



OPEN

Clinical and pathological characteristics associated with the presence of the IS6110 *Mycobacterium tuberculosis* transposon in neoplastic cells from non-small cell lung cancer patients

Oscar Arrieta¹✉, Camilo Molina-Romero¹, Fernanda Cornejo-Granados², Brenda Marquina-Castillo³, Alejandro Avilés-Salas⁴, Gamaliel López-Leal², Andrés F. Cardona^{5,6,7}, Alette Ortega-Gómez⁸, Mario Orozco-Morales¹, Adrián Ochoa-Leyva² & Rogelio Hernandez-Pando³✉

Lung cancer (LC) and pulmonary tuberculosis (TB) are the deadliest neoplastic and bacterial infectious diseases worldwide, respectively. Clinicians and pathologists have long discussed the co-existence of LC and TB, and several epidemiologic studies have presented evidence indicating that TB could be associated with the development of LC, particularly adenocarcinoma. Nonetheless, this data remains controversial, and the mechanism which could underlie the association remains largely unexplored. Some bioinformatic studies have shown that human cancer biopsies have a very high frequency of bacterial DNA integration; since *Mycobacterium Tuberculosis* (MTb) is an intracellular pathogen, it could play an active role in the cellular transformation. Our group performed an exploratory study in a cohort of 88 LC patients treated at the Instituto Nacional de Cancerología (INCan) of Mexico City to evaluate the presence of MTb DNA in LC tissue specimens. For the first time, our results show the presence of the MTb IS6110 transposon in 40.9% (n = 36/88) of patients with lung adenocarcinomas. Additionally, through in-situ PCR we identified the presence of IS6110 in the nuclei of tumor cells. Furthermore, shotgun sequencing from two samples identified traces of MTb genomes present in tumor tissue, suggesting that similar Mtb strains could be infecting both patients.

Two critical diseases that are a frequent cause of morbidity and mortality are tuberculosis (TB) and lung cancer (LC)^{1,2}. TB is caused by *Mycobacterium tuberculosis* (MTb) and constitutes the leading cause of death by a single pathogen bacterial agent. Annually, TB causes approximately 1.2 million deaths among HIV negative individuals, according to the World Health Organization (WHO), and incident cases were estimated at 10 million cases

¹Thoracic Oncology Unit and Laboratory of Personalized Medicine, Instituto Nacional de Cancerología (INCan), San Fernando #22, Section XVI, Tlalpan, 14080 Mexico City, Mexico. ²Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico. ³Experimental Pathology Laboratory, Department of Pathology, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico. ⁴Department of Pathology, INCan, Mexico City, Mexico. ⁵Luis Carlos Sarmiento Angulo Cancer Treatment and Research Center (CTIC), Bogotá, Colombia. ⁶Foundation for Clinical and Applied Cancer Research (FICMAC), Bogotá, Colombia. ⁷Molecular Oncology and Biology Systems Research Group (FOX-G/ONCOLGroup), Universidad El Bosque, Bogotá, Colombia. ⁸Translational Medicine Laboratory, Instituto Nacional de Cancerología (INCan), Mexico City, Mexico. ✉email: ogar@unam.mx; rhdezpando@hotmail.com

(8.9–11.0) in 2019 alone¹. Moreover, the WHO estimates that approximately 23% of the world population is infected by MTb, though only 1 in 10 individuals will develop active disease^{3,4}. The importance of this infectious agent cannot be understated, and although several world regions have gained control through public health measures, other areas are still under a considerable burden by pulmonary TB¹.

Another relevant disease in terms of global mortality is LC. LC is currently the first cause of cancer-related deaths worldwide, and although some encouraging data emerged in the last years, the survival rate for this neoplasm remains merely at 14–15%². Each year, approximately 1.76 million deaths are attributed to LC, as well as 2.09 million incident cases, most of these diagnosed in the advanced setting, and therapies lack curative intent in most cases^{5,6}. Non-small cell lung cancer (NSCLC) accounts for approximately 85% of all LCs, and the most common subtype is lung adenocarcinoma⁷. Although tobacco smoke plays an undeniable role in the development of LC, there is still an essential proportion of patients who develop this disease without a history of smoking⁸. Non-smoker patients comprise approximately 10–20% of lung cancer cases globally, although this number can increase in specific subgroups; for example, among female patients in Mexico, only 30% of cases have a positive smoking history⁹. Several other risk factors have been well-described and causally linked to the development of LC, including radon exposure, asbestos, arsenic, and others^{9,10}. Infections also seem to play a role, although their specific mechanism remains elusive. In this regard, a history of TB has been associated with LC development in several epidemiologic studies, particularly for the development of adenocarcinoma¹¹.

Animal models have presented experimental evidence regarding the association between TB and LC, in such cases the increased inflammatory process from chronic TB infection induces cell dysplasia and squamous cell carcinoma. Nonetheless, the association in epidemiological studies has been observed in humans between TB and lung adenocarcinoma, rather than squamous cell carcinoma, which would suggest other mechanistic pathways¹². Interestingly, although the International Agency for Research on Cancer Monographs has identified eleven biological agents as group 1 carcinogens, MTb has not been included in this list^{13,14}.

TB and LC co-occurrence have been frequently reported in the literature, though the nature of this observation remains undescribed^{11,15}. Although results have been inconclusive, one large cohort in China identified that risk of LC was significantly increased in subjects with a previous TB history^{16,17}. Similarly, a prospective cohort in Asia showed an increased incidence of LC in TB patients, a risk which was also increased by presence of other comorbidities including COPD and risk factors such as smoking^{18,19}. Interestingly, results also indicate that adenocarcinoma is most frequently associated with a TB history, as reported by a systematic review¹¹. Furthermore, Wong et al. found an association between TB and lung adenocarcinoma (OR 1.31, 95% CI 1.03–1.66, $p = 0.027$) among never-smoking Asian women in a genome-wide association study using Mendelian randomization and pathway analysis²⁰. Last, a recent meta-analysis concluded that pre-existing TB increases the risk of LC (RR 2.170 (1.833–2.569)). The results emphasize the importance of LC screening in this patient subgroup, as there could be a need for a considerable follow-up after the infection has been treated²¹.

Having a history of TB appears not only to increase the risk of developing LC, but it can also negatively affect prognosis. In a cohort study conducted in the Netherlands which included over 8000 persons and had a follow-up of 18 years a total of 214 cases of LC were found, of which 13 had a history of pulmonary TB. The overall survival of patients with LC and a history of pulmonary TB was significantly lower than patients without a history of TB (HR 2.36, 95% CI 1.1–4.9), with an average result of 311 days difference between the two groups²². The role of a previous TB infection in terms of prognosis is scarcely understood, though it might be related to specific molecular alterations and response to treatment. For example, a correlation has been observed between TB and *EGFR* mutations in patients with LC²³. Further, considerable differences have been observed in response to treatment using tyrosine kinase inhibitors (TKIs)²⁴, and could be related to a high expression of Epiregulin in *EGFR* positive tumors²⁵. Interestingly, inducible nitric oxide synthase (iNOS) and Epiregulin are highly expressed in chronic inflammatory processes^{12,24}.

Previous preclinical and clinical studies have provided information regarding the possible mechanisms underlying the relationship between TB and LC. Mtb induces a strong inflammatory response in the lung tissue of infected hosts, and this can in turn promote cancer development and progression^{26–29}. Moreover, the infection is also characterized by the formation of TB-induced scars, a process which has been suggested to play an etiological role in LC development due to the cellular proliferation which occurs during tissue repair^{11,27–30}. The role of angiogenesis has also been explored, which is characteristic during repair processes³¹.

Considering the extensive evidence of LC and TB's relationship, we sought to explore the presence of MTb genetic material within tissue samples of patients with LC treated at the Instituto Nacional de Cancerología (INCan) in Mexico City.

Results

Patient characteristics. The baseline characteristics of the eighty-eight patients included in this study (Fig. 1) are summarized in Table 1. The median age of patients was 60 (21–92) years, and the majority were women 69.3% (61/88). In total, 61.4% (54/88) of the patients did not have tobacco smoke exposure, 56.8% (50/88) had no wood-smoke exposure, 84.1% (74/88) were stage IV at diagnosis, and 93.2% (82/88) had adenocarcinoma. Most patients had an acinar histology subtype (27/88; 30.7%). Only 58% (51/88) of the patients were tested for *EGFR* mutations (*EGFRm*), and among these 47.1% ($n = 24/51$) had an *EGFRm*. The most commonly reported *EGFRm* was a deletion in exon 19, present in 54.2% of the cases (13/24) (Table 1).

Patient characteristics according to the presence of the IS6110 transposon detected by End-Point PCR. To evaluate the presence of the IS6110 MTb transposon, End-Point PCR was carried out using DNA isolated from the tumors of 88 patients. Among all samples included, 40.9% (36/88) had a positive End-Point PCR result for the IS6110 transposon. The mean age of positive patients was 54.75 (± 16.07) years.

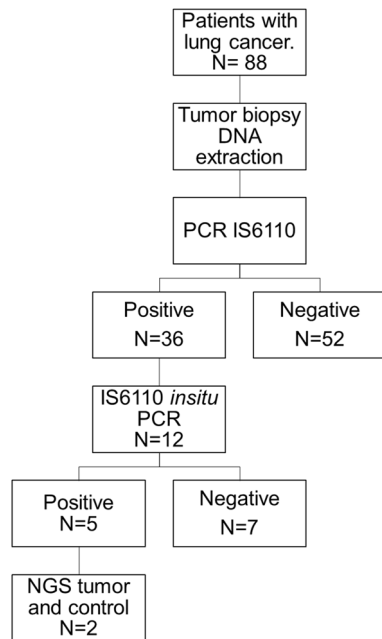


Figure 1. Consort diagram showing the flow of tests and patients included in the study.

Approximately 63.9% (23/36) of patients positive for IS6110 were younger than 60 years old. In contrast, only 36.1% (13/36) of patients ≥ 61 years old, were positive for IS6110 ($P=0.03$) (Table 1). Moreover, 86.1% (31/36) of female patients were IS6110 positive, compared with male patients, among who 13.9% (5/36) were positive for the IS6110 transposon ($P=0.004$). Additionally, 41.7% (15/36) of the patients with brain metastases at diagnosis were positive for IS6110 ($P=0.006$). Interestingly, the presence of an *EGFR* mutation was not associated with IS6110 ($P=0.71$) (Table 1).

Overall survival according to patient characteristics. The median OS for all patients included in the study was 23.1 months (95% CI 7.6–23.2). OS was longer for female patients compared with male patients (20.5 [95% CI 10.9–30.0] vs. 7.4 [95% CI 0.9–13.8]), however the difference did not reach statistical significance ($P=0.08$). The survival analysis revealed that the only characteristics that were significantly associated with a reduction in OS were histology when comparing adenocarcinoma vs. other (17.1 [95% CI 8.6–25.7] vs. 2.1 [95% CI 1.1–3.2]; $P=0.001$) and adenocarcinoma subtype papillary vs. acinar vs. solid vs. other (25.6 [95% CI 10.6–42.4] vs. 23.0 [95% CI 14.8–31.1] vs. 12.3 [95% CI 2.6–22.0] vs. 4.5 [0.0–10.2]; $P=0.02$) (Table 2).

Overall survival according to the presence of the IS6110 transposon. When comparing patients with positive and negative IS6110 PCR results, median OS is 15.3 (95% CI 5.5–25.3) vs. 15.5 (95% CI 4.1–26.9) months ($P=0.58$). Among patients with a positive IS6110 result, OS was longer for female compared with male patients (17.1 [95% CI 5.6–28.7] vs. 3.8 [95% CI 0.1–7.4]; $P=0.28$), though this difference did not reach statistical significance. Moreover, patients with adenocarcinoma histology had a prolonged OS compared with other histologic types (15.4 [95% CI 5.6–25.2] vs. 0.9 [95% CI NR]; $P=0.001$). Interestingly, among patients with a positive IS6110 result, those with an *EGFRm* had a longer OS compared with patients without *EGFRm* (56.0 months [95% CI 14.2–98.0] vs. 21.4 [95% CI 0.0–44.1]; $P=0.16$), though this did not reach statistical significance. No other associations were identified among baseline characteristics for patients with a positive IS6110 (Table 3). Baseline characteristics associated with patients with a negative IS6110 transposon result are summarized in Table 4.

Presence of the IS6110 transposon detected by in-situ PCR. The samples that resulted positive for IS6110 End-Point PCR were selected for in-situ PCR using the same primers for both experiments. Among 36 tumor samples positive for IS6110, 12 had enough paraffin-embedded tumor tissue to perform the *in-situ* PCR test. Among these, 5 were positive (41.6%; $n=5/12$), and two of them had the positive signal at the tumor cell nucleus area (Fig. 2). Since the End-Point PCR results for IS6110 suggested the presence of *Mtb* genome in the tumor, these 2 samples were chosen for tumor sequencing.

Total shotgun DNA sequencing. To identify the *MTb* strain present in the tumor tissues of both sequenced samples (TB31 and TB34), we mapped all sequenced reads against 23 *MTb* reference genomes (Table 3). After mapping the total reads, we identified 12,191 reads and 10,200 reads that mapped to the reference *MTb* genomes for TB31 and TB34, respectively. An average of 82.66% of the reads mapped against three strains: *Mycobacterium tuberculosis* KZN 1435 (TB31: 5675 reads and TB34: 4364 reads), *Mycobacterium tuberculosis* str.

Characteristics	Total	IS6110 ^a PCR		P-value
	N = 88	Negative n = 52	Positive n = 36	
	% (n/N)	% (n/N)	% (n/N)	
Sex				
Male	30.7 (27/88)	42.3 (22/52)	13.9 (5/36)	0.004
Female	69.3 (61/88)	57.7 (30/52)	86.1 (31/36)	
Age at diagnosis (years)				
Mean (SD)	58.48 ± 15.06	61.06 ± 13.89	54.75 ± 16.07	0.050
Age (years)				
< 60	50 (44/88)	40.4 (21/52)	63.9 (23/36)	0.030
≥ 61	50 (44/88)	59.6 (31/52)	36.1 (13/36)	
Age category				
20–39	10.2 (9/88)	7.7 (4/52)	13.9 (5/36)	0.608
40–69	64.8 (57/88)	65.4 (34/52)	63.9 (23/36)	
70+	25 (22/88)	26.9 (14/52)	22.2 (8/36)	
Body-mass Index (kg/m²)				
Mean (SD)	25.57 ± 4.18	24.71 ± 4.51	25.96 ± 4.01	0.977
Median (IQR)	24.69 (16.2–34)	24.35 (18.5–32.2)	25.96 (23.1–28.8)	0.916
Tobacco exposure				
Absent	61.4 (54/88)	61.5 (32/52)	61.1 (22/36)	0.968
Present	38.6 (34/88)	38.5 (20/52)	38.9 (14/36)	
Tobacco Index				
Mean (SD)	28.27 ± 38.31	24.5 ± 31.43	27.5 ± 37.4	0.726
Median (IQR)	20 (1–202)	15 (2–86)	27.5 (1–54)	0.506
Wood-smoke exposure				
Absent	56.8 (50/88)	55.8 (29/52)	58.3 (21/36)	0.811
Present	43.2 (38/88)	44.2 (23/52)	41.7 (15/36)	
Wood-smoke exposure Index				
Mean (SD)	102.53 ± 156.39	112.16 ± 126.47	15.5 ± 10.6	0.149
Median (IQR)	39 (5–816)	58 (5–300)	15.5 (8–23)	0.731
Asbestos exposure				
Absent	90.9 (80/88)	90.4 (47/52)	91.7 (33/36)	0.837
Present	9.1 (8/88)	9.6 (5/52)	8.3 (3/36)	
Type of disease				
Loco-regional	40.9 (36/88)	44.2 (23/52)	36.1 (13/36)	0.692
Oligometastatic	55.7 (49/88)	51.9 (27/52)	61.1 (22/36)	
Metastatic	3.4 (3/88)	3.8 (2/52)	2.8 (1/36)	
Histology				
Adenocarcinoma	93.2 (82/88)	90.4 (47/52)	97.2 (35/36)	0.211
Other	6.8 (6/88)	9.6 (5/52)	2.8 (1/36)	
Histology subtype				
Lepidic	6.8 (6/88)	3.8 (2/52)	11.1 (4/36)	0.502
Acinar	30.7 (27/88)	26.9 (14/52)	36.1 (13/36)	
Papilar	11.4 (10/88)	11.5 (6/52)	11.1 (4/36)	
Solid	26.1 (23/88)	28.8 (15/52)	22.2 (8/36)	
Not specified	25 (22/88)	28.8 (15/52)	19.4 (7/36)	
Clinical stage				
I–II	5.7 (5/88)	7.7 (4/52)	2.8 (1/36)	0.420
III	10.2 (9/88)	13.5 (7/52)	5.6 (2/36)	
IV	84.1 (74/88)	78.8 (41/52)	91.7 (33/36)	
ECOG status				
0–1	85.2 (75/88)	84.6 (44/52)	86.1 (31/36)	0.98
2+	14.8 (13/88)	15.4 (8/52)	13.9 (5/36)	
EGFR mutation				
Wildtype		16	11	0.71
Mutated		13	11	
Not determined (37)		–	–	

Table 1. Baseline characteristics of NSCLC patients. Significant values are in bold. ^aIS6110: *Mycobacterium tuberculosis* insertion sequence. NSCLC Non-small cell lung cancer.

Characteristic	OS (95% CI)	P
PCR 6110		
Negative	15.5 (4.1–26.9)	0.58
Positive	15.3 (5.5–25.3)	
Age		
< 60	15.5 (5.5–25.5)	0.92
≥ 60	13.8 (3.6–24.3)	
Gender		
Male	7.4 (0.9–13.8)	0.08
Female	20.5 (10.9–30.0)	
Smoking history		
No	15.5 (6.5–24.5)	0.31
Yes	15.4 (6.0–24.7)	
Woodsmoke exposure		
No	17.1 (3.9–30.3)	0.92
Yes	12.3 (5.3–19.3)	
Asbestos exposure		
No	17.1 (8.1–16.2)	0.24
Yes	9.6 (0.0–22.9)	
ECOG		
0–1	17.1 (8.7–25.5)	0.06
≥ 2	7.6 (0.7–14.4)	
Diabetes		
No	15.5 (7.1–23.9)	0.97
Yes	13.8 (0.0–30.2)	
EGFR		
Wildtype	22.4 (15.0–29.7)	0.93
Mutated	23.0 (8.4–37.5)	
Histology		
Adenocarcinoma	17.1 (8.6–25.7)	0.001
Other	2.1 (1.1–3.2)	
Adenocarcinoma subtype		
Papillary	26.5 (10.6–42.4)	0.02
Acinar	23.0 (14.8–31.1)	
Solid	12.3 (2.6–22.0)	
Lepidic	NA	
Not specified/other histology	4.5 (0.0–10.2)	

Table 2. Overall survival according to baseline characteristics. Significant values are in bold.

Haarlem/NITR202 (TB31: 2758 reads and TB34: 3399 reads) and *Mycobacterium tuberculosis* RGTB327 (TB31: 1378 reads and TB34: 891 reads). Interestingly, in both stances reads mapped mainly against MTb strain KZN 1435, suggesting that similar strains could be infecting both patients. It is important to note that the mapped reads were distributed across all the MTb genome (Fig. 3). This result also suggests that the strain present in both patients is closely related to KZN 1435.

Discussion

Despite the well documented epidemiological association between pulmonary Tb infection and LC, the molecular mechanisms by which Tb promotes the development and progression of LC remain unknown¹⁹. Nonetheless, the association has been documented even in large population-based studies, highlighting a considerable risk also for secondary lung cancer even after adjusting for important covariates³². TB appears to be a significant factor in LC development, though the mechanism behind this association has seldom been explored. It is well known that TB induces an inflammatory response, and that chronic inflammation is conducive to neoplastic processes through several pathways^{26–29,31}. Furthermore, other infectious agents have been well characterized for their oncogenic properties, accounting for a considerable proportion of worldwide cancers³³. Therefore, it is not unreasonable to explore new hypotheses where MTb actively participates in the cellular transformation that leads to cancer. In this study, we evaluated the presence of the IS6110 MTb transposon on the primary tumor DNA of eighty-eight patients through End-Point PCR, and results show that 40.9% (36/88) had a positive End-Point PCR for the IS6110 transposon. The IS6110 transposon is an insertion sequence that is only present in the pathogens of the MTb complex, and therefore the presence of this element in DNA recovered from lung adenocarcinomas,

Characteristic	OS (95% CI)	P
PCR 6110		
Negative	15.5 (4.1–26.9)	0.58
Positive	15.3 (5.5–25.3)	
Age		
< 60	15.5 (5.5–25.5)	0.92
≥ 60	13.8 (3.6–24.3)	
Gender		
Male	7.4 (0.9–13.8)	0.08
Female	20.5 (10.9–30.0)	
Smoking history		
No	15.5 (6.5–24.5)	0.31
Yes	15.4 (6.0–24.7)	
Woodsmoke exposure		
No	17.1 (3.9–30.3)	0.92
Yes	12.3 (5.3–19.3)	
Asbestos exposure		
No	17.1 (8.1–16.2)	0.24
Yes	9.6 (0.0–22.9)	
ECOG		
0–1	17.1 (8.7–25.5)	0.06
≥ 2	7.6 (0.7–14.4)	
Diabetes		
No	15.5 (7.1–23.9)	0.97
Yes	13.8 (0.0–30.2)	
Histology		
Adenocarcinoma	17.1 (8.6–25.7)	0.001
Other	2.1 (1.1–3.2)	
Adenocarcinoma subtype		
Papillary	26.5 (10.6–42.4)	0.02
Acinar	23.0 (14.8–31.1)	
Solid	12.3 (2.6–22.0)	
Lepidic	NA	
Not specified/other histology	4.5 (0.0–10.2)	

Table 3. Overall survival according to baseline characteristics among patients with (+) IS6110. Significant values are in bold.

together with the identification of this element through in situ PCR in the nuclei of cancer cells constitutes irrefutable evidence which indicates the presence of the genetic material of MTb in these tumor samples³⁴.

This result is higher compared with the reported rate of latent TB infection worldwide (23%)³, nevertheless, it must be noted that this cohort of LC patients represents a selected population and therefore a selection bias could be responsible for this difference, compared with open population and other epidemiology reports which have shown an association between LC and TB^{26–29,31}. This is not the first time MTb DNA is found in the absence of a TB histological reaction. We previously showed that mycobacterial DNA was identified in 38% and 35% of normal lung tissue samples from Ethiopia and Mexico, respectively, by conventional PCR, in subjects without a previous TB history³⁵.

Another interesting observation is the difference in age of patients who tested positive for IS6110 compared with those who tested negative. The mean age of positive patients was 54.75 (\pm 16.07) years. Among them, 63.9% (23/36) were younger than 60 years old (P = 0.03). Our group recently reported that the mean age of patients with LC in Latin America was 62.2 years (\pm 12.3)³⁶ and the median age for this cohort is similar to the one reported for patients with no known risk factors for LC⁹. It is essential to highlight that previous studies have already shown this important association between young age and increased risk of lung cancer among subjects with a TB history. In a study which compared 3776 pulmonary TB patients with 18,880 matched controls, the authors identified that the risk for lung cancer increased as a function of younger age. Compared to patients < 50 years of age, the risks for lung cancer were HR 9.85, 7.1, 3.32, and 2.57 in patients aged 50–59, 60–69, and \geq 70 years, respectively³⁷.

Another critical factor to consider pertaining to the baseline characteristics of patients with a positive IS6110 result is the significant predominance of female sex patients with this characteristics. In a study by Chang et al. the authors perform a retrospective cohort among patients with lung cancer receiving *EGFR* targeted therapy, with the objective of assessing whether TB affects the outcome of patients with NSCLC. In this study, the authors

Characteristic	OS (95% CI)	P
Age		
< 60	22.4 (10.3–34.4)	0.57
≥60	21.3 (0.0–62.7)	
Gender		
Male	7.4 (0.0–16.5)	0.12
Female	22.4 (9.9–34.8)	
Smoking history		
No	15.5 (0.46–30.6)	0.7
Yes	15.4 (0.0–30.8)	
Woodsmoke exposure		
No	13.8 (0.0–29.7)	0.61
Yes	15.5 (4.2–26.8)	
Asbestos exposure		
No	18 (4.5–31.4)	0.57
Yes	9.6 (2.6–16.7)	
ECOG		
0–1	20.5 (10.5–30.5)	0.01
≥2	6.8 (0.0–15.7)	
Diabetes		
No	15.5 (2.8–28.2)	0.66
Yes	13.8 (0.0–29.7)	
EGFR		
Wildtype	26.5 (18.1–34.8)	0.32
Mutated	15.5 (1.6–29.4)	
Histology		
Adenocarcinoma	17.9 (5.9–30.0)	0.01
Other	2.1 (0.5–1.2)	
Adenocarcinoma subtype		
Papillary	22.4 (5.6–39.2)	< 0.001
Acinar	25.3 (20–30.6)	
Solid	20.5 (8.8–32.2)	
Lepidic	1.2 (NR)	
Not specified/other histology	4.0 (1.8–6.1)	

Table 4. Overall survival (OS) according to baseline characteristics in patients with (–) MTb 6110. Significant values are in bold.

identify that a history of TB has a gender-dependent impact²⁴. This could be a hypothesis worth exploring, given the association between female sex and *EGFR* mutations, and the possible mechanistic pathway for TB-induced lung adenocarcinoma, which some hypothesize could involve the *EGFR* pathway. In a study which included 477 patients with pulmonary adenocarcinoma, 39% had *EGFR*-mutated tumors, while 21% had pre-existing TB lesions. In this same study, the authors report that the frequency of *EGFR* mutations is significantly higher in the subgroup of patients with TB lesions, and multivariate analysis revealed that pre-existing TB lesions were an independent factor associated with *EGFR* mutations³⁸. Although our study included a small number of *EGFR* patients due to the fact that many samples were not tested for this alteration, and tissue was not available to perform it at the time. It is important to highlight that among positive IS6110 patients, those with an *EGFR* had a numerically higher median OS compared with wild-type subjects (56.0 vs. 21.4 months), although the difference is considerable, this did not reach statistical significance, likely due to sample size and patients lost to follow-up. Interestingly, this tendency was not observed for IS6110 negative patients with *EGFR* mutations.

To confirm that MTb was present in the lung cancer tissue without histological evidence of TB lesions, in-situ PCR was performed in samples which had a positive End-Point PCR for the IS6110 transposon. Twelve patients had enough paraffin-embedded tumor tissue to perform in-situ PCR, with a positive result in 41.6% (5/12) of the cases. Among these, 2 samples showed nuclear labeling in neoplastic cells (Fig. 2) and were selected to perform whole-genome sequencing. We found a low number of sequencing reads mapping to MTb genomes and an average of 82.33% were from three strains: *Mycobacterium tuberculosis* KZN 1435, *Mycobacterium tuberculosis* str. Haarlem/NITR202, *Mycobacterium tuberculosis* RGTB327. The reads of both patients were mainly associated with the *Mycobacterium tuberculosis* KZN 1435 strain, suggesting that the MTb strain present in both patients is closely related. Interestingly, MTb KZN 1435 was isolated for the first time from patients in KwaZulu-Natal, South Africa, and is considered as a multiple drug resistant (MDR) strain (resistance to isoniazid and rifampicin)³⁹,

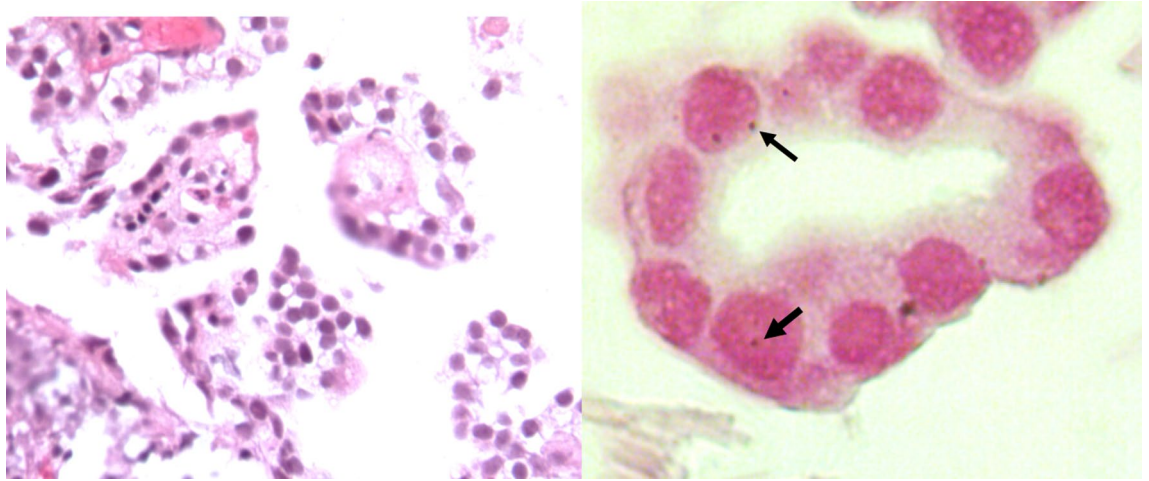


Figure 2. Representative micrographs of pulmonary adenocarcinoma and its in-situ PCR detection of mycobacterial DNA. Left figure shows well differentiated adenocarcinoma stained with haematoxylin/eosin, neoplastic cells show big hyperchromatic basophilic nucleus. The right figure shows high power magnification of neoplastic gland, some cells show positivity for mycobacterial DNA (IS-6110) detected by in situ RT-PCR (dark blue dots, arrows) in the nucleus stained by nuclear fast red.

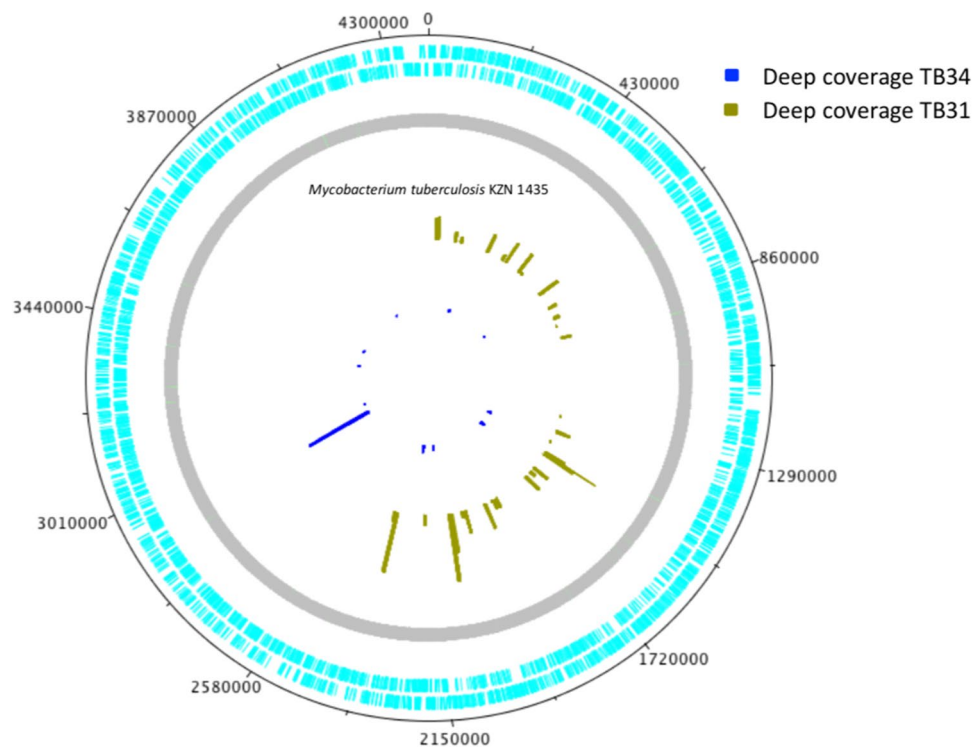


Figure 3. Mapped reads from NGS sequences.

it belongs to the LAM family of MTb. The observation of bacterial sequences in cancer genomes is not novel; recently, Riley et al. performed a bioinformatic analysis of the nuclear and mitochondrial genome in normal and tumor tissue from publicly available sequence data. They found frequent incidences of LGT involving Acinetobacter- and Pseudomonas-like DNA integrated in the human genomes from samples of acute myeloid leukemia. The cancer samples had a 210-fold higher frequency of integration of bacterial DNA compared with the normal tissue⁴⁰. Although, the recruitment fragments and the End-Point PCR results suggested the presence of the MTb genomes in the tumor, more studies are needed to describe the impact of MTb DNA as a causal role in cancer or is a simple byproduct of carcinogenesis.

The information from this study adds to the body of knowledge which seeks to explore the relationship between TB and LC. Nonetheless, the association is difficult to determine due to several challenges, including

the subclinical nature resulting from the primary infection, which makes it difficult to identify when it occurred, and several confounding environmental and host factors which can modulate pathogenesis³³. Despite the considerable challenges, this is an association worth exploring further due to the considerable impact both in terms of follow-up and screening.

Conclusion

Results from this study show that in patients with lung adenocarcinoma from Latin America, a large proportion of tumor samples have MTb DNA sequence IS6110, additionally we report that these sequences can be identified in the nuclear area of neoplastic cells by in situ PCR. The shotgun sequencing effort suggests that genotypes of the two sequenced patients could be related. The presence of MTb DNA by PCR in this cohort is significantly associated with sex and younger age. Although chronic inflammation is increasingly implicated as a cancer development mechanism following bacterial infection, proto-oncogene disruption by MTb DNA could provide another mechanism in lung oncogenesis. However, further controlled experiments should be undertaken to assess the possible mechanisms by which TB participates in the development of LC.

Materials and methods

Experimental design. The present work was a clinical, longitudinal, prospective, observational and analytical study, using a cohort of lung cancer patients to select a non-probabilistic sample type.

Patients and tissue samples. From January 2015 to December 2017, patients admitted to the Instituto Nacional de Cancerología (INCan) with a pulmonary lesion suggestive of primary lung carcinoma were biopsied prospectively. Lung tissues were obtained by computer tomography-guided tru-cut (CareFusion, San Diego, CA, USA) in the clinically suspected primary tumor after informed consent was obtained.

Patients with histologic confirmed locally advanced and metastatic lung cancer (stages III B and IV) were eligible for inclusion in the present study. If there was a histologic report that did not indicate primary lung cancer, the patient was excluded. A complete clinical-medical history was included, and all lung tumor specimens were collected after confirmation of diagnosis. This study was conducted according to the principles of the World Medical Association Declaration of Helsinki "Ethical Principles for Medical Research Involving Human Subjects"⁴¹. All experimental protocols were approved by the Scientific and Bioethical committees at INCan (014/009/ICI, CEI/870/14). All patients provided a signed written consent to participate in genotyping/genomic studies. Primary tumor core biopsy was performed before any treatment, and the specimen was snap-frozen in liquid nitrogen for DNA extraction. A total of 88 tumor samples were included in the study (Consort).

Statistical analysis. Continuous variables were tabulated as medians with ranges, or as means with standard deviations (SDs), depending on data's distribution. The distribution was assessed using the Shapiro–Wilk test with a P-value greater than 0.05 considered as normally distributed. Two group comparisons were tested using Student's t test or Mann–Whitney U depending on data's distribution. Nominal data was analysed using the chi square (X^2) test. All data were analyzed using the SPSS package v. 20 (SPSS, Inc., Chicago, Ill, US) following methods previously reported by this research group⁴².

DNA isolation. DNA was extracted from frozen tumor biopsies, weighted, and cryo-fractured in liquid nitrogen. The procedure for extraction and purification of total DNA from tissue (up to 5 mg tissue) was performed using QIAGEN QIAamp DNA UCP Micro Kit (Cat. 5204).

End-point and in-situ PCR. End-point PCR evaluated DNA from tumor tissue for the MTb transposon IS6110. The DNA concentration was assessed using ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and the quality by agarose gel. End-Point PCR was performed in a 25 μ l mixture containing 100 ng of DNA, HotStar Taq Master Mix (Qiagen) and the following IS6110 primers IS-1 5'CCTGCG AGCGTAGGCGTCGG'3 and IS-2 5'CTCGTCCAGCGCCGCTTCGG'3, as previously described⁴³. The PCRs were performed with the following cycling conditions: hold at 95 °C for 15 min; complete 40 cycles of denaturation at 95 °C for 30 s; annealing at 70 °C for 30 s followed by extension at 72 °C for 45 s and a final extension at 72 °C for 10 min. The presence of the transposon product of 123 bp was evaluated in a 2% agarose gel. As a positive control, we used MTb H37Rv DNA^{35,44}. The samples that resulted positive for IS6110 End-Point PCR were selected for in-situ PCR with the same primers for IS6110, as previously reported³⁵. Briefly, 4 μ m sections were obtained using a microtome apparatus from paraffin-embedded tumor tissue and placed on electrostatically charged slides. Incubation removed paraffin at 60 °C for 20 min and then hydrated gradually in xylene, alcohol, and DEPC-water. Tissue permeabilization was performed with chlorhydric acid (0.02 M) for 10 min, then digested with proteinase K (1 mg/Lt) at 37 °C for 30 min and fixed with 20% acetic acid. The reaction mix contained the FastStart Taq DNA polymerase (Roche), dNTPs couples with digoxigenin (PCR DIG labeling Mix, Roche) and the following primers IS-1 5'CCTGCGAGCGTAGGCGTCGG'3 and IS-2 5'CTCGTCCAGCGC CGCTTCGG'3. The in-situ PCR was performed using Amplicover discs and AmpliClips system from Applied Biosystems in a Touchgene thermo cycler, and the cycling conditions were the same as the end-point PCR mention above. The PCR products were detected with an anti-digoxigenin antibody (1:500) incubated for 30 min in a wet chamber; NBT/BCIP (1:50) was used as a substrate incubated for 30 min in a wet chamber and Nuclear Fast Red as counterstaining³⁵.

Shotgun sequencing. Two out of 5 samples that were positive for IS6110 transposon by end-point and in-situ PCRs were subjected to massive sequencing of the total DNA. To this end, the Truseq Nano DNA HT kit (Illumina USA, Cat. No. 20015965) was used to construct each library, starting from 1 µg of tumor genomic DNA, according to the manufacturer's instructions. The massive sequencing was performed using the Illumina HiSeq 4000 platform at Novogene (CA, USA).

Bioinformatic analysis. The pair-end reads were quality checked and trimmed (>Q20) using fastx toolkit and trimmomatic software's. Then, the quality-filtered reads were mapped against 23 complete genomes of *Mycobacterium tuberculosis* strains downloaded from NCBI (Supplementary Table 1) using SMALT (<https://sourceforge.net/projects/smalt/>). The read count to each genome was obtained using Samtools and Artemis tools, respectively^{45,46}. Deep coverage plots were generated using DNAPlotter⁴⁷.

Received: 5 June 2021; Accepted: 17 January 2022

Published online: 09 February 2022

References

1. World Health Organization. *Global Tuberculosis Report 2020: Executive Summary*. 2020. <https://www.who.int/publications/i/item/9789240013131>. Accessed 20 Jan 2021.
2. Siegel, R. L. *et al.* Cancer statistics, 2021. *CA Cancer J. Clin.* **71**(1), 7–33 (2021).
3. OMS. *Tuberculosis*. 2018 September 18, 2018 [cited 2018 November]. <http://www.who.int/mediacentre/factsheets/fs104/es/index.html>. Accessed 20 Jan 2021.
4. Houben, R. M. & Dodd, P. J. The global burden of latent tuberculosis infection: A re-estimation using mathematical modelling. *PLoS Med.* **13**(10), e1002152 (2016).
5. Arrieta, O. *et al.* National consensus of diagnosis and treatment of non-small cell lung cancer. *Rev. Investig. Clin.* **65**(Suppl 1), S5–S84 (2013).
6. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **136**(5), E359–E386 (2015).
7. Arrieta, O. *et al.* Genotyping non-small cell lung cancer (NSCLC) in Latin America. *J. Thorac. Oncol.* **6**(11), 1955–1959 (2011).
8. Corrales, L. *et al.* Lung cancer in never smokers: The role of different risk factors other than tobacco smoking. *Crit. Rev. Oncol. Hematol.* **148**, 102895 (2020).
9. Arrieta, O. *et al.* Clinical and pathological characteristics, outcome and mutational profiles regarding non-small-cell lung cancer related to wood-smoke exposure. *J. Thorac. Oncol.* **7**(8), 1228–1234 (2012).
10. Alberg, A. J. *et al.* Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* **143**(5 Suppl), e1S–e29S (2013).
11. Liang, H. Y. *et al.* Facts and fiction of the relationship between preexisting tuberculosis and lung cancer risk: A systematic review. *Int. J. Cancer* **125**(12), 2936–2944 (2009).
12. Nalbandian, A. *et al.* Lung carcinogenesis induced by chronic tuberculosis infection: The experimental model and genetic control. *Oncogene* **28**(17), 1928–1938 (2009).
13. Bouvard, V. *et al.* A review of human carcinogens—Part B: Biological agents. *Lancet Oncol.* **10**(4), 321–322 (2009).
14. International Agency for Research on Cancer. *Monographs on the Evaluation of Carcinogenic Risks to Humans. A Review of Carcinogen—Part B: Biological Agents*. (International Agency for Research on Cancer, 2011).
15. Cicens, S. & Vencevicius, V. Lung cancer in patients with tuberculosis. *World J. Surg. Oncol.* **5**, 22 (2007).
16. Littman, A. J. *et al.* Prior lung disease and risk of lung cancer in a large prospective study. *Cancer Causes Control* **15**(8), 819–827 (2004).
17. Engels, E. A. *et al.* Tuberculosis and subsequent risk of lung cancer in Xuanwei, China. *Int. J. Cancer* **124**(5), 1183–1187 (2009).
18. Yu, Y. H. *et al.* Increased lung cancer risk among patients with pulmonary tuberculosis: A population cohort study. *J. Thorac. Oncol.* **6**(1), 32–37 (2011).
19. Molina-Romero, C., Arrieta, O. & Hernandez-Pando, R. Tuberculosis and lung cancer. *Salud Publica Mex* **61**(3), 286–291 (2019).
20. Wong, J. Y. Y. *et al.* Tuberculosis infection and lung adenocarcinoma: Mendelian randomization and pathway analysis of genome-wide association study data from never-smoking Asian women. *Genomics* **112**, 1223–1232 (2019).
21. Abdeahad, H. *et al.* Previous pulmonary tuberculosis enhances the risk of lung cancer: Systematic reviews and meta-analysis. *Infect. Dis. (Lond)* 1–14. <https://doi.org/10.1080/23744235.2021.2006772> (2021).
22. Heuvers, M. E. *et al.* History of tuberculosis as an independent prognostic factor for lung cancer survival. *Lung Cancer* **76**(3), 452–456 (2012).
23. Luo, Y. H. *et al.* Association between tumor epidermal growth factor receptor mutation and pulmonary tuberculosis in patients with adenocarcinoma of the lungs. *J. Thorac. Oncol.* **7**(2), 299–305 (2012).
24. Chang, C. H. *et al.* Gender-based impact of epidermal growth factor receptor mutation in patients with nonsmall cell lung cancer and previous tuberculosis. *Medicine (Baltimore)* **94**(4), e444 (2015).
25. Zhang, J. *et al.* Intratumoral epiregulin is a marker of advanced disease in non-small cell lung cancer patients and confers invasive properties on EGFR-mutant cells. *Cancer Prev. Res. (Phila)* **1**(3), 201–207 (2008).
26. Coussens, L. M. & Werb, Z. Inflammation and cancer. *Nature* **420**(6917), 860–867 (2002).
27. Engels, E. A. Inflammation in the development of lung cancer: Epidemiological evidence. *Expert Rev. Anticancer Ther.* **8**(4), 605–615 (2008).
28. Dheda, K. *et al.* Lung remodeling in pulmonary tuberculosis. *J. Infect. Dis.* **192**(7), 1201–1209 (2005).
29. Cooper, A. M. & Khader, S. A. The role of cytokines in the initiation, expansion, and control of cellular immunity to tuberculosis. *Immunol. Rev.* **226**, 191–204 (2008).
30. Auerbach, O., Garfinkel, L. & Parks, V. R. Scar cancer of the lung: increase over a 21 year period. *Cancer* **43**(2), 636–642 (1979).
31. Ardies, C. M. Inflammation as cause for scar cancers of the lung. *Integr. Cancer Ther.* **2**(3), 238–246 (2003).
32. Ho, L. J. *et al.* Increased risk of secondary lung cancer in patients with tuberculosis: A nationwide, population-based cohort study. *PLoS ONE* **16**(5), e0250531 (2021).
33. Pagano, J. S. *et al.* Infectious agents and cancer: Criteria for a causal relation. *Semin. Cancer Biol.* **14**(6), 453–471 (2004).
34. Gunisha, P. *et al.* Polymerase chain reaction using IS6110 primer to detect *Mycobacterium tuberculosis* in clinical samples. *Indian J. Pathol. Microbiol.* **44**(2), 97–102 (2001).
35. Hernandez-Pando, R. *et al.* Persistence of DNA from *Mycobacterium tuberculosis* in superficially normal lung tissue during latent infection. *Lancet* **356**(9248), 2133–2138 (2000).

36. Arrieta, O. *et al.* Updated frequency of EGFR and KRAS mutations in nonsmall-cell lung cancer in Latin America: The Latin-American Consortium for the Investigation of Lung Cancer (CLICaP). *J. Thorac. Oncol.* **10**(5), 838–843 (2015).
37. An, S. J. *et al.* Effects of age on the association between pulmonary tuberculosis and lung cancer in a South Korean cohort. *J. Thorac. Dis.* **12**(3), 375–382 (2020).
38. Hwang, I. K., Paik, S. S. & Lee, S. H. Impact of pulmonary tuberculosis on the EGFR mutational status and clinical outcome in patients with lung adenocarcinoma. *Cancer Res. Treat.* **51**(1), 158–168 (2019).
39. Ioerger, T. R. *et al.* Genome analysis of multi- and extensively-drug-resistant tuberculosis from KwaZulu-Natal, South Africa. *PLoS One* **4**(11), e7778 (2009).
40. Riley, D. R. *et al.* Bacteria-human somatic cell lateral gene transfer is enriched in cancer samples. *PLoS Comput. Biol.* **9**(6), e1003107 (2013).
41. General Assembly of the World Medical Association. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *J. Am. Coll. Dent.* **81**(3), 14–18 (2014).
42. Barrera, L. *et al.* Levels of peripheral blood polymorphonuclear myeloid-derived suppressor cells and selected cytokines are potentially prognostic of disease progression for patients with non-small cell lung cancer. *Cancer Immunol. Immunother.* **67**(9), 1393–1406 (2018).
43. Donoghue, H. D. *et al.* *Mycobacterium tuberculosis* complex DNA in calcified pleura from remains 1400 years old. *Let. Appl. Microbiol.* **27**(5), 265–269 (1998).
44. Cornejo-Granados, F. *et al.* Secretome prediction of two *M. tuberculosis* clinical isolates reveals their high antigenic density and potential drug targets. *Front. Microbiol.* **8**, 128 (2017).
45. Carver, T. *et al.* Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**(4), 464–469 (2012).
46. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**(16), 2078–2079 (2009).
47. Carver, T. *et al.* DNAPlotter: Circular and linear interactive genome visualization. *Bioinformatics* **25**(1), 119–120 (2009).

Author contributions

Conception and design: O.A., A.O.L., R.H.P. Data collection: C.M.R., F.C.G., B.M., A.O.G., M.O.M. Contributed data or analysis tools: C.M.R., O.A., A.F.C., B.M., A.O.L., R.H.P. Performed the analysis: C.M.R., F.C.G., B.M., A.A.S., G.L.L., A.O.G. Manuscript writing: C.M.R., A.F.C., O.A. Manuscript editing: C.M.R., A.F.C., O.A. Final approval: all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-05749-z>.

Correspondence and requests for materials should be addressed to O.A. or R.H.-P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022