Clinical and molecular characterization of a large Brazilian lung cancer cohort: a real-world observational study

Rodrigo de Oliveira Cavagna,[a] Flávia Escremim de Paula,[b] Gustavo Noriz Berardinelli,[b] Murilo Bonatelli,[b] Iara Santana,[c] Monise Tadin Reis,[c] Eduardo Caetano Albino da Silva,[c] Beatriz Garbe Zaniolo,[a,l] Josiane Mourão Dias,[d] Flávio Augusto Ferreira da Silva,[d] Carlos Eduardo Baston Silva,[d] Roberta Martins Queiroz Barbosa,[d] Flávio Henrique de Queiroz,[d] Ingrid Marriel Ramos Novais,[d] Rodrigo Sampaio Chiarantano,[e] Gabryella Severino Bueno,[f] José Favoreto Neto,[f] Lucas Ferreira,[f] Lucas Veras,[g] Alexandre Arthur Jacinto,[h] Rachid Eduardo Noleto da Nóbrega Oliveira,[i] Welinton Yoshio Hirai,[j] Miguel A. Molina-Vila,[k] Gustavo Teixeira,[c,l] Letícia Ferro Leal,[a,l,**] and Rui Manuel Reis[a,b,m,*]

[a]Molecular Oncology Research Center, R. Antenor Duarte Viléla, 1331 - Dr. Paulo Prata, Barretos, São Paulo, 14784-400, Brazil
[b]Molecular Diagnostic Laboratory, R. Antenor Duarte Viléla, 1331 - Dr. Paulo Prata, Barretos, São Paulo, 14784-400, Brazil
[c]Department of Pathology, R. Antenor Duarte Viléla, 1331 - Dr. Paulo Prata, Barretos, São Paulo, 14784-400, Brazil
[d]Department of Medical Oncology, R. Antenor Duarte Viléla, 1331 - Dr. Paulo Prata, Barretos, São Paulo, 14784-400, Brazil
[e]Department of Radiology, Barretos Cancer Hospital, R. Antenor Duarte Viléla, 1331 - Dr. Paulo Prata, Barretos, São Paulo, 14784-400, Brazil
[f]Department of Medical Oncology, Barretos Cancer Hospital - Amazônia, BR-364, KM 15, Porto Velho, Rondônia, 76834-899, Brazil
[g]Department of Pathology, Barretos Cancer Hospital - Amazônia, BR-364, KM 15, Porto Velho, Rondônia, 76834-899, Brazil
[h]Department of Radiation Therapy, R. Antenor Duarte Viléla, 1331 - Dr. Paulo Prata, 14784-400, Barretos, São Paulo, Brazil
[i]Deparment of Thoracic Surgery, R. Antenor Duarte Viléla, 1331 - Dr. Paulo Prata, 14784-400, Barretos, São Paulo, Brazil
[j]Department of Epidemiology and Biostatistics, Barretos Cancer Hospital, R. Antenor Duarte Viléla, 1331 - Dr. Paulo Prata, 14784-400, Barretos, São Paulo, Brazil
[k]Laboratory of Oncology/Pangaea Oncology, Dexeus University Hospital, Carrer de Sabino Arana, 5, Les Corts, Barcelona, 08028, Spain
[l]Barretos School of Health Sciences Dr. Paulo Prata – FACISB, Av. Loja Maçonica Renovadora N° 68, n° 100 - Barretos, São Paulo, 14785-002, Brazil
[m]Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Campus de Gualtar, Universidade do Minho, Braga, 4710-057, Portugal

The Lancet Regional Health - Americas 2026;56: 101429

Published Online xxx https://doi.org/10.1016/j.lana.2026.101429

Summary

Background Driver alterations influence lung cancer management and vary among ethnicities. Patients from Latin America remain underrepresented in genomic studies. We characterized the molecular and ancestry profiles of Brazilian patients with lung cancer using real-world data and evaluated associations with clinicopathological features.

Methods We retrospectively analyzed 1131 patients with lung cancer from a referral center, using DNA/RNA-based Next-generation Sequencing (NGS), immunohistochemistry, and ancestry-informative markers to assess molecular profiles and associate them with clinical features.

Findings Oncogenic alterations were detected in 988 (88%) of patients, mainly at *TP53 (Tumor Protein p53)* [656 (58%)], *KRAS* (*Kirsten Rat Sarcoma Viral*) [289 (25.6%)], and *EGFR (Epidermal Growth Factor Receptor)* [228 (20.6%)] genes. *TP53* mutations were associated with former smoking (OR: 2.04, 95% CI: 1.43–2.90), current smoking (OR: 3.79, 95% CI: 2.65–5.41), central nervous system (CNS) metastases (OR: 1.75, 95% CI: 1.18–2.58), and higher African ancestry (OR: 1.53, 95% CI: 1.08–2.18). *KRAS* mutations were associated with former smoking (OR: 6.47, 95% CI: 3.59–11.68) and current smoking (OR: 7.42, 4.13–13.31). *EGFR* mutations were associated with never smoking (OR: 12.87, 95% CI 7.12–23.2). Worse cancer-specific survival was observed among patients who currently smoke (HR: 1.38, 95% CI: 1.07–1.78), with poor performance status (HR: 1.99, 1.64–2.41), and those with CNS metastases (HR: 6.38, 4.80–8.47). In patients with *TP53* mutations, treatment with chemotherapy was associated with longer survival (15.0 *vs.* 2.0 months; log-rank p < 0.0001). In the subset of patients harboring *EGFR* mutations and treated with targeted inhibitors, *TP53* co-mutations were associated with shorter cancer-specific survival (24.0 *vs.* 61.0 months; log-rank p = 0.003).

*Corresponding author. Molecular Oncology Research Center, Rua Antenor Duarte Villela, 1331, CEP, 14784 400, Barretos, S. Paulo, Brazil.
**Corresponding author. Molecular Oncology Research Center, Rua Antenor Duarte Villela, 1331, CEP, 14784 400, Barretos, S. Paulo, Brazil.
E-mail addresses: ruireis.hcb@gmail.com, rreis@med.uminho.pt (R.M. Reis), leticiaferro@hotmail.com, leticia.leal@facisb.edu.br (L.F. Leal).

Interpretation Most Brazilian patients with lung cancer harbor actionable genomic alterations. *TP53* status is prognostically relevant, including in the *EGFR*-mutant disease, and should be routinely incorporated. It provides region-specific evidence to inform equitable access to molecular diagnostics and targeted therapies in Latin America.

Funding Public Ministry of Labor Campinas (Research, Prevention, and Education of Occupational Cancer—15th zone, Campinas, Brazil), PRONON—PRONON/MS (Abordagens móveis e de tecnologia para prevenção primária e secundária de câncer—NUP: 25000.015000/2019-53), and Barretos Cancer Hospital.

Keywords: Lung cancer; Brazil; Molecular profile; Mutations; Fusions; Ancestry

Research in context

Evidence before this study

We searched PubMed and Web of Science from Oct 2020 to Jan 2025, with periodic updates during the study period, using the terms "lung cancer", "non-small cell lung cancer", "Brazil", "Latin America", "molecular profiling", "next-generation sequencing", "gene fusions", "genetic ancestry", "driver mutations", "actionable alterations", "smoking", "survival", and "clinicopathological features". We prioritized studies reporting molecular profiling and/or outcomes in Brazilian or Latin American populations and, when available, analyses incorporating genetic ancestry. Previous evidence shows that lung cancer is molecularly heterogeneous and that the distribution of actionable alterations varies across populations. In Brazil, available studies have most often focused on the alteration frequency of selected targets *(Epidermal Growth Factor Receptor [EGFR], Kirsten Rat Sarcoma Virus [KRAS],* and *Anaplastic Lymphoma Kinase [ALK])*, often limited by small sample sizes and restricted to single regions of the country. Evidence linking comprehensive genomic profiles with genetic ancestry and outcomes in Brazilian real-world scenarios remains scarce.

Added value of this study

This study provides a large real-world molecular characterization of lung cancer in Brazil, integrating next-generation sequencing–based profiling, gene fusion assessment, Programmed death-ligand 1 (PD-L1) immunohistochemistry, and genetic ancestry estimates in a nationwide referral-center cohort. By combining genomic data with clinicopathological features and survival, we describe regional variation in alteration frequencies and clarify associations between genetic ancestry, smoking exposure, and key oncogenic alterations. Treatment with chemotherapy showed better prognosis in patients harboring *Tumor Protein p53 (TP53)* mutations. Moreover, in patients with *EGFR*-mutant tumors treated with tyrosine kinase inhibitors, *TP53* co-mutations were associated with poorer cancer-specific survival, suggesting a potential role for treatment stratification in this subgroup in routine practice.

Implications of all the available evidence

Together with existing evidence, our findings support the need for comprehensive molecular profiling in lung cancer care in Brazil and other highly admixed populations. The consistent association between *TP53* alterations and poorer outcomes in *EGFR*-mutant disease, alongside the improved prognosis observed in *TP53*-mutant patients receiving chemotherapy, suggests that *TP53* status may provide clinically relevant information to support therapeutic decision-making and potential treatment stratification in routine practice. These data also highlight the importance of strengthening molecular diagnostic capacity within publicly funded health systems to reduce regional disparities and promote equitable implementation of precision oncology. Prospective studies are needed to validate these real-world observations and to determine the most effective way to integrate expanded molecular testing into routine care in resource-constrained settings.

Introduction

Lung cancer is the most common and deadliest malignancy worldwide.[1] Non-small cell lung cancer (NSCLC) is the most frequent histological type, accounting for 85% of cases.[2] Molecularly, NSCLC is a highly heterogeneous disease, characterized by alterations in multiple driver genes, including the most common mutations in *Epidermal Growth Factor Receptor* (*EGFR*), *V-raf Murine Sarcoma Viral Oncogene Homolog B* (*BRAF*), *Erb-B2 Receptor Tyrosine Kinase 2* (*ERBB2*), and *Kirsten Rat Sarcoma Virus* (*KRAS*), as well as gene rearrangements involving *Anaplastic Lymphoma Kinase* (*ALK*), *Rearranged During Transfection* (*RET*), *ROS Proto-Oncogene 1, Receptor Tyrosine Kinase* (*ROS1*), *and Neurotrophic Receptor Tyrosine Kinase* (NTRK)1, *2, and 3*.[2–7] Of note, this molecular profile can vary significantly based on factors such as smoking history and the patient's ethnicity.[8–10] Clinically, in recent decades, targeted therapies such as TKIs (tyrosine kinase inhibitors) and other specific small molecules targeting additional transmembrane and intracellular kinases have revolutionized the clinical management of

patients diagnosed with lung cancer.[2] This treatment is designed to target cancer cells harboring specific actionable alterations.[2] Consequently, international guidelines for molecular testing have emerged as crucial for identifying actionable mutations in NSCLC, guiding targeted therapies.[2,7,11]

EGFR mutations and *ALK* fusions are successful examples of targeted therapies, as patients harboring these alterations are eligible for treatment with anti-EGFR and anti-ALK agents, resulting in a significant improvement in overall survival.[7,12] Despite the efficacy of TKIs, the presence of co-occurring alterations may modify the efficacy of targeted therapies or even cause resistance to them.[2,7,13] Therefore, beyond identifying actionable mutations, understanding the broader mutational context is essential to anticipate resistance mechanisms and optimize treatment options.

The Brazilian population is highly admixed, reflecting its diverse migration origins and distinct ancestral contributions, including Europeans, Africans, Asians, and Native Americans.[14,15] This high admixture may lead to a distinctive mutational profile, since many mutations may be associated with a specific genetic ancestry background.[9,16–18] However, the molecular profile of lung tumors in Brazil remains poorly understood, and a comprehensive, large-scale molecular profile, along with its clinical impact, is critically needed.

Therefore, in this study, we evaluated the mutational status of key driver alterations in a cohort of 1131 Brazilian patients with lung cancer, profiled their genetic ancestry, and investigated associations with clinicopathological features in a real-world setting.

Methods

Study population

A retrospective series of 1131 consecutive Brazilian patients with lung cancer, diagnosed between 2018 and 2023, was evaluated from two units of Barretos Cancer Hospital, a tertiary philanthropic referral center operating within the Brazilian Public Health System (SUS), with nationwide patient referral pathways: the central unit in Barretos, São Paulo State (93.4%; n = 1056), and the regional unit in Porto Velho-Rondônia, in the Ocidental Amazonia region (6.6%; n = 75). Inclusion criteria were a confirmed diagnosis of lung cancer and participation in routine molecular testing at the Laboratory of Molecular Diagnostics of Barretos Cancer Hospital, using a predefined, sequenced reflex test as reported.[19] Exclusion criteria included patients with biopsies containing less than 20% tumor content, DNA quantification below 2ng/µl, and those with inconclusive next-generation sequencing (NGS) results (Supplementary Figure S1). Briefly, all cases were evaluated for single-nucleotide variants using an NGS panel, followed by mRNA fusion panels in cases without actionable mutations (*EGFR*, *KRAS*, *BRAF*, and *ERB2*). ALK and PD-L1 (programmed death-ligand 1) immunohistochemistry (IHC) were performed for the majority of cases, and, in a subset, fluorescence in situ hybridization (FISH) and the cobas® EGFR Mutation Test v2 assay were also used as validation methods (Supplementary Figure S1).

A sample size calculation was performed to ensure that the cohort had sufficient statistical power to detect rare variants (occurring at a frequency of 1% or less). The calculation was based on the standard sample size formula $n = \left(Z^2 \cdot p(1-p)\right)/E^2$, where n is the required sample size, p is the expected proportion of the variant (0.01), E is the acceptable margin of error (5%), and Z is the z-score corresponding to the desired confidence level (typically 1.96 for 95% confidence). The calculation also considered the mutual exclusivity of alterations in key oncogenes, including *EGFR, KRAS, ALK, ERBB2, BRAF, RET,* and *ROS1*, with a significance level of 0.05. Based on these parameters, a minimum of 1000 patients was deemed sufficient to detect rare variants with adequate power. Therefore, the final cohort of 1131 patients met the sample size requirements established to identify meaningful molecular associations.

Sociodemographic, clinicopathological, and molecular data were collected from patients' medical records. Ethnicity was treated as a sociocultural construct and recorded based on patients' self-reported skin color, according to the official criteria of the Brazilian Institute of Geography and Statistics (IBGE).[20] Associations and survival analyses by ethnicity were performed for descriptive and exploratory purposes only and were not interpreted as reflecting biological differences. Genetic ancestry estimates were prioritized when addressing population structure.

The Institutional Review Board approved the study protocol (CAAE 05744712.3.0000.5437) and waived written informed consent due to the study's retrospective nature.

DNA and RNA isolation

DNA and RNA isolation were performed from FFPE (formalin-fixed paraffin-embedded) tumor tissue sections as reported.[19] Briefly, one slide was stained with hematoxylin and eosin (H&E) and evaluated by a pathologist for sample adequacy and selection of the tumor area. DNA was isolated using the commercial kit QIAamp DNA Micro Kit (Qiagen, Hilden, Germany) according to the manufacturer's specifications. DNA concentration and purity were evaluated by NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA) and by Qubit 2.0 Fluorometer (Thermo Fisher Scientific) with Qubit dsDNA HS assay kit (Thermo Fisher Scientific). RNA was isolated using the RNeasy FFPE Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. RNA concentration and purity were measured using the NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA).

Mutation detection and classification

Mutational analysis (n = 1131) was performed with the TruSight Tumor 15 (TST15) panel (Illumina, San Diego, CA, USA) using NGS on the MiSeq instrument, following the manufacturer's instructions as previously described.[21] The read alignment and variant calling were performed with BaseSpace BWA Enrichment version 2.1 (Illumina, San Diego, CA, USA) and Sophia DDM® software version 4.2 (Sophia Genetics SA, Saint Sulpice, Switzerland). Variants were filtered out according to the following criteria: intronic (except splicing variants), synonymous single nucleotide variants (SNVs), population frequency >1%, low read depth <500, allele frequency <5%, and non-pathogenic variants according to Sophia DDM®.

The retained variants were classified based on oncogenicity, using the scoring system proposed by the most recent guidelines as follows: Oncogenic, Likely-Oncogenic, VUS (variant of undetermined significance), Likely-benign, and Benign.[22] These variants were then classified according to the clinical impact into four categories (TIERs): actionable oncogenic and likely-oncogenic variants with targeted therapies (TIER I), oncogenic and likely-oncogenic variants with potential clinical significance (TIER II), VUS variants with unknown clinical significance (TIER III), and likely benign and benign variants (TIER IV).[23] Finally, only the oncogenic and likely-oncogenic variants were retained.

A subset of patients (n = 81/1131), the NGS results of *EGFR* status were also validated by real-time PCR using the cobas® EGFR Mutation Test v2 (Roche Diagnostics, Basileia, Switzerland) as previously described.[24]

Gene fusions and *METex14* detection

Distinct methodologies (Supplementary Figure S1) were used to assess the presence of *ALK/RET/ROS1/*

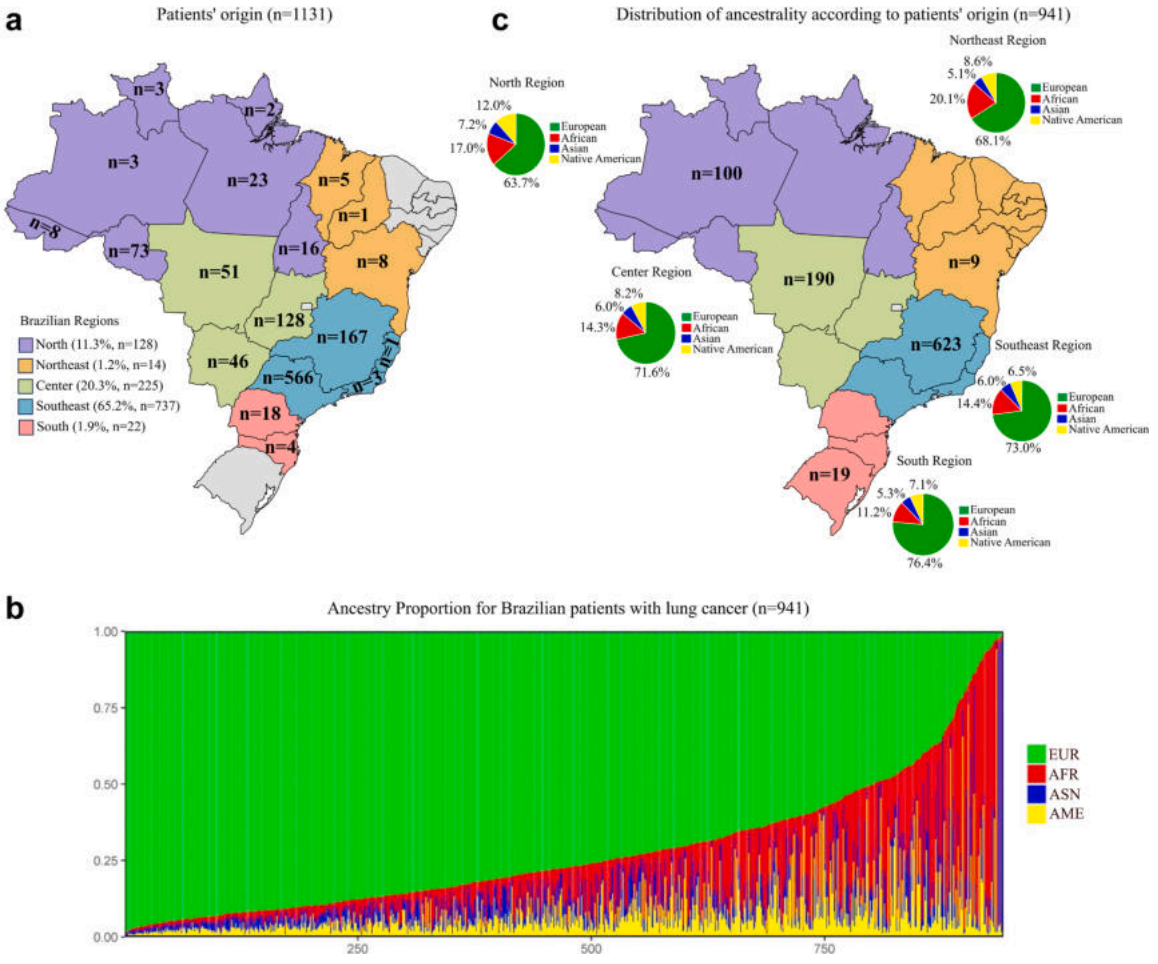

Fig. 1: Geographic distribution and ancestry analysis of patients. (a) Distribution by state and region of Brazil, based on the address provided at diagnosis (n = 1131). States without patients are shown in gray; (b) Distribution of ancestry proportions for each ancestry group (n = 941). The X-axis represents each patient, while the Y-axis indicates the proportion of ancestry; (c) Distribution of ancestry proportions for each Brazilian region, based on the address provided at diagnosis (n = 941).

NTRK1,2,3 fusions, and *METex14* (*MET* exon 14 skipping).

mRNA fusion transcripts were analyzed in 40.3% of the samples (n = 456/1131) using RNA-based platforms, including the Archer FusionPlex Custom Solid Panel with Anchored Multiplex PCR (ArcherDX, Boulder, CO, USA), the Idylla™ Gene Fusion Assay RT-PCR kit (Biocartis, Mechelen, Belgium), and a custom nCounter Elements XT panel (NanoString Technologies, Seattle, WA, USA). The NanoString panel was designed to detect *ALK*, *RET*, *ROS1*, and *NTRK1/2/3* fusions by identifying specific fusion partners and assessing 3'/5' expression imbalance, as well as *METex14*.[6] Data obtained from the nCounter custom panel were analyzed using the R environment (version 3.4.1).

In addition to RNA-based assays, *ALK* fusions were also evaluated by immunohistochemistry in 81.4% of cases (n = 921/1131) using the Ventana ALK (D5F3) CDx Assay (Roche, Basel, Switzerland) on an automated BenchMark Ventana Ultra™ platform. Furthermore, FISH was used to evaluate *RET* and *ROS1* fusions in 0.2% (n = 2/1131) and 17.8% (n = 201/1131) of cases, respectively, employing commercial break-apart probes (ZytoLight SPEC RET and ROS1 Dual Color Break Apart; ZytoVision) and the FISHView 7.0 software (Applied Spectral Imaging).[4]

PD-L1 expression by immunohistochemistry

In 790 patients, PD-L1 expression was evaluated by immunohistochemistry using the 22C3 clone (PharmaDx antibody) as described.[25] PD-L1 expression was reported as tumor proportional score (TPS), with results categorized as no expression (<1%), low expression (1–49%), and high expression (≥50%).[25]

Genetic ancestry analysis

In the tumor DNA, the genetic ancestry background was evaluated using a set of 46 ancestry-informative markers (AIMs), including INDELs for European (EUR), African (AFR), Asian (ASN), and Native American (AME) ancestry, as previously reported.[16]

Statistical analysis

For statistical analysis, percentages were used to describe categorical variables, and medians were used to describe continuous variables. As previously described, individual ancestry proportions were defined as categorical variables and divided into tertiles (low, intermediate, and high).[17] The Mann–Whitney U test was used to assess associations involving continuous variables. For univariate analysis, Fisher's exact test and the Chi-square test were used to assess the association between the clinicopathological and molecular data in categorical variables. Multivariate analysis was

Characteristic	Parameters	n	%
Age at diagnosis (years)	Median (Q1–Q3)	64.0 (57–71)	
	≤64	587	51.9
	>64	544	48.1
Sex	Female	508	44.9
	Male	623	55.1
Self-declared skin color[a]	White	708	62.6
	Non-White	231	20.4
	N.I.	192	17.0
Smoking status	Never	297	26.2
	Quitter	372	32.9
	Current	419	37.0
	N.I.	44	3.9
Weight loss within 6 months prior to diagnosis	No	385	34.0
	Yes	520	46.0
	N.I.	226	20.0
ECOG Performance Status	0–1	672	59.4
	2–4	313	27.7
	N.I.	140	12.9
Histology	Adenocarcinoma	909	80.4
	Squamous	65	5.7
	NSCLC NOS	119	10.5
	Others[b]	38	3.4
Stage at diagnosis	I/II	192	17.0
	III	181	16.0
	IV	727	64.3
	N.I.	31	2.7
Metastasis at diagnosis	No	374	33.1
	Yes, CNS	256	22.6
	Yes, others	473	41.8
	No	28	2.5
Evolution after first treatment	No	243	21.5
	Locoregional recurrence	30	2.7
	Systemic recurrence	39	3.4
	Locoregional and systemic recurrence	15	1.3
	Locoregional progression	88	7.8
	Systemic progression	200	17.7
	Locoregional and systemic progression	74	6.5
	Death by cancer-specific	319	28.2
	Lost of follow-up/Missing	123	10.9
Chemotherapy 1st line	No	481	42.5
	Yes, curative	172	15.2
	Yes, palliative	399	35.3
	Lost of follow-up/Missing	79	7.0
Chemotherapy 2nd line	No	173	15.3
	Yes, palliative	262	23.2
	Not applicable	562	49.7
	Lost of follow-up/Missing	134	11,8
Radiotherapy 1st line	No	601	53.1
	Yes, curative	152	13.4
	Yes, palliative	298	26.3
	Lost of follow-up/Missing	80	7.2

(Table 1 continues on next page)

Characteristic	Parameters	n	%
(Continued from previous page)			
Radiotherapy 2nd line	No	270	23.9
	Yes, palliative	166	14.7
	Not applicable	562	49.7
	Lost of follow-up/Missing	133	11.7
Surgery 1st line	No	844	74.6
	Yes, curative	176	15.6
	Yes, palliative	31	2.7
	Lost of follow-up/Missing	80	7.1
Surgery 2nd line	No	424	37.5
	Yes, palliative	12	1.1
	Not applicable	562	49.7
	Lost of follow-up/Missing	133	11.7
TKI 1st line	No	903	79.8
	Yes, curative	2	0.2
	Yes, palliative	146	12.9
	Lost of follow-up/Missing	80	7.1
TKI 2nd line	No	374	33.1
	Yes, palliative	62	5.5
	Not applicable	562	49.7
	Lost of follow-up/Missing	133	11.7
Immunotherapy 1st line	No	993	87.8
	Yes, curative	9	0.8
	Yes, palliative	49	4.3
	Lost of follow-up/Missing	80	7.1
Immunotherapy 2nd line	No	410	36.3
	Yes, palliative	26	2.3
	Not applicable	562	49.7
	Lost of follow-up/Missing	133	11.7

Q1–Q3, interquartile range; N.I., no information available; NSCLC NOS, Non-Small Cell Lung Cancer not otherwise specified; CNS, Central Nervous System. [a]According to the IBGE (Brazilian Institute of Geography and Statistics). [b]Includes Large Cell Carcinoma (n = 4), Adenosquamous Carcinoma (n = 18), Sarcomatoid Carcinoma (n = 1), small cell carcinoma/oat cell (n = 1), atypical carcinoid (n = 1), Large Cell Neuroendocrine Carcinoma (n = 8), No information (n = 2), Undifferentiated Carcinoma (n = 3).

Table 1: **Clinicopathological characteristics of the series (n = 1131).**

conducted using logistic regression, including only variables with a p-value <0.20 in the univariate analysis, to identify independent associations between clinicopathological and molecular features. The log-rank test and Kaplan–Meier method were used to estimate cancer-specific survival (CSS) and time to event (TTE) after first treatment, and Cox proportional hazards regression was applied to assess the association between clinicopathological variables and outcomes. For CSS, the time origin was the date of diagnosis confirmed by biopsy. Cancer-specific death was considered the event. Patients were censored at the date of last follow-up if alive, at the date of loss to follow-up, or at the time of death from causes unrelated to lung cancer. For TTE, the time origin was the date on which patients initiated first-line systemic therapy. Events were defined as any disease progression, recurrence, or cancer-related death. Patients were censored at the date of last follow-up without evidence of progression or recurrence, or at the time of death due to causes other than lung cancer. Missing data were handled using complete case analysis. All analyses were conducted in IBM SPSS Statistics Version 25 (IBM, Armonk, New York, USA) with a limit of statistical significance of 0.05. Kaplan–Meier curves were generated using the function survfit from the package survival and the forest plots were designed using the function forestplot from the package forestplot using R environment version 4.4.2.

Role of the funding source

The funding source had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Results

Demographic, clinicopathological, and genetic ancestry features

The study included a total of 1131 patients diagnosed with lung cancer who were routinely evaluated for a set of key driver genes.

Demographically, based on patients' addresses provided at diagnosis, all Brazilian regions were represented in our cohort, with the majority from the Southeast (65.2%) (Fig. 1a). Genetic ancestry was reliably determined in 941 patients, with the following proportions: 71.7% European, 14.6% African, 7.4% Native American, and 6.1% Asian (Fig. 1b). When stratifying patients by their genetic ancestry and region of residence (n = 941), we observed a difference between the North and other regions of Brazil (Fig. 1c), with a higher proportion of Native American ancestry (12%) and a lower proportion of European ancestry (63.7%) compared to the other regions. Additionally, varying proportions of ancestry were observed across all regions (Fig. 1c).

The main clinicopathological features of the study population are summarized in Table 1. The median age at diagnosis was 64 years (interquartile range [57–71]), most were male (55.1%; n = 623/1131), self-identified as White (62.6%; n = 708/1131), and were current or former smokers (69.9%; n = 790/1131) (Table 1). The great majority of cases (80.4%) were adenocarcinoma. Clinically, 46.0% (n = 520/1131) of patients experienced weight loss within six months before diagnosis, 59.4% (n = 671/1131) had an ECOG (Eastern Cooperative Oncology Group) performance status of 0 or 1 at diagnosis, 64.3% (n = 727/1131) were stage IV at diagnosis, and 22.6% (n = 256/1131) presented with central nervous system (CNS) metastases. Most patients (67.6%; n = 765/1131) experienced progression or recurrence of disease after first treatment, and the great majority received systemic therapies, including chemotherapy, TKIs, and immunotherapy (Table 1).

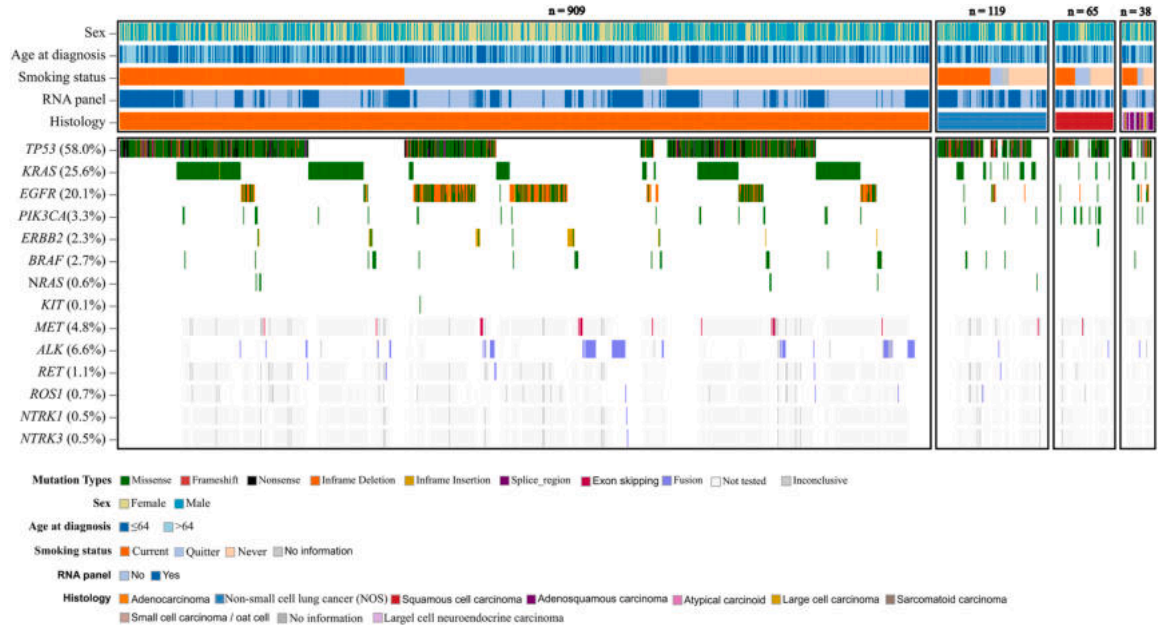

Fig. 2: Molecular profile of Brazilian patients with lung cancer (n = 1131). Only variants classified as oncogenic or likely oncogenic were considered. Patients are classified by histology as follows: Adenocarcinoma, NSCLC NOS (Non-small Cell Lung Cancer, Not Otherwise Specified), squamous cell carcinoma, and Other histologies.

Molecular profile

Molecularly, 88.0% of patients exhibited at least one oncogenic alteration (Fig. 2), and 58.4% carried at least one actionable mutation (TIER I) (Supplementary Table S1). The *TP53* gene was the most commonly mutated, observed in 58.0% (n = 656/1131; Supplementary Figure S2) of cases, followed by *KRAS* mutations in 25.6% (n = 289/1131; Supplementary Figure S3), *EGFR* mutations in 20.1% (n = 228/1131; Supplementary Figure S4), and *ALK* fusions in 6.1% (n = 69/1049). Alterations in *Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha* (*PIK3CA*), *ERBB2*, *BRAF*, *METex14*, *Neuroblastoma Rat Sarcoma Virus* (*NRAS*), *KIT Proto-Oncogene, Receptor Tyrosine Kinase* (*KIT*), *RET*, *ROS1*, *NTRK1*, and *NTRK3* were present in fewer than 5% of cases. No mutations or fusions were found in the *AKT Serine/Threonine Kinase 1* (*AKT1*), *Forkhead Box* L2 (*FOXL2*), *G Protein Subunit Alpha 11* (*GNA11*), *G Protein Subunit Alpha Q* (*GNAQ*), *Platelet Derived Growth Factor Receptor Alpha* (*PDGFRA*), and *NTRK2* genes.

In patients with *EGFR* mutations (n = 228), the most common co-occurring mutations were in the *TP53* gene (58.8%; n = 134/228), followed by *PIK3CA* mutations (3.5%; n = 8/228). Similarly, among patients with *KRAS* mutations (n = 289), the most frequent co-occurring mutations were in *TP53* and *PIK3CA* (50.5%; n = 146/289 and 3.5%; n = 10/289, respectively). *BRAF* mutations (n = 31) commonly co-occurred with *TP53* and *PIK3CA* (35.5%; n = 11/31 and 9.7%; n = 3/31, respectively). *ERBB2* alterations (n = 33) were most commonly concomitant with *TP53* mutations (38.5%; n = 10/26).

For patients harboring *METex14*, *TP53* mutations (52.9%; n = 9/17) were the most frequent co-occurring alteration, and one case also harbored the *KRAS* mutation p.Gly12Cys. *ALK* fusions co-occurred only with *TP53* mutations (26.1%; n = 18/69), and one case co-occurred with the *KRAS* mutation p.Gly12Ala. *ROS1* and *RET* fusions co-occurred with *TP53* mutations (n = 2/4 and n = 4/6, respectively). *NTRK* fusions (n = 2) had no concurrent mutations. A detailed description of co-occurring mutations is provided in Supplementary Table S1.

Geographical tobacco consumption and association with mutational distribution

The frequency of *EGFR* and *KRAS* mutations, *ALK* fusions, and tobacco consumption was evaluated across the five Brazilian regions (Fig. 3a–b). *EGFR* mutations were more frequent in the Northeast (28.6%) and North (25.8%), and *ALK* fusions were more frequent in the South (18.2%). Similarly, the Northeast, South, and North had higher proportions of patients who never smoked (71%, 59%, and 39%, respectively). *KRAS* mutations showed similar frequencies across the Southeast, North, and Center regions (25.4%, 24.2%, and 27.8%, respectively), while a higher frequency was observed in the Northeast (35.7%) and a lower frequency in the South (9.1%).

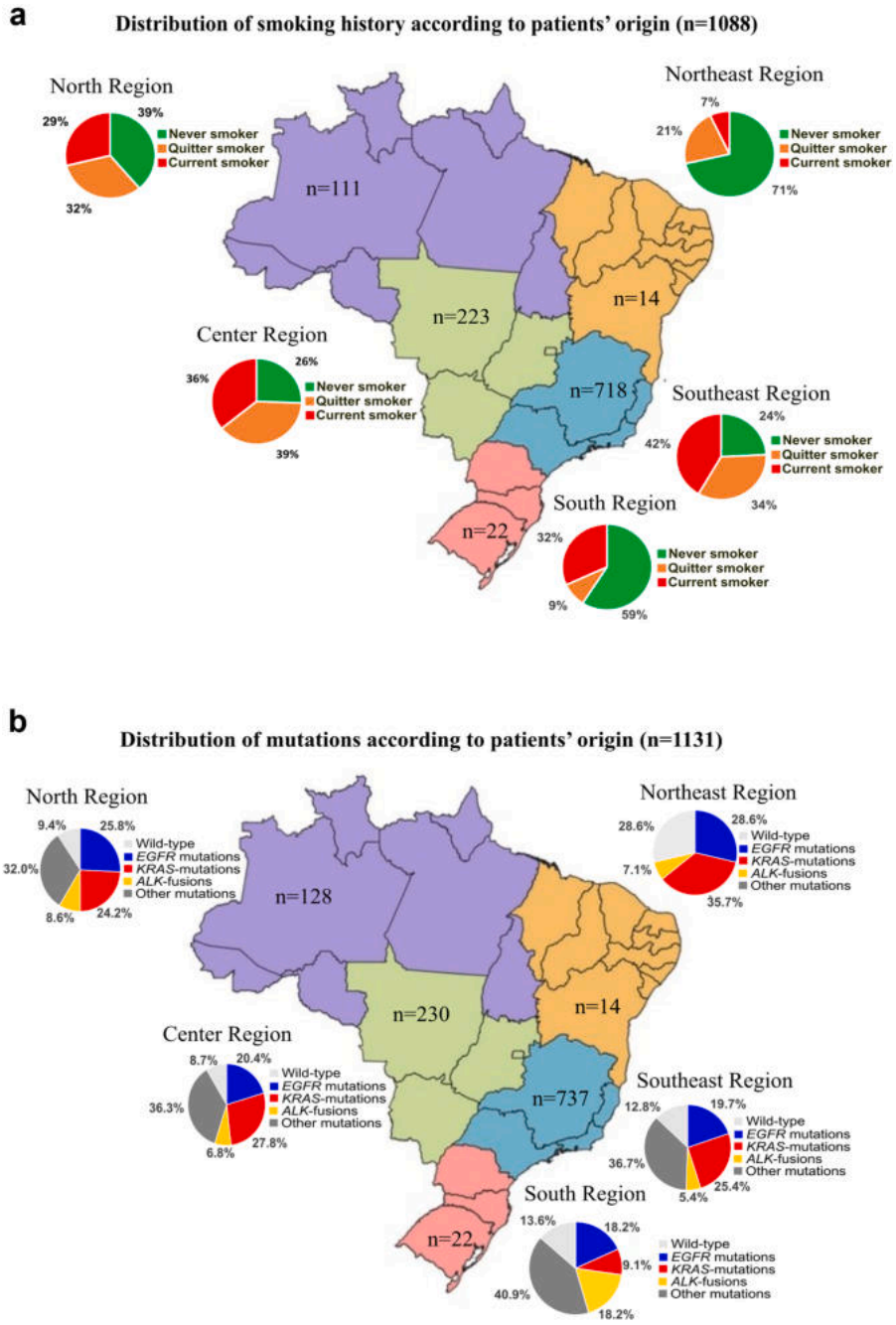

Fig. 3: Distribution of (a) smoking history (n = 1088) and (b) mutations (n = 1131) based on the address provided at diagnosis.

Clinicopathological, ancestry, and molecular associations

We next examined the clinicopathological characteristics and genetic ancestry of patients harboring gene alterations present in more than 2% of cases. Univariate analysis identified several significant associations, which are summarized in Supplementary Table S2. To assess the independence of these associations, a multivariate analysis was performed on the variables found to be significant in the univariate analysis (Fig. 4 and Table 2).

TP53 mutations were independently associated with patients with the following features: male sex, tobacco consumption, non-adenocarcinoma histology, CNS metastasis, and higher African ancestry (Fig. 4a; Table 2). *EGFR* mutations were independently

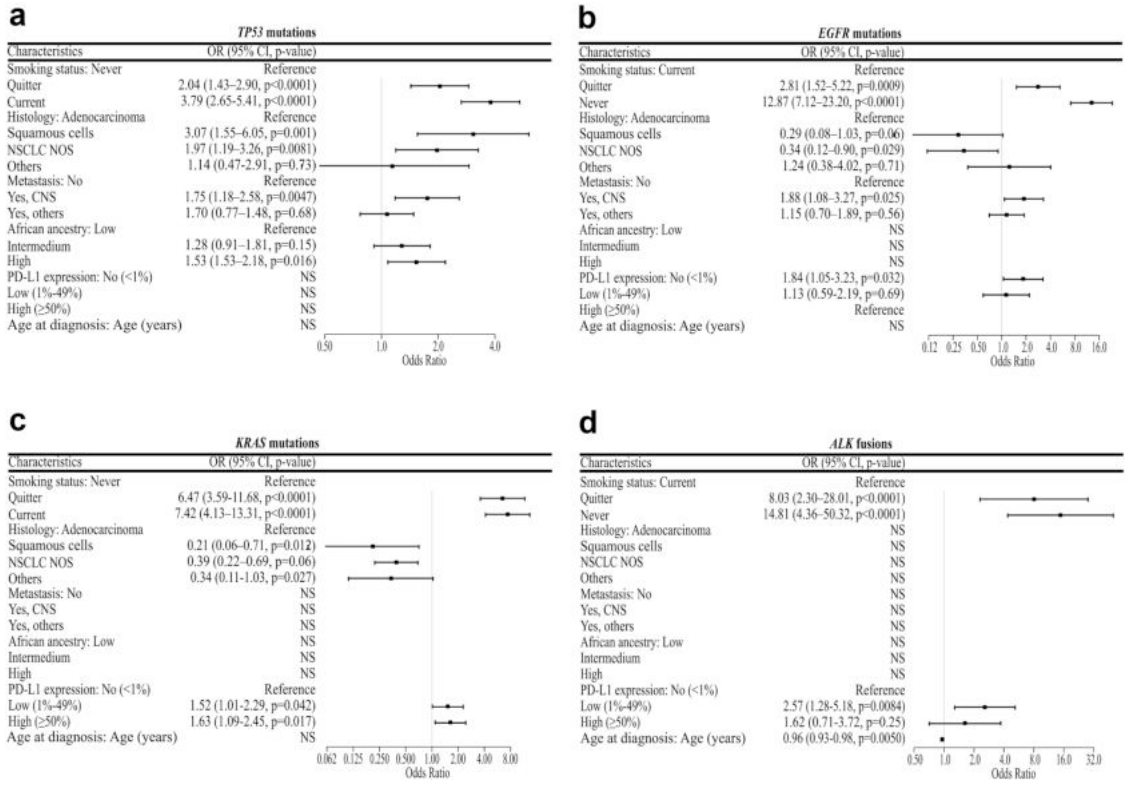

Fig. 4: Forest plot illustrating the multivariate logistic regression analysis of associations between patient features and molecular alterations. Significant associations are indicated for alterations in the following genes: *TP53*(a), *EGFR*(b), *KRAS*(c), and *ALK*(d). OR, odds ratio; NS, Not Significant, in the final multivariate logistic regression model; CI, confidence interval; NSCLC NOS, Non-Small Cell Lung Cancer not otherwise specified.

associated with patients who never smoked or were former smokers, adenocarcinoma histology, CNS metastasis, and no PD-L1 expression (Fig. 4b; Table 2). *KRAS* mutations were independently associated with ever-smoker patients, adenocarcinoma histology, and PD-L1 expression (Fig. 4c; Table 2). *ALK* fusions were associated with younger patients at diagnosis, patients who never smoked, and low PD-L1 expression (Fig. 4d; Table 2). *PIK3CA* mutations were independently associated with older patients at diagnosis and with squamous histology (Supplementary Figure S5a; Table 2). *BRAF* mutations were independently associated with older and female patients (Supplementary Figure S5b; Table 2). Finally, *METex14* mutations were independently associated with patients who never smoked and high expression of PD-L1 (Supplementary Figure S5c; Table 2). *ERBB2* mutations were independently associated with patients who never smoked and low PD-L1 expression (Supplementary Figure S5d; Table 2).

Impact of clinicopathological and molecular features on patient outcomes

The median CSS for the entire patient population was 20.0 months (95% CI: 16.95–23.05; Supplementary Figure S6a), with a median follow-up of 12.0 months (0.0–161.0), and a median TTE of 9.0 months (95% CI: 7.99–10.00; Supplementary Figure S6b). Univariate analysis identified clinical and molecular factors associated with CSS (Table 3). Poorer survival was observed for male sex, tobacco consumption, weight loss prior to diagnosis, higher ECOG score, advanced stage, and CNS metastases (Table 3; Supplementary Figure S7a–f). Conversely, adenocarcinoma or squamous histology, along with systemic therapies (TKIs, chemotherapy, immunotherapy), was associated with improved survival (Table 3; Supplementary Figure S7g–j). At the molecular level, mutations in *TP53*, *KRAS*, and *BRAF* were associated with worse outcomes, whereas *EGFR* mutations and *ALK* fusions were associated with better survival (Table 3; Fig. 5a–d; Supplementary Figure S7k).

Moreover, we applied a multivariate Cox regression analysis to identify patient characteristics independently associated with worse outcomes. Current smoking ($p = 0.011$; HR: 1.38, 95% CI: 1.07–1.78), higher ECOG performance status ($p < 0.0001$; HR: 1.99, 95% CI: 1.64–2.41), non-adenocarcinoma histology ($p = 0.028$; HR: 1.34, 95% CI: 1.03–1.76; and $p = 0.021$;

Characteristics	Parameters	*TP53* mutations				*EGFR* mutations				*KRAS* mutations				*ALK-fusions*			
		OR	95% CI		p-value	OR	95% CI		p-value	OR	95% CI		p-value	OR	95% CI		p-value
			Min	Max			Min	Max			Min	Max			Min	Max	
Sex	Female	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Male	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Smoking status	Never	Ref	Ref	Ref	Ref	12.87	7.12	23.2	**<0.0001**	Ref	Ref	Ref	Ref	14.81	4.36	50.32	**<0.0001**
	Quitter	2.04	1.43	2.90	**<0.0001**	2.81	1.52	5.22	**0.0009**	6.47	3.59	11.68	**<0.0001**	8.03	2.30	28.01	**<0.0001**
	Current	3.79	2.65	5.41	**<0.0001**	Ref	Ref	Ref	Ref	7.42	4.13	13.31	**<0.0001**	Ref	Ref	Ref	Ref
Histology	Adenocarcinoma	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	–	–	–	–
	Squamous	3.07	1.55	6.05	**0.001**	0.29	0.08	1.03	0.06	0.21	0.06	0.71	**0.012**	–	–	–	–
	NSCLC NOS	1.97	1.19	3.26	**0.0081**	0.34	0.12	0.90	**0.029**	0.39	0.22	0.69	0.06	–	–	–	–
	Others[a]	1.14	0.47	2.91	0.73	1.24	0.38	4.02	0.71	0.34	0.11	1.03	**0.027**	–	–	–	–
Metastasis at diagnosis	No	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	–	–	–	–	–	–	–	–
	Yes, CNS	1.75	1.18	2.58	**0.0047**	1.88	1.08	3.27	**0.025**	–	–	–	–	–	–	–	–
	Yes, Others	1.07	0.77	1.48	0.68	1.15	0.70	1.89	0.56	–	–	–	–	–	–	–	–
African ancestry	Low	Ref	Ref	Ref	Ref	–	–	–	–	–	–	–	–	–	–	–	–
	Intermedium	1.28	0.91	1.81	0.15	–	–		–	–	–	–	–	–	–	–	–
	High	1.53	1.08	2.18	**0.016**	–	–	–	–	–	–	–	–	–	–	–	–
PD-L1 expression	No (<1%)	–	–	–	–	1.84	1.05	3.23	**0.032**	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
	Low (1%–49%)	–	–	–	–	1.13	0.59	2.19	0.69	1.52	1.01	2.29	**0.042**	2.57	1.28	5.18	**0.0084**
	High (≥50%)	–	–	–	–	Ref	Ref	Ref	Ref	1.63	1.09	2.45	**0.017**	1.62	0.71	3.72	0.25
Age at diagnosis	Age (years)[b]	–	–	–	–	–	–	–	–	–	–	–	–	0.96	0.93	0.98	**0.0050**

Characteristics	Parameters	*PIK3CA* mutations				*BRAF* mutations				*METΔex14* mutation				*ERBB2* mutations			
		OR	95% CI		p-value	OR	95% CI		p-value	OR	95% CI		p-value	OR	95% CI		p-value
			Min	Max			Min	Max			Min	Max			Min	Max	
Sex	Female	–	–	–	–	2.54	1.19	5.4	**0.015**	–	–	–	–	–	–	–	–
	Male	–	–	–	–	Ref	Ref	Ref	Ref	–	–	–	–	–	–	–	–
Smoking status	Never	–	–	–	–	–	–	–	–	13.05	2.36	72.01	**0.0032**	4.43	1.37	14.30	**0.013**
	Quitter	–	–	–	–	–	–	–	–	3.37	0.57	19.73	0.17	0.84	0.18	3.81	0.82
	Current	–	–	–	–	–	–	–	–	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Histology	Adenocarcinoma	Ref	Ref	Ref	Ref	–	–	–	–	–	–	–	–	–	–	–	–
	Squamous	7.43	3.33	16.16	**<0.0001**	–	–	–	–	–	–	–	–	–	–	–	–
	NSCLC NOS	1.00	0.29	3.41	0.99	–	–	–	–	–	–	–	–	–	–	–	–
	Others[a]	2.38	0.53	10.63	0.25	–	–	–	–	–	–	–	–	–	–	–	–
Metastasis at diagnosis	No	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Yes, CNS	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Yes, Others	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
African ancestry	Low	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Intermedium	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	High	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
PD-L1 expression	No (<1%)	–	–	–	–	–	–	–	–	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
	Low (1%–49%)	–	–	–	–	–	–	–	–	1.69	0.10	27.90	0.71	3.39	1.17	9.79	**0.024**
	High (≥50%)	–	–	–	–	–	–	–	–	31.76	3.84	262.34	**0.0013**	1.34	0.32	5.54	0.67
Age at diagnosis	Age (years)[b]	1.04	1.01	1.08	**0.010**	1.04	1.00	1.07	0.024	–	–	–	–	–	–	–	–

METex14, MET exon 14 skipping; *OR, Odds Ratio, CI, Confidence Interval, Min, Minimum value, Max, Maximum value.* Significant statistical values (p < 0.05) are in bold. [a]Includes Large Cell Carcinoma, Adenosquamous Carcinoma, Sarcomatoid Carcinoma, and Large Cell Neuroendocrine Carcinoma. [b]Continuous variable.

Table 2: **Multivariate logistic regression test with patients' features for mutations in *TP53, EGFR, KRAS, PIK3CA, BRAF, ERBB2, METex14,* and fusions in the *ALK* gene.**

HR: 1.73, 95% CI: 1.08–2.76), and the presence of metastases at diagnosis (p < 0.0001; HR: 5.33, 95% CI: 4.10–6.93), particularly CNS metastases (p < 0.0001; HR: 6.38, 95% CI: 4.80–8.47), were independently associated with an increased risk of death (Table 3; Fig. 6). In contrast, systemic treatments such as TKIs (p < 0.0001; HR: 0.35, 95% CI: 0.26–0.46), immunotherapies (p < 0.0001; HR: 0.49, 95% CI: 0.36–0.66), and chemotherapies (p < 0.0001; HR: 0.53, 95% CI: 0.44–0.65), were independently associated with a

Characteristics	Parameters	n	Univariable analysis					Multivariable analysis[d]		
			Event	Censored	Time, months		p-value[a]	HR	95% CI	p-value
					Median	95% CI				
Age	≤64 years	584	330	254	18.0	13.97–22.03	0.65			
	>64 years	543	310	233	20.0	15.56–24.43				
Sex	Female	508	266	242	24.0	18.82–29.18	**0.001**			
	Male	619	374	245	15.0	11.73–18.27				
Self-declared race[b]	White	706	409	297	20.0	16.58–23.41	0.71			
	Nonwhite	230	134	96	17.0	12.21–21.78				
Smoking status	Never	296	140	156	32.0	25.04–38.95	**<0.0001**	Ref	Ref	Ref
	Quitter	372	207	144	18.0	11.05–29.94		1.20	0.94–1.54	0.13
	Current	418	274	165	12.0	9.46–14.53		1.38	1.07–1.78	**0.011**
Loss of weight	No	384	170	214	41.0	25.48–56.51	**<0.0001**			
	Yes	520	352	168	10.0	7.81–12.18				
ECOG PS	0–1	672	339	333	30.0	23.99–36.00	**<0.0001**	Ref	Ref	Ref
	2–4	313	230	83	5.0	3.44–6.55		1.99	1.64–2.41	**<0.0001**
Histology	Adenocarcinoma	906	494	412	21.0	17.11–24.88	**<0.0001**	Ref	Ref	Ref
	Squamous Cell	65	37	28	25.0	1.02–48.97		1.46	0.98–2.15	0.06
	NSCLC NOS	118	83	35	6.0	3.87–8.12		1.34	1.03–1.76	**0.028**
	Other	38	26	12	15.0	5.00–24.99		1.73	1.08–2.76	**0.021**
Stage at diagnosis	I/II	192	35	157	146.0	55.72–236.27	**<0.0001**			
	III	181	86	95	32.0	22.69–41.30				
	IV	726	508	218	11.0	9.34–12.65				
Metastasis at diagnosis	No	374	119	255	82.0	50.03–113.96	**<0.0001**	Ref	Ref	Ref
	Yes, CNS	255	192	63	7.0	4.10–9.89		6.38	4.80–8.47	**<0.0001**
	Yes, others[c]	473	318	155	12.0	9.99–.14.00		5.33	4.10–6.93	**<0.0001**
Asian ancestry	Low	308	171	137	22.0	15.24–28.75	0.21			
	Intermedium	318	184	134	16.0	12.28–19.71				
	High	311	192	119	15.0	10.96–19.03				
African ancestry	Low	306	165	141	23.0	14.65–31.34	0.32			
	Intermedium	322	193	129	16.0	12.14–19.85				
	High	309	189	120	16.0	12.10–19.89				
European ancestry	Low	311	187	124	16.0	11.73–20.26	0.17			
	Intermedium	313	192	121	15.0	11.25–18.74				
	High	313	168	145	22.0	15.02–28.97				
Native American ancestry	Low	295	168	127	19.0	12.44–25.55	0.14			
	Intermedium	328	185	143	18.0	13.00–22.99				
	High	314	194	120	15.0	11.18–18.81				
TP53 status	Wild-type	475	231	244	32.0	24.77–39.22	**<0.0001**			
	Mutant	652	409	243	15.0	12.52–17.47				
KRAS status	Wild-type	839	474	365	20.0	16.58–23.41	**0.042**			
	Mutant	288	166	122	14.0	9.71–18.28				
EGFR status	Wild-type	900	532	368	15.0	12.69–17.30	**<0.0001**			
	Mutant	227	108	119	36.0	28.93–43.06				
ALK-fusion status	Negative	978	575	403	18.0	15.11–20.88	**0.0027**			
	Positive	69	28	41	56.0	Not reached				
ERBB2 status	Wild-type	1101	626	475	19.0	16.00–21.99	0.96			
	Mutant	26	14	12	28.0	3.25–52.75				
BRAF status	Wild-type	1097	618	479	20.0	16.83–23.16	**0.0060**			
	Mutant	30	22	8	7.0	2.14–11.85				
PIK3CA status	Wild-type	1090	616	474	19.0	15.82–22.17	0.83			
	Mutant	37	24	13	20.0	13.45–26.54				
*MET*Δex14 status	Not detected	355	185	170	18.0	13.23–22.76	0.16			
	Detected	17	11	6	7.0	0.74–13.26				

(Table 3 continues on next page)

Characteristics	Parameters	n	Univariable analysis					Multivariable analysis[d]		
			Event	Censored	Time, months		p-value[a]	HR	95% CI	p-value
					Median	95% CI				
(Continued from previous page)										
PD-L1 expression	No (<1%)	408	218	190	25.0	19.19–30.80	**0.0032**			
	Low (1%–49%)	193	99	94	23.0	11.59–34.40				
	High (>50%)	188	119	69	14.0	8.70–19.29				
TKI treatment	No	861	519	342	15.0	12.50–17.49	**<0.0001**	Ref	Ref	Ref
	Yes	189	89	100	36.0	25.75–46.24		0.35	0.26–0.46	**<0.0001**
Immunotherapy treatment	No	945	549	396	18.0	14.77–21.22	**0.013**	Ref	Ref	Ref
	Yes	106	59	47	31.0	23.86–38.13		0.49	0.36–0.66	**<0.0001**
Chemotherapy treatment	No	408	236	172	8.0	2.64–13.35	**0.0013**	Ref	Ref	Ref
	Yes	643	372	271	22.0	18.23–25.76		0.53	0.44–0.65	**<0.0001**

NSCLC NOS, Non-Small Cell Lung Cancer not otherwise specified; CNS, Central Nervous System; HR, Hazard Ratio; CI, Confidence Interval. Significant statistical values ($p < 0.05$) are in bold. [a]p-value from Log-rank test. [b]According to the IBGE (Brazilian Institute of Geography and Statistics). [c]Includes Large Cell Carcinoma, Adenosquamous Carcinoma, Sarcomatoid Carcinoma, small-cell carcinoma/oat cell and Large Cell Neuroendocrine Carcinoma. [d]Only significant variables in the final model are shown (Total, n = 897; Event, n = 529; Censored, n = 368).

Table 3: **Cancer-specific survival analysis for patients diagnosed with lung cancer.**

reduced risk of death (Table 3; Fig. 6). Likewise, survival analyses stratified by metastatic status at diagnosis confirmed the prognostic relevance of ECOG performance status in both non-metastatic and metastatic groups. At the same time, systemic therapies were predominantly associated with outcomes in metastatic disease, whereas tobacco exposure and histological subtype were primarily associated with prognosis in non-metastatic patients (Supplementary Tables S3 and S4).

TP53 mutations as predictors of shorter cancer-specific survival in patients treated with TKI

We then evaluated CSS and TTE survival in patients with metastatic disease harboring *EGFR* mutations or *ALK* fusions who were treated with TKIs (anti-EGFR and anti-ALK; Supplementary Table S5; Supplementary Figure S8a–d).

In this cohort, patients with *TP53* mutations had significantly shorter CSS (Fig. 5a). Among patients with *EGFR*-mutant tumors treated with TKIs, *TP53* co-mutations were associated with markedly worse outcomes, including shorter CSS ($p = 0.0032$; 24.0 months [95% CI: 19.48–28.51] vs. 61.0 months [95% CI: 29.00–92.99]; Fig. 7a) and a trend toward shorter TTE ($p = 0.11$; 13.0 months [95% CI: 9.63–16.36] vs. 20.0 months [95% CI: 13.81–26.19]; Fig. 7b). Moreover, when considering patients with *TP53*-mutated tumors, those receiving chemotherapy had significantly longer CSS ($p < 0.0001$; 15.0 months [95% CI: 12.41–17.58] vs. 2.0 months [95% CI: 1.58–2.41]; Fig. 7c) and TTE ($p < 0.0001$; 7.0 months [95% CI: 5.63–8.37] vs. 2.0 months [95% CI: 1.38–2.62]; Fig. 7d) compared with patients not treated with chemotherapy. Statistical analyses were not conducted for patients with concurrent *ALK* fusions and *TP53* mutations treated with TKIs due to the very limited sample size (n = 4).

Discussion

In this genomic analysis of 1131 patients with lung cancer across Brazil, we present a comprehensive view of the molecular landscape shaped by demographic, clinical, and ancestry-related factors. Beyond confirming mutation patterns observed in global cohorts, such as the predominance of *TP53*, *KRAS*, and *EGFR* mutations, our study also highlights the impact of *TP53* status on survival in patients with *EGFR* mutations and the features unique to the highly admixed Brazilian population.

Ancestry estimation in our cohort revealed meaningful regional and tumor molecular differences in genetic backgrounds. The North region had the highest proportion of Native American ancestry, whereas the South region had the highest proportion of European ancestry. This regional ancestry diversity aligns with studies from our group and others.[14,18,26] Recently, Nunes et al. reported a slightly distinct ancestry distribution in the general Brazilian population, with a lower proportion of European ancestry (59% vs. 71% in our cohort) and higher African (27% vs. 22%) and Native American contributions (13% vs. 7%).[14] These differences likely stem from both population-specific factors and methodological factors. Nunes et al.'s cohort included healthy individuals evenly distributed across Brazilian regions, while our cohort contains a higher proportion of patients from the Southeast.[14] Additionally, the authors estimated ancestry using whole-genome sequencing, whereas our study employed a 46–AIM panel.[14] Despite methodological and population differences across studies, these findings consistently highlight the extensive genetic admixture characterizing the Brazilian population. Importantly, our analysis reflects a clinical lung cancer cohort rather than a general population sample, underscoring that ancestry-informed molecular profiling may have

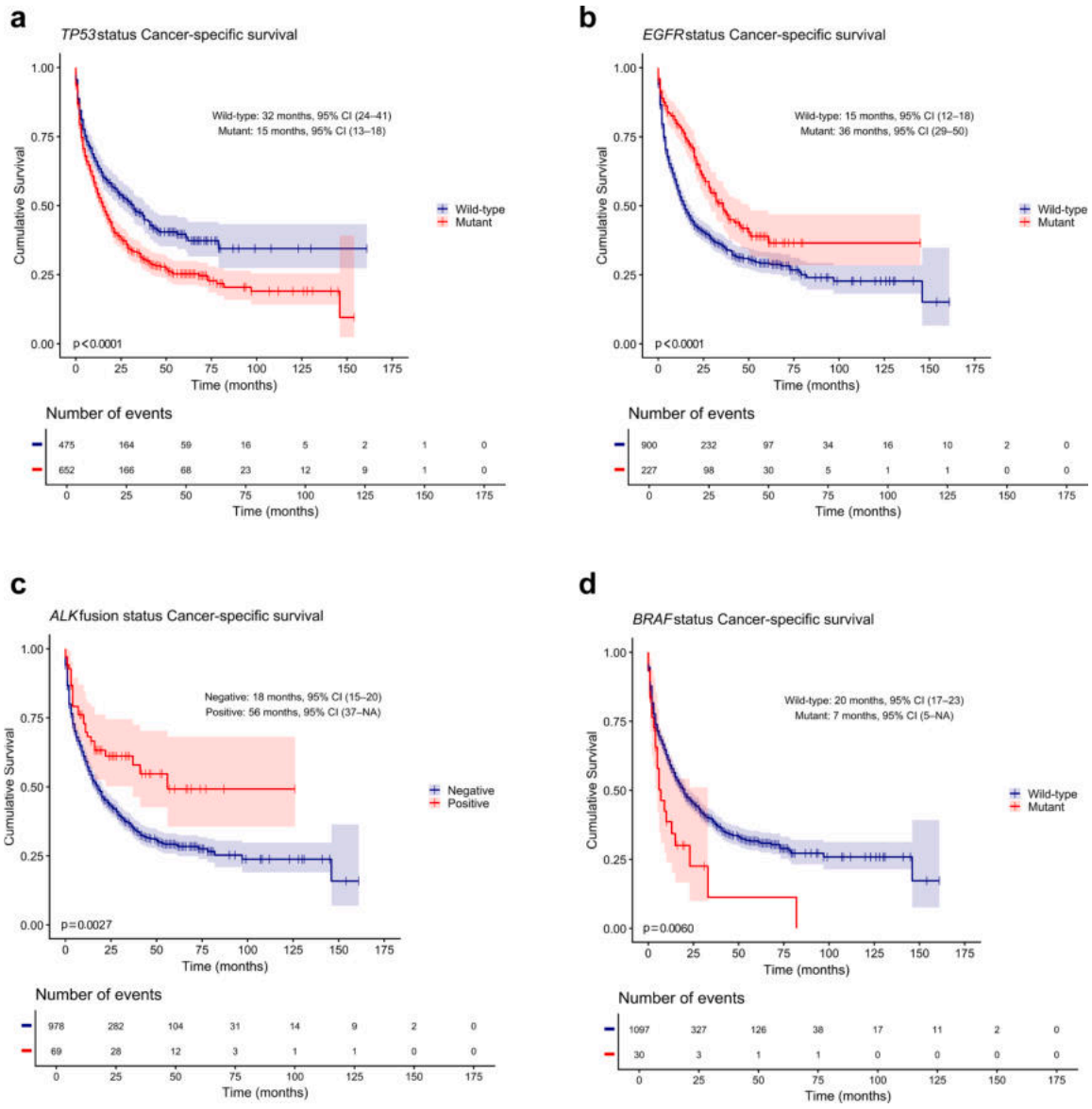

Fig. 5: Representation of cancer-specific survival (CSS) analysis for patients by molecular alterations: (a) *TP53* status CSS; (b) *EGFR* status CSS; (c) *ALK* fusion status CSS; (d) *BRAF* status CSS.

direct implications for diagnostic strategies and therapeutic decision-making in real-world oncology practice.

We observed that 88% of cases harbored oncogenic molecular alterations. The prevalence of *EGFR* mutations (20.5%) and their higher frequencies in Brazilian regions with higher proportions of Native American or Asian ancestry indicate the influence of genetic ancestry on actionable genomic alterations.[9] *EGFR* mutations were associated with patients who never smoked, CNS metastasis, adenocarcinoma histology, and no expression of PD-L1, which is in agreement with previous studies.[8,17,19,24,27,28] Similarly, our finding that *KRAS* mutations (25.6%) were more common among smokers but were not influenced by ancestry supports prior literature from TCGA and Latin American studies, although the ancestry-neutral distribution in our cohort suggests population-specific nuances.[5,17,29] In addition to the tobacco association, we observed that *KRAS* mutations were associated with adenocarcinoma histology, and PD-L1 expression, as previously reported in the literature.[2,3,5,17] The frequency of *ALK* fusions (6.1%) and *MET*ex14 (4.8%) aligns with international estimates and affirms the importance of including

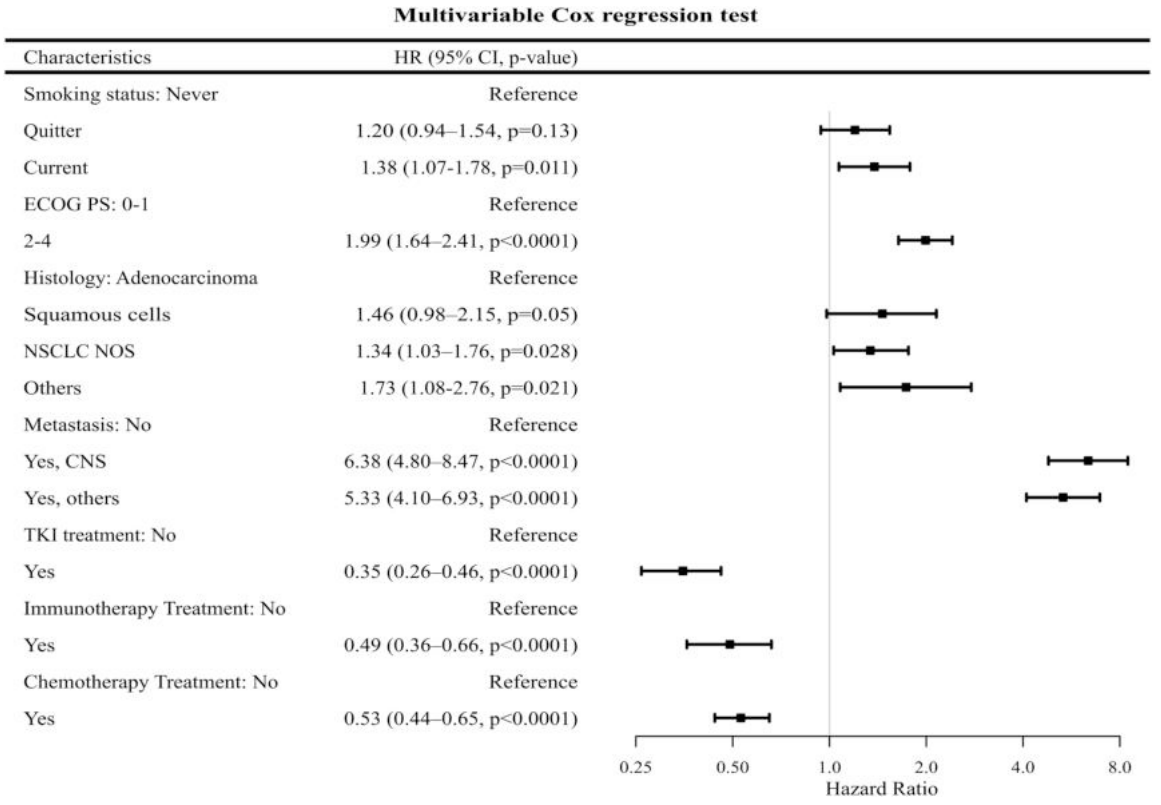

Fig. 6: Forest plot of the multivariate Cox regression analysis for association with an increased risk of death due to cancer-specific causes. Only significant associations in the final model are shown. HR stands for hazard ratio; CI stands for confidence interval; ECOG PS refers to Eastern Cooperative Oncology Group Performance Status; NSCLC NOS indicates Non-Small Cell Lung Cancer not otherwise specified; CNS denotes central nervous system; and TKI stands for tyrosine kinase inhibitor.

these targets in routine diagnostic panels.[4,29–31] Moreover, *ALK* fusions were associated with younger patients, never-smokers, and low expression of PD-L1, whereas *METex14* alteration were associated with patients who never smoked and high expression of PD-L1 in agreement with previous studies.[4,8,19,32]

Fusions were observed in less than 1% of patients in the genes *RET*, *ROS1*, *NTRK1*, and *NTRK3*, while mutations in the genes *BRAF*, *PIK3CA*, *NRAS*, *KIT*, and *ERBB2* were found in less than 3% of patients, a finding consistent with previous studies worldwide.[4,6,8,19,31] In contrast, *PIK3CA* is described as more frequent (6%) in the TCGA (The Cancer Genome Atlas), and in Brazil, a previous study reported *PIK3CA* at a higher frequency (8.8%) in lung tumors, such as *ERBB2* mutations (4.9%), which may be explained by distinct methodologies applied in the studies.[3,29]

The most frequently mutated gene was *TP53*, observed in 58.0% of the patients, and was associated with tobacco consumption, squamous histology, CNS metastasis, and high African ancestry, in accordance with the literature.[3,16,33–36] Our study also indicates the frequent co-occurrence of *TP53* mutations with other driver alterations, notably *EGFR* and *KRAS*. This is consistent with several studies, including those from Asian cohorts, which report *TP53* co-mutations in 30–50% of *EGFR*-mutated NSCLCs. The clinical implications of this genomic background are substantial: *EGFR/TP53* co-mutation was associated with poorer CSS in our cohort, supporting growing evidence that *TP53* mutations contribute to therapeutic resistance, disease aggressiveness, and a worse prognosis. Interestingly, recent clinical trials such as NEJ009 and FLAURA2, explored combination of chemotherapy and TKI approaches to overcome resistance in this subgroup.[37,38]

Although ancestry proportions in Brazil have been linked to differences in mutation profiles and clinical outcomes, most genomic datasets still disproportionately represent individuals from North America, Europe, and Asia, leaving Latin American populations markedly underrepresented. In Brazil, individuals with higher African ancestry, who are more frequently classified as Black or Pardo, tend to have lower incomes, reduced educational access, and higher smoking prevalence, factors that may influence both molecular

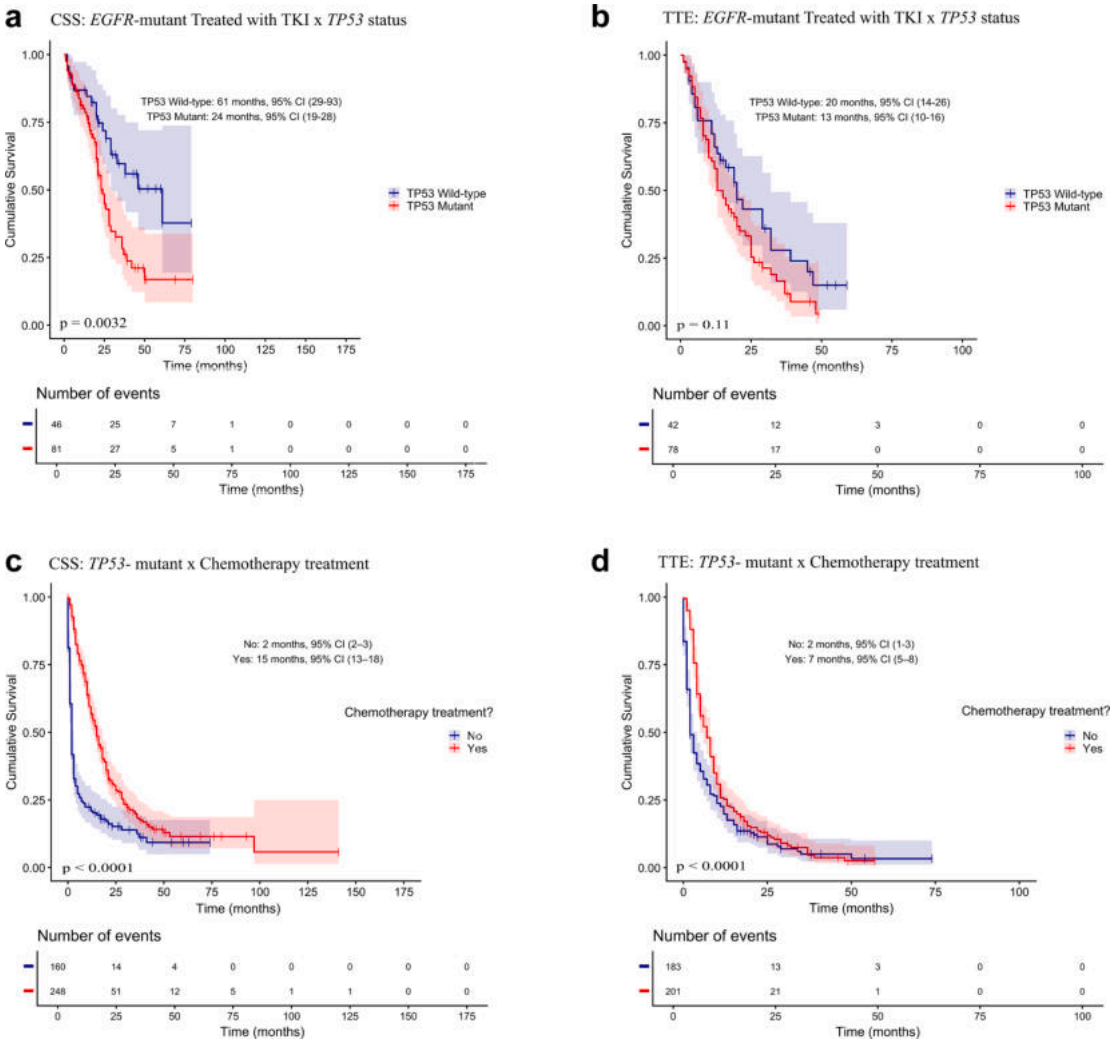

Fig. 7: Analyses of cancer-specific survival (CSS) and time to event (TTE) by treatment and molecular alterations in patients diagnosed with metastatic disease. (a) CSS of patients with *EGFR* mutations who received TKI treatment, comparing those with *TP53* mutations to those with wild-type *TP53*; (b) TTE of patients with *EGFR* mutations who received TKI treatment, comparing those with *TP53* mutations to those with wild-type *TP53*; (c) CSS of patients with *TP53* mutations, comparing those treated with chemotherapy at any line to those who were not; (d) TTE of patients with *TP53* mutations, comparing those who received first-line chemotherapy to those who did not. CSS, cancer-specific survival; TTE, time to event; CI, confidence interval; TKI, tyrosine kinase inhibitor.

patterns and survival outcomes.[39,40] Moreover, socioeconomic disparities often translate into unequal access to timely diagnosis and treatment, contributing to poorer outcomes in groups with higher African ancestry.[41]

Our study addresses this critical gap by including diverse and admixed populations in molecular studies to ensure that biomarker discoveries and therapeutic advances are applicable and equitable across ancestries. Our cohort derives from a national referral cancer center that attends only patients from the public Brazilian health system, with approximately 20% of participants self-identifying as non-White (Black, Yellow, Pardo, translated as Brown/Mixed, and Indigenous—in accordance with the Brazilian IBGE census), consistent with expected demographic patterns in the Brazilian populations.[20] Moreover, the genetic ancestry estimates were broadly concordant with self-reported skin color and varied across regions, reflecting the country's highly admixed population. Taken together, these observations suggest that, although not population-based, the cohort provides insight into geographic and demographic representativeness within a real-world clinical context and highlights the importance of equity inclusion in precision oncology research.

Despite our findings, some limitations should be acknowledged, most of which are inherent to retrospective real-world analyses conducted in resource-constrained settings. These include convenience sampling, incomplete follow-up data in some regions, and the predominance of patients from the Southeast region. Moreover, the limited gene content of our DNA NGS panel, followed by gene fusion assessment, together with the heterogeneity of molecular methods used in routine care, may have contributed to the underestimation of rare alterations, gene fusions, or complex co-mutation patterns. Additionally, therapeutic data were not always uniform across sites, which may confound treatment-related outcomes and is an inherent limitation of retrospective studies. Nevertheless, these limitations reflect the complexity of implementing precision oncology in routine clinical practice and highlight the challenges faced by publicly funded healthcare systems in ensuring equitable access to molecular diagnostics in Brazil and across Latin America.

Conclusion

Our study indicates that most Brazilian patients diagnosed with lung cancer have at least one oncogenic alteration, with over half showing actionable genomic targets. These results emphasize the potential benefits of precision oncology in this group. The association between *TP53* co-mutations and worse cancer-specific survival in metastatic *EGFR*-mutated patients treated with EGFR-TKIs highlights the need to include *TP53* status in molecular assessments and treatment planning. Notably, better outcomes seen with combined EGFR-TKI and chemotherapy strategies are consistent with emerging clinical trial data and support a more personalized approach for this high-risk group. Overall, our research offers real-world molecular evidence from a large national referral center serving a socioeconomically and geographically diverse population. These findings add to the regional evidence base required to guide policy, improve access to molecular diagnostics, and optimize targeted therapy use. Ultimately, such efforts are vital to enhancing lung cancer outcomes across Latin America.

Contributors

Conceptualization: R.O.C., L.F.L., and R.M.R.; Supervision: R.M.R. and L.F.L.; Writing – original draft: R.O.C., L.F.L., and R.M.R.; Formal Analysis: R.O.C., W.Y.H.; Methodology: F.E.P, G.N.B., M.B., M.A.M.; Data Collection: B.G.Z., R.M.Q.B., F.H.Q., I.M.R.N., G.S.B., J.F.N.; Data Curation: I.S., M.T.R., G.T., E.C.S.; J.M.D., F.A.F.S., C.E.B.S., R.S.C., L.F., L.V., A.A.J., R.E.N.N.O.; Writing review and editing: R.O.C., L.F.L., R.M.R., M.A.M., J.M.D.; R.M.R. made the decision to submit the manuscript; All authors have reviewed the original draft.

Data sharing statement

The data that support the findings of this study are available from Dr. Rui Manuel Reis, but restrictions apply to the availability of these data, due to patients' personal data. De-identified data are, however, available from the authors upon resonable request.

Editor note

The Lancet Group takes a neutral position with respect to territorial claims in published maps and institutional affiliations.

Declaration of interests

J.M.D. - Payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events from AstraZeneca, Johnson & Johnson, and Merck Sharp and Dohme; Support for attending meetings and/or travel from Amgen, AstraZeneca, and Johnson & Johnson; Funding for clinical research – to Institution from Amgen, AstraZeneca, BeOne, Bristol-Myers Squibb, Daiichi-Sankyo, GlaxoSmithKline, Incyte Corporation, Ipsen, Johnson & Johnson, Lilly, Merck, Merck Sharp and Dohme, Novartis, Pfizer, Regeneron, Roche, Sanofi, Takeda, and Xcovery. L.V. - Payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events from AstraZeneca. M.A.M. Grants or contracts from Sumitomo Inc, AstraZeneca, and Merck KGaA; Consulting fees from Athenum; Payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events from Spanish Melanoma Group, Fundación Ricky Rubio, Liquid Biopsy Symposium, and University of Naples Federico II; Support for attending meetings and/or travel from European Thoracic Oncology Platform (ETOP). R.O.C., W.Y.H., F.E.P., G.N.B., M.B., B.G.Z., R.M.Q.B., F.H.Q., I.M.R.N., G.S.B., J.F.N., I.S., M.T.R., G.T., E.C.S.; J.M.D., F.A.F.S., C.E.B.S., R.S.C., L.F., L.V., A.A.J., R.E.N.N.O., L.F.L., R.M.R. have declared no conflicts of interest.

Acknowledgements

This study was funded by the Public Ministry of Labor Campinas (Research, Prevention, and Education of Occupational Cancer—15th zone, Campinas, Brazil), PRONON—PRONON/MS (Abordagens móveis e de tecnologia para prevenção primária e secundária de câncer—NUP: 25000.015000/2019-53), and Barretos Cancer Hospital. L.F.L. were supported by Public Ministry of Labor Campinas (Research, Prevention, and Education of Occupational Cancer—15th zone, Campinas, Brazil). R.O.C. was the recipient of a CAPES Fellowship and PRONON—PRONON/MS; B.G.Z. was supported by São Paulo Research Foundation (FAPESP). R.M.R. and L.F.L are recipients of a CNPq Productivity Fellowship.

Funding sources have no contribution to authorship for the present study. We thank all members of the GTOP group (Translational Group of Pulmonary Oncology-Barretos Cancer Hospital, Brazil) for scientific discussion and suggestions.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.lana.2026.101429.

References

1 Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74:229–263.
2 Hendriks LELL, Remon J, Faivre-Finn C, et al. Non-small-cell lung cancer. *Nat Rev Dis Primers*. 2024;10(1):71. https://www.nature.com/articles/s41572-024-00551-9.
3 Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014;511(7511):543–550. https://doi.org/10.1038/nature13385.
4 Novaes LAC, da Silva LS, de Marchi P, et al. Simultaneous analysis of ALK, RET, and ROS1 gene fusions by NanoString in Brazilian lung adenocarcinoma patients. *Transl Lung Cancer Res*. 2021;10(1):292–303. https://tlcr.amegroups.com/article/view/48020/html.
5 Cavagna R, Escremim de Paula F, Sant'Anna D, et al. Frequency of KRAS p.Gly12Cys mutation in Brazilian patients with lung cancer. *JCO Glob Oncol*. 2021;7:639–645. https://ascopubs.org/doi/10.1200/GO.20.00615.
6 de Oliveira Cavagna R, de Andrade ES, Tadin Reis M, et al. Detection of NTRK fusions by RNA-based nCounter is a feasible diagnostic methodology in a real-world scenario for non-small cell lung cancer assessment. *Sci Rep*. 2023;13(1):21168. https://doi.org/10.1038/s41598-023-48613-4.

7 Hendriks LE, Kerr KM, Menis J, et al. Oncogene-addicted metastatic non-small-cell lung cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Ann Oncol*. 2023;34(4):339–357. https://doi.org/10.1016/j.annonc.2022.12.009.

8 Mack PC, Klein MI, Ayers KL, et al. Targeted next-generation sequencing reveals exceptionally high rates of molecular driver mutations in never-smokers with lung adenocarcinoma. *Oncologist*. 2022;27(6):476–486.

9 Carrot-Zhang J, Soca-Chafre G, Patterson N, et al. Genetic ancestry contributes to somatic mutations in lung cancers from admixed Latin American populations. *Cancer Discov*. 2021;11(3):591–598. http://cancerdiscovery.aacrjournals.org/lookup/doi/10.1158/2159-8290.CD-20-1165.

10 Calvayrac O, Pradines A, Pons E, Mazières J, Guibert N. Molecular biomarkers for lung adenocarcinoma. *Eur Respir J*. 2017;49(4):1–17. https://doi.org/10.1183/13993003.01734-2016.

11 Riely GJ, Wood DE, Ettinger DS, et al. Non–Small cell lung cancer, version 4.2024. *J Natl Compr Cancer Netw*. 2024;22(4):249–274. https://jnccn.org/view/journals/jnccn/22/4/article-p249.xml.

12 Solomon BJ, Mok T, Kim DW, et al. First-line crizotinib versus chemotherapy in ALK -Positive lung cancer. *N Engl J Med*. 2014;371(23):2167–2177.

13 Lan B, Zhao N, Du K, Leng B. Concurrent TP53 mutations predict a poor prognosis of EGFR-mutant NSCLCs treated with TKIs: an updated systematic review and meta-analysis. *Oncol Lett*. 2022;24(5):384. http://www.spandidos-publications.com/10.3892/ol.2022.13504.

14 Nunes K, Araújo Castro E Silva M, Rodrigues MR, et al. Admixture's impact on Brazilian population evolution and health. *Science*. 2025;388(6748):eadl3564. http://www.ncbi.nlm.nih.gov/pubmed/40373151.

15 Naslavsky MS, Scliar MO, Yamamoto GL, et al. Whole-genome sequencing of 1,171 elderly admixed individuals from Brazil. *Nat Commun*. 2022;13(1):1004. https://www.nature.com/articles/s41467-022-28648-3.

16 Cavagna RO, Pinto IA, Escremim de Paula F, et al. Disruptive and truncating TP53 mutations are associated with African-ancestry and worse prognosis in Brazilian patients with lung adenocarcinoma. *Pathobiology*. 2023;90:344–355. Available from: http://www.ncbi.nlm.nih.gov/pubmed/37031678.

17 Leal LF, de Paula FE, De Marchi P, et al. Mutational profile of Brazilian lung adenocarcinoma unveils association of EGFR mutations with high Asian ancestry and independent prognostic role of KRAS mutations. *Sci Rep*. 2019;9(1):1–10.

18 da Costa Vieira RA, Sant'Anna D, Laus AC, et al. Genetic ancestry of 1127 Brazilian breast cancer patients and its correlation with molecular subtype and geographic region. *Clin Breast Cancer*. 2023;23(5):527–537.

19 Cavagna RO, Escremim de Paula F, Berardinelli GN, et al. Molecular profile of driver genes in lung adenocarcinomas of Brazilian patients who have never smoked: implications for targeted therapies. *Oncologist*. 2024;29(10):e1419–e1424. http://www.ncbi.nlm.nih.gov/pubmed/38944844.

20 IBGE IB de G e E. *Desigualdades Sociais por Cor ou Raça no Brasil*. Rio de Janeiro; 2022. https://biblioteca.ibge.gov.br/visualizacao/livros/liv101972_informativo.pdf.

21 Campanella NC, Silva EC, Dix G, et al. Mutational profiling of driver tumor suppressor and oncogenic genes in Brazilian malignant pleural mesotheliomas. *Pathobiology*. 2020;87(3):208–216. https://www.karger.com/Article/FullText/507373.

22 Horak P, Griffith M, Danos AM, et al. Standards for the classification of pathogenicity of somatic variants in cancer (oncogenicity): joint recommendations of Clinical Genome Resource (ClinGen), Cancer Genomics Consortium (CGC), and Variant Interpretation for Cancer Consortium (VICC). *Genet Med*. 2022;24(5):986–998. https://linkinghub.elsevier.com/retrieve/pii/S1098360022000016.

23 Li MM, Datto M, Duncavage EJ, et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a joint consensus recommendation of the Association for molecular pathology, American Society of Clinical Oncology, and College of American Pathologists. *J Mol Diagnostics*. 2017;19(1):4–23. https://doi.org/10.1016/j.jmoldx.2016.10.002.

24 Alves Pinto I, de Oliveira Cavagna R, Virginio da Silva AL, et al. EGFR mutations and PD-L1 expression in early-stage non-small cell lung cancer: a real-world data from a single center in Brazil. *Oncologist*. 2022;27:e899–e907. https://academic.oup.com/oncolo/advance-article/doi/10.1093/oncolo/oyac167/6696863.

25 De Marchi P, Leal LF, Duval da Silva V, da Silva ECA, Cordeiro de Lima VC, Reis RM. PD-L1 expression by Tumor Proportion Score (TPS) and Combined Positive Score (CPS) are similar in non-small cell lung cancer (NSCLC). *J Clin Pathol*. 2021;74(11):735–740. https://jcp.bmj.com/lookup/doi/10.1136/jclinpath-2020-206832.

26 Durães RO, Berardinelli GN, da Costa AM, et al. Role of genetic ancestry in 1,002 Brazilian colorectal cancer patients from Barretos Cancer Hospital. *Front Oncol*. 2020;10:145.

27 Sun H, Ren P, Chen Y, et al. Optimal therapy for concomitant EGFR and TP53 mutated non-small cell lung cancer: a real-world study. *BMC Cancer*. 2023;23(1):1–11.

28 Han G, Bi J, Tan W, et al. A retrospective analysis in patients with EGFR-mutant lung adenocarcinoma: is EGFR mutation associated with a higher incidence of brain metastasis? *Oncotarget*. 2016;7(35):56998–57010. http://www.ncbi.nlm.nih.gov/pubmed/27486770.

29 Mascarenhas E, Gelatti AC, Araújo LH, et al. Comprehensive genomic profiling of Brazilian non-small cell lung cancer patients (GBOT 0118/LACOG0418). *Thorac Cancer*. 2021;12(5):580–587. https://onlinelibrary.wiley.com/doi/10.1111/1759-7714.13777.

30 Liu L, Liu J, Shao D, et al. Comprehensive genomic profiling of lung cancer using a validated panel to explore therapeutic targets in East Asian patients. *Cancer Sci*. 2017;108(12):2487–2494. http://www.ncbi.nlm.nih.gov/pubmed/28949084.

31 Takeuchi K, Soda M, Togashi Y, et al. RET, ROS1 and ALK fusions in lung cancer. *Nat Med*. 2012;18(3):378–381. http://www.nature.com/articles/nm.2658.

32 Hofman P. ALK in non-small cell lung cancer (NSCLC) pathobiology, epidemiology, detection from tumor tissue and algorithm diagnosis in a daily practice. *Cancers (Basel)*. 2017;9:107. MDPI AG.

33 Saleh MM, Scheffler M, Merkelbach-Bruse S, et al. Comprehensive analysis of TP53 and KEAP1 mutations and their impact on survival in localized and advanced stage non-small cell lung cancer. *J Thorac Oncol*. 2021;17:1–13.

34 Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. 2012;489(7417):519–525. https://pubmed.ncbi.nlm.nih.gov/22960745/.

35 Toyooka S, Tsuda T, Gazdar AF. The TP53 gene, tobacco exposure, and lung cancer. *Hum Mutat*. 2003;21(3):229–239.

36 Wang H, Ou Q, Li D, et al. Genes associated with increased brain metastasis risk in non–small cell lung cancer: comprehensive genomic profiling of 61 resected brain metastases versus primary non–small cell lung cancer (Guangdong Association Study of Thoracic Oncology 1036). *Cancer*. 2019;125(20):3535–3544.

37 Yang JC, Robichaux J, Planchard D, et al. MA12.03 FLAURA2: resistance, and impact of baseline TP53 alterations in patients treated with 1L osimertinib ± platinum-pemetrexed. *J Thorac Oncol*. 2024;19(10):S101–S102. https://linkinghub.elsevier.com/retrieve/pii/S155608642401058X.

38 Hou X, Li M, Wu G, et al. Gefitinib plus chemotherapy vs gefitinib alone in untreated EGFR -Mutant non–small cell lung cancer in patients with brain metastases. *JAMA Netw Open*. 2023;6(2):e2255050. https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2801106.

39 Moreira ALO da SC de LMRAPSMNJC. The colors of smoking : relationship between race and tobacco use in Brazil. *Rev Bras Cancerol*. 2022;68(1):1–4. https://doi.org/10.32635/2176-9745.RBC.2022v68n1.1552.

40 Lima-Costa MF, Macinko J, Vaz De Melo Mambrini J, et al. Genomic ancestry, self-rated health and its association with mortality in an admixed population: 10 year follow-up of the Bambui-Epigen (Brazil) cohort study of ageing. *PLoS One*. 2015;10(12):1–14.

41 Davis MB, Cardona F, Arrieta O, et al. *Genomics and cancer Disparities : the Justice and Power of Inclusion*. 2021:805–809.