



Research article

Telomere-based risk models for the early diagnosis of lung cancer

Sonia Molina-Pinelo ^{a,b,c}, Irene Ferrer Sánchez ^{c,d}, Pilar Najarro ^e, Luis Paz-Ares ^{c,f}, Luis Fernández ^e, Nila Castelló ^e, Luis Alberto Richart López ^g, Juan Diego Rodríguez Gambarte ^g, Máximo Sanz García ^h, Ana Salinas ^{a,b}, Rocío Suárez ^{c,d}, Beatriz Romero-Romero ⁱ, José Martín-Juan ⁱ, María Eugenia Viñuela ⁱ, Ray G. Butler ^j, Nuria de Pedro ^{e,*}

^a Department of Medical Oncology, University Hospital Virgen del Rocío, Sevilla, Spain

^b Institute of Biomedicine de Seville (CSIC), University of Seville, Seville, Spain

^c Spanish Center for Biomedical Research Network in Oncology (CIBERONC), Madrid, Spain

^d H120-CNIO Lung Cancer Clinical Research Unit, Health Research Institute Hospital 12 de Octubre (i+12)/Spanish National Cancer Research Center (CNIO), Madrid, Spain

^e Life Length SL, Madrid, Spain

^f Department of Medical Oncology, Hospital Universitario 12 de octubre, Madrid, Spain

^g Centro de Transfusión de la Comunidad de Madrid, Madrid, Spain

^h Department of Anaesthesiology, Hospital Universitario Príncipe de Asturias, Alcalá de Henares, Spain

ⁱ Department of Pneumology, University Hospital Virgen del Rocío, Seville, Spain

^j Butler Scientifics, Barcelona, Spain

ARTICLE INFO

Keywords:

Telomere length
Lung cancer
Cancer risk
Cancer screening
Biomarker

ABSTRACT

Background: The objective of this study was to evaluate the use of telomere length measurements as diagnostic biomarkers during early screening for lung cancer in high-risk patients.

Methods: This was a prospective study of patients undergoing lung cancer diagnosis at two Spanish hospitals between April 2017 and January 2020. Telomeres from peripheral blood lymphocytes were analysed by Telomere Analysis Technology, which is based in high-throughput quantitative fluorescent in situ hybridization. Analytical predictive models were developed using Random Forest from the dataset of telomere-associated variables (TAV). Receiver Operating Characteristic curves were used to characterize model performance.

Findings: From 233 patients undergoing lung cancer diagnosis, 106 patients aged 55–75 with lung cancer or lung cancer and COPD were selected. A control group (N = 453) included individuals of similar age with COPD or healthy. Telomere analysis showed that patients in the cancer cohort had a higher proportion of short telomeres compared to the control cohort. A TAV-based predictive model assuming a prevalence of 5 % of lung cancer among screened subjects showed an AUC of 0.98 %, a positive predictive value of 0.60 (95 % CI, 0.49–0.70) and a negative predictive value of 0.99 (95 % CI, 0.98–0.99) for prediction of lung cancer.

Interpretation: The results of this study suggest that TAV analysis in peripheral lymphocytes can be considered a useful diagnostic tool during screening for lung cancer in high-risk patients. TAV-based models could improve the predictive power of current initial diagnostic pathways, but further work is needed to integrate them into routine clinical evaluation.

Funding: Life Length SL.

* Corresponding author.

E-mail address: ndepedro@lifelength.com (N. de Pedro).

1. Research in context

1.1. Evidence before this study

We searched PubMed for reviews, observational studies, clinical trials, and cohort studies in the use of telomeres as biomarkers for diagnosis and prognosis of lung cancer up to October 1st, 2022, using the terms “telomere” or “telomere length” in combination with each of the following terms: “lung cancer”, “NSCLC”, or “chronic obstructive pulmonary disease”. After reviewing the abstracts, we identified 9 articles directly related to the study of telomere length and lung cancer. None of the studies used high-throughput quantitative fluorescent in situ hybridization to assess telomere length. Other studies reviewing evidence of telomere length and cancer were also considered. A total of 20 reviews were found on this topic of telomere biology and cancer or lung disease.

1.2. Added value of this study

In this study investigating the use of highly accurate and sensitive techniques for the analysis of telomeres, we found that patients in the cancer cohort had a higher proportion of short telomeres compared to the control cohort, suggesting that technology could be considered a useful diagnostic tool during early screening for lung cancer in high-risk patients.

1.3. Implications of all the available evidence

Based on the current evidence, the study of telomeres in peripheral blood cells should be further explored in clinical studies to define its potential use in diagnostic pathways of lung cancer. Telomere evaluation in blood is an attractive potential biomarker as it is non-invasive, inexpensive, and quantifiable. This is especially relevant for the screening of high-risk patients, such as smokers or those with chronic obstructive pulmonary disease, as it has been estimated that underscreening in this population leads to increased cancer-related mortality.

2. Introduction

Lung cancer is the leading cause of cancer death among both men and women worldwide, making up almost 25 % of all cancer deaths [1]. In Spain it was estimated that 21,918 patients died from lung cancer in 2020, and 30,948 new cases were diagnosed in 2022 [2]. Although the overall 5-year survival rate of lung cancer patients in the US was only 20.5 % in 2020, if the cancer is localized at diagnosis the 5-year survival rate raises to 59.0 % [3]. For this reason, early detection through screening programs has been strongly recommended for asymptomatic patients at high risk for lung cancer, such as those aged 55–80 years who have a smoking history of 30 packs or more per year and currently smoke or have quit within the past 15 years. Patients with chronic obstructive pulmonary disease (COPD) and smoking habit are at higher risk of lung cancer [4]. It has been demonstrated that early screening by low-dose computed tomography (LDCT), as compared to chest X-rays, can strongly reduce mortality among these high-risk patients [5]. However, LCDT screening remains low and it has been estimated that underscreening could be related to approximately 12,000 lung cancer-related deaths per year in the US alone [6].

Telomeres, the nucleoprotein assemblages that protect the chromosome ends, have emerged in recent years as potential biomarkers for risk evaluation in a variety of cancers [7–14]. Shorter telomeres have been associated to increased mortality in some cancer types [7,11,15,16]. However, studies of telomere length variation and lung cancer have been ambiguous, as increased risk or increased mortality have been found associated with both short [17–19], and long telomeres [13,20–24]. It is likely that biases in the population studied, in the study design, or in the accuracy and sensitivity of the techniques used to measure telomere length, could be obscuring the relationship between telomere length and lung cancer risk. Although telomere evaluation in peripheral blood lymphocytes is an attractive potential biomarker, as it is non-invasive, inexpensive, and quantifiable, high accuracy and standardization of measurements are essential for it to be useful [25].

Of the various technologies developed to measure telomere length, high-throughput quantitative fluorescent in situ hybridization (HT Q-FISH) has been demonstrated to generate highly accurate and sensitive measurements of telomere length [26,27]. Recently an extension of this technique which includes image capture and processing into an integrated Telomere Analysis Technology (TAT®) was validated [28], and a study showed that TAT could be used to create predictive models for the early diagnosis of prostate cancer [29]. One of the advantages of TAT is that it can evaluate many telomere-associated variables (TAV) in the cell population, thus enriching the interpretation of the telomeric status of the sample. Among other parameters, TAV can describe the full distribution of telomere lengths, the proportions of short and long telomeres, or the percentage of cells with a specific telomere length average. The combination of these values generates unique profiles, or ‘TAV signatures’, which can be used to define diagnostic patterns and prognostic use [28].

Here we use TAT with the main objective of evaluating the possibility of using telomere length measurements in peripheral blood lymphocytes as potential diagnostic biomarkers in patients at high risk of lung cancer. To do so, we prospectively identified high risk patients that were in the process of lung cancer diagnosis and analysed their telomeres by TAT. Based on the results, we developed analytical predictive models to evaluate patients from a high-risk population with an increased risk of developing lung cancer.

3. Materials and methods

The collection of blood samples from high-risk patients in the process of being diagnosed for possible lung cancer was conducted prospectively at the University Hospital Virgen del Rocío (Seville, Spain) and the University Hospital 12 de Octubre (Madrid, Spain) from April 2017 to January 2020. The samples from healthy subjects that were used as controls were collected at the Centre for blood transfusion of the Autonomous Community of Madrid (Madrid, Spain) from subjects aged >18 and < 65 years, and at the Hospital Príncipe de Asturias Hospital (Alcalá de Henares, Spain) from subjects aged >65 years. All samples were analysed in the laboratories of Life Length (Madrid, Spain). This study was approved by the Ethics Committee of the CEIC Hospital Universitario 12 de Octubre (code 17/026). The researchers informed each patient about the study to, its objectives, methods, duration scheme, expected benefits, discomforts and possible risks. Each patient were given an information sheet and provided written informed consent. This study was carried out in accordance with the Standards of Good Clinical Practice (GCP), the Declaration of Helsinki, and the Spanish Law 14/2007 on Biomedical Research.

3.1. Patients and samples

The inclusion criteria for the patients at high risk of lung cancer were: age above 40 years; who smoked more than 400 cigarettes (20 packs) per year; presented nodules after x-ray radiography and/or haemoptysis; were having a bronchoalveolar lavage (BAL) and blood extraction performed following standard clinical practice for diagnosis; and gave informed consent to participate. After clinical diagnosis of the subjects with possible lung cancer, they were divided into 4 groups: patients with lung cancer; patients with COPD; patients with lung cancer and COPD, and patients without lung cancer or COPD.

The exclusion criteria for participation in this study for the patients at high risk of lung cancer were previous diagnosis of lung cancer or any other type of oncological disease (except for basal cell carcinoma); active pulmonary tuberculosis; previous lung surgery; presence of any chronic or acute inflammatory disease (except for COPD); and presence of risks derived from a blood extraction or the performance of a BAL.

For the healthy subjects in the control cohort, the only exclusion criterion was not being diagnosed for any chronic disease.

Tobacco use by all subjects was segmented into 4 categories from no smokers (never smoke) to heavy/high smokers (>30 pack-year smoking history). An electronic notebook was used to register clinical data associated with the subjects, including age and smoking habits. For patients in hospitals, clinical data (spirometry, Eastern Cooperative Oncology Group [ECOG] status, family history of lung cancer), and final diagnosis of lung cancer and/or COPD was also collected.

3.2. Telomere analysis

The study of the telomeres from blood samples was carried out in the laboratories of Life Length SL (Madrid, Spain) within the scope of CLIA (99D2112462) and ISO 15189 quality standards. Telomere analyses were performed blindly with respect to all clinical data and the final diagnosis of the patient. The telomeres were analysed by Telomere Analysis Technology (TAT®), a novel, laboratory-developed, validated, high-throughput platform for the comprehensive study of telomere variables based on high-throughput quantitative fluorescence in situ hybridization (HT-Q-FISH) [26,28]. In short, peripheral blood lymphocytes were collected from the blood and fixed in 384 well plates. Individual telomeres in the cell nucleus were visualized by hybridization with a fluorescent Peptide Nucleic Acid probe (PNA) that recognizes telomere repeats (sequence: Alexa488-OO-CCCTAACCTAACCTAA, Panagene Inc, South Korea). Quantitative image acquisition was then performed, and the fluorescent intensities translated to base pair through a standard regression curve which is generated using control cell lines with known telomere lengths. The data generated a TAV profile with descriptive statistics of telomere length, values for each percentile of telomere length, percentages of telomere length values (ShortTel), percentages of cells with specific telomere values (ShortCell) and dispersion parameters for each sample. A summary list of TAV is presented in [Supplementary Table 1](#).

3.3. Data analysis and model generation

The software KNIME (Zürich, Switzerland) was used for data processing [30]. Numeric variables were normalized using Z-score for model generation and principal component analysis (PCA). Random Forest was used to generate the models for the full TAV set, also using a PCA approach. The distribution of the classifications of samples for each set was forced to resemble the distribution of the classifications of samples observed in the total set (both total and validation set had similar percentage of cancer cases), but the exact samples for each classification on each set were selected at random among those from said classification. There were no set thresholds pre-defined for evaluation of the models and, by default, it was 0.5.

Initial performance of each model was determined assessing firstly the Kappa of Cohen and the differential (Δ K Cohen) between the lower and the highest K-Cohen obtain for each model set analysed. Likewise, the differential accuracy (Δ accuracy) was calculated for each model set analysed. The differential values obtained in all cases were low (mostly below 0.05), which underscored the fitness of the models for each analysis.

The selected TAV set used was obtained through a guided automated exploratory data analysis, using AutoDiscovery® methodology (Butler Scientifics, S.L.; Barcelona, Spain) followed by an expert clinician screening to study clinically valuable associations in order to select the best predictor variables. The AutoDiscovery® methodology included multiple pairwise statistical comparisons, effect size analysis (strength), and the Benjamini-Hochberg procedure for false-discovery rate calculation.

Receiver Operating Characteristic (ROC) curves were used to characterize the performance of the models.

Role of the funding source

The study sponsor, Life Length SL, was fully involved in study design, data collection, analysis, and interpretation. It also participated in the writing of the report and in the decision to submit the paper for publication. Full details of funding sources can be found in 'Funding'.

4. Results

The participating hospitals initially identified 233 patients who were undergoing lung cancer screens (Supplementary Fig. 1). Of these, 52 patients were excluded due to lack of biopsy data or not meeting clinical criteria. After the diagnosis process a total of 181 subjects were distributed into four groups: patients with lung cancer (67 patients, 36 %), patients with COPD (11 patients, 6 %), patients with lung cancer and COPD (81 patients, 42 %) and patients with no COPD or cancer (22 patients, 11 %). Healthy subjects were selected in an independent process from two ongoing studies and included blood donors from two groups aged 40–65 years and >65 years. The 'cancer cohort' (N = 148) included patients with lung cancer and with lung cancer and COPD, while the 'control cohort' (N = 1378) included patients presenting COPD, no COPD/lung cancer, and the healthy subjects (Supplementary Fig. 1). The main characteristics of the patients in both cohorts are shown in Table 1. The median (IQR) ages of the patients were 67 (59–72) years and 53 (46–68) years in the cancer and in the control cohorts, respectively. For the rest of the study, patients aged 55–75 were selected (Table 1).

4.1. Telomere length analysis

Blood samples from all patients in both cohorts were processed by TAT to analyse telomere length from peripheral lymphocytes. Telomere length was analysed for the four groups of patients aged 55–75 under study, and the results of three variables are shown in Fig. 1. The results show that patients with lung cancer and COPD present the longest telomere lengths in all percentiles and patients with COPD the shortest (Fig. 1A). Significant differences are found between patients with COPD and patients without lung cancer or COPD in percentiles >40. Additionally, patients with lung cancer and COPD and patients with lung cancer show the highest percentage of short telomere length (<7,000 bp) and lowest percentage of long telomere length (>14,000 bp) (Fig. 1B). The profile in patients with no lung cancer or COPD was different to the rest of the patients. Finally, patients with lung cancer and patients with lung cancer and COPD present the highest percentage of cells with an average telomere length of <7,000 bp, and lowest percentage of cells with an average telomere length of >14,000 bp (Fig. 1C). The profile in patients with no lung cancer and no COPD was different to the rest of the patients, having a smaller percentage of cells with short telomeres.

Since the some of the groups were too small to allow further analysis, we grouped the patients in two cohorts: patients with lung cancer (including those with cancer only and those with cancer and COPD, N = 106) and patients without cancer (patients with COPD only and healthy subjects, N = 453). The distribution of telomere lengths revealed that the patients in the two cohorts presented distinct profiles (Fig. 2). Patients in the cancer cohort presented a lower proportion of long telomeres compared with the control cohort (Fig. 2A). Similarly, patients in the cancer cohort generated distributions of the telomeres with a given length (Fig. 2B) or of cells with telomeres with a given length (Fig. 2C) that was significantly different from that of the control cohort. As there was a higher proportion of males versus females in both cohorts (Table 1), a complementary analysis was conducted to evaluate if there were differences in

Table 1
Patient characteristics.

	N	Age, years, median (IQR)	Male %	Smoking habits ^a , N (%)				
				NA	No	Low	Moderate	Heavy
Total								
Cancer cohort	148	67 (59–72)	81.8	9 (6.1)	0	8 (5.4)	32 (21.6)	99 (66.9)
Control cohort	1381	53 (46–68)	66.8	8 (0.6)	1128 (81.7)	115 (8.3)	119 (8.6)	11 (0.8)
All patients								
LC	67	64 (57–71)	76.1	6 (9.0)	0	5 (7.5)	16 (23.9)	40 (59.7)
LC + COPD	81	69 (62–75)	86.4	3 (3.7)	0	3 (3.7)	16 (19.8)	59 (72.8)
COPD	11	66 (58–71)	72.7	3 (27.3)	0	2 (18.2)	1 (9.1)	5 (45.5)
No LC diagnosis	22	59 (54–72)	77.3	5 (22.7)	0	4 (18.2)	7 (31.8)	6 (27.3)
Healthy	1345	53 (46–68)	66.6	0	1128 (83.9)	107 (8.0)	110 (8.2)	0
Patients aged 55–75								
LC	44	65 (60–70)	77.3	3 (6.8)	0	4 (9.1)	8 (18.2)	29 (65.9)
LC + COPD	62	65 (58–71)	85.5	1 (1.6)	0	1 (1.6)	10 (16.1)	50 (80.6)
COPD	10	66.5 (60–70)	70.0	3 (30.0)	0	1 (10.0)	1 (10.0)	5 (50.0)
No LC diagnosis	9	58 (56–61)	77.7	2 (22.2)	0	2 (22.2)	3 (33.3)	2 (22.2)
Healthy	434	64 (58–70)	60.1	0	362 (83.4)	33 (7.6)	39 (9.0)	0

Abbreviations: COPD, chronic obstructive pulmonary disease; IQR, interquartile range; LC, lung cancer; NA, not available; SD, standard deviation.

^a No = never smoke; low = ≤20 cigarette packs/year; moderate = 20–40 packs/year; heavy = ≥40 packs/year.

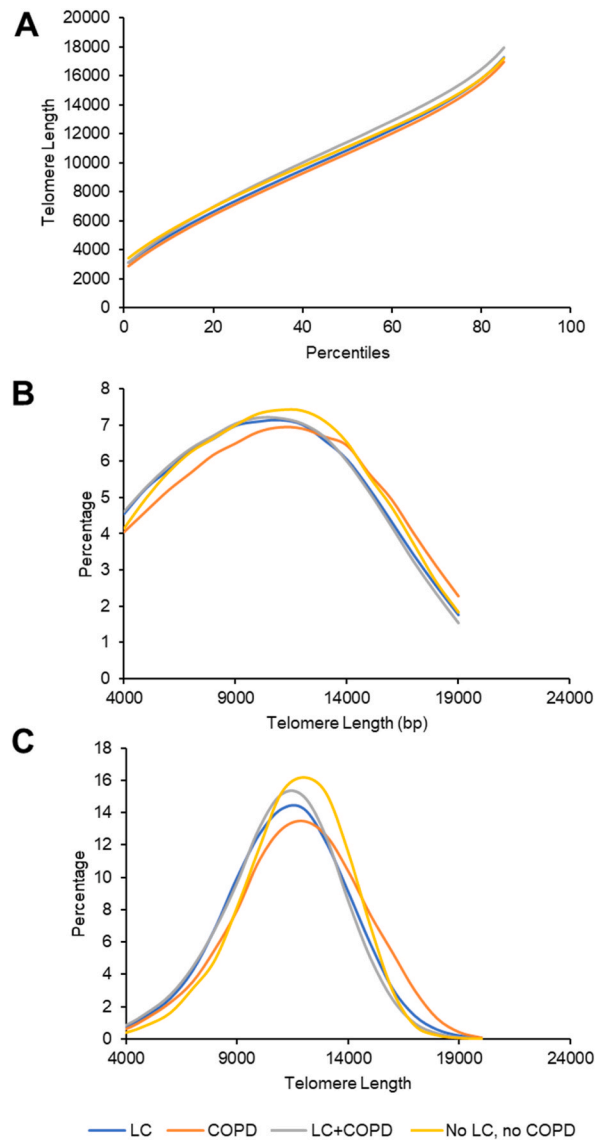


Fig. 1. Distribution of telomere lengths in patients with lung cancer (N = 44), patients with COPD (N = 10), patients with lung cancer and COPD (N = 62), and patients without lung cancer or COPD (N = 9). A, percentiles of telomere length. B, percentage of telomeres of a sample with a telomere length under a given threshold (Short Tel). C, percentage of cells with a specific average telomere length. Abbreviations: COPD, chronic obstructive pulmonary disease; LC, lung cancer.

telomere lengths according to sex (Supplementary Figs. S2–S4). No differences between sexes could be found ($p > 0.05$). Further, the observed effects of lung cancer on telomere length were comparable for both sexes.

Discreet TAV that analyse the differences between the 1st and the 99th percentiles of the telomere lengths of a sample (P1-99), or the interquartile of the telomere lengths of a sample (p25-75) also showed significant differences in the cancer versus control cohorts ($p < 0.0001$ in both cases) (Fig. 3A and B). Thus, TAT analysis indicated that in the cancer cohort shorter telomeres were over-represented and longer telomeres underrepresented if compared with the control cohort. Taken together, these results suggest that telomere length measurements and TAV analysis could be used as potential biomarkers to distinguish between patients with or without lung cancer.

4.2. Model construction and validation

TAV-based models were developed, and their performance was evaluated by ROC probability curves. Two different approaches were used for the analysis: one that included an age- and gender-matched group of cancer and control patients in an approximately 1:1 proportion (lung cancer prevalence of 48 %), and one in which the control cohort subjects were overrepresented (lung cancer

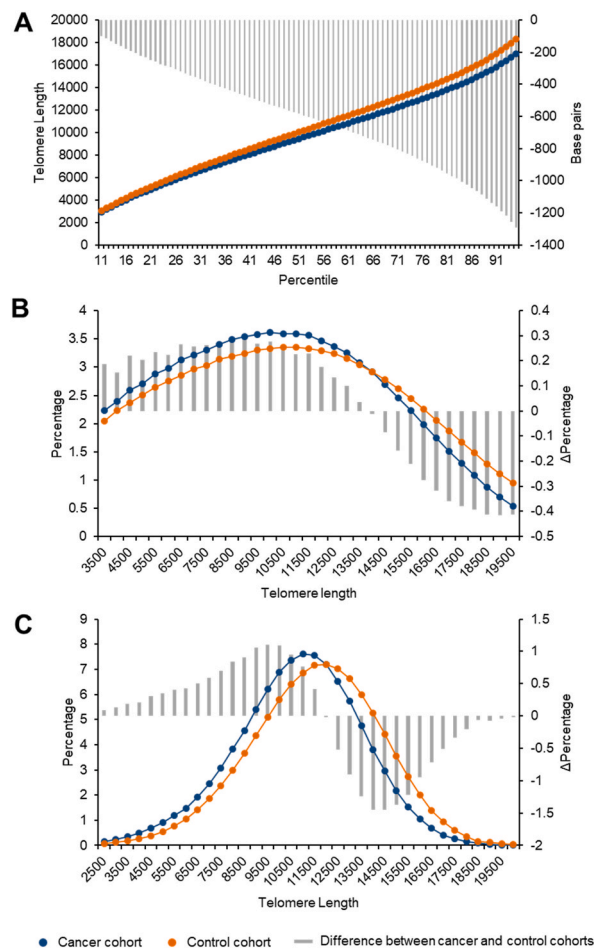


Fig. 2. Distribution of telomere lengths in patient samples of the cancer ($N = 106$) and control ($N = 453$) cohorts. A, percentiles of telomere length. B, percentage of telomeres of a sample with a telomere length under a given threshold (Short Tel). C, percentage of cells of a sample whose telomeres average a length under a given threshold (Short Cell). The differences between the curves are indicated by the grey bars and quantified in the right axis.

prevalence of 5 %). The results showed that in the model with a lung cancer prevalence of 48 %, the area under the curve (AUC) was of 0.79 % (Fig. 4A). This model demonstrated a positive predictive value (PPV) of 0.79 (95 % CI, 0.65–0.88) and a negative predictive value (NPV) of 0.83 (95 % CI, 0.63–0.94) for prediction of lung cancer (Table 2). When the model was generated with a lung cancer prevalence of 5 %, the AUC was 0.98 % (Fig. 4B) and in this case the PPV = 0.60 (95 % CI, 0.49–0.70) and the NPV = 0.99 (95 % CI, 0.98–0.99) for prediction of lung cancer (Table 2).

5. Discussion

In this study we evaluated the use of telomere length measurements derived from peripheral blood lymphocytes as potential diagnostic biomarkers in lung cancer. We showed that a cohort of patients diagnosed with lung cancer presented generally shorter telomeres compared with the control cohort without cancer (including patients at high risk without cancer and healthy individuals). This was reflected in a lower proportion of long telomeres, in the overall distribution of telomere lengths, and in the overrepresentation of short telomeres over long telomeres in these patients. Based on these results, TAV-based predictive models were developed which could be useful during the screening and diagnosis process of high-risk patients.

In our study shorter telomeres in the population of patients with lung cancer, compared with those of the control population, was observed. An association between abnormally shorter telomeres and increased risk or increased mortality in lung cancer had been reported [17,18,31]. In a large prospective study of heavy smokers who developed lung cancer, short telomere length determined before diagnosis was associated with increased risk of death in patients with small cell lung cancer (SCLC) [19]. A random-effects meta-analysis showed that patients with lung cancer had shorter telomere length than controls, with an odds ratio of 1.13 (95 % CI: 0.82–1.81, $p = 0.46$) [32]. In contrast, some genetic association and prospective cohort studies have shown an association between longer telomeres and adenocarcinoma histology in lung cancer [23,24,33–35]. This suggests the possibility that short telomeres could

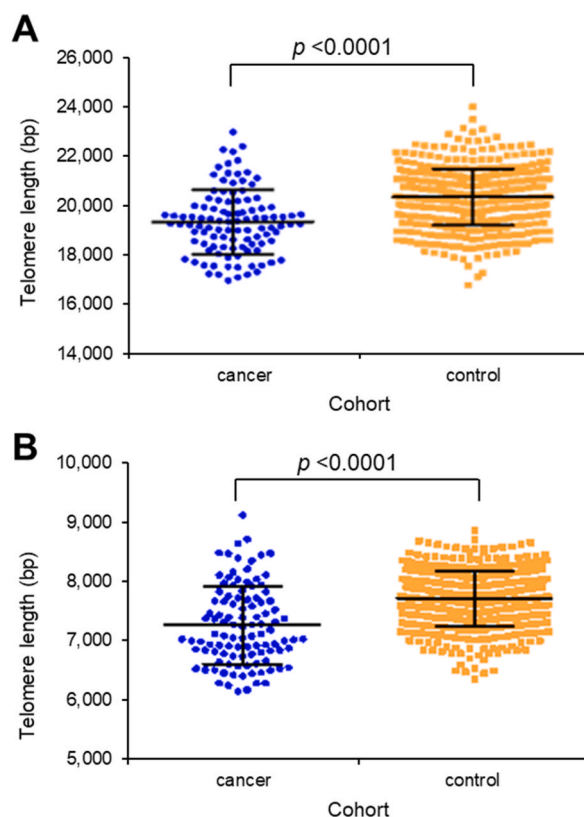


Fig. 3. Distribution of telomere lengths in patients with lung cancer (N = 106) and controls (N = 453). A, difference between the 99th percentile of the telomere lengths of a sample and the 1st percentile of the telomere lengths of that sample ($p < 0.0001$). B, interquartile of the telomere lengths of a sample ($p < 0.0001$). bp, base pairs.

be associated with SCLC and long telomeres with risk of adenocarcinoma [19]. Telomere dysfunction could be at the basis of the cancer and a variety of lung diseases [31,36,37]. Since our study was aimed at early screening of all patients at risk, no attempt at stratification by histology was made. However, in our cancer cohort 89 % of patients were moderate or heavy smokers, and smoking, although associated with all lung cancers, is most strongly related to risk of SCLC histology [38]. It is therefore possible that in our study the population of heavy smokers was enriched with patients with SCLC and shorter telomeres. Further, COPD is also associated with accelerated aging and shortening of telomeres [39,40], and COPD patients comprised 58 % of the cancer cohort.

Two models with different baseline prevalence of lung cancer were developed and evaluated. A prevalence of 48 % of lung cancer patients was chosen as it represented the population of patients that start the diagnostic pathway and already present some symptoms. In this setting a telomere evaluation would help clinicians in the diagnosis of the patient, complementary to other tests. The prevalence of 5 % was chosen to represent the population of lung cancer patients within the total population at risk (e.g., asymptomatic smokers aged >55 years). In this setting the telomere evaluation could contribute to evaluate the patient during early screening of asymptomatic lung cancer. While the AUC was 0.79 for the high-prevalence model, it was 0.98 for the low-prevalence model, suggesting that telomere length determination, together with TAV analysis, could help and could be of use in the screening of high-risk patients. This is further supported by the high specificity (96 %) and high NPV (99 %) determined for the model with low prevalence, both parameters especially relevant in a screening setting.

Although LDCT screening has proven to be highly effective due to its simplicity and high sensitivity, independent biomarkers could greatly help to support the LDCT risk assessment [41]. For example, biomarkers could help reduce screening costs by refining the screening selection criteria and making it independent of age and tobacco exposure. Also, in some patients, biomarkers could inform on the prognosis of the disease at an early stage, or help assess the risk of indeterminate pulmonary nodules which are beyond the reach of biopsy techniques [41]. Unfortunately, although many potentially useful biomarkers of various types have been explored and described, to date none have been clinically validated for early lung cancer diagnosis [42–44]. The main hurdle for biomarker development is the need for resource intensive, long term, longitudinal studies to demonstrate clinical utility. In this regard, it has been suggested that a combination of multiple techniques and biomarkers, perhaps aided with computational models based on machine learning, could be the best approach [45]. The TAV described here can generate unique profiles, or TAV signatures, that can be used to define patterns of diagnostic and prognostic use [28,29]. When used to develop predictive models such as those described in this study, they could be helpful tools during the screening of asymptomatic high-risk patients.

There are several limitations in our study that must be considered when interpreting the data. In addition to the possible bias in the

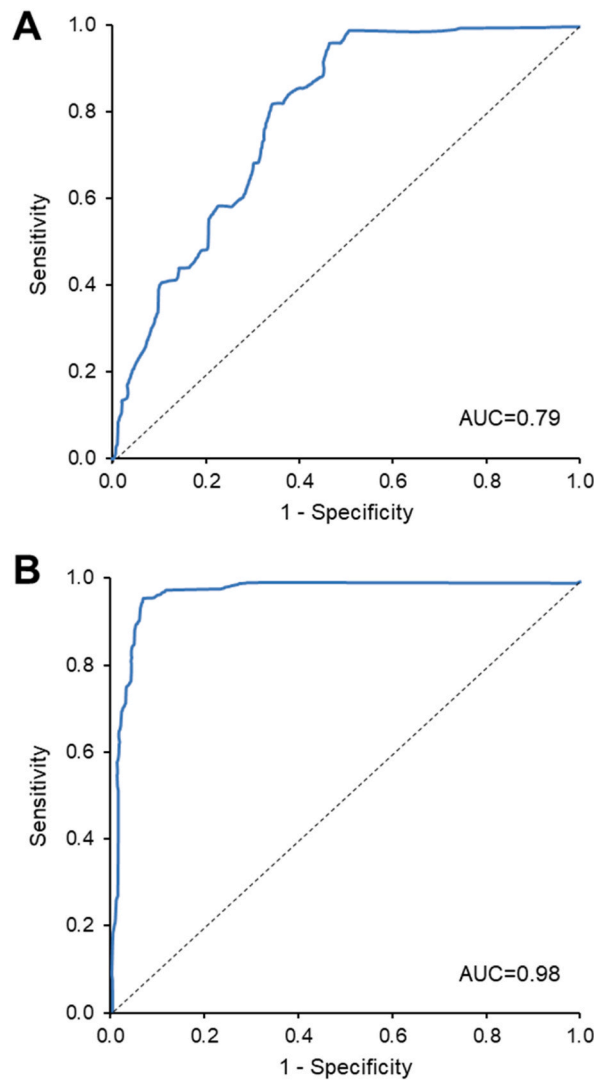


Fig. 4. Performance of the TAV-based models using two separate modelling approaches. Receiver operating characteristic (ROC) curves are shown to compare performances of the validation cohorts at LC prevalences of 48 % in TAV model 1 (A) and of 5 % in TAV model 2 (B). AUC, area under the curve.

Table 2
Performance of the TAV models in the validation cohort.

TAV model 1 (LC prevalence = 48 %)				
	Diagnosis		Total	Sensitivity: 0.88 (95 % CI, 0.68–0.97)
	LC	No LC		
Predicted LC	22	6	28	Specificity: 0.71 (95 % CI, 0.48–0.89)
Predicted No LC	3	15	18	PPV: 0.79 (95 % CI, 0.65–0.88)
Total	25	21	46	NPV: 0.83 (95 % CI, 0.63–0.94)
TAV model 2 (LC prevalence = 5 %)				
	Diagnosis		Total	Sensitivity: 0.80 (95 % CI, 0.65–0.91)
	LC	No LC		
Predicted LC	33	22	55	Specificity: 0.96 (95 % CI, 0.95–0.98)
Predicted No LC	8	582	590	PPV: 0.60 (95 % CI, 0.49–0.70)
Total	41	604	645	NPV: 0.99 (95 % CI, 0.98–0.99)

Abbreviations: CI, confidence interval; LC, lung cancer; NPV, negative predictive value; PPV, positive predictive value; TAV, telomere-associated variables.

lung cancer patients included in the cancer cohort mentioned before, which could have favoured some histologic types over others, other aspects of the samples analysed could have impacted in the results obtained (e.g., ethnic origin, other comorbidities). Also, although all patients in the cancer cohort were smokers, telomere length is determined by other environmental or genetic factors. Further studies with a larger population are needed to validate the findings presented.

In conclusion, the preliminary results of this study suggest that telomere length measurements in peripheral lymphocytes could be useful as diagnostic biomarkers during lung cancer screening in high-risk patients. TAV-based models could improve the predictive power of current initial diagnostic pathways, but further work is needed to integrate them into routine clinical evaluation of patients at risk.

CRediT authorship contribution statement

Sonia Molina-Pinelo: Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Investigation, Data curation, Conceptualization. **Irene Ferrer Sánchez:** Writing – review & editing, Writing – original draft, Resources, Project administration, Investigation, Data curation, Conceptualization. **Pilar Najarro:** Writing – original draft, Investigation, Funding acquisition, Data curation, Conceptualization. **Luis Paz-Ares:** Writing – review & editing, Project administration, Investigation, Conceptualization. **Luis Fernández:** Formal analysis, Data curation, Conceptualization. **Nila Castelló:** Writing – original draft, Project administration, Data curation, Conceptualization. **Luis Alberto Richart López:** Writing – review & editing, Resources. **Juan Diego Rodríguez Gambarte:** Writing – review & editing, Resources. **Máximo Sanz García:** Writing – review & editing, Resources. **Ana Salinas:** Resources, Investigation. **Rocío Suárez:** Investigation, Conceptualization. **Beatriz Romero-Romero:** Writing – review & editing, Resources, Investigation. **José Martín-Juan:** Resources, Investigation. **María Eugenia Viñuela:** Resources, Investigation. **Ray G. Butler:** Methodology, Formal analysis, Data curation, Conceptualization. **Nuria de Pedro:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Data curation, Conceptualization.

Ethics statement

This study was approved by the Ethics Committee of the CEIC Hospital Universitario 12 de Octubre (code 17/026). All patients provided written informed consent for the publication of their anonymised case details.

Data availability statement

Data will be available from the corresponding author upon reasonable request.

Funding

This work was partially funded with the Horizon2020 ONCOCHECK project grant number 738707 and Life Lengths's own funds and R&D budget. SMP was funded by the Ministry of Health and Social Welfare of the Junta de Andalucía (Nicolas Monardes Program RC-0004-2020, and PECART-0091-2020), Andalusian Research, Development and Innovation Plan (PY20_00992), and ISCIII (PI20/01109) and co-funded by FEDER from Regional Development European Funds (European Union). LP-A was funded by a grant from the Comunidad de Madrid (CAMB2017/BMD3884), AECC (TRNSC18004PAZ), Fundación CRIS contra el cancer (Unidad Integral CRIS de Inmuno-oncología, 2018/0043), ISCIII (PI17/00778, PI20/00870, AC20/0070), and CIBERONC (CD16/12/00442), and co-funded by FEDER from Regional Development European Funds (European Union). IF was funded by i+12 and ISCIII (PI16/01311; PI19/00320; CP21/00052) and co-funded by FEDER from Regional Development European Funds (European Union).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Life Length SL reports financial support was provided by European Commission. Sonia Molina-Pinelo reports financial support was provided by Life Length S. Luis Paz-Ares reports financial support was provided by Community of Madrid. Luis Paz-Ares reports financial support was provided by Asociación Española contra el Cáncer. Sonia Molina-Pinelo reports financial support was provided by Andalusian Research, Development and Innovation Plan. Sonia Molina-Pinelo reports financial support was provided by European Regional Development Fund. Luis Paz-Ares reports financial support was provided by European Regional Development Fund. Sonia Molina-Pinelo reports financial support was provided by Carlos III Health Institute. Luis Paz-Ares reports financial support was provided by Carlos III Health Institute. Luis Paz-Ares reports was provided by Fundación CRIS contra el cancer. Luis Paz-Ares reports financial support was provided by Biomedical Research Network Centre in Cancer. Irene Ferrer Sanchez reports financial support was provided by Carlos III Health Institute. Irene Ferrer Sanchez reports financial support was provided by European Regional Development Fund. PN, LF, NC, and NdP are or have been employees and/or investors of Life Length SL. LP-A has received honoraria for scientific advice and speaker fees from Lilly, Merck Sharp & Dohme, Bristol-Myers Squibb, Roche, PharmaMar, Merck, AstraZeneca, Novartis, Boehringer Ingelheim, Celgene, Servier, Sysmex, Amgen, Incyte, Pfizer, Ipsen, Adacap, Sanofi, Bayer and Blueprint, and participates as an external member on the board of Genómica. He is founder and board member of Altum sequencing and has received institutional support for contracted research from Merck Sharp & Dohme, Bristol-Myers Squibb, AstraZeneca and Pfizer. Other authors declare no potential conflicts of interest.

Acknowledgments

The authors wish to thank Laura Esteban, Sheila González, Jorge García, and Marta Blanco for their clinical and data analysis input. Francisco López de Saro, PhD (Trialance SCCL) provided medical writing support, funded by Life Length SL.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e41040>.

References

- [1] American Cancer Society, *Facts & Figures 2020*, American Cancer Society, Atlanta, Ga, 2020.
- [2] Sociedad Española de Oncología (SEOM), *Las cifras del cáncer en España 2022*, available at: 2022 https://seom.org/images/LAS_CIFRAS_DEL_CANCER_EN_ESPANA_2022.pdf.
- [3] National Cancer Institute, *Surveillance, epidemiology, and end results (SEER) program, cancer stat facts: lung and bronchus cancer*, National Cancer Institute Bethesda, MD, <https://seer.cancer.gov/statfacts/html/lungb.html>, 2020.
- [4] C. Mourante-Roibás, V. Leiro-Fernández, A. Fernández-Villar, M. Botana-Rial, C. Ramos-Hernández, A. Ruano-Ravina, COPD, emphysema and the onset of lung cancer. A systematic review, *Cancer Lett.* 382 (2016) 240–244.
- [5] D.R. Aberle, A.M. Adams, C.D. Berg, et al., Reduced lung-cancer mortality with low-dose computed tomographic screening, *N. Engl. J. Med.* 365 (2011) 395–409.
- [6] A. Jemal, S.A. Fedewa, Lung cancer screening with low-dose computed tomography in the United States-2010 to 2015, *JAMA Oncol.* 3 (2017) 1278–1281.
- [7] M. Weischer, B.G. Nordestgaard, R.M. Cawthon, J.J. Freiberg, A. Tybjaerg-Hansen, S.E. Bojesen, Short telomere length, cancer survival, and cancer risk in 47102 individuals, *J Natl Cancer Inst* 105 (2013) 459–468.
- [8] D. Campa, B. Mergarten, I. De Vivo, et al., Leukocyte telomere length in relation to pancreatic cancer risk: a prospective study, *Cancer Epidemiol. Biomarkers Prev.* 23 (2014) 2447–2454.
- [9] C. Duggan, R. Risques, C. Alfano, et al., Change in peripheral blood leukocyte telomere length and mortality in breast cancer survivors, *J Natl Cancer Inst* 106 (2014) dju035.
- [10] L. Hou, B.T. Joyce, T. Gao, et al., Blood telomere length attrition and cancer development in the normative aging study cohort, *EBioMedicine* 2 (2015) 591–596.
- [11] L. Rode, B.G. Nordestgaard, S.E. Bojesen, Peripheral blood leukocyte telomere length and mortality among 64,637 individuals from the general population, *J Natl Cancer Inst* 107 (2015) djv074.
- [12] C. Naing, K. Aung, P.K. Lai, J.W. Mak, Association between telomere length and the risk of colorectal cancer: a meta-analysis of observational studies, *BMC Cancer* 17 (2017) 24.
- [13] X. Zhang, Q. Zhao, W. Zhu, et al., The association of telomere length in peripheral blood cells with cancer risk: a systematic review and meta-analysis of prospective studies, *Cancer Epidemiol. Biomarkers Prev.* 26 (2017) 1381–1390.
- [14] Q. Wang, Y. Zhan, N.L. Pedersen, F. Fang, S. Hagg, Telomere length and all-cause mortality: a meta-analysis, *Ageing Res. Rev.* 48 (2018) 11–20.
- [15] P. Willeit, J. Willeit, A. Mayr, et al., Telomere length and risk of incident cancer and cancer mortality, *JAMA* 304 (2010) 69–75.
- [16] X. Xu, K. Qu, Q. Pang, Z. Wang, Y. Zhou, C. Liu, Association between telomere length and survival in cancer patients: a meta-analysis and review of literature, *Front. Med.* 10 (2016) 191–203.
- [17] J.S. Jang, Y.Y. Choi, W.K. Lee, et al., Telomere length and the risk of lung cancer, *Cancer Sci.* 99 (2008) 1385–1389.
- [18] Y. Qian, T. Ding, L. Wei, S. Cao, L. Yang, Shorter telomere length of T-cells in peripheral blood of patients with lung cancer, *OncoTargets Ther.* 9 (2016) 2675–2682.
- [19] J.A. Doherty, L. Grieshaber, J.R. Houck, et al., Telomere length and lung cancer mortality among heavy smokers, *Cancer Epidemiol. Biomarkers Prev.* 27 (2018) 829–837.
- [20] M. Shen, R. Cawthon, N. Rothman, et al., A prospective study of telomere length measured by monochrome multiplex quantitative PCR and risk of lung cancer, *Lung Cancer* 73 (2011) 133–137.
- [21] Q. Lan, R. Cawthon, Y. Gao, et al., Longer telomere length in peripheral white blood cells is associated with risk of lung cancer and the rs2736100 (CLPTM1L-TERT) polymorphism in a prospective cohort study among women in China, *PLoS One* 8 (2013) e59230.
- [22] W.J. Seow, R.M. Cawthon, M.P. Purdue, et al., Telomere length in white blood cell DNA and lung cancer: a pooled analysis of three prospective cohorts, *Cancer Res.* 74 (2014) 4090–4098.
- [23] J.A. Doherty, L. Grieshaber, J.R. Houck, et al., Nested case-control study of telomere length and lung cancer risk among heavy smokers in the β -Carotene and Retinol Efficacy Trial, *Br. J. Cancer* 118 (2018) 1513–1517.
- [24] J.M. Yuan, K.B. Beckman, R. Wang, et al., Leukocyte telomere length in relation to risk of lung adenocarcinoma incidence: findings from the Singapore Chinese Health Study, *Int. J. Cancer* 142 (2018) 2234–2243.
- [25] M.S. Pepe, Z. Feng, H. Janes, P.M. Bossuyt, J.D. Potter, Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design, *J Natl Cancer Inst* 100 (2008) 1432–1438.
- [26] A. Canela, E. Vera, P. Klatt, M.A. Blasco, High-throughput telomere length quantification by FISH and its application to human population studies, *Proc Natl Acad Sci U S A* 104 (2007) 5300–5305.
- [27] A.J. Montpetit, A.A. Alhareeri, M. Montpetit, et al., Telomere length: a review of methods for measurement, *Nurs. Res.* 63 (2014) 289–299.
- [28] N. de Pedro, M. Díez, I. García, et al., Analytical validation of Telomere Analysis Technology for the high-throughput analysis of multiple telomere-associated variables, *Biol. Proced. Online* 22 (2020) 2.
- [29] Galisteo JM, Rubio, L. Fernandez, E. Gomez Gomez, et al., Telomere-based risk models for the early diagnosis of clinically significant prostate cancer, *Prostate Cancer Prostatic Dis.* 24 (2021) 88–95.
- [30] M.R. Berthold, N. Cebon, F. Dill, et al., *KNIME: the Konstanz Information Miner*, Springer, Berlin, Heidelberg, 2008.
- [31] Y. Xue, X. Guo, X. Huang, et al., Shortened telomere length in peripheral blood leukocytes of patients with lung cancer, chronic obstructive pulmonary disease in a high indoor air pollution region in China, *Mutat. Res. Genet. Toxicol. Environ. Mutagen* 858–860 (2020) 503250.
- [32] B. Karimi, M. Yunesian, R. Nabizadeh, P. Mehdipour, A. Aghaie, Is leukocyte telomere length related with lung cancer risk?: a meta-analysis, *Iran. Biomed. J.* 21 (2017) 142–153.
- [33] M.J. Machiela, C.A. Hsiung, X.-O. Shu, et al., Genetic variants associated with longer telomere length are associated with increased lung cancer risk among never-smoking women in Asia: a report from the female lung cancer consortium in Asia, *Int. J. Cancer* 137 (2015) 311–319.
- [34] P.C. Haycock, S. Burgess, A. Nounu, et al., Association between telomere length and risk of cancer and non-neoplastic diseases: a mendelian randomization study, *JAMA Oncol.* 3 (2017) 636–651.
- [35] L. Kachuri, O. Saarela, S.E. Bojesen, et al., Mendelian Randomization and mediation analysis of leukocyte telomere length and risk of lung and head and neck cancers, *Int. J. Epidemiol.* 48 (2019) 751–766.

- [36] S. Vodenkova, M. Kroupa, Z. Polivkova, et al., Chromosomal damage and telomere length in peripheral blood lymphocytes of cancer patients, *Oncol. Rep.* 44 (2020) 2219–2230.
- [37] A. Ruiz, J. Flores-Gonzalez, I. Buendia-Roldan, L. Chavez-Galan, Telomere shortening and its association with cell dysfunction in lung diseases, *Int. J. Mol. Sci.* 23 (2021) 425.
- [38] S.A. Khuder, Effect of cigarette smoking on major histological types of lung cancer: a meta-analysis, *Lung Cancer* 31 (2001) 139–148.
- [39] J.M.J. Houben, E.M. Mercken, H.B. Ketelslegers, et al., Telomere shortening in chronic obstructive pulmonary disease, *Respir. Med.* 103 (2009) 230–236.
- [40] E. Córdoba-Lantús, S. Cazorla-Rivero, A. Espinoza-Jiménez, et al., Telomere shortening and accelerated aging in COPD: findings from the BODE cohort, *Respir. Res.* 18 (2017) 59.
- [41] L.M. Seijo, N. Peled, D. Ajona, et al., Biomarkers in lung cancer screening: achievements, promises, and challenges, *J. Thorac. Oncol.* 14 (2019) 343–357.
- [42] M. Hassanein, J.C. Callison, C. Callaway-Lane, M.C. Aldrich, E.L. Grogan, P.P. Massion, The state of molecular biomarkers for the early detection of lung cancer, *Cancer Prev. Res.* 5 (2012) 992–1006.
- [43] K. Marschollek, A. Brzecka, A. Pokryszko-Dragan, New biochemical, immune and molecular markers in lung cancer: diagnostic and prognostic opportunities, *Adv. Clin. Exp. Med.* (2022), <https://doi.org/10.17219/acem/152349> published online Aug 24.
- [44] X.-G. Liu, M. Li, S.-J. Mai, R.-J. Cai, Telomere length-related signature as a novel biomarker of prognosis and immune response in non-small cell lung cancer, *Eur. Rev. Med. Pharmacol. Sci.* 26 (2022) 1304–1319.
- [45] J. Benzaquen, J. Boutros, C. Marquette, H. Delingette, P. Hofman, Lung cancer screening, towards a multidimensional approach: why and how? *Cancers* 11 (2019). DOI:E212 [pii] 10.3390/cancers11020212 cancers11020212 [pii].