SA, Klinge DM, Dekker JD, et al. Gene promoter methylation in plasma and sputum increases with lung cancer risk. *Clin Cancer Res*. 2005;11(18):6505-6511. doi:10.1158/1078-0432.CCR-05-0625

30. Li W, Liu J-B, Hou L-K, et al. Liquid biopsy in lung cancer: significance in diagnostics, prediction, and treatment monitoring. *Mol Cancer*. 2022;21(1):25. doi:10.1186/s12943-022-01505-z

31. Wei Y, Xia W, Zhang Z, et al. Loss of trimethylation at lysine 27 of histone H3 is a predictor of poor outcome in breast, ovarian, and pancreatic cancers. *Mol Carcinog*. 2008;47(9):701-706. doi:10.1002/mc.20413

32. Mitani Y, Oue N, Hamai Y, et al. Histone H3 acetylation is associated with reduced p21(WAF1/CIP1) expression by gastric carcinoma$^{WAF1/CIP1}$ expression by gastric carcinoma. *J Pathol*. 2005;205(1):65-73. doi:10.1002/path.1684

33. Ellinger J, Kahl P, Von Der Gathen J, et al. Global histone H3K27 methylation levels are different in localized and metastatic prostate cancer. *Cancer Investig*. 2012;30(2):92-97. doi:10.3109/07357907.2011.636117

34. Tzao C, Tung H-J, Jin J-S, et al. Prognostic significance of global histone modifications in resected squamous cell carcinoma of the esophagus. *Mod Pathol*. 2009;22(2):252-260. doi:10.1038/modpathol.2008.172

35. Cai M-Y, Tong Z-T, Zhu W, et al. H3K27me3 protein is a promising predictive biomarker of patients' survival and chemoradioresistance in human nasopharyngeal carcinoma. *Mol Med*. 2011;17(11-12):1137-1145. doi:10.2119/molmed.2011.00054

36. Cai M-Y, Hou J-H, Rao H-L, et al. High expression of H3K27me3 in human hepatocellular carcinomas correlates closely with vascular invasion and predicts worse prognosis in patients. *Mol Med*. 2011;17(1-2):12-20. doi:10.2119/molmed.2010.00103

37. Chervona Y, Costa M. Histone modifications and cancer: biomarkers of prognosis? *Am J Cancer Res*. 2012;2(5):589-597.

38. Chen X, Song N, Matsumoto K, et al. High expression of trimethylated histone H3 at lysine 27 predicts better prognosis in non-small cell lung cancer. *Int J Oncol*. 2013;43(5):1467-1480. doi:10.3892/ijo.2013.2062

39. Grolleau E, Candiracci J, Lescuyer G, et al. Circulating H3K27 methylated nucleosome plasma concentration: synergistic information with circulating tumor DNA molecular profiling. *Biomolecules*. 2023;13(8):1255. doi:10.3390/biom13081255

40. Shin H, Oh S, Hong S, et al. Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes. *ACS Nano*. 2020;14(5):5435-5444. doi:10.1021/acsnano.9b09119

41. Scimia M, Du J, Pepe F, et al. Evaluation of a novel liquid biopsy-based ColoScape assay for mutational analysis of colorectal neoplasia and triage of FIT+ patients: a pilot study. *J Clin Pathol*. 2018;71(12):1123-1126. doi:10.1136/jclinpath-2018-205412

42. Cohen JD, Javed AA, Thoburn C, et al. Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers. *Proc Natl Acad Sci USA*. 2017;114(38):10202-10207. doi:10.1073/pnas.1704961114

43. Klein EA, Richards D, Cohn A, et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Ann Oncol*. 2021;32(9):1167-1177. doi:10.1016/j.annonc.2021.05.806

44. Zhang Y, Sun B, Yu Y, et al. Multimodal fusion of liquid biopsy and CT enhances differential diagnosis of early-stage lung adenocarcinoma. *npj Precis Oncol*. 2024;8(1). doi:10.1038/s41698-024-00551-8

45. Henschke CI, Yankelevitz DF, Mirtcheva R, McGuinness G, McCauley D, Miettinen OS, CT. Screening for lung cancer. *AJR Am J Roentgenol*. 2002;178:1053-1057.

Supplementary Table S1.

No.	Abbreviation	Full name
1	H3.1	Nucleosome containing H3.1 variant (H3.1-nucleosome)
2	H3K9Ac	Nucleosome containing Histone H3 acetyl Lys9 (H3K9Ac-nucleosome)
3	H3K9Me3	Nucleosome containing Histone 3 lysine 9 trimethylation (H3K9Me3-nucleosome)
4	H3K27Me3	Nucleosome containing Histone 3 lysine 27 trimethylation (H3K27Me3-nucleosome)
5	H3K36Me3	Nucleosome containing Histone 3 lysine 36 trimethylation (H3K36Me3-nucleosome)