Translational Oncology 15 (2022) 101276

Contents lists available at ScienceDirect

Translational Oncology

journal homepage: www.elsevier.com/locate/tranon

Original Research

p.G12C KRAS mutation prevalence in non-small cell lung cancer: Contribution from interregional variability and population substructures among Hispanics

Alejandro Ruiz-Patiño [a,b,*,1], July Rodríguez [a,b,1], Andrés F. Cardona [a,b,c,*,1], Jenny Ávila [a,b,1], Pilar Archila [a,b,1], Hernán Carranza [a,b,c], Carlos Vargas [a,b,c], Jorge Otero [a,b,c], Oscar Arrieta [d], Lucia Zatarain-Barrón [d], Carolina Sotelo [a,b], Camila Ordoñez [a,b], Juan Esteban García-Robledo [e], Leonardo Rojas [b,c,f], Maritza Bermúdez [a,b], Tatiana Gámez [a,b], Diana Mayorga [a,b], Luis Corrales [g], Claudio Martín [h], Gonzalo Recondo [i], Luis Mas [j], Suraj Samtani [k], Luisa Ricaurte [l], Umberto Malapelle [m], Alessandro Russo [n], Feliciano Barrón [d], Nicolas Santoyo [a,b], Christian Rolfo [o], Rafael Rosell [p], on behalf of ONCOLGroup [∫], and CLICaP [Ψ]

[a] *Foundation for Clinical and Applied Cancer Research (FICMAC), Calle 116 No. 9 – 72, c. 718, Bogotá, Colombia*
[b] *Molecular Oncology and Biology Systems Research Group (Fox-G/ONCOLGroup), Universidad el Bosque, Bogotá, Colombia*
[c] *Clinical and Traslational Oncology Group, Clínica del Country, Bogotá, Colombia*
[d] *Thoracic Oncology Unit, National Cancer Institute (INCan), México, Mexico*
[e] *Division of Hematology/Oncology, Mayo Clinic, Scottsdale, United States*
[f] *Oncology Department, Clínica Colsanitas, Bogotá, Colombia*
[g] *Oncology Department, Hospital San Juan de Dios, San José Costa Rica, Costa Rica*
[h] *Medical Oncology Group, Fleming Institute, Buenos Aires, Argentina*
[i] *Thoracic Oncology Section, Centro de Educación Médica e Investigaciones Clínicas – CEMIC, Buenos Aires, Argentina*
[j] *Thoracic Oncology Unit, Instituto de Enfermedades Neoplásicas, Lima, Perú*
[k] *Medical Oncology Service, Clinica Bradford Hill, Santiago, Chile*
[l] *Pathology Department, Mayo Clinic, Rochester, MN, United States*
[m] *Department of Public Health, University of Naples Federico II, Naples, Italy*
[n] *Medical Oncology Unit, A.O. Papardo, Messina, Italy*
[o] *Center for Thoracic Oncology, Tisch Cáncer Center, Mount Sinai Hospital System & Icahn School of Medicine, Mount Sinai, New York, NY, United States*
[p] *Cancer Biology and Precision Medicine Program, Catalan Institute of Oncology, Barcelona, Spain*
[∫] *Colombian Group for Clinical and Translational Cancer Research - ONCOLGroup*
[Ψ] *Latin American Consortium for the Investigation of Lung Cancer - CLICaP*

ARTICLE INFO

Keywords:
Non small cell lung cancer
Molecular epidemiology
KRAS
Population markers

ABSTRACT

Background: The KRAS exon 2 p. G12C mutation in patients with lung adenocarcinoma has been increasing in relevance due to the development and effectiveness of new treatment medications. Studies around different populations indicate that regional variability between ethnic groups and ancestries could play an essential role in developing this molecular alteration within lung cancer.

Methods: In a prospective and retrospective cohort study on samples from lung adenocarcinoma from 1000 patients from different administrative regions in Colombia were tested for the *KRAS* p.G12C mutation. An analysis of STR populations markers was conducted to identify substructure contributions to mutation prevalence.

Results: Included were 979 patients with a national mean frequency for the KRAS exon 2 p.G12C mutation of 7.97% (95%CI 6.27–9.66%). Variation between regions was also identified with Antioquia reaching a positivity value of 12.7% (95%CI 9.1–16.3%) in contrast to other regions such as Bogota DC (Capital region) with 5.4% (2.7–8.2%) and Bolivar with 2.4% (95%CI 0–7.2%) (p-value = 0.00262). Furthermore, Short tandem repeat

* Corresponding authors at: Foundation for Clinical and Applied Cancer Research (FICMAC), Calle 116 No. 9 – 72, c. 718, Bogotá, Colombia.
E-mail addresses: alejandro.ruiz.pat@gmail.com (A. Ruiz-Patiño), andres.cardona@clinicadelcountry.com (A.F. Cardona).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.tranon.2021.101276
Received 1 November 2021; Accepted 4 November 2021

population substructures were found for eight markers that strongly yielded association with KRAS exon 2 p.G12C frequency reaching an adjusted R2 of 0.945 and a *p*-value of < 0.0001.
Conclusions: Widespread identification of KRAS exon 2 p.G12C mutations, especially in cases where NGS is not easily achieved is feasible at a population based level that can characterize regional and national patterns of mutation status. Furthermore, this type of mutation prevalence follows a population substructure pattern that can be easily determined by population and ancestral markers such as STR.

Introduction

Non-small-cell lung cancer (NSCLC) accounts for ~85% of primary lung cancers, the most common subtypes being adenocarcinoma (ADC) and squamous cell carcinoma (SCC) [1]. Recently, genetic profiling has identified driver mutations, believed to contribute to early carcinogenesis, in over 80% of ADC cases and ~47% of SCC cases. In the last two decades, understanding driver mutations have evolved, from reports of *KRAS* mutations in ADC in 1987 [2], to reports of fusions between *KIF3B* and *RET* in 2012 [3]. Therapies targeting driver mutations have provided promising outcomes in relevant populations, including those with mutations in *EGFR, BRAF, Her2*, and *MET*, as well as fusion genes involving *ALK, ROS1*, and *RET*, among others [4]. It is also important to consider mutations in tumor suppressor genes (TSGs) associated with NSCLC, including *TP53, STK11*, and *PTEN*, since they frequently coincide with other oncogenic mutations, affect prognosis, and potentially impact the response to immunotherapy [5].

KRAS is one of the most frequently mutated genes in NSCLC. *KRAS* mutations are associated with smokers and with a more aggressive phenotype [6–10], and the frequency varies among distinct populations, accounting for approximately 25% in Whites and <10% in East Asians [6]. Recently, Liu et al. characterized 17,113 NSCLC specimens using next-generation sequencing (NGS) and classified them based on presence and types of *KRAS*; across the entire cohort, 27% of samples harbored a *KRAS* mutation, being the most frequent variants the exon 2 p.G12C (40%), followed by exon 2 p.G12V (19%) and exon 2 p.G12D (15%) [7]. The prevalence of *KRAS* mutations was 37.2% among adenocarcinoma and only 4.4% in squamous cell carcinomas [11]. In addition, most ADCs carrying *KRAS* mutations present a High Tumor Mutation Burden (TMB, defined by> 10 mts/Mb), which varies across the different *KRAS* variants, being most common in exon 2 p.G13X (68.3%) and least common in exon 2 p.G12D (43.2%) [7]. In the same dimension, PD-L1 expression also varied among *KRAS* NSCLC, finding that the exon 2 p.G12C alteration was the most likely to be PD-L1 positive, with 65.5% TPS >1%, and the most likely to be PD-L1 high, with 41.3% TPS >50%. *KRAS* is associated with various commutations; *STK11* is especially relevant, which is altered in 36% of cases (more in the exon 2 p.G13X variant 36.2% and less in the exon 2 p.G12D 14.2%) [7,11]. In contrast, *TP53* mutations were more frequent in *KRAS* wild-type NSCLC (73.6%) [7].

In less than five years, the availability of novel therapies has transformed the landscape of *KRAS* mutant NSCLC, and a target that was once considered "undruggable" now has several possibilities tested in clinical trials and many others in progress. There are essentially four novel approaches to target KRAS that can be divided into targeted therapies and immunotherapy, potentially combined [12–14]. A novel class of KRAS inhibitors that specifically block the exon 2 p.G12C mutation through direct interaction with cysteine residue and a somewhat broader approach using a pan-KRAS inhibitor that two do not attempt to inhibit specific mutations. The compounds AMG-510 (Sotorasib, Amgen, Thousand Oaks, CA) and MRTX849 (Adagrasib, Mirati Therapeutics, San Diego, CA) was developed to target the *KRAS* exon 2 p.G12C mutation [14,15]. These specific inhibitors locked exon 2 p.G12C mutation in an inactive state, hampering the oncogenic signals and allowed the normal function of remained wild-type *KRAS* [12,15,16]. In a phase I study, 32.2% (19 out of 59) of sotorasib-treated patients presented objective response, and 88.1% (52 out of 59) presented disease control [12]. In phase I and II study, 94% (17 out of 18) of adagrasib-treated NSCLC patients presented with disease control, and the objective response was not yet available (KRYSTAL-1 study; ClinicalTrials.gov identifier: NCT03785249) [14].

Although the differences in somatic mutation patterns between ADCs from patients of different ethnicities, the landscape of ancestry effects on the lung cancer genomes for the Latin American populations has not been comprehensively described, recently, Carrot-Zhang et al. performed an extensive ancestry analysis from tumor samples in an admixed population of NSCLC patients from México and Colombia [16], finding that native ancestry was positively correlated with mutations *EGFR*, and anti-correlated with mutations in *KRAS* and *STK11*. Furthermore, the ancestry effect on *KRAS* mutation was independent of smoking, and the frequency of mutations in this gene was only 12%. In contrast, Cavagna et al. recently reported a prevalence of *KRAS* mutations of 25.3% from the analysis of 844 Brazilian patients with NSCLC [17].

To establish the population differences among Hispanics, we present the analysis of *KRAS* mutations among 979 Colombian NSCLC patients from 19 regions of the country, which allowed us to assess the contribution from interregional variability and population substructures.

Material and methods

Study design and patients

The methods of the study were performed in two stages. In the first one, the conditions of the method for detecting the exon 2 p.G12C mutation of the *KRAS* gene by Digital PCR (ddPCR) were standardized. Likewise, Next Generation Sequencing (NGS) methods were also standardized for the detection of the exon 2 p.G12C mutation in the *KRAS* gene using DNA samples extracted from paraffin-embedded tissue (FFPE) and *cf*DNA extracted from plasma (Supplementary Fig. 1). On the other hand, in the second stage, the *KRAS* exon 2 p.G12C mutation was detected in DNA samples extracted from paraffin-embedded tissue (FFPE) in 979 duplicate samples from patients with NSCLC selected from the biobank of the Foundation for Clinical and Applied Cancer Research (FICMAC, Bogotá, Colombia) (Supplementary Fig. 2). Tumor DNA from tumor FFPE and *cf*DNA samples from 61 patients with NSCLC cancer were retrospectively and prospectively collected; each case required a paired sample from tumor and plasma. Samples from patients were considered negative for *EGFR, ALK*, and *ROS1*, have enough tumor tissue for analysis, and availability of the liquid sample. Once the samples were collected, the standardization of the method for detecting the exon 2 p.G12C mutation of the *KRAS* gene by ddPCR and NGS techniques was carried out following the methods described below.

Sample preparation

In FFPE samples, a minimum of 2 sections of 4–6 μm with a minimum of 20% tumor representation was taken for DNA isolation with the ReliaPrep™ FFPE gDNA Miniprep System - Promega kit, and the manufacturer's recommendations were followed. In plasma samples, 2 ml

were taken for cfDNA isolation with the MagMAX™ Total Nucleic Acid Isolation kit -Invitrogen- and the manufacturer's recommendations were followed. Tumor DNA and cfDNA samples are already isolated or isolated de novo. All samples were quantified by spectrophotometric methods with the NanoDrop 2000 / 2000c and Qubit® 3.0 Fluorometer Catalog Number Q33216 kits; by the same token, the manufacturer's recommendations were followed.

Detection of KRAS exon 2 p.G12C mutation

The digital droplet PCR (ddPCR) technique was performed with the PrimePCR™ ddPCR™ Mutation Detection Assay Kit: KRAS G12C - Biorad in the QX200 Droplet Digital PCR System, then the data were analyzed with the QuantaSoft™ Software. Subsequently, the DNA samples that had the exon 2 p.G12C mutation detected by the digital PCR technique also were determined by NGS technique with the preparation of the library with Oncomine™ Focus Assay - Thermo Fisher Scientific, following the manufacturer's recommendations. The data obtained by this technique were analyzed by Oncomine™ Reporter software and an in-house bioinformatics pipeline. Comparisons between ddPCR and NGS were conducted on 60 samples.

Furthermore, and in order to determine nation prevalence and variation, we retrospectively analyzed 1000 tumor DNA samples obtained from FFPE of patients with NSCLC archived in the Foundation for Clinical and Applied Cancer Research (Bogotá, Colombia). Samples and information were collected from January 15, 2020, to December 31, 2020. All patients met the following inclusion criteria: histological confirmed NSCLC, locally advanced or advanced disease (stage IIIB/IV), age >18 years, and adequate FFPE tissue available to detect *KRAS* mutations. We also obtained some demographic characteristics, histology, tumor grade, and site of metastases. All included patients provided signed informed consent and consented to the analysis of archival diagnostic tissue. An Institutional Review Board and Privacy Board waiver was obtained to facilitate collection of clinical-pathologic and molecular data (Lung Cancer-FICMAC/CLICaP Platform - Registration No. 2020/048, Kayre, Clínica del Country, Bogotá, Colombia). In addition, all cases should be negative for *EGFR* and *ALK* and had a known status for PD-L1 expression.

Detection of KRAS exon 2 p.G12C mutation in tumor DNA samples

The tumor DNA was quantified, the digital PCR technique was performed with the PrimePCR™ ddPCR™ Mutation Detection Assay Kit: KRAS G12C - Biorad in the QX200 Droplet Digital PCR System, then the data were analyzed with the QuantaSoft™ Software, and the manufacturer's recommendations were followed.

Statistical analysis

For statistical analysis, the percentage was used to describe categorical variables, and medians were used to describe continuous variables. Fisher's exact test and χ^2 test were used for the association between *KRAS* mutations and the clinicopathologic data. A stepwise multiple linear regression was conducted. *KRAS* exon 2 p.G12C regional prevalence were the dependent variable and, as the independent variable, the allelic frequencies of regionally reported CODIS short tandem repeats (STR), including all 23 markers and their alleles. These markers correspond to 23 STR that are located across the human genome, are not related to the KRAS gene, and have been amply used in the forensic sciences for individual identification and population characterizations. They are amplified by PCR and identified by capillary electrophoresis. The largest known study to characterize these markers among different regions in Colombia, included over 11.000 samples and was the data source for the model [18]. Collinearities were manually excluded from the model. All analyses were performed with R (version 4.02, The R Foundation).

Results

Validation of mutation detection methods

Next-generation sequencing was performed paired with ddPCR in 60 tissue samples for the validation of the ddPCR diagnostic test. In the NGS group, a total of 8 samples revealed a *KRAS* exon 2 p.G12C mutation compared to 10 in the ddPCR samples. Discordant samples were subjected to Sanger sequencing to evaluate possible false negatives. Considering NGS and Sanger sequencing and a mutation prevalence of 13.3%, ddPCR in tissue samples revealed a sensitivity of 100% (95%CI 63 - 100%) and a specificity of 96,2% (95%CI 88.6 - 99.5%). A positive predictive value of 80% (95%CI 44 - 97%) and a negative predictive value of 100% (95%CI 93 – 100%) were also estimated.

Detection in patient samples

A total of 979 samples among 19 regions were included in the analysis. Regions and samples were excluded from the initially screened 1000 samples if the region of collection lacked enough samples to be representative or the region did not contribute to any number of samples. Eight samples were considered the threshold value, considering that one sample was expected to be detected among those 8 considering an estimated prevalence of 12.5%. Sample and region inclusions and exclusions are presented in Fig. 1. 78 occurrences of a *KRAS* exon 2 p.G12C mutation were detected among sampled NSCLC patients, yielding a national prevalence of 7.97% (95%CI 6.27–9.66%). Differences in several regions were also observed. Antioquia, revealed a mutational prevalence of 12.7% (95%CI 9.1–16.3%) in contrast to other regions such as Bogota DC (Capital region) with 5.4% (2.7–8.2%) and Bolivar with 2.4% (95%CI 0–7.2%) (p value = 0.00262 for Bolivar and Bogota compared to Antioquia). Fig. 2 depicts regional prevalence and geographic distribution of regional positivity and national average. These results indicate a variety of mutational prevalence among regions and highlight a possible population substructure that presents with enriched or depleted *KRAS* exon 2 p.G12C mutational prevalence.

Population markers and KRAS exon 2 p.G12C mutations

Multiple linear regression identified a series of markers that yielded a relevant R2 value and p-value. The regional allelic frequencies of D13S317 12, D19S433 13.2, D1S1656 11, D1S1656 16, D22S1045 15, D22S1045 17, D2S1338 21, D2S441 12, D2S441 12.2, D7S820 11, D8S1179 13, FGA 19, TH01 6, TP0 × 9, Via 20. A trade-off between these two parameters was taken to construct a model with relatively few predictors and good predictive performance. Fig. 3 represents a relationship between the number of predictors and adjusted R2 and their combined association and the same metric, with an adjusted R2 of 0.945 and a p-value of < 0.0001, a model of 8 predictors was chosen. Table 1 depicts chosen markers and their corresponding coefficients in the constructed model.

Since the representation of all regions could not be achieved, *KRAS* exon 2 p.G12C mutation prevalence in non-represented regions could not be obtained; allelic frequencies for chosen STR were imputed in the model to obtain these values. Fig. 4 presents measured frequencies and estimated ones across all the national territories. This model indicates a high intranational and interregional variation with an estimated

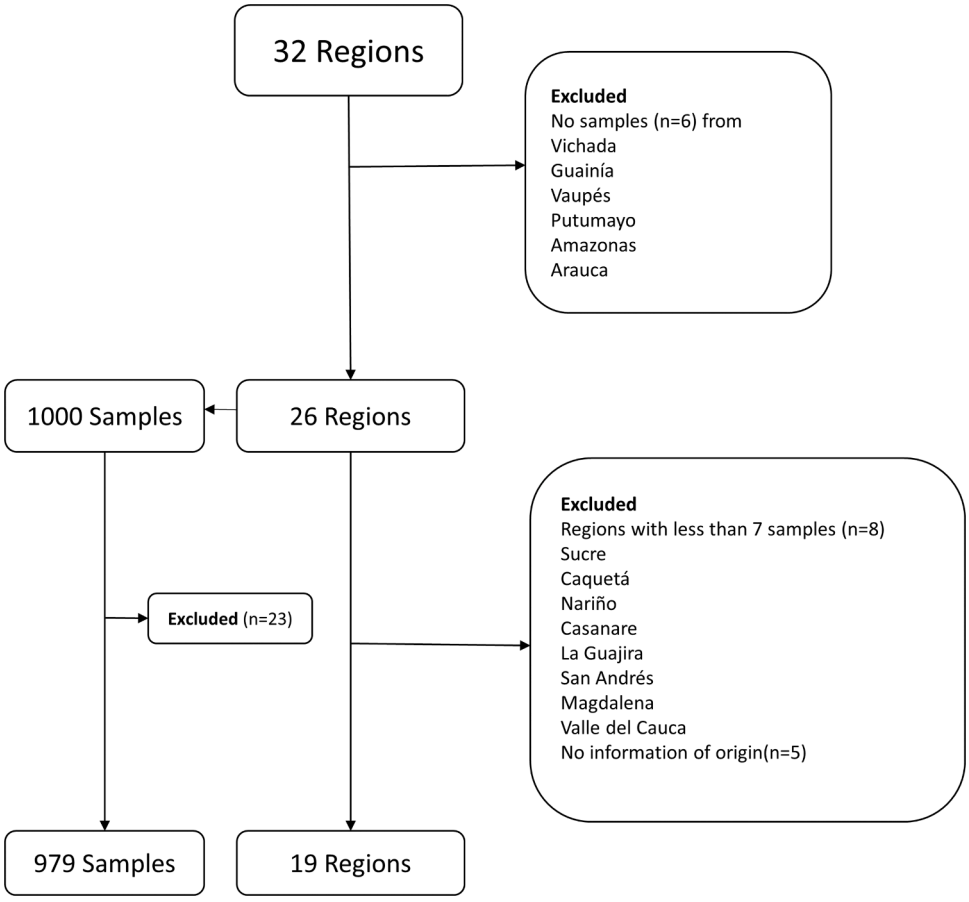

Fig. 1. CONSORT Diagram of analyzed samples and regions.

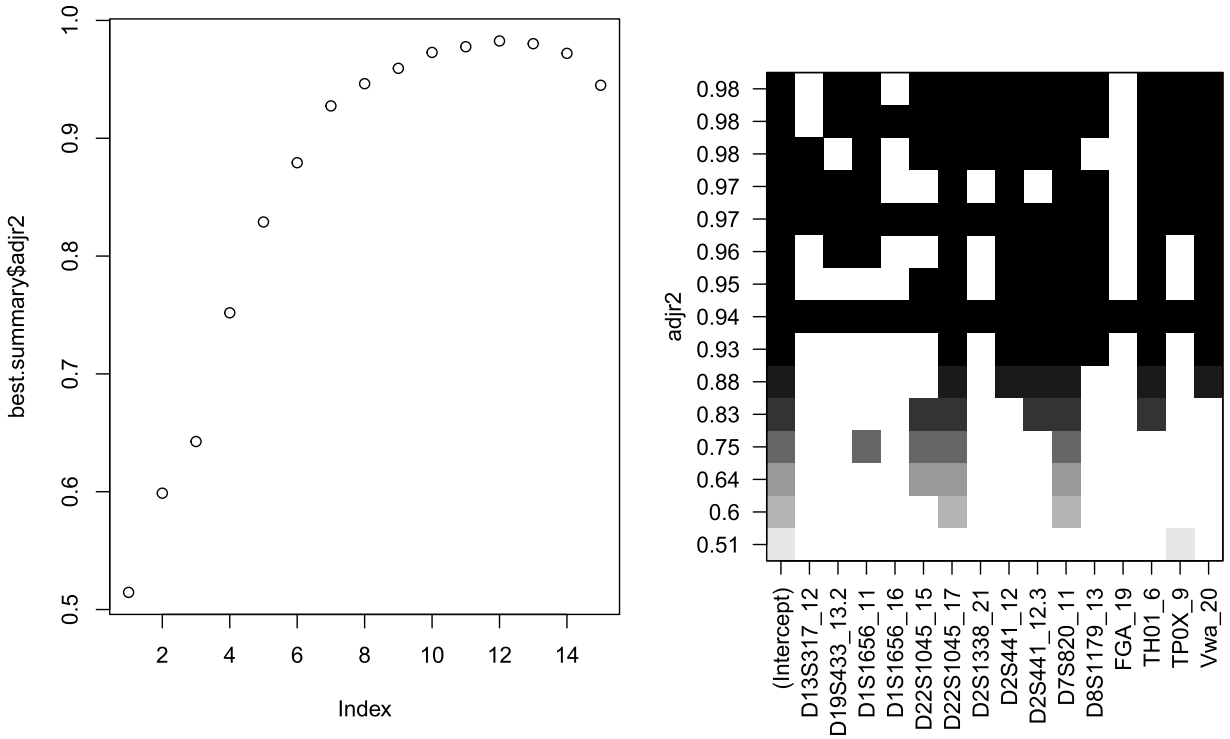

Fig. 2. Regression Model selection based on adjusted R2 in relationship to number and combinatory of predictors. The left panel depics adjusted R2 in relationship with number of markers. The right panel identifies as a black square including several combinations of markers and their impact on adjusted R2, represented in the Y axis.

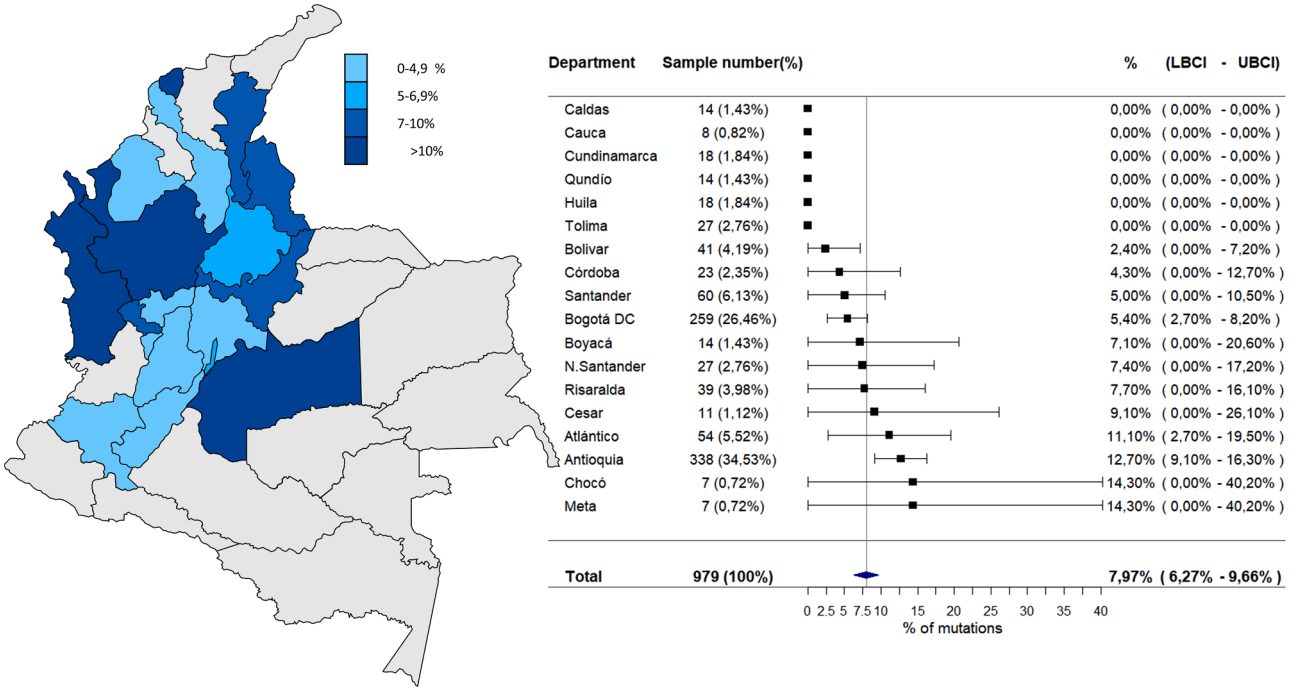

Fig. 3. Geographic distribution and regional mutation prevalence.

Table 1
Multiple regression model and performance.

Residuals				
Min	1q	Median	3q	Max
−1.001	−0.66	−0.052	0.578	1.586
Coefficients	**β**	**Standard error**	***T* value**	***P*-value**
Intercept	−7.971	10.48	−0.761	0.4687
D22S1045_15	−26.755	13.11	−2.04	0.0756
D22S1045_17	172.169	16.27	10.581	< 0.0001
D2S441_12	95.645	19.65	4.867	0.001
D2S441_12.3	−324.392	52.49	−6.180	< 0.0001
D7S820_11	−111.804	15.20	−7.352	< 0.0001
D8S1179_13	−26.201	12.53	−2.091	0.069
TH01_6	127.062	18.21	6.975	< 0.0001
Vwa_20	540.786	139.33	3.882	0.004
Residual standard error		1.21 in 8 degrees of freedom		
Multiple R2		0.973		
Adjusted R2		0.946		
F statistic		36.26 in 8 degrees of freedom		
p-value		< 0.0001		

prevalence reaching up to 20.04% in La Guajira. Other regions behaved around the median value, others tending to be low, behaving similarly as seen in the measured regions.

Discussion

The genomic landscape of LUADs is strikingly varied in Latin American patients with mixed ancestries. Previous studies demonstrated that native ancestry was correlated with somatic driver alterations, including *EGFR* and *KRAS* mutations that small molecule inhibitors can effectively target to prolong survival [19,20], and TMB and *STK11* are potential prognostic biomarkers in lung cancer patients [21,22]. Cavagna et al. recently reported a frequency of *KRAS* mutations in an open population of patients with NSCLC from Brazil, with the exon 2 p.G12C variant being 9% of the cases. They also found that *KRAS* mutations were associated with smoking status (current or quitter) and worse overall survival [17]. Our study demonstrated a national prevalence of 7.97% of *KRAS* exon 2 p.G12C mutations, like that previously reported in Brazil. We also identified that certain regions presented with lower [2.4% (95% CI 0–7.2%); 5.4% (95% CI2.7 - 8.2%)] and higher frequencies [12.7% (95% CI 9.1 −16.3%)] than the national average. In addition, evaluating the association of these prevalence and population markers based on short tandem repeats (CODIS STR), we also identified that eight markers and their allelic frequencies were highly associated with regional positivity.

A recent review reported that *KRAS* mutations are present in 18%−32% of lung adenocarcinoma, 12.8% of large cell carcinoma, 10% of adenosquamous carcinomas, and 1.6%−7.1% of squamous-cell carcinomas in White patients [23]. Moreover, African-American patients with NSCLC are more frequently identified with *KRAS* mutations than White patients. The frequency of *KRAS* mutations in Western populations with lung adenocarcinoma is about 26% and about 6% in the squamous-cell carcinoma population. In Asian patients, the frequency of *KRAS* mutations is 11.2% of patients with NSCLC [24]. According to The Cancer Genome Atlas, *KRAS* mutations are found in 33% of lung adenocarcinoma [4]. Our previous study involving 5738 NSCLC cases reported 14% of *KRAS*-mutated cases in Latin American except for Brazil (Argentina, Mexico, Colombia, Peru, Costa Rica, and Panama) [19].

The detection of the *KRAS* exon 2 p.G12C mutation is not homogenous for all diagnostic tests. Our results highlight that different performances can be obtained from different methodologies. PCR is a cost-saving alternative that can be implemented in low to middle-income countries where the costs of conducting NGS on all newly diagnosed patients remain prohibitive. In addition, taking into consideration a negative predictive value of 100%, it would be safe to assume that it can be used as a screening methodology. If required, confirmatory sequence analysis is left for a minority of patients, optimizing resources while identifying candidates for newly available medications [12,14].

This study also indicates that the prevalence of *KRAS* exon 2 p.G12C prevalence follows a pattern that relies on population composition and explains to some degree, the variation in different regions around Latin America and the world. It was also conducted on a nation with high ancestral population diversity, leading to different population substructures within the same geographical regions. In addition, administrative regions tend to follow a pattern of ancestral migration and ethnic composition that serve as the perfect comparative scenario, thus serving

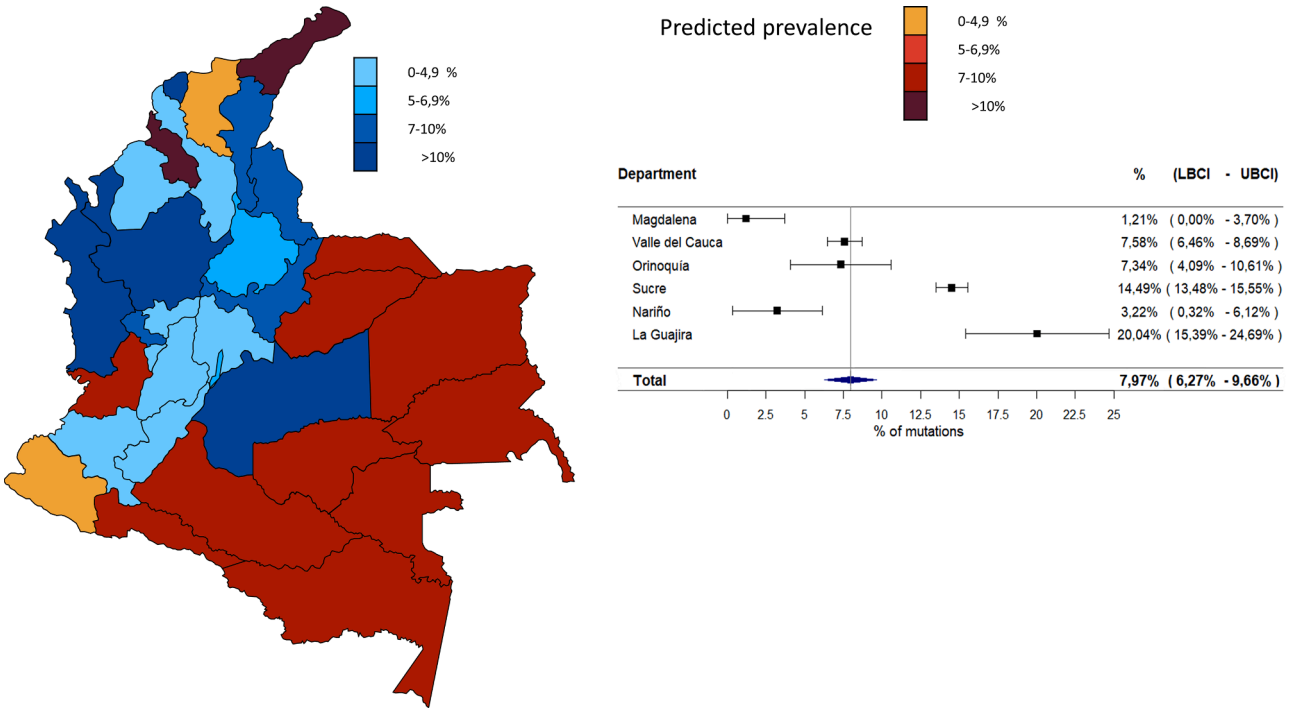

Fig. 4. Geographical representation of estimated prevalence in regions compared to measured regions. Shades of blue indicate measured *KRAS* exon 2 p.G12C prevalence, in a similar manner as observed in Fig. 2. Shades of red and yellow as well as forest plot, indicate estimated prevalence for each missing region using the multiple regression model.

as comparative micro ethnic scenarios [25]. Since phenotypes for mixed individuals fall outside the traditional Asian, Black, White, or Hispanic characterizations, molecular ancestral and population markers offer a higher degree of substructure identification. Four contributors for ancestral populations for Colombia have been identified. These are Native America, European, Mediterranean and African. Among these four, each has varying degree of contribution to certain regions. In addition, there are also representation from different subpopulations among these large contributors. In the case of native American, in whom Asian ancestry has been identified, Andes Piedmont contributions, that originated from Ecuador and northern Peru have been identified in the southern and central regions of Colombia. The northern regions tend to represent Chibcha and Paez ancestries that are indigenous to the region. Considering European ancestries, Central and southern Spain markers have been identified across all sampled regions. Minor contributions from Portuguese and Western Spain were found in the northern and western region of the country. Moreover, Eastern Mediterranean identifiers are present in the central region, whereas Sephardic in the eastern and northern region, coinciding with Antioquia, the region with one of the highest number of positive p.G12C mutants. Finally, African contribution comprised in the vast majority of West African descent. Interestingly, African markers are vastly represented in the coastal regions and are rather scarce in the center of the nation. These findings indicate that a clear population substructure from different ancestries exists in Colombia [25].

These substructures are important in determining the impact on molecular phenotypes in lung cancer. Regional variation of *KRAS* exon 2 p.G12C and STR was identified within regions confirming the hypotheses that population substructures that better characterize average prevalence exist. Population factors such as tobacco consumption has been associated with *KRAS* positivity [2]. Interestingly this variable was not taken into consideration since tobacco consumptions rates are relatively stable among included administrative regions in Colombia [26], suggesting a higher contribution of genetic aspects.

Since the prevalence of *KRAS* exon 2 p.G12C mutations and their variation within other nations has not widely been determined, the estimation of the individual population using STR allelic frequencies becomes useful. These frequencies have been widely studied since they are used in individual identification and paternity testing [27–29].

This approach of identifying associations with surrogate markers that are more available than *KRAS* exon 2 p.G12C mutations for specific populations allows for health care systems around the world to estimate their regional positivity rate with the aims of resource allocation.

Since representative samples of certain regions could not be obtained in this study, estimation of *KRAS* exon 2 p.G12C mutations yielded an exciting picture of mutation distributions in these regions. Sucre, located in the northern and costal region, yielded a similar prevalence as Antioquia, closing on 14%. Since regional closeness could potentially correspond to migratory patterns, it would be expected that the prevalence would fall nearby. La Guajira follows a similar pattern reaching close to 20% with a much wider confidence interval that encompasses Sucre and Antioquia, indicating that similar prevalence would be expected.

This study is not free of limitations. Since the characterization of *KRAS* exon 2 p.G12C mutations was conducted with ddPCR, a possibility for error in diagnosis, with an increased overestimation, is worth mentioning. Furthermore, model selection could have followed fewer STRs, warranting less overfitting by sacrificing performance. This, in turn, warrants that these results should be validated in other populations in which a similar analysis is conducted. Finally, the decision to

determine regional variability based on administrative regions could have potentially decreased population precision by including more individuals, but due to data availability, especially STR frequencies, this compromise had to be made.

Conclusion

In conclusion, implementing a screening program of *KRAS* exon 2 p.G12C mutation in LUAD patients with ddPCR could be considered based on good diagnostic performance metrics compared to NGS. In addition, this type of mutation follows a population substructure distribution pattern that determines their prevalence. Furthermore, by denoting these substructures with population markers such as STRs', the prevalence of *KRAS* exon 2 p.G12C mutations can be inferred.

CRediT authorship contribution statement

Alejandro Ruiz-Patiño: Conceptualization, Data curation, Formal analysis, Validation, Writing – original draft, Writing – review & editing. **July Rodríguez:** Data curation, Methodology, Writing – review & editing. **Andrés F. Cardona:** Conceptualization, Formal analysis, Supervision, Validation, Writing – original draft. **Jenny Ávila:** Data curation, Methodology, Writing – review & editing. **Pilar Archila:** Data curation, Methodology, Writing – review & editing. **Hernán Carranza:** Data curation, Methodology, Writing – review & editing. **Carlos Vargas:** Data curation, Methodology, Writing – review & editing. **Jorge Otero:** Data curation, Methodology, Writing – review & editing. **Oscar Arrieta:** Supervision, Project administration, Writing – original draft. **Lucia Zatarain-Barrón:** Supervision, Project administration, Writing – original draft. **Carolina Sotelo:** Supervision, Project administration, Writing – review & editing. **Camila Ordoñez:** Supervision, Project administration, Writing – original draft. **Juan Esteban García-Robledo:** Supervision, Project administration, Writing – review & editing. **Leonardo Rojas:** Data curation, Writing – review & editing. **Maritza Bermúdez:** Methodology, Writing – review & editing. **Tatiana Gámez:** Methodology, Writing – review & editing. **Diana Mayorga:** Methodology, Writing – review & editing. **Luis Corrales:** Validation, Writing – review & editing. **Claudio Martín:** Validation, Writing – review & editing. **Gonzalo Recondo:** Validation, Writing – review & editing. **Luis Mas:** Validation, Writing – review & editing. **Suraj Samtani:** Validation, Writing – review & editing. **Luisa Ricaurte:** Validation, Writing – review & editing. **Umberto Malapelle:** Validation, Writing – review & editing. **Alessandro Russo:** Validation, Writing – review & editing. **Feliciano Barrón:** Validation, Writing – review & editing. **Nicolas Santoyo:** Validation, Writing – review & editing. **Christian Rolfo:** Conceptualization, Project administration, Writing – original draft. **Rafael Rosell:** Conceptualization, Project administration, Writing – original draft.

CRediT authorship contribution statement

Alejandro Ruiz-Patiño: Conceptualization, Data curation, Formal analysis, Validation, Writing – original draft, Writing – review & editing. **July Rodríguez:** Data curation, Methodology, Writing – review & editing. **Andrés F. Cardona:** Conceptualization, Formal analysis, Supervision, Validation, Writing – original draft. **Jenny Ávila:** Data curation, Methodology, Writing – review & editing. **Pilar Archila:** Data curation, Methodology, Writing – review & editing. **Hernán Carranza:** Data curation, Methodology, Writing – review & editing. **Carlos Vargas:** Data curation, Methodology, Writing – review & editing. **Jorge Otero:** Data curation, Methodology, Writing – review & editing. **Oscar Arrieta:** Supervision, Project administration, Writing – original draft. **Lucia Zatarain-Barrón:** Supervision, Project administration, Writing – original draft. **Carolina Sotelo:** Supervision, Project administration, Writing – review & editing. **Camila Ordoñez:** Supervision, Project administration, Writing – original draft. **Juan Esteban García-Robledo:** Supervision, Project administration, Writing – review & editing. **Leonardo Rojas:** Data curation, Writing – review & editing. **Maritza Bermúdez:** Methodology, Writing – review & editing. **Tatiana Gámez:** Methodology, Writing – review & editing. **Diana Mayorga:** Methodology, Writing – review & editing. **Luis Corrales:** Validation, Writing – review & editing. **Claudio Martín:** Validation, Writing – review & editing. **Gonzalo Recondo:** Validation, Writing – review & editing. **Luis Mas:** Validation, Writing – review & editing. **Suraj Samtani:** Validation, Writing – review & editing. **Luisa Ricaurte:** Validation, Writing – review & editing. **Umberto Malapelle:** Validation, Writing – review & editing. **Alessandro Russo:** Validation, Writing – review & editing. **Feliciano Barrón:** Validation, Writing – review & editing. **Nicolas Santoyo:** Validation, Writing – review & editing. **Christian Rolfo:** Conceptualization, Project administration, Writing – original draft. **Rafael Rosell:** Conceptualization, Project administration, Writing – original draft.

Acknowledgements

This work was financed in cooperation with Amgen Biotecnológica S. A.S.

Disclaimer

For administrative region reference, please refer to the following link: https://upload.wikimedia.org/wikipedia/commons/6/67/Departments_of_colombia.svg

Declaration of Competing Interest

Andrés F. Cardona discloses financial research support from Merck Sharp & Dohme, Boehringer Ingelheim, Roche, Bristol-Myers Squibb, Foundation Medicine, Roche Diagnostics, Termo Fisher, Broad Institute, Amgen, Flatiron Health, Teva Pharma, Rochem Biocare, Bayer, INQBox and The Foundation for Clinical and Applied Cancer Research – FICMAC. Additionally, he was linked and received honoraria as an advisor, participate in speakers' bureau. He gave expert testimony to EISAI, Merck Serono, Jannsen Pharmaceutical, Merck Sharp & Dohme, Boehringer Ingelheim, Roche, Bristol-Myers Squibb, Pfizer, Novartis, Celldex Therapeutics, Foundation Medicine, Eli Lilly, Guardant Health, Illumina, and Foundation for Clinical and Applied Cancer Research – FICMAC.

Oscar Arrieta reports personal fees from Pfizer, grants and individual fees from Astra Zeneca, grants and individual fees from Boehringer-Ingelheim, Lilly, Merck, Bristol Myers Squibb, Roche, outside the submitted work.

Christian Rolfo reports relation with Mylan, Archer Biosciences, Oncopass, Inivata, Merck Serono Novartis, MSD, Boehringer Ingelheim, Guardant Health, etc. AstraZeneca as part of Speakers' Bureau. Also, he received research funding from Pfizer and had uncompensated Relationships with OncoDNA, Biomark, and Guardant Health.

Gonzalo Recondo discloses travel, accommodations, and expenses from AstraZeneca and Pfizer, and consulting advisory role with Roche, Amgen, and Pfizer.

Umberto Malapelle reports personal fees from Boehringer Ingelheim, AstraZeneca, Roche, MSD, Amgen, Merck, BMS for participation in a speaker bureau; personal fees from Boehringer Ingelheim, MSD, Amgen, Merck, BMS for acting in an advisory role, financial support from Boehringer Ingelheim and Amgen that was paid directly to his institution outside the submitted work.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tranon.2021.101276.

References

[1] D.R. Youlden, S.M. Cramb, P.D. Baade, The International Epidemiology of Lung Cancer: geographical distribution and secular trends, J. Thorac. Oncol. 3 (8) (2008) 819–831. Aug.

[2] S. Rodenhuis, M.L. van de Wetering, W.J. Mooi, S.G. Evers, N. van Zandwijk, J. L. Bos, Mutational activation of the K-ras oncogene. A possible pathogenetic factor in adenocarcinoma of the lung, N. Engl. J. Med. 317 (15) (1987) 929–935. Oct 8.

[3] Y.S. Ju, W.C. Lee, J.Y. Shin, S. Lee, T. Bleazard, J.K. Won, et al., A transforming KIF5B and RET gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing, Genome Res. 22 (3) (2012) 436–445. Mar.

[4] Cancer Genome Atlas Research Network, Comprehensive molecular profiling of lung adenocarcinoma, Nature 511 (7511) (2014) 543–550. Jul 31.

[5] C. Zhang, M. Jiang, N. Zhou, H. Hou, T. Li, H. Yu, et al., Use tumor suppressor genes as biomarkers for diagnosis of non-small cell lung cancer, Sci. Rep. 11 (1) (2021) 3596. Feb 12.

[6] T. Kohno, T. Nakaoku, K. Tsuta, K. Tsuchihara, S. Matsumoto, K. Yoh, et al., Beyond ALK-RET, ROS1 and other oncogene fusions in lung cancer, Transl. Lung Cancer Res. 4 (2) (2015) 156–164. Apr.

[7] S.V. Liu, A.M. Vanderwalde, H. Mamdani, L.E. Raez, Y. Baca, J. Xiu, et al., Characterization of KRAS mutations (mt) in non-small cell lung cancer (NSCLC), JCO 38 (15_suppl) (2020). May 20 9544–9544.

[8] L.F. Leal, F.E. de Paula, P. De Marchi, L. de Souza Viana, G.D.J. Pinto, C.D. Carlos, 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. 9 (1) (2019) 3209. Mar 1.

[9] M.L. Johnson, C.S. Sima, J. Chaft, P.K. Paik, W. Pao, M.G. Kris, et al., Association of KRAS and EGFR mutations with survival in patients with advanced lung adenocarcinomas, Cancer 119 (2) (2013) 356–362. Jan 15.

[10] E. Bange, M.E. Marmarelis, W.T. Hwang, Y.X. Yang, J.C. Thompson, J. Rosenbaum, et al., Impact of KRAS and TP53 Co-mutations on outcomes after first-line systemic therapy among patients with STK11-mutated advanced non-small-cell lung cancer, JCO Precis Oncol. 3 (2019).

[11] La Fleur L., Falk-Sörqvist E., Smeds P., Berglund A., Sundström M., Mattsson J.S., et al. Mutation patterns in a population-based non-small cell lung cancer cohort and prognostic impact of concomitant mutations in KRAS and TP53 or STK11. Lung Cancer. 2019 Apr;130:50–8.

[12] D.S. Hong, M.G. Fakih, J.H. Strickler, J. Desai, G.A. Durm, G.I. Shapiro, et al., KRASG12C Inhibition with sotorasib in advanced solid tumors, N. Engl. J. Med. 383 (13) (2020) 1207–1217. Sep 24.

[13] M. Nagasaka, Y. Li, A. Sukari, S.H.I. Ou, M.N. Al-Hallak, A.S. Azmi, KRAS G12C game of thrones, which direct KRAS inhibitor will claim the iron throne? Cancer Treat. Rev. 84 (2020), 101974. Mar.

[14] P.A. Jänne, I.I. Rybkin, A.I. Spira, G.J. Riely, K.P. Papadopoulos, J.K. Sabari, et al., KRYSTAL-1: activity and safety of Adagrasib (MRTX849) in advanced/metastatic non–small-cell lung cancer (NSCLC) harboring KRAS G12C Mutation [Internet], in: Proceedings of the 32nd Symposium ENA 2020 EORTC NCI AACR, 2020 [cited 2021 Aug 3]. Available from, https://cm.eortc.org/cmPortal/Searchable/ENA2020/config/normal#!abstractdetails/0000902150.

[15] J. Canon, K. Rex, A.Y. Saiki, C. Mohr, K. Cooke, D. Bagal, et al., The clinical KRAS (G12C) inhibitor AMG 510 drives anti-tumour immunity, Nature 575 (7781) (2019) 217–223. Nov.

[16] J. Carrot-Zhang, G. Soca-Chafre, N. Patterson, A.R. Thorner, A. Nag, J. Watson, et al., Genetic ancestry contributes to somatic mutations in lung cancers from admixed latin american populations, Cancer Discov. 11 (3) (2021) 591–598. Mar.

[17] R. Cavagna, F. Escremim de Paula, D. Sant'Anna, I. Santana, V.D. da Silva, E.C. A. da Silva, et al., Frequency of KRAS p.Gly12Cys Mutation in brazilian patients with lung cancer, JCO Glob. Oncol. 7 (2021) 639–645. Apr.

[18] Galeano Molina M.D.S. Nuevos datos estadísticos genético-poblacionales para microsatélites en Colombia. 2019.

[19] O. Arrieta, A.F. Cardona, C. Martín, L. Más-López, L. Corrales-Rodríguez, G. Bramuglia, 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) (2015) 838–843. May.

[20] N.T. Gimbrone, B. Sarcar, E.R. Gordian, J.I. Rivera, C. Lopez, S.J. Yoder, et al., Somatic mutations and ancestry markers in Hispanic lung cancer patients, J. Thorac. Oncol. 12 (12) (2017) 1851–1856. Dec.

[21] K.C. Arbour, E. Jordan, H.R. Kim, J. Dienstag, H.A. Yu, F. Sanchez-Vega, et al., Effects of Co-occurring genomic alterations on outcomes in patients with KRAS-mutant non-small cell lung cancer, Clin. Cancer Res. 24 (2) (2018) 334–340. Jan 15.

[22] D.B. Johnson, G.M. Frampton, M.J. Rioth, E. Yusko, Y. Xu, X. Guo, et al., Targeted next generation sequencing identifies markers of response to PD-1 blockade, Cancer Immunol. Res. 4 (11) (2016) 959–967. Nov.

[23] O. Calvayrac, A. Pradines, E. Pons, J. Mazières, N. Guibert, Molecular biomarkers for lung adenocarcinoma, Eur. Respir. J. 49 (4) (2017), 1601734. Apr.

[24] S. Dearden, J. Stevens, Y.L. Wu, D. Blowers, Mutation incidence and coincidence in non small-cell lung cancer: meta-analyses by ethnicity and histology (mutMap), Ann. Oncol. 24 (9) (2013) 2371–2376. Sep.

[25] J.C. Chacón-Duque, K. Adhikari, M. Fuentes-Guajardo, J. Mendoza-Revilla, V. Acuña-Alonzo, R. Barquera, et al., Latin Americans show wide-spread Converso ancestry and imprint of local Native ancestry on physical appearance, Nat. Commun. 9 (1) (2018) 5388. Dec 19.

[26] DANE. Encuesta nacional de consumo de sustancias psicoactivas (ENCSPA) [Internet]. 2020 [cited 2021 Nov 1]. Available from: https://www.dane.gov.co/files/investigaciones/boletines/encspa/bt-encspa-2019.pdf.

[27] V.R.C. Aguiar, A.M. de Castro, V.C.O. Almeida, F.S.V. Malta, A.C.S. Ferreira, I. D. Louro, New CODIS core loci allele frequencies for 96,400 Brazilian individuals, Forensic Sci. Int. Genet. 13 (2014) e6–12. Nov.

[28] C.M. Ruitberg, D.J. Reeder, J.M. Butler, STRBase: a short tandem repeat DNA database for the human identity testing community, Nucleic Acids. Res. 29 (1) (2001) 320–322. Jan 1.

[29] T.R. Moretti, L.I. Moreno, J.B. Smerick, M.L. Pignone, R. Hizon, J.S. Buckleton, et al., Population data on the expanded CODIS core STR loci for eleven populations of significance for forensic DNA analyses in the United States, Forensic Sci. Int. Genet. 25 (2016) 175–181. Nov.