



Validation of Lung EpiCheck, a novel methylation-based blood assay, for the detection of lung cancer in European and Chinese high-risk individuals

Mina Gaga ¹, Joanna Chorostowska-Wynimko ², Ildikó Horváth³, Martin C. Tammemagi⁴, David Shitrit⁵, Vered H. Eisenberg⁶, Hao Liang⁷, David Stav⁸, Dan Levy Faber^{9,10}, Maarten Jansen¹¹, Yael Raviv¹², Vasileios Panagoulas¹³, Piotr Rudzinski², Gabriel Izbiccki¹⁴, Ohad Ronen ¹⁵, Adiv Goldhaber ¹⁶, Rawia Moalem¹⁷, Nadir Arber¹⁸, Ilana Haas¹⁹ and Qinghua Zhou ⁷



@ERSpublications

Lung EpiCheck, a simple blood test, detected 85% of early-stage lung cancers with specificity of 64% in high-risk population, reaching AUC of 0.942 when combined with risk factors. This could improve efficiency of implementing lung cancer screening. <https://bit.ly/3jWhL0n>

Cite this article as: Gaga M, Chorostowska-Wynimko J, Horváth I, *et al.* Validation of Lung EpiCheck, a novel methylation-based blood assay, for the detection of lung cancer in European and Chinese high-risk individuals. *Eur Respir J* 2021; 57: 2002682 [<https://doi.org/10.1183/13993003.02682-2020>].

ABSTRACT

Aim: Lung cancer screening reduces mortality. We aim to validate the performance of Lung EpiCheck, a six-marker panel methylation-based plasma test, in the detection of lung cancer in European and Chinese samples.

Methods: A case-control European training set (n=102 lung cancer cases, n=265 controls) was used to define the panel and algorithm. Two cut-offs were selected, low cut-off (LCO) for high sensitivity and high cut-off (HCO) for high specificity. The performance was validated in case-control European and Chinese validation sets (cases/controls 179/137 and 30/15, respectively).

Results: The European and Chinese validation sets achieved AUCs of 0.882 and 0.899, respectively. The sensitivities/specificities with LCO were 87.2%/64.2% and 76.7%/93.3%, and with HCO they were 74.3%/90.5% and 56.7%/100.0%, respectively. Stage I nonsmall cell lung cancer (NSCLC) sensitivity in European and Chinese samples with LCO was 78.4% and 70.0% and with HCO was 62.2% and 30.0%, respectively. Small cell lung cancer (SCLC) was represented only in the European set and sensitivities with LCO and HCO were 100.0% and 93.3%, respectively. In multivariable analyses of the European validation set, the assay's ability to predict lung cancer was independent of established risk factors (age, smoking, COPD), and overall AUC was 0.942.

Conclusions: Lung EpiCheck demonstrated strong performance in lung cancer prediction in case-control European and Chinese samples, detecting high proportions of early-stage NSCLC and SCLC and significantly improving predictive accuracy when added to established risk factors. Prospective studies are required to confirm these findings. Utilising such a simple and inexpensive blood test has the potential to improve compliance and broaden access to screening for at-risk populations.

This article has supplementary material available from erj.ersjournals.com

This study is registered at clinicaltrials.gov as NCT02373917

This article has an editorial commentary: <https://doi.org/10.1183/13993003.04367-2020>

Received: 9 July 2020 | Accepted after revision: 7 Oct 2020

Copyright ©ERS 2021. This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0.

Introduction

Lung cancer is the leading cause of death from cancer, with 1.76 million deaths worldwide in 2018 [1]. Risk factors include age, smoking, family history and occupational/asbestos exposure. 5-year survival rate for lung cancer is only 18.6%, mainly due to diagnosis at late stages [2]. Screening with low-dose computed tomography (LDCT) has been proven to reduce lung cancer mortality in high-risk populations [3, 4]. However, LDCT has a significant rate of false positives and overdiagnosis, involves radiation hazard, is reader dependent and requires substantial infrastructure. In the USA, up to 14% of the eligible population undergo lung cancer screening [5]. Current barriers are infrastructure and knowledge and awareness gaps among referring physicians and the public. Importantly, lung cancer screening is targeting a very high-risk population, representing merely a quarter of lung cancer patients [6].

Several types of tumour-derived biomarkers have been assessed for lung cancer detection, including circulating tumour cells, exosomes, mutations and methylation changes in cell-free (cf)DNA, microRNA and proteins [7, 8]. Genome-wide hypomethylation and hypermethylation changes are found in lung cancer and could potentially serve as markers [9].

EpiCheck is a simple ultrasensitive PCR-based assay that detects cancer-associated hypermethylation changes in a selected panel of markers from any body fluid or tissue. The urine-based Bladder EpiCheck demonstrated sensitivity of 92% for high-grade urothelial carcinoma with specificity of 88% in bladder cancer patients undergoing surveillance [10].

The purpose of this study is to validate the performance of Lung EpiCheck®, a six-methylation-marker blood test, in lung cancer detection.

Methods

Study samples

Training set samples were used to select the markers for the panel using Nucleix's proprietary bioinformatics techniques (Nucleix, Rehovot, Israel). Six markers were selected based on their synergistic information and an algorithm calculating the EpiScore was developed and locked down (supplementary figure S1). Two cut-offs were defined to allow for different clinical scenarios, a low cut-off (LCO) of EpiScore ≥ 60 , favouring high sensitivity, and a high cut-off (HCO) of EpiScore ≥ 70 , favouring high specificity. The European validation set was a new set of samples used to validate the performance of the assay using the pre-defined algorithm and cut-offs.

The training and the European validation sets were obtained by applying a single protocol: a case-control study performed on samples from sequential recruitment in 18 departments and clinics in 16 healthcare organisations, and three biobanks in Europe and Israel (supplementary table S2). Samples were collected from July 2016 to March 2018. The initial series of cases and controls were used for training and the subsequent series was used for validation. Cases were recruited from pulmonology, thoracic surgery and oncology departments and clinics in Europe and Israel. Present and past smokers, serving as controls, were recruited from blood collection stations in primary care clinics and from general surgery departments in Israel. Potential participants were randomly approached as they came to perform a blood test for any reason (table 1). Sample processing was performed at the Nucleix laboratory (Rehovot). Disease staging of the cases was according to the American Joint Committee on Cancer staging manual (AJCC)7 and AJCC8. Adenocarcinomas were included only if classified as invasive adenocarcinomas according to International

Affiliations: ¹7th Respiratory Medicine Dept, Athens Chest Hospital, Athens, Greece. ²National Institute of Tuberculosis and Lung Diseases, Warsaw, Poland. ³National Korányi Institute of Pulmonology, Budapest, Hungary. ⁴Brock University, St Catharines, ON, Canada. ⁵Pulmonary Dept, Meir Medical Center, Kfar Saba, Israel. ⁶Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel. ⁷Lung Cancer Center/Lung Cancer Institute, West China Hospital, Sichuan University, Chengdu, China. ⁸Lung Institute, Maccabi Health Services Hashalom, Tel-Aviv, Israel. ⁹Dept of Cardiothoracic Surgery, Lady Davis Carmel Medical Center, Haifa, Israel. ¹⁰Ruth and Bruce Rappaport Faculty of Medicine, Technion – Israel Institute of Technology, Haifa, Israel. ¹¹Pulmonary Dept, Ziekenhuisgroep Twente, Hengelo, The Netherlands. ¹²Dept of Medicine, Pulmonology Institute, Soroka Medical Center, Ben-Gurion University, Beer-Sheva, Israel. ¹³2nd Respiratory Medicine Dept, Athens Chest Hospital, Athens, Greece. ¹⁴Pulmonary Institute, Shaare Zedek Medical Center, Jerusalem, Israel. ¹⁵Dept of Otolaryngology – Head and Neck Surgery, Galilee Medical Center affiliated with Azrieli Faculty of Medicine, Safed, Israel. ¹⁶Clalit Health Services, Raanana, Israel. ¹⁷Gastroenterology Institute, The Holy Family Hospital, Nazareth, Israel. ¹⁸Integrated Cancer Prevention Center, Tel Aviv Sourasky Medical Centre, Sackler Faculty of Medicine, Tel Aviv University, Tel-Aviv, Israel. ¹⁹Breast Unit, Meir Medical Center, Kfar Saba, Israel.

Correspondence: Mina Gaga, 7th Resp. Med. Dept and Asthma Center, 152 Mesogion Ave Athens 11527, Greece. E-mail: minagaga@yahoo.com

TABLE 1 Eligibility criteria

	Inclusion criteria	Exclusion criteria
Cases	Subjects with pathologically proven primary lung cancer (NSCLC and SCLC) or Subjects with suspected primary lung cancer undergoing a diagnostic procedure. Patients enrolled with suspicion of lung cancer were included in the analysis if diagnosis was pathologically confirmed primary lung cancer (SCLC and NSCLC)	Subjects with cancer, other than lung cancer Subjects with history of cancer of any kind (except for fully resected nonmelanoma skin cancer)
Controls	European sets: Age \geq 50 years Current or former smoker Chinese set: Healthy individuals willing to donate blood for the study	Subjects with current diagnosis or history of cancer of any kind (except for fully resected nonmelanoma skin cancer)

NSCLC: nonsmall cell lung cancer; SCLC: small cell lung cancer.

Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society classification [11].

The Chinese validation set was a small feasibility study assessing the applicability of Lung EpiCheck for lung cancer detection in a Chinese population. This was a blinded, case-control, single-centre study performed in the Lung Cancer Center/Lung Cancer Institute at the West China Hospital (Sichuan University, Chengdu, China). Samples were collected from January 2018 to November 2018. Patients suspected or confirmed to have lung cancer arriving for lung surgery were enrolled. Healthy volunteers were enrolled sequentially as controls (table 1). Sample processing was performed on site. Disease staging for the cases was according to AJCC8.

Relevant medical, smoking and family history data were collected prior to study-related procedures. The study was approved by the ethics committees of the various institutions involved, and all subjects provided signed informed consent. The study registration number is NCT02373917.

Lung EpiCheck testing

Lung EpiCheck (Nucleix) is a blood test that detects lung cancer-associated hypermethylation in six markers in cfDNA. Plasma is separated from a 10 mL EDTA tube within 4 h of blood draw by two consecutive centrifugations at 1500 \times g for 10 min and stored at -20°C to -80°C until DNA extraction. Lung EpiCheck's reagents and methylation-sensitive enzymes are used for DNA extraction, digestion and amplification in real-time PCR (ABI 7500 Fast Dx; Thermo Fisher Scientific, Carlsbad, CA, USA). Three PCR wells are amplified for the markers and one for an internal control to verify the quality of plasma separation by detecting leukocyte-derived DNA. Lung EpiCheck software analyses the PCR output calculating an EpiScore, a numerical score (0–100) reflecting the overall methylation level in the assay's markers.

Statistical analysis

The groups' baseline characteristics were compared using Chi-squared test for categorical parameters and Wilcoxon's rank-sum test for continuous parameters. Sensitivity and specificity were calculated for the entire sample and for different subgroups of interest along with 95% exact binomial confidence intervals. The predictive ability of the continuous EpiScore was evaluated *via* logistic regression and the corresponding area under the receiver operator characteristic curve (AUC) was calculated. Positive likelihood ratio (LR+) and negative likelihood ratio (LR-) were calculated for the entire sample and for different subgroups of interest along with 95% exact binomial confidence intervals (LR+ = sensitivity/(1-specificity), LR- = (1-sensitivity)/specificity). A multivariable logistic regression was used to examine the relationship between the true patient status (lung cancer case or control) and their EpiScore result. The contribution of the EpiScore result was examined adjusting for the patient's personal characteristics and known risk factors for lung cancer. An additional multivariable logistic regression analysis was performed to examine whether the EpiScore outcome is affected by a patient's personal characteristics or known risk factors for lung cancer. Both analyses used the

subset of patients for whom all relevant information was available. The contribution of each predictor in the model was evaluated *via* odds ratio and the overall prediction ability of the model was evaluated *via* AUC.

Results

The training set included 367 subjects (102 cases and 265 controls). Cases were significantly older with a higher number of pack-years compared to controls, while sex and number of years since quitting smoking were similar (table 2). Smoking status was also significantly different, as half of the cases were missing this information. A balanced distribution of histological subtypes and stages was achieved with 28% of nonsmall lung cancer (NSCLC) patients having stage I disease (table 3). The AUC (95% CI) was 0.890 (0.848–0.932) (figure 1a), the sensitivity/specificity combinations were 84.3% (75.8–90.8%)/77.7% (72.2–82.6%) with LCO and 73.5% (63.9–81.8%)/93.6%(89.9–96.2%) with HCO (table 3). Likelihood ratios are reported in supplementary table S5.

The European validation set included 363 subjects, out of which 316 subjects were eligible and had EpiScore results (179 cases and 137 controls; supplementary figure S3a). Cases were heavier smokers (median 41 *versus* 20 pack-years), slightly older (median age 65 *versus* 63 years) and were more likely to be male compared to controls (74% *versus* 63%); all statistically significant (table 2). Despite similar rates of current smokers (42%), smoking status was significantly different, as a third of the cases were missing this information. Years since quitting smoking were similar between the groups (18 *versus* 17 years). Adenocarcinoma was the most common histological subtype (46%), followed by squamous cell carcinoma (41%) (table 3). Small cell lung cancer (SCLC) was underrepresented compared to the incidence reported in the literature (8% *versus* 13% [12]). A balanced distribution of NSCLC stages was achieved and 26% of NSCLC patients had stage I disease (table 3). The AUC was 0.882 (0.846–0.918) (figure 1b) overall and 0.797 (0.704–0.889), 0.830 (0.764–0.895) and 0.862 (0.813–0.910) in stage I, stages I and II and early stages (stages I, II and IIIA), respectively (supplementary figure S4). Overall sensitivity/specificity combinations were 87.2% (81.3–91.7%)/64.2% (55.6–72.2%) with LCO and 74.3% (67.2–80.5%)/90.5% (84.3–94.9%) with HCO. Applying the LCO/HCO cut-offs, Lung EpiCheck detected 85.1% (76.7–91.4%)/70.3% (60.4–79.0%) of early-stage NSCLC, 78.4% (61.8–90.2%)/62.2% (44.8–77.5%) of stage I NSCLC and 57.1% (18.4–90.1%)/42.9% (9.9–81.6%) of stage I NSCLC ≤ 20 mm. Lung EpiCheck demonstrated high sensitivity of 100.0% (54.1–100.0%)/100.0% (54.1–100.0%) in limited SCLC and 100.0% (66.4–100.0%)/88.9% (51.8–99.7%) in extensive SCLC with LCO/HCO; however, the numbers of SCLC were small. Sensitivities were similar ($p > 0.05$) across histological subtypes, NSCLC stages, NSCLC early-/late-stage groups and limited/extensive SCLC for each cut-off. The only factor significantly impacting sensitivity in univariate analyses were tumour size ($p < 0.001/p < 0.0001$ in LCO/HCO) and tumour size of stage I ($p = 0.003/p = 0.020$).

TABLE 2 Patient demographics

	Training set			European validation set			Chinese validation set		
	Cases	Controls	p-value	Cases	Control	p-value	Cases	Control	p-value
Subjects	102	265		179	137		30	15	
Age years	67 [51–83]	62 [49–82]	$p < 0.0001$	65 [23–89]	63 [50–87]	0.0130	64 [40–79]	33 [23–51]	$p < 0.0001$
Sex			0.182			0.037			0.053
Male	70 [68.6]	162 [61.1]		132 [73.7]	86 [62.8]		21 [70.0]	6 [40.0]	
Female	32 [31.4]	103 [38.9]		47 [26.3]	51 [37.2]		9 [30.0]	9 [60.0]	
Smoking status			$p < 0.0001^{\S}$			$p < 0.0001^{\S}$			$p < 0.001^{\S}$
Current smoker [#]	36 [35.3]	117 [44.2]		75 [41.9]	57 [41.6]		11 [36.7]	0 [0.0]	
Former smoker	7 [6.9]	148 [55.8]		35 [19.6]	80 [58.4]		4 [13.3]	1 [6.7]	
Never-smoker	6 [5.9]			9 [5.0]			14 [46.7]	14 [93.3]	
Unknown	53 [52.0]			60 [33.5]			1 [3.3]	0 [0.0]	
Smoking pack-years	40 [2–129]	20 [1–138]	0.010	41 [4–182]	20 [1–120]	$p < 0.0001$	30 [15–80]	10 [NA]	<i>f</i>
Years since quitting smoking[†]	13 [2–41]	17 [1–50]	0.267	18 [1–42]	17 [1–58]	0.394	7 [3–10]	10 [NA]	<i>f</i>

Data are presented as n, median [range] or n (%), unless otherwise stated. [#]: smokers reporting quitting within 1 year prior to study inclusion were considered to be current smokers; ^{||}: reported in smokers only; this information was missing for five cases in the training set; [†]: reported in former smokers only; this information was missing for five controls in the training set, and in five patients (n=4 cases, n=1 control) in the European validation set; [§]: this analysis included only patients in the current and former smoker categories; ^f: not calculated as there was only one control subject with history of smoking in this set.

TABLE 3 Lung EpiCheck performance by sensitivity and specificity

	Training set				European validation set				Chinese validation set			
	LCO EpiScore=60		HCO EpiScore=70		LCO EpiScore=60		HCO EpiScore=70		LCO EpiScore=60		HCO EpiScore=70	
Cases/controls	102/265				179/137				30/15			
AUC (95% CI)	0.890 (0.848–0.932)				0.882 (0.846–0.918)				0.899 (0.809–0.989)			
Overall sensitivity	86/102	84.3	75/102	73.5	156/179	87.2	133/179	74.3	23/30	76.7	17/30	56.7
		(75.8–90.8)		(63.9–81.8)		(81.3–91.7)		(67.2–80.5)		(59.1–88.2)		(39.2–72.6)
Overall specificity	206/265	77.7	248/265	93.6	88/137	64.2	124/137	90.5	14/15	93.3	15/15	100.0
		(72.2–82.6)		(89.9–96.2)		(55.6–72.2)		(84.3–94.9)		(68.0–99.9)		(78.1–100.0)
Sensitivity by histological subtype		Overall [§]		Overall [§]		Overall [§]		Overall [§]		p=0.029		p<0.001
		p=0.075		p=0.051		p=0.250		p=0.311				
		NSCLC versus SCLC		NSCLC versus SCLC		NSCLC versus SCLC		NSCLC versus SCLC				
		p=0.213		p=0.278		p=0.224		p=0.120				
Adenocarcinoma	34/45	75.6	27/45	60.0	73/82	89.0	59/82	72.0	12/19	63.2	6/19	31.6
		(60.5–87.1)		(44.3–74.3)		(80.2–94.9)		(60.9–81.3)		(38.4–83.7)		(12.6–56.6)
Squamous cell carcinoma [#]	35/38	92.1	32/38	84.2	61/74	82.4	53/74	71.6	11/11	100.0	11/11	100.0
		(78.6–98.3)		(68.7–94.0)		(71.8–90.3)		(59.9–81.5)		(71.5–100.0)		(71.5–100.0)
Other NSCLC	2/3	66.7	2/3	66.7	5/6	83.3	5/6	83.3				
		(9.4–99.2)		(9.4–99.2)		(35.9–99.6)		(35.9–99.6)				
All NSCLC	71/86	82.6	61/86	70.9	139/162	85.8	117/162	72.2	23/30	76.7	17/30	56.7
		(72.9–89.9)		(60.1–80.2)		(79.5–90.8)		(64.7–79.0)		(57.7–90.1)		(37.4–74.5)
Small cell carcinoma	10/10	100.0	9/10	90.0	15/15	100.0	14/15	93.3				
		(69.2–100.0)		(55.5–99.7)		(78.2–100.0)		(68.1–99.8)				
Other/unknown	5/6	83.3	5/6	83.3	2/2	100.0	2/2	100.0				
		(35.9–99.6)		(35.9–99.6)		(15.8–100.0)		(15.8–100.0)				
Sensitivity by NSCLC stage		p=0.075		p=0.012		p=0.089		p=0.187		p=0.193		p=0.018
Stage I	18/26	69.2	13/26	50.0	29/37	78.4	23/37	62.2	7/10	70.0	3/10	30.0
		(48.2–85.7)		(29.9–70.1)		(61.8–90.2)		(44.8–77.5)		(34.8–93.3)		(6.7–65.2)
Stage II	17/21	81.0	14/21	66.7	24/28	85.7	20/28	71.4	3/6	50.0	2/6	33.3
		(58.1–94.6)		(43.0–85.4)		(67.3–96.0)		(51.3–86.8)		(11.8–88.2)		(4.3–77.7)
Stage III	17/20	85.0	16/20	80.0	53/59	89.8	44/59	74.6	10/11	90.9	10/11	90.9
		(62.1–96.8)		(56.3–94.3)		(79.2–96.2)		(61.6–85.0)		(58.7–99.8)		(58.7–99.8)
Stage IV	17/17	100.0	16/17	94.1	33/37	89.2	30/37	81.1	3/3	100.0	2/3	66.7
		(80.5–100.0)		(71.3–99.9)		(74.6–97.0)		(64.8–92.0)		(29.2–100.0)		(9.4–99.2)
Unstaged	2/2	100.0	2/2	100.0	0/1	0.0	0/1	0.0				
		(15.8–100.0)		(15.8–100.0)		(0.0–97.5)		(0.0–97.5)				

Continued

TABLE 3 Continued

	Training set		European validation set				Chinese validation set					
	LCO EpiScore=60	HCO EpiScore=70	LCO EpiScore=60	HCO EpiScore=70	LCO EpiScore=60	HCO EpiScore=70	LCO EpiScore=60	HCO EpiScore=70				
Sensitivity by tumour size (largest diameter), NSCLC only					p<0.001 ^f		p<0.0001 ^f		p=0.584		p=0.043	
≤20 mm			11/16	68.8 (41.3–89.0)	7/16	43.8 (19.8–70.1)	3/5	60.0 (14.7–94.7)	1/5	20.0 (0.5–71.6)		
21–30 mm			14/22	63.6 (40.7–82.8)	10/22	45.5 (24.4–67.8)	4/6	66.7 (22.3–95.7)	2/6	33.3 (4.3–77.7)		
31–50 mm			46/49	93.9 (83.1–98.7)	38/49	77.6 (63.4–88.2)	11/14	78.6 (49.2–95.3)	9/14	64.3 (35.1–87.2)		
>50 mm			47/50	94.0 (83.5–98.7)	44/50	88.0 (75.7–95.5)	5/5	100.0 (47.8–100.0)	5/5	100.0 (47.8–100.0)		
Unknown [¶]			23/27	85.2 (66.3–95.8)	20/27	74.1 (53.7–88.9)						
Sensitivity by tumour size (largest diameter), stage I NSCLC only					p=0.003		p=0.020		p=0.700		p=0.200	
≤20 mm			4/7	57.1 (18.4–90.1)	3/7	42.9 (9.9–81.6)	2/4	50.0 (6.8–93.2)	0/4	0.0 (0.0–60.2)		
21–30 mm			6/11	54.5 (23.4–83.3)	4/11	36.4 (10.9–69.2)	3/3	100.0 (29.2–100.0)	1/3	33.3 (0.8–90.6)		
>30 mm			19/19	100.0 (82.4–100.0)	16/19	84.2 (60.4–96.6)	2/3	66.7 (9.4–99.2)	2/3	66.7 (9.4–99.2)		
Sensitivity by stage group, NSCLC only					p=0.058		p=0.008		p=0.290		p=0.196	
Early stages (I, II and IIIA)	45/59	76.3 (63.4–86.4)	36/59	61.0 (47.4–73.5)	86/101	85.1 (76.7–91.4)	71/101	70.3 (60.4–79.0)	17/24	70.8 (48.9–87.4)	12/24	50.0 (29.1–70.9)
Advanced stages (IIIB and IV) [*]	24/25	96.0 (79.6–99.9)	23/25	92.0 (74.0–99.0)	53/60	88.3 (77.4–95.2)	46/60	76.7 (64.0–86.6)	6/6	100.0 (54.0–100.0)	5/6	83.3 (35.9–99.6)
Sensitivity by SCLC stage					p=1.000		p=0.300		p=1.000		p=1.000	
Limited	3/3	100.0 (29.2–100.0)	2/3	66.7 (9.4–99.2)	6/6	100.0 (54.1–100.0)	6/6	100.0 (54.1–100.0)				
Extensive	7/7	100.0 (59.0–100.0)	7/7	100.0 (59.0–100.0)	9/9	100.0 (66.4–100.0)	8/9	88.9 (51.8–99.7)				

Data are presented as n/N or % [95% CI], unless otherwise stated. LCO: low cut-off; HCO: high cut-off; AUC: area under the curve; NSCLC: nonsmall cell lung cancer; SCLC: small cell lung cancer. In the cases of complete or quasi-complete separation, exact p-values were calculated. #: patients with adenocarcinoma histology were grouped with the squamous cell carcinoma histology subtype; ¶: tumour sizes were missing or unmeasurable for n=3 stage II, n=14 stage III and n=10 stage IV; *: stage III tumours without stage IIIA/IIIB classification were included in the advanced stages group; §: p-value calculation for the comparison of histological subtypes did not include the “unknown” group; ^f: p-value calculation for the comparison by tumour size did not include the “unknown” group.

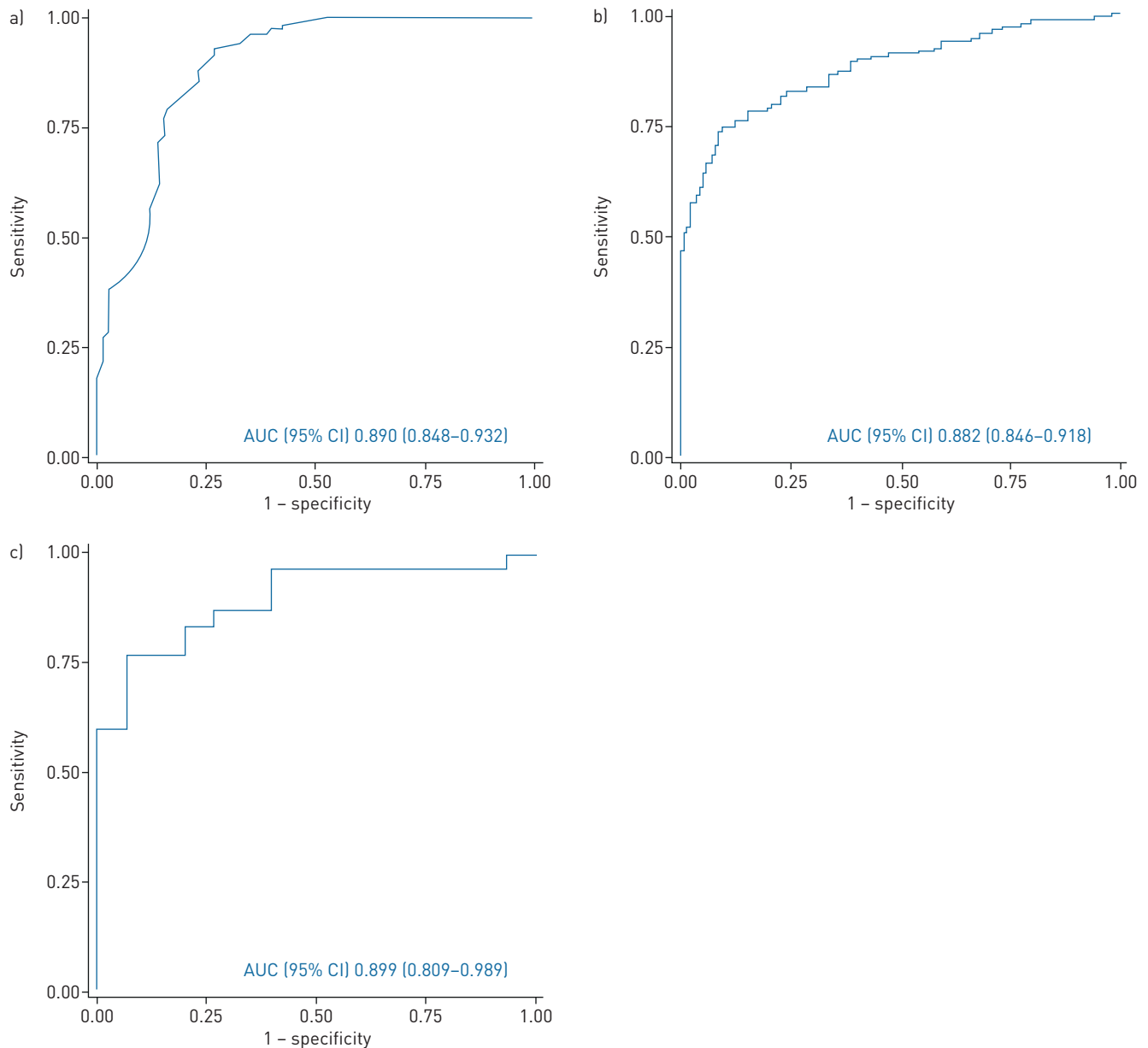


FIGURE 1 Receiver operating characteristic curves for a) training set; b) European validation set; c) Chinese validation set. AUC: area under the curve.

In a multivariable analysis of patients with smoking information ($n=242$), established risk factors for lung cancer (age, smoking status, pack-years and quit years) and sex did not influence having a positive Lung EpiCheck result in either cut-off (figure 2). Presence of COPD significantly decreased the chance of having a positive Lung EpiCheck result at LCO. A trend of higher EpiScores in patients without COPD *versus* patients with the condition was maintained when looking at various statistical measures of EpiScore (mean, median, 1st and 3rd quartile) of cases and in controls separately, but when combining the two groups, this trend was reversed (data not shown). The only factor driving a positive result was the group (case/control) with odds ratio (95% CI) of 18.2 (7.2-45.7), $p<0.0001$ with LCO and 23.7 (10.1-55.5), $p<0.0001$ with HCO. Likelihood ratios are reported in supplementary table S5.

A multivariable analysis was performed to assess the accuracy of lung cancer prediction based on risk factors alone, or in combination with EpiScore (figure 3). In our data, age, sex, smoking status, quit years, pack-years and COPD together yielded an AUC (95% CI) of 0.852 (0.805-0.900). Adding EpiScore significantly increased the AUC to 0.942 (0.913-0.971), $p<0.0001$. This analysis was performed on a subset of 242 patients with full smoking history, and the Lung EpiCheck AUC was similar to that of the entire

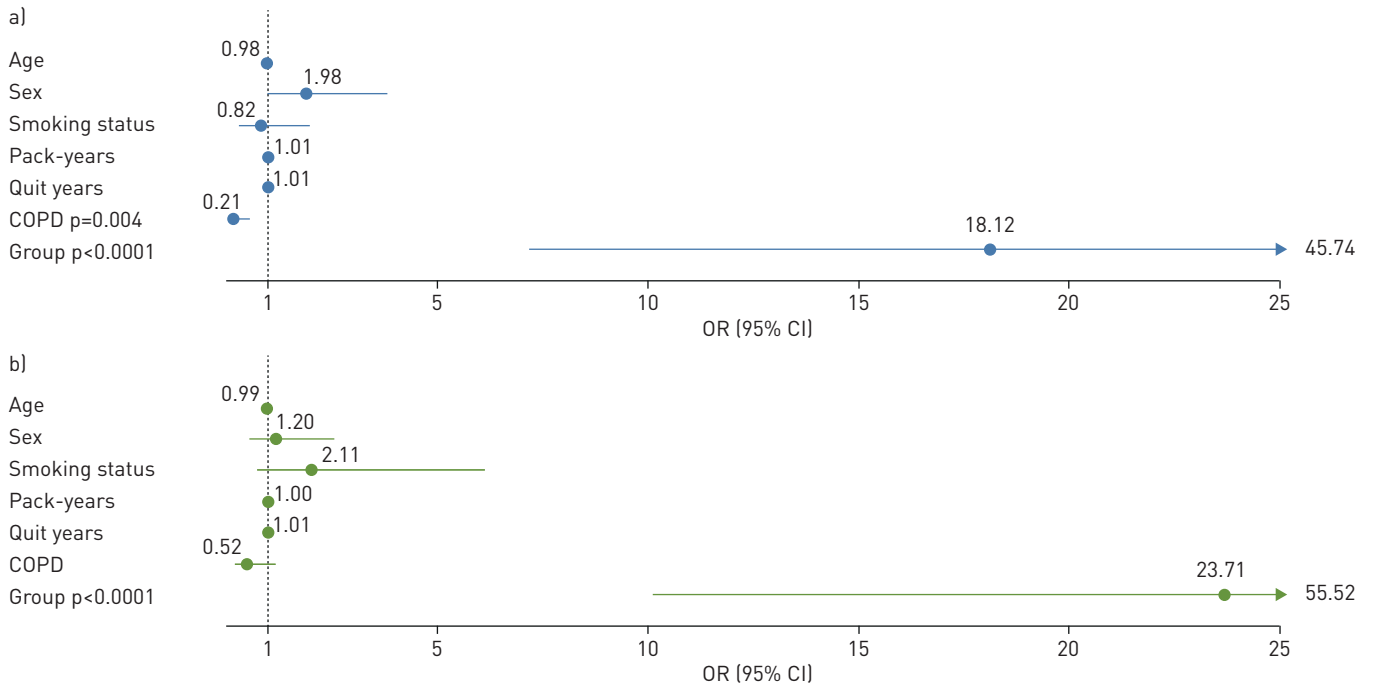


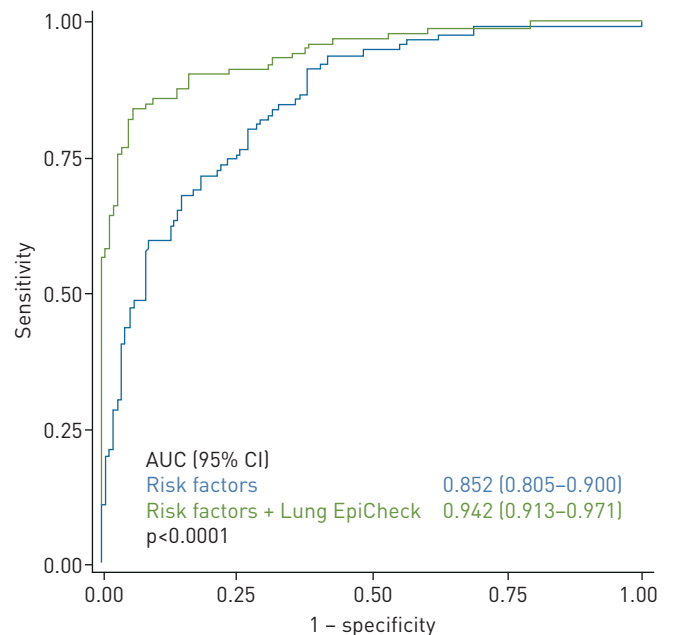
FIGURE 2 Multivariable logistic regression of factors potentially impacting Lung EpiCheck positive result, by cut-off: a) low cut-off EpiScore ≥ 60 ; b) high cut-off EpiScore ≥ 70 . This analysis included only patients with history of smoking and full smoking data, n=242 (n=106 cases, n=136 controls). Risk factors included age, pack-years and quit years as continuous measures; sex (female versus male), smoking status (former versus current smoker), COPD (yes versus no), and group (cases versus controls). For current smokers, quit years were counted as 0.

set (0.881 (0.843–0.918) versus 0.882 (0.846–0.918)), which suggests that this subsample is representative of the entire set; however, these results should be interpreted with caution.

10 (3.1%) of the tests in the validation sets failed to yield a result; two due to insufficient amount of DNA in the sample, and eight due to failed plasma separation.

The Chinese validation set enrolled 92 sequential cases and 15 controls. 41 cases were eligible, out of which 30 were selected to ensure good representation of all stages (n=10 stage I, n=6 stage II, n=11 stage III and n=3 stage IV) (supplementary figure S3b). As expected, age and female/male ratio were not

FIGURE 3 Multivariable logistic regression analysis of predictive factors for lung cancer. This analysis included only patients with history of smoking and full smoking data, n=242 (n=106 cases, n=136 controls). Risk factors: age, pack-years, quit years (continuous), sex (male/female), smoking status (current/past), COPD (yes/no). For current smokers, quit years were counted as 0. AUC: area under the curve.



comparable between cases and controls, as the latter were healthy volunteers (table 2). The AUC was 0.899 (0.809–0.989) (figure 1c), the sensitivity/specificity combinations were 76.7% (59.1–88.2%)/93.3% (68.0–99.9%), respectively, with LCO and 56.7% (39.2–72.6%)/100.0% (78.1–100.0%), respectively, with HCO (table 3). Lung EpiCheck with LCO/HCO detected 70.8% (48.9–87.4%)/50% (29.1–70.9%) of early-stage cancers, 70% (34.8–93.3%)/30% (6.7–65.2%) of stage I cancers and 60% (14.7–94.7%)/20% (0.5–71.6%) of stage I NSCLC \leq 20 mm. Likelihood ratios are reported in supplementary table S5.

No tests failed in the Chinese set.

Discussion

Our data demonstrate that Lung EpiCheck achieved performance characteristics suggesting that, if prospectively validated, may be suitable for clinical use in early detection of lung cancer. The predictive performance of Lung EpiCheck in the European validation data was very high, with AUC of 0.882. While maximising sensitivity (87.2%) with the LCO, the specificity remained good (64.2%), and while maximising specificity (90.5%) using the HCO, the sensitivity remained high (74.3%). In the Chinese set the AUC of 0.899 yielded high specificity in both cut-offs (LCO 93.3% and HCO 100.0%), with good sensitivity with LCO (76.7%). Differences between the validation sets are probably due to including young nonsmoking controls and surgical patients with small resectable tumours in the Chinese set. Detection at early stages is the key performance factor for Lung EpiCheck to ensure patients are detected in time for curative treatment. In the European set, the AUCs were consistently high in stage I NSCLC (0.797), stages I and II NSCLC (0.830) and early-stage (stages I, II and IIIA) NSCLC (0.862). In the validation sets, Lung EpiCheck detected 70–80% of stage I NSCLC with LCO, detecting tumours as small as 8 mm (adenocarcinoma in the Chinese set). With HCO, European results remained strong with stage I NSCLC sensitivity of 62.2%, but Chinese performance deteriorated to 30.0%. Early-stage performance should be interpreted with caution, as the controls were not scanned or followed-up to ensure no asymptomatic cancer existed. These results compare favourably with published results of other blood tests for lung cancer detection, reporting stage I sensitivity \sim 40% (22–71%), many of which are from training sets [13–17]. Sensitivity of NSCLC tumours was significantly impacted by size, but even in stage I NSCLC \leq 20 mm, there is a good signal of effectiveness detecting \geq 50% of tumours with LCO (four out of seven and two out of four in the European and Chinese validations sets, respectively). With HCO this stage I NSCLC \leq 20 mm sensitivity was similar in the European set (three out of seven), but all were missed in the Chinese set. Likelihood ratios can be used to simply and quickly estimate the post-test probability; however, there is no established gold-standard threshold for determining an acceptable likelihood ratio in the developing field of biomarkers for lung cancer screening. Regardless, we believe that the likelihood ratios achieved by Lung EpiCheck appear to be in a good range.

Both the Centers for Medicare & Medicaid Services [18] and the U.S. Preventive Services Task Force [19] recommend lung cancer screening with LDCT for high-risk populations, but national screening rates are very low, up to 14% [5]. Obstacles to lung cancer screening uptake are probably due to patient and primary care provider concerns about costs, inconvenience and possible risks associated with radiation and false-positive results. Additional limiting factors are absence of efficient proven programmes or lack of programme infrastructure. Offering a simple blood test to noncompliant eligible patients, as a tool to motivate them to get LDCT, could help overcome some of these barriers. Ease and safety of a blood test could encourage patients to be tested, and a positive blood test result could potentially convince patients to participate in LDCT screening programmes. If performance is confirmed in a prospective clinical study, prioritising eligible patients for LDCT based on such a test could alleviate systems restraints by reducing the number of unnecessary procedures and providing effective patient selection. Reducing the number of LDCTs could also indirectly impact on the number of false-positive findings, and their adverse outcomes and costs.

Alternatively, such a test can be used to better identify high-risk people who should undergo LDCT. Currently, high-risk populations are defined by demographic and exposure factors (age and smoking history) with very limited discrimination of AUC 0.6–0.7 [20]. Subsequently, in the USA, a mere 2.51 cancers are detected per 1000 LDCT scans [21] and $>$ 50% of lung cancer patients will not be considered eligible [7]. More elaborate risk models, such as PLCOm2012 (Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial), report better performance (AUC 0.7–0.8) as they include other surrogate markers such as COPD and history of cancer, but they are more cumbersome and harder to implement in the clinical routine. In our analysis, relationship of cases and positive Lung EpiCheck did not vary substantially by value of other risk factors (age, pack-years, quit years, sex, smoking status and COPD). Presence of these risk factors did not change or impact the strong relationship between lung cancer and the test result. Moreover, combining Lung EpiCheck with risk factors achieved a very high discrimination of 94.2%, allowing for optimal selection of high-risk populations for lung cancer screening. Further

validation is required to confirm Lung EpiCheck predictive performance and to define the best way to combine Lung EpiCheck and risk factors.

Similar to published evidence, showing correlation of cfDNA levels in the blood with tumour burden from NSCLC [22] and other solid tumours [23–26], Lung EpiCheck sensitivity correlated with tumour size. Two studies found cfDNA signal to inversely correlate with survival in patients with newly diagnosed lung cancer [27, 28], suggesting that it is also a prognostic marker for aggressiveness of the tumour. Further investigations are needed to inform whether there is a lower size limit of detectability by cfDNA, and if lack of cfDNA signal is an independent prognostic factor, or potentially a sign of overdiagnosis. Either way, a liquid biopsy for early detection must be very sensitive, in order to pick up the signal in the blood of early curable cancers. Lung EpiCheck's good preliminary results in early cancer classification can be explained by an analytical sensitivity of 1:200 000 [29] which is 20–200-fold higher than other liquid biopsy products available [30–32].

Cost-effectiveness is an important consideration in the applicability of screening tests. In a recently published model, in order to maintain the cost-effectiveness threshold of USD 50 000 per life-year gained, a marker added to risk factors to improve selection of patients for lung cancer screening could cost up to USD 300 [33]. Therefore, the next-generation sequencing based liquid biopsy tests, common in advanced settings, are irrelevant for this field, as their running costs alone are currently much higher at USD 1000–2000 per test. Similarly, other available lung cancer biomarker tests have sensitivity and specificity below screening requirements, lowering further the price level they can charge to be cost-effective. In contrast, Lung EpiCheck, with its high preliminary performance and its simple PCR-based technology, appear to be potentially well situated to be cost effective and commercially viable.

Mutations are established key factors of cancer development (*e.g.* driver mutations, acquired therapy resistance) as well as important targets for treatment, and could be potential markers for lung cancer detection. However, mutations in genes such as p53 [34] drive clonal haematopoiesis [35] in up to 21% of healthy elderly people. This can pose as a serious confounder and can generate false-positive results when using these mutations in blood tests for early detection of cancer. Alternatively, and unhampered by such problems, methylation changes have recently emerged as promising markers for cancer detection [16, 36].

Limitations

Case-control studies are prone to selection bias, as cases and controls do not necessarily come from the same population and are not truly comparable. This is reflected in our study by the significant difference between the groups in parameters such as age, sex, smoking status and pack-years. Cases were patients diagnosed with lung cancer due to any reason (symptoms, incidental finding, screening), and do not reflect a screening population. Controls did not receive LDCT screening, nor were they followed-up after blood draw; therefore, it is not known whether lung cancer cases were among them and were missed. Therefore, a prospective study in high-risk individuals undergoing LDCT with follow-up for lung cancer incidence is essential to confirm the current study findings. Staging was performed locally according to local standard of care in each site, so in the European validation there was a mix between AJCC7 and AJCC8. This probably translates to a potential overlap between large stage I to small stage II cancers. Collection of some data were limited in biobank samples, leading to missing smoking histories in 19% of the European validation set, limiting the multivariable analyses to a subpopulation of that set. Personal history of cancer is a known risk factor for lung cancer [37]; however, to ensure that the signal emerges from lung cancer, such patients were excluded. The Chinese set was a small single-centre study that included surgical patients only, therefore not representative of the Chinese lung cancer population; additionally, the controls were young, healthy and mostly nonsmokers, not representative of patients at risk. A larger prospective study in China is warranted to confirm the performance of Lung EpiCheck in this population.

Our current findings need to be validated in prospective trials.

Conclusions

Lung EpiCheck demonstrated strong suggestive performance in lung cancer prediction in case-control European and Chinese samples, detecting up to 78% of stage I tumours, up to 100% of SCLC and significantly improving predictive accuracy when added to established risk factors. Prospective studies are required to confirm these findings. Utilising such a simple and inexpensive blood test to select people for lung cancer screening has the potential to improve compliance and broaden access to screening for high-risk populations.

Acknowledgements: The results published here are in whole or part based upon data generated by the Cancer Genome Atlas managed by the National Cancer Institute and National Human Genome Research Institute. Information about the Cancer Genome Atlas can be found at <http://cancergenome.nih.gov>.

Support statement: This study was funded by Nucleix Ltd (Rehovot, Israel), and was partly supported by the grant from the National Key Research and Development Plan of China (number 2016YFE0103400). This publication is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme, under grant agreement number 829218. Funding information for this article has been deposited with the Crossref Funder Registry.

Conflict of interest: M. Gaga has nothing to disclose. J. Chorostowska-Wynimko reports grants, personal fees and non-financial support from Grifols, AstraZeneca, Pfizer, CSL Behring and CelonPharma, grants and personal fees from Boehringer Ingelheim, personal fees and non-financial support from MSD and BMS, personal fees from Amgen, GSK, Novartis, Chiesi, Roche and Lekam, outside the submitted work. I. Horváth reports personal fees from AstraZeneca, Novartis, CSL Behring, Boehringer Ingelheim, GSK and Berlin-Chemie, outside the submitted work. M.C. Tammemagi has served as consultant to Johnson & Johnson/Janssen, Medial EarlySign, Nucleix, bioAffinity Technologies and AstraZeneca. D. Shitrit has nothing to disclose. V.H. Eisenberg has nothing to disclose. H. Liang has nothing to disclose. D. Stav has nothing to disclose. D. Levy Faber has nothing to disclose. M. Jansen has nothing to disclose. Y. Raviv has nothing to disclose. V. Panagoulas has nothing to disclose. P. Rudzinski has nothing to disclose. G. Izbicki has nothing to disclose. O. Ronen has nothing to disclose. A. Goldhaber has nothing to disclose. R. Moalem has nothing to disclose. N. Arber has nothing to disclose. I. Haas has nothing to disclose. Q. Zhou has nothing to disclose.

References

- 1 International Agency for Research on Cancer. GLOBOCAN2018 Global Cancer Observatory. <https://gco.iarc.fr>. Last accessed July 2020.
- 2 National Cancer Institute Surveillance, Epidemiology, and End Results Program. Lung Cancer Stat Facts: Lung and Bronchus Cancer. www.seer.cancer.gov/statfacts/html/lungb.html Date last accessed: December 2018.
- 3 National Lung Screening Trial Research Team, Aberle DR, Adams AM, *et al*. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; 365: 395–409.
- 4 de Koning HJ, van der Aalst CM, de Jong PA, *et al*. Reduced lung-cancer mortality with volume CT screening in a randomized trial. *N Engl J Med* 2020; 382: 503–513.
- 5 Zahnd WE, Eberth JM. Lung cancer screening utilization: a behavioral risk factor surveillance system analysis. *Am J Prev Med* 2019; 57: 250–255.
- 6 Pinsky PF, Berg CD. Applying the National Lung Screening Trial eligibility criteria to the US population: what percent of the population and of incident lung cancers would be covered? *J Med Screen* 2012; 19: 154–156.
- 7 Rijavec E, Coco S, Genova C, *et al*. Liquid biopsy in non-small cell lung cancer: highlights and challenges. *Cancers* 2019; 12: 17.
- 8 Hanash SM, Ostrin EJ, Fahrman JF. Blood based biomarkers beyond genomics for lung cancer screening. *Transl Lung Cancer Res* 2018; 7: 327–335.
- 9 Li L, Fu K, Zhou W, *et al*. Applying circulating tumor DNA methylation in the diagnosis of lung cancer. *Precis Clin Med* 2019; 2: 45–56.
- 10 Witjes JA, Morote J, Cornel EB, *et al*. Performance of the Bladder EpiCheck™ methylation test for patients under surveillance for non-muscle-invasive bladder cancer: results of a multicenter, prospective, blinded clinical trial. *Eur Urol Oncol* 2018; 1: 307–313.
- 11 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: 244–285.
- 12 American Cancer Society. Cancer Facts & Figures 2019. Atlanta, American Cancer Society, 2019. www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2019/cancer-facts-and-figures-2019.pdf
- 13 Liu MC, Jamshidi A, Venn O, *et al*. Genome-wide cell-free DNA (cfDNA) methylation signatures and effect on tissue of origin (TOO) performance. *J Clin Oncol* 2019; 37: 3049.
- 14 Weiss G, Schlegel A, Kottwitz D, *et al*. Validation of the SHOX2/PTGER4 DNA methylation marker panel for plasma-based discrimination between patients with malignant and nonmalignant lung disease. *J Thorac Oncol* 2017; 12: 77–84.
- 15 Sullivan FM, Mair FS, Anderson W, *et al*. Earlier diagnosis of lung cancer in a randomised trial of an autoantibody blood test followed by imaging. *Eur Respir J* 2020; in press doi.org/10.1183/13993003.00670.
- 16 Doseeva V, Colpitts T, Gao G, *et al*. Performance of a multiplexed dual analyte immunoassay for the early detection of non-small cell lung cancer. *J Transl Med* 2015; 13: 55.
- 17 Cohen JD, Li L, Wang Y, *et al*. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* 2018; 359: 926–930.
- 18 Centers for Medicare & Medicaid Services. LungCancer Screenings. www.medicare.gov/coverage/lung-cancer-screenings. Date last accessed July 2020.
- 19 U.S. Preventive Services Task Force. Lung Cancer Screening. Final Recommendation Statement. www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/lung-cancer-screening. Date last accessed July 2020.
- 20 Ten Haaf K, Jeon J, Tammemägi MC, *et al*. Risk prediction models for selection of lung cancer screening candidates: a retrospective validation study. *PLoS Med* 2017; 14: e1002277.
- 21 National Radiology Data Registry. Facility Report, Dec 2019. Public Facility (Facility ID: 100853). Available from: <http://nrdrsupport.acr.org/support/solutions/articles/11000039783-lcsr-available-reports> Date last accessed: September 23, 2020.
- 22 Mondaca S, Offin M, Borsu L, *et al*. Lessons learned from routine, targeted assessment of liquid biopsies for EGFR T790M resistance mutation in patients with EGFR mutant lung cancers. *Acta Oncol* 2019; 58: 1634–1639.
- 23 Valpione S, Gremel G, Mundra P, *et al*. Plasma total cell-free DNA (cfDNA) is a surrogate biomarker for tumour burden and a prognostic biomarker for survival in metastatic melanoma patients. *Eur J Cancer* 2018; 88: 1–9.
- 24 Strijker M, Soer EC, de Pastena M, *et al*. Circulating tumor DNA quantity is related to tumor volume and both predict survival in metastatic pancreatic ductal adenocarcinoma. *Int J Cancer* 2020; 146: 1445–1456.

- 25 Egyud M, Tejani M, Pennathur A, *et al.* Detection of circulating tumor DNA in plasma: a potential biomarker for esophageal adenocarcinoma. *Ann Thorac Surg* 2019; 108: 343–349.
- 26 Hamfjord J, Guren TK, Dajani O, *et al.* Total circulating cell-free DNA as a prognostic biomarker in metastatic colorectal cancer before first-line oxaliplatin-based chemotherapy. *Ann Oncol* 2019; 30: 1088–1095.
- 27 Chabon JJ, Hamilton EG, Kurtz DM, *et al.* Integrating genomic features for non-invasive early lung cancer detection. *Nature* 2020; 580: 245–251.
- 28 Oxnard GR, Chen X, Fung ET, *et al.* Prognostic significance of blood-based cancer detection in plasma cell-free DNA (cfDNA): evaluating risk of overdiagnosis. *J Clin Oncol* 2019; 37: 1545.
- 29 Frumkin D, Shuali A, Savin O, *et al.* A new ultrasensitive assay for detection of hypermethylated tumor DNA in liquid biopsies. 11th CNAPS (Circulating Nucleic Acids in Plasma and Serum) International Symposium. Poster P-21.
- 30 Zou H, Allawi H, Cao X, *et al.* Quantification of methylated markers with a multiplex methylation-specific technology. *Clin Chem* 2012; 58: 375–383.
- 31 Cottrell SE, Distler J, Goodman NS, *et al.* A real-time PCR assay for DNA-methylation using methylation-specific blockers. *Nucleic Acids Res* 2004; 32: e10.
- 32 Kristensen LS, Hansen LL. PCR-based methods for detecting single-locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment. *Clin Chem* 2009; 55: 1471–1483.
- 33 Larose TL, Meheus F, Brennan P, *et al.* Assessment of biomarker testing for lung cancer screening eligibility. *JAMA Netw Open* 2020; 3: e200409.
- 34 Chen S, Wang Q, Yu H, *et al.* Mutant p53 drives clonal hematopoiesis through modulating epigenetic pathway. *Nat Commun* 2019; 10: 5649.
- 35 Shlush LI. Age-related clonal hematopoiesis. *Blood* 2018; 131: 496–504.
- 36 Parikh AR, Van Seventer EE, Boland GM, *et al.* A plasma-only integrated genomic and epigenomic circulating tumor DNA (ctDNA) assay to inform recurrence risk in colorectal cancer (CRC). *J Clin Oncol*; 37: 3602.
- 37 Tammemägi MC, Katki HA, Hocking WG, *et al.* Selection criteria for lung-cancer screening. *N Engl J Med* 2013; 368: 728–736.