Article

A Pilot Study: Contrasting Genomic Profiles of Lung Adenocarcinoma Between Patients of European and Latin American Ancestry

Bertha Rueda-Zarazua [1,2], Humberto Gutiérrez [3], Humberto García-Ortiz [4], Lorena Orozco [4], Gustavo Ramírez-Martínez [5], Luis Jiménez-Alvarez [5], Francina V. Bolaños-Morales [6], Joaquín Zuñiga [5,7], Federico Ávila-Moreno [8,9] and Jorge Melendez-Zajgla [2,*]

[1] Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico; bgrueda@inmegen.edu.mx
[2] Laboratorio de Genómica Funcional del Cáncer, Instituto Nacional de Medicina Genómica, Mexico City 14610, Mexico
[3] Instituto Nacional de Medicina Genómica, Mexico City 14610, Mexico; hgutierrez@inmegen.gob.mx
[4] Laboratorio de Inmunogenómica y Enfermedades Metabólicas, Instituto Nacional de Medicina Genómica, Mexico City 14610, Mexico; hgarcia@inmegen.gob.mx (H.G.-O.); lorozco@inmegen.gob.mx (L.O.)
[5] Laboratorio de Inmunología y Genética, Instituto Nacional de Enfermedades Respiratorias, Mexico City 14080, Mexico; grmunam@yahoo.com.mx (G.R.-M.); jimenez.alvarez.l@gmail.com (L.J.-A.); joaquin.zuniga@iner.gob.mx (J.Z.)
[6] Subdirección de Cirugía, Instituto Nacional de Enfermedades Respiratorias, Mexico City 14080, Mexico; francina.valezka@iner.gob.mx
[7] Tecnologico de Monterrey, Escuela de Medicina y Ciencias de la Salud, Mexico City 14380, Mexico
[8] Lung Diseases and Functional Epigenomics Laboratory (LUDIFE), Biomedicine Research Unit (UBIMED), Facultad de Estudios Superiores-Iztacala, Universidad Nacional Autónoma de México, Tlalnepantla 54090, Mexico; f.avila@unam.mx
[9] Research Tower, Subdirección de Investigación Básica, Instituto Nacional de Cancerología (INCan), Mexico City 14080, Mexico
* Correspondence: jmelendez@inmegen.gob.mx; Tel.: +52-(55)-53501900 (ext. 1901)

Abstract: Lung cancer remains as the leading cause of cancer mortality worldwide. However, while current evidence suggests the existence of genomic differences between populations, indicating different risk factors associated with population-level genetic backgrounds, most studies have concentrated on populations of European ancestry, and more research is needed on non-European populations. We analyzed whole-exome sequencing data from 25 Mexican lung adenocarcinoma patients and compared them with a TCGA-PanCancer cohort enriched with patients of European ancestry as reference. Clinically relevant germline variants in cancer susceptibility genes are more frequent in our cohort (32% vs. 6.4%) than in the reference. Several mutational signatures (SBS32, SBS85, SBS12, SBS19) occurred at significantly higher frequencies in the Mexican cohort compared to the reference ($p < 0.0001$). Interestingly, the smoking-associated signature SBS4, present in 67.6% of smokers in the reference cohort, was absent in smoking Mexican patients ($p < 0.01656$). Somatic variant frequencies in SLC36A4 (20%; $p < 0.00002$), AP1S1 (8%; $p < 0.00002$), and TP53 (16%; $p = 0.00005$) showed significant differences from the European reference cohort. We demonstrate that all these observed biases were independent of the sample size. This study uncovers distinct genomic biases in lung cancer carcinogenesis in this population, compared to a European ancestry reference population, suggesting implications for precision medicine strategies in Latin American populations.

Keywords: NSCLC; lung adenocarcinoma; genomic profile; Mexico; exome

1. Introduction

Lung cancer continues to be the most common and deadly type of cancer worldwide. Lung adenocarcinoma (LUAD) stands as the most prevalent subtype of lung cancer in the world [1,2]. The genomic landscape of lung adenocarcinoma has been extensively examined, revealing notable differences across various populations. Among the most scrutinized aspects is the variation in the frequency of EGFR mutations. These mutations are predominantly associated with individuals of Asian ancestry, females, and nonsmokers, and are often associated with a more favorable prognosis [3,4]. In contrast, KRAS is frequently mutated in populations of European ancestry and is mainly associated with smoking. In Latin American countries, observed frequencies of mutations in both genes occupy an intermediate position between those of the previously mentioned populations. This has sparked vigorous debates, pointing towards the influence of genetic backgrounds or the impact of distinct risk factors tied to each population as cultural and environmental variations [5,6]. This pattern underscores the importance of considering genetic diversity in the diagnosis and treatment of lung adenocarcinoma.

The underrepresentation of certain populations in large-scale genomic studies, particularly those from Latin American countries, often leads to the unwarranted overgeneralization of findings derived predominantly from European or Asian cohorts. These generalized conclusions can have profound implications for clinical decision-making, potentially misguiding the understanding of disease progression and therapeutic responses. Latin American populations are uniquely characterized by their rich genetic diversity, shaped by complex migration patterns, cultural heterogeneity, and distinct sociodemographic factors. This lack of representation not only delays the equitable distribution of benefits from medical research to these populations but also impedes the advancement of precision medicine approaches tailored to their specific genetic and environmental contexts. Addressing this gap is crucial to fostering inclusive research and ensuring that genomic medicine serves global health needs effectively [7–9].

Research in Mexican populations has identified several clinically significant genetic variants and has also highlighted specific risk factors for lung cancer, such as exposure to wood smoke, which significantly impacts a large segment of the population [10,11]. However, previous studies have only included a limited number of genes of interest, and there is still much to be understood about the peculiarities that may exist within this population. This study aims to identify, using whole exome sequencing, differences in the mutational profiles of lung adenocarcinoma among Mexican patients, compared to those typically observed in European ancestry patients.

2. Results

2.1. Study Population

In the present study, 25 patients with confirmed diagnosis of lung adenocarcinoma were included. All patients were from the central region of Mexico and were diagnosed at the National Institute of Respiratory Diseases in Mexico City. Denominators correspond to the patient's clinical information where available. The study included 11 males and 14 females. The majority of the patients (9 out of 14, 64%) were at stage IV and the mean age was 58.8 ± 14.5 years. A total of 66% (8 out of 12) of the participants reported a family history of cancer. Among the risk factors, 60% (10 out of 17) reported a history of smoking, while 28% (4 out of 14) reported exposure to wood smoke as a risk factor, all of whom were women. Three patients reported exposure to pesticides, fine clay dust, and asbestos as other risk factors (Figure 1A).

In order to determine the ancestry of our patients, we employed SNPs masked from three distinct base populations: African (YRI), northern European (CEU), and indigenous

Mexican from the García-Ortiz's previous study [12]. Our cohort's primary contribution is indigenous Mexican ancestry, accounting for 53.27% of the total. This is followed by European ancestry at 42.38% and African ancestry at 4.8% (Figure 1B). The population structure of our cohort is comparable to that observed in a group of open population controls of Mexicans from central Mexico, where contributions of Indigenous Mexican, European, and African ancestry were observed to be 69.52%, 25.56%, and 2.68%, respectively, using the same ancestry references [13]. All patients have some proportion of indigenous Mexican ancestry.
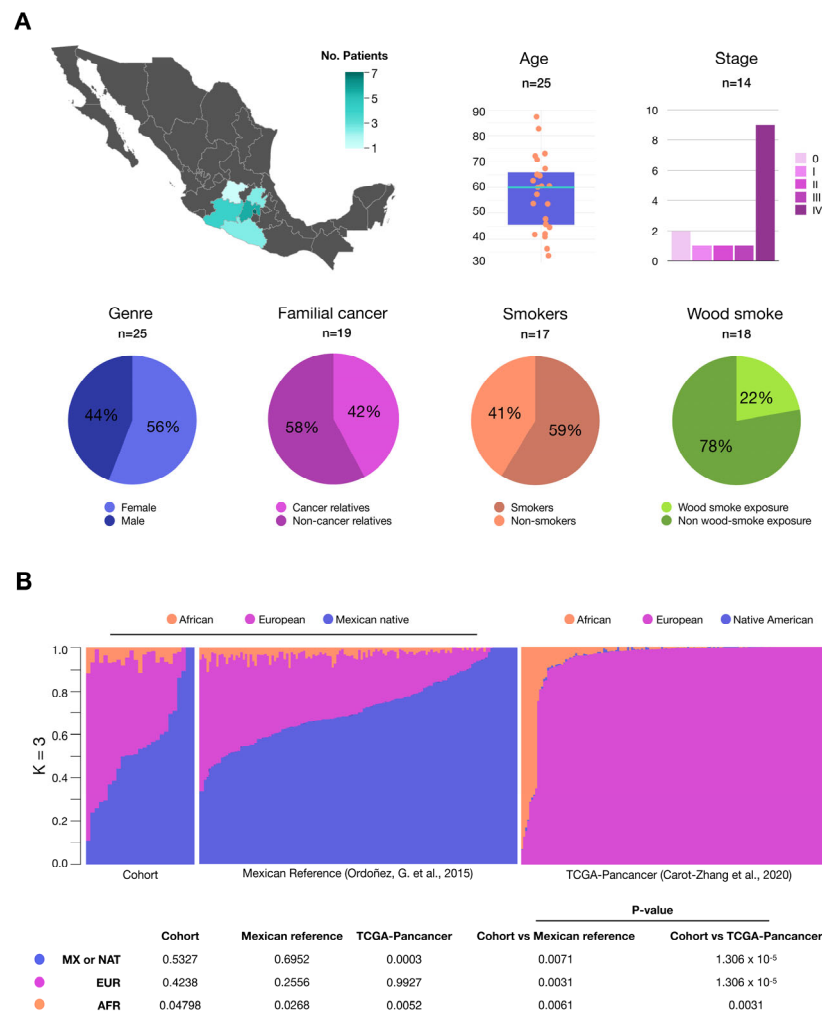

Figure 1. General description of the cohort of Mexican patients with lung adenocarcinoma used in this study. (**A**) Clinical features and risk factors of our Mexican patients. (**B**) Population ancestry structure of our cohort of Mexican mestizos contrasted with an open population control group from central [13] Mexico and the TCGA-PanCancer reference cohort. The Mexican cohorts used genetic markers of European, African, and indigenous Mexican ancestry, while the TCGA-PanCancer cohort [14] used markers of Native American ancestry, which is closely related to indigenous Mexican populations. Estimated individual proportions of ancestral backgrounds revealed a significant contribution from Native Mexican ancestry in the Mexican cohorts. In contrast, the TCGA reference cohort exhibited a predominant contribution from European ancestry. Comparison between cohorts was performed with one-sample Wilcoxon tests using medians.

We use the TCGA-PanCancer cohort (syn1759475) as a reference, which predominantly consists of patients of European ancestry, reflecting the most extensively studied population. This cohort, comprising over 90% European ancestry (Figure 1B) [14], is approximately 10 times larger than our own and is well-documented. It includes a higher

proportion of smokers (85% vs. 59%) and exhibits a more uniform distribution of disease stages, which may introduce slight biases in smoking-related and stage-related features. Nevertheless, these differences enable us to effectively compare and identify variations in the genomic landscape.

2.2. Germline Variants

To evaluate potential genetic susceptibility factors, we identified germline gene variants present in our cohort. We only considered variants that were part of the list of 152 genes previously associated with cancer susceptibility [15] and classified as "Pathogenic" or "Likely pathogenic" according to the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) recommendations [16]. Furthermore, we explored Uncertain Significance variants using the results from pathogenicity prediction tools (see Methods). A total of 20 variants were identified (Table S1); some of them affect Homologous Recombination-related genes (Figure 2A).
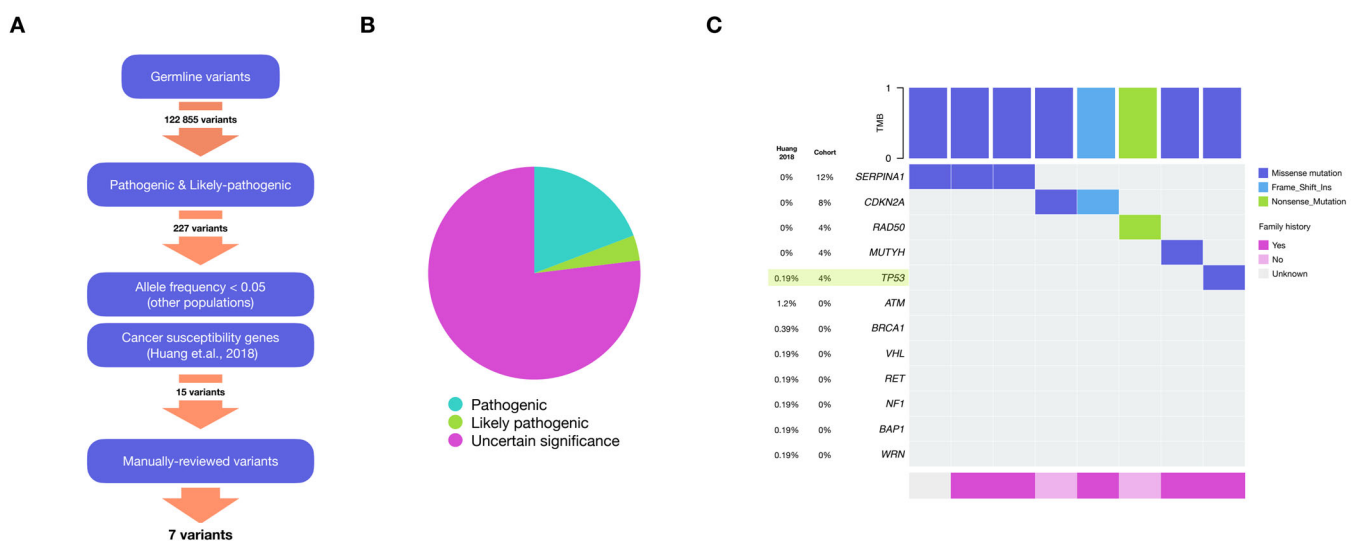

Figure 2. Pathogenic and Likely pathogenic germline variants. (**A**) Workflow of pathogenic and likely pathogenic germline short variants discovery. (**B**) Distribution of Uncertain significance, pathogenic and likely pathogenic germline variants (**C**) Details of genes affected by pathogenic and likely pathogenic germline variants in our cohort and the contrast with results reported by (Huang et al., 2018 [15]). In our cohort, 32% of patients exhibit germline variants, compared to only 6.4% in the reference cohort. In both cohorts, variants in TP53 were present (*p*-value = 0.09004), whereas SERPINA1 is the most affected gene in our cohort and was not reported to be affected in the reference cohort (*p*-value = 8.66×10^{-5}).

We use the study by Huang et al., 2018 [15] as a reference cohort because they used the data from patients of the PanCancer cohort with a method similar to ours for the identification of germline variants. We found that 32% of patients had pathogenic or likely pathogenic germline variants, compared to 6.4% identified in the reference cohort (*p* < 0.00002). Among the patients with any germline variant, 62.5% (5 out of 8) reported a family history of cancer. Both cohorts identified variants in TP53 (*p* = 0.09004). Additionally, we observed variants in SERPINA1, CDKN2A, RAD50, and MUTYH, which were not identified in Huang's study, even though these genes were interrogated [15] (Figure 2B).

Whereas in our cohort, 12% (3 out of 25) of our patients have variants in SERPINA1 there are no reports of variants in this gene in the reference cohort. This absence in the reference cohort is statistically significant (*p* = 8.66×10^{-5}), suggesting this gene's potential relevance for our Mexican patients. Although this variant has been reported to be more prevalent in Latin American populations, further investigation is required

to assess its potential clinical significance. Overall, these results show a significant bias in the proportion of patients with known pathogenic germline variants in our cohort compared to the reference cohort, as well as variants not previously found in the Huang study (Figure 2B).

2.3. Somatic Mutations

2.3.1. Mutational Signatures

Single base substitutions (SBS) refer to the replacement of a single nucleotide, taking into account not only the mutated base but also the immediate 5′ and 3′ flanking bases. Each signature displays a unique distribution pattern of these combinations. These somatic mutation patterns have been linked to distinct carcinogenic processes—such as tobacco exposure or UV radiation—and have recently emerged as promising biomarkers for cancer and predictors of treatment response [17].

We identified the contribution of a total of 36 single-base substitution (SBS) mutational signatures in our cohort, of which 29 are shared with the 64 signatures present in the reference TCGA-PanCancer cohort (syn1759475) (Figure 3A). To control for the effect of tumor mutational burden (TMB) differences, we used the mean proportions for each signature per individual. Signatures SBS40c (Unknown, suggested aging), SBS41 (Unknown, suggested colibactin), SBS59 (Artifact), and SBS85 (Cytidine deaminase) were only observed in our cohort, being absent in the reference cohort. When comparing the contributions of signatures shared between the two cohorts, the largest differences were found in SBS32, SBS5, SBS85, SBS12, and SBS95 (Bonferroni adjusted p-value < 0.0001).
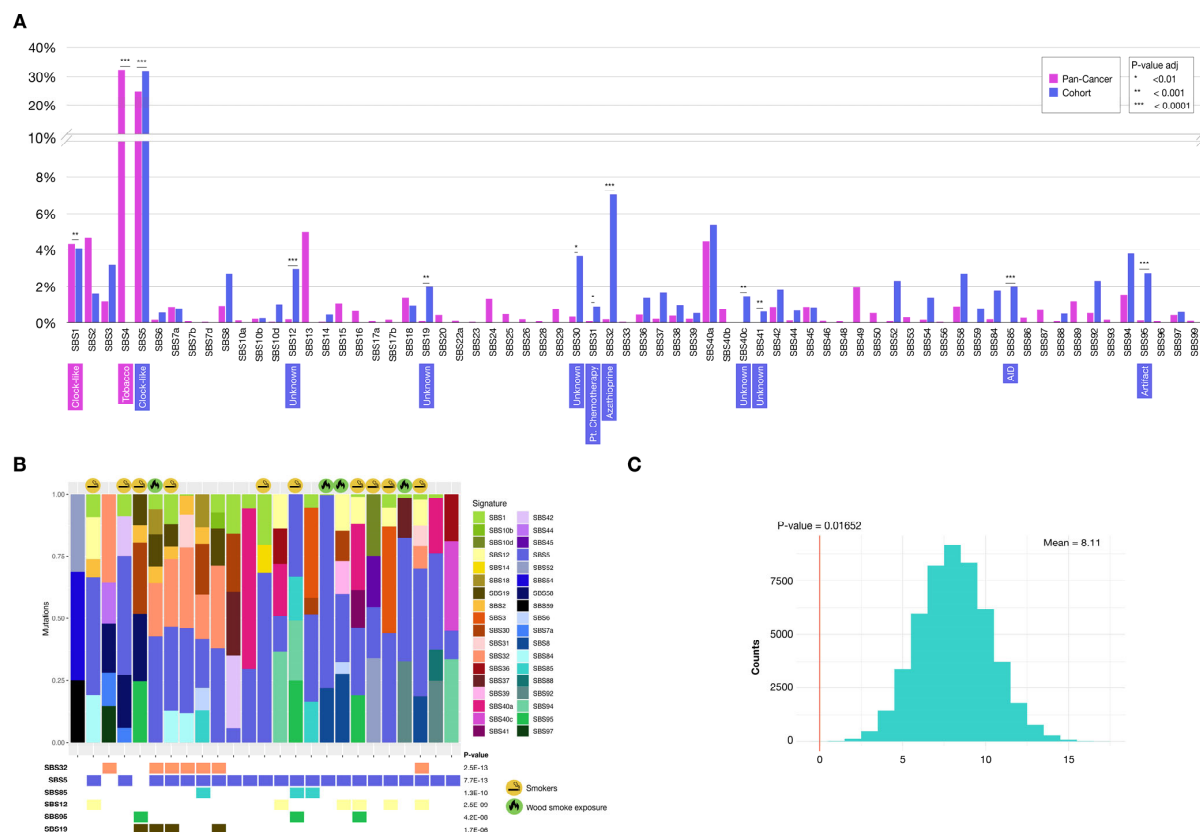

Figure 3. Mutational signature patterns. (**A**) Comparison of mutational signature profiles between our cohort and the TCGA-PanCancer cohort (syn1759475). (**B**) Proportion of mutational signatures per patient. (**C**) The SBS4 signature, associated with tobacco, does not appear in the cohort, even in patients with a history of smoking, and this absence was statistically significant (inset red vertical line).

The SBS5 signature is a clock-like signature, associated with aging, and is widely represented in all types of adult cancers. In our cohort the representation of SBS5 is higher compared to the reference cohort, even though the median age in our sample is lower (60 vs. 67 years, *p*-value = 0.007) [18]. Signature SBS32 is associated with immunosuppressive treatment with Azathioprine, and SBS85 is related to indirect effects of induced activation of cytidine deaminase. The etiology of the SBS12 signature is unknown, and SBS95 is associated with sequencing artifacts (Figure 3B). Most notably, the SBS4 signature, present in 67.6% of smokers in the reference cohort, was absent in our cohort, even in patients with a history of smoking (*p*-value = 0.01652) (Figure 3C). Taken together, these results show statistically significant differences in the somatic mutational signature profiles of our cohort when compared to a reference cohort originally enriched in European ancestry.

2.3.2. Genes with Potential Positive Selection (Drivers)

A total of 22,927 single nucleotide variants (SNVs) and 2044 insertions and deletions (indels) were identified with a median TMB of 6.35 somatic SNVs per megabase. In order to identify mutated genes with possible positive selection in tumors (driver candidates), we used a combination of three methods (dNdScv, MutSigCV, and Oncodrivefml). These three tools are commonly used to identify genes under positive selection in tumors, each employing a slightly different methodology. Essentially, the higher a gene ranks, the more likely it is to be under positive selection in the tumor, and therefore, the greater its likelihood of being a driver gene.

However, their reliability in small samples may be compromised by potential random variation. To overcome this limitation, we incorporated bootstrapping and adopted a gene population-based approach to enhance statistical robustness. We used these tools to look for significant coincidences among the top 10% most highly ranked genes by all three methods when applied to both. First, we asked if a significant coincidence is observed in the gene ranking generated by all three tools when compared with chance expectations in our cohort, as assessed using a bootstrap analysis consisting of 50,000 equally sized random gene samples (see methods). This analysis showed a highly statistically significant coincidence for the top 10% ranking when comparing all three methods despite the small number of patients.

We next compared the 10% most highly ranked genes jointly detected by at least two of all three tools with those of the TCGA-PanCancer reference cohort (syn1759475), and found an overlap of 49 genes out of 631, a number of genes significantly higher than expected by chance (*p*-value $< 2 \times 10^{-5}$, as numerically assessed using 50,000 random samples of equal size drawn from each cohort and looking at the intersection) (Figure 4A).

We next assessed potential functional associations among the ranking 631 genes in our cohort by assessing the frequency of protein–protein interactions among the associated gene products. This analysis was conducted to ensure that the ranking does not prioritize random genes. It showed an enrichment of interactions among these genes above chance expectations (*p* = 0.0379). These results demonstrate that the top most highly ranked genes are statistically enriched in genes that are also detected in the top ranking genes in the reference cohort, and that the associated protein products of these genes engage in a number of protein–protein interactions higher than expected by chance, suggesting an effective functional association between these genes, possibly linked to tumor progression, in our cohort of Mexican patients (Figure 4B,C).

We wanted to compare the gene ranking of the reference data generated by our workflow and the ranking reported by the authors using their own workflow (syn1661643). To this end, we selected the top 10% of genes ranked by each approach and calculated the intersection between the resulting gene sets. As shown in Figure 5A, 40% of genes detected

are not at the top in the PanCancer reference list. In our cohort, 88% of detected genes also fall outside of this top (Figure 5B). This result suggests the existence of a large number of potential drivers not previously identified in populations with European ancestry.
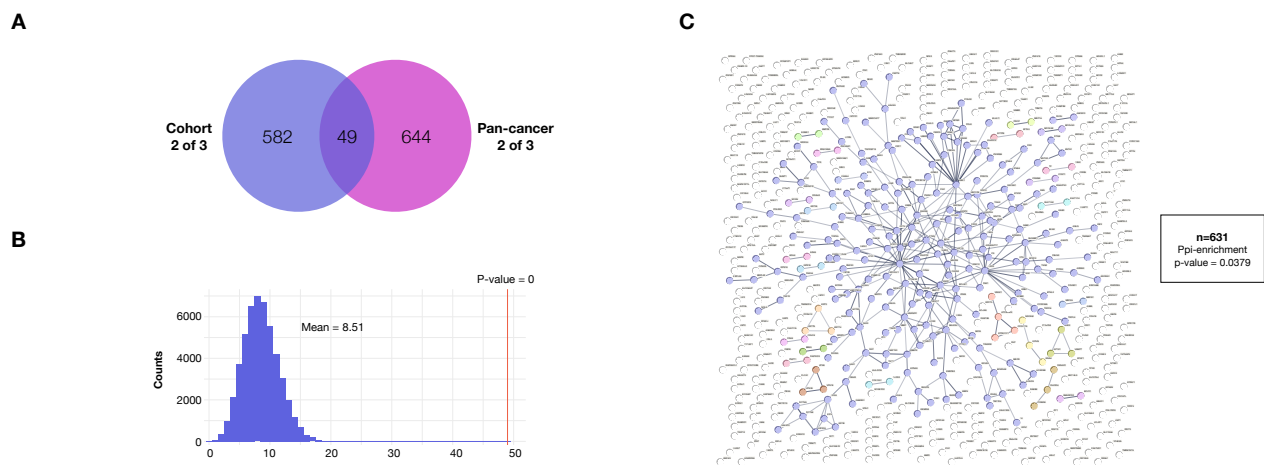

Figure 4. Enrichment of protein–protein interactions among genes most highly ranked for potential positive selection using three independent driver gene tools. (**A**) Enrichment of significant driver genes identified in the TCGA-PanCancer reference cohort (syn1759475), within the first decile of genes with most potential positive selection in our cohort. (**B**) Histogram showing the numerically assessed (bootstrap) expected intersection between the genes in the top decile of the PanCancer reference cohort and the top ranking (potential drivers) in our cohort. Actual intersections indicated (red line). (**C**) Protein–protein interaction network of the top-ranking genes in the cohort of Mexican patients (*p*-value = 0.0379).
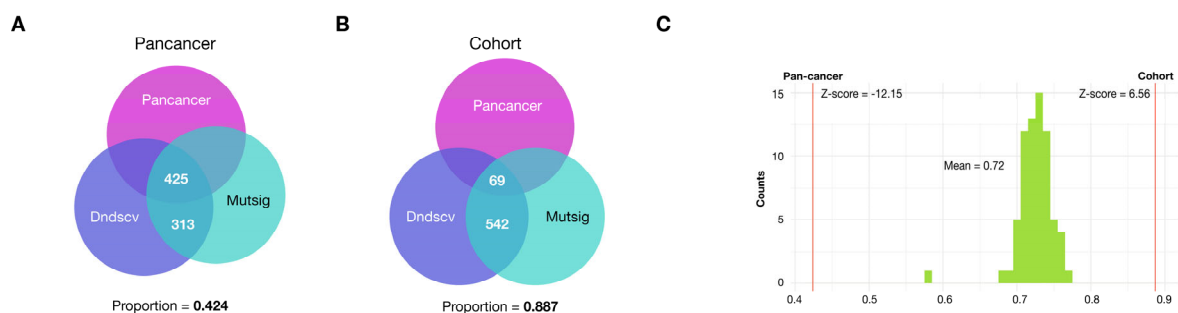

Figure 5. The small sample size does not explain observed bias in potential drivers. (**A**) Genes jointly called by Dndscv and Mutsig in the Pan-cancer cohort and their intersection with the Pan-cancer reference top 10% ranked list (syn1661643). (**B**) Genes jointly called by the two tools in our cohort and their intersection with the Pan-cancer reference top 10% ranked list (syn1661643). (**C**) A sample size effect is demonstrated by taking 25 random samples from the reference cohort and assessing the intersections of the first decile with a reported list of driver genes from the same cohort. The proportions of genes non-shared with the reference list is larger in our cohort, even when the size effect is considered.

To confirm that this result was not an artifact of the small sample size of our cohort, we conducted a bootstrap analysis using random samples of 25 individuals drawn from the whole set of samples from the Pancancer cohort and obtained the expected distribution of intersections between them and the PanCancer reference list. As shown in Figure 5C, there is an effect of the sample size but the proportion of non-PanCancer genes among the top 10% was significantly higher than expected, demonstrating that the bias observed in our cohort is not an artifact of the reduced sample size of our cohort.

From this point, we divided the genes with the highest potential for positive selection or potential drivers (from the intersection of the first deciles) into those shared with the

TCGA-PanCancer reference cohort and those not shared, which appear to be unique to our cohort.

The genes that showed the highest number of somatic mutations that are also present in the reference cohort are NRCAM (24%), EGFR (24%), TET1 (20%). However, the genes that have significantly different mutation frequency in our cohort compared to the reference cohort are SLC36A4 ($p < 0.00002$), AP1S1 ($p < 0.00002$), and TP53 ($p = 0.00005$). The mutation frequencies in EGFR (24% vs. 30%, $p = 0.6572$) and KRAS (16% vs. 10%, $p = 0.3101$), on the other hand, are similar to those reported in previous gene panel studies in the Mexican patients.

Oncoplots shown in Figure 6 correspond to either the top 25 most highly ranked genes using all three driver detection tools shared between our cohort and the reference cohort (Figure 6A), or to the 25 highest-ranked genes that are unique to our cohort (Figure 6B). Functional overrepresentation analyses were performed.
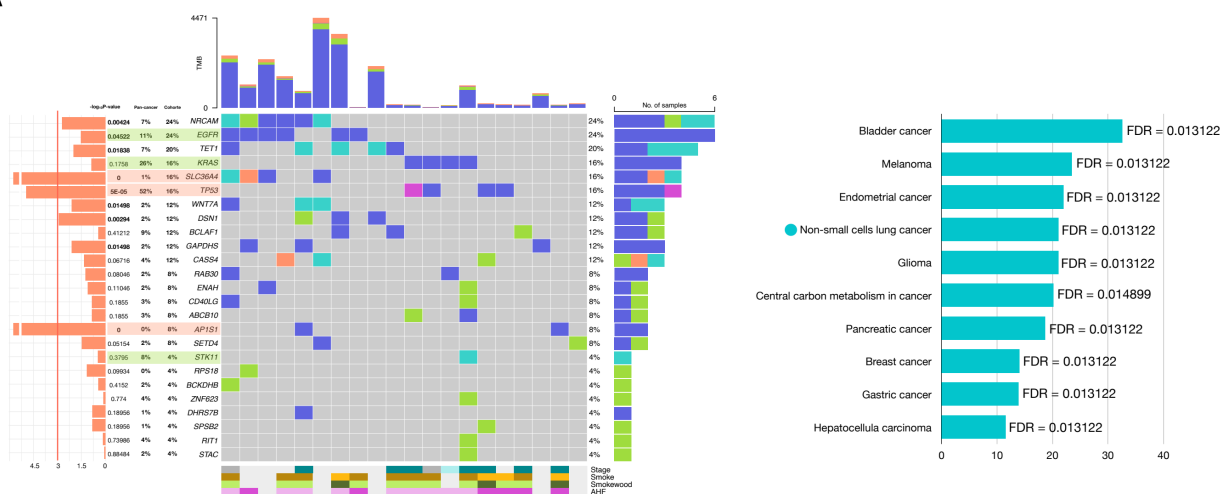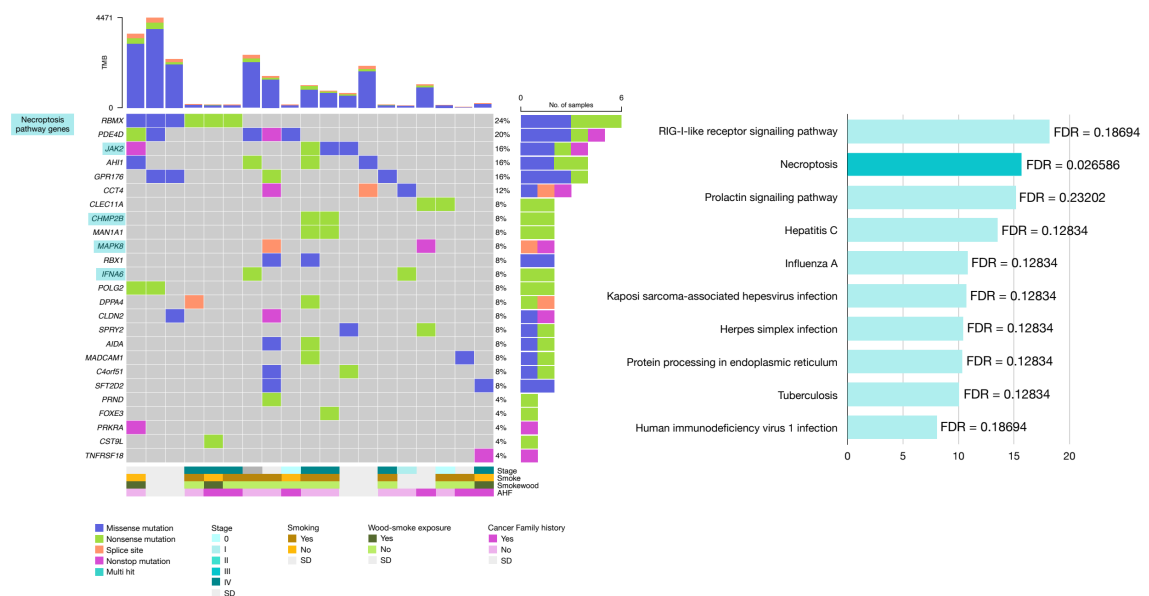

Figure 6. SNVs are among top ranking genes for their highest potential for positive selection. (**A**) Top 25 genes with the highest potential for positive selection shared with the TCGA-PanCancer reference cohort (syn1759475) and its corresponding pathway enrichment result. (**B**) Top 25 genes with the highest potential for positive selection specific to our cohort and its enriched pathways.

Among the top 25 genes shared within the Pan-cancer cohort, a significant enrichment of non-small cells lung cancer was observed, as expected. Although the pathway labels refer to different cancer types, the genes contributing to these enrichments are largely shared and have been described as playing important roles across multiple cancer types. As for the genes with the highest potential positive selection that were not shared with the reference cohort, RBMX (24%), PDE4D (20%), JAK2 (16%) were the most frequently mutated (Figure 6A). Among the top 25 genes unique to our cohort, the Necroptosis pathway was found significantly enriched (FDR = 0.0266), suggesting a potential functional role of this pathway. The mutated genes in our cohort contributing to this pathway were JAK2 (16%), CHMP2B (8%), MAPK8 (8%), and IFNA6 (8%) (Figure 6B). Taken together, our results reveal distinct genomic biases in lung cancer carcinogenesis in this Mexican cohort, when compared to a European ancestry reference population.

3. Discussion

Previous research has indicated that genomic variations among populations may reflect distinct carcinogenic processes in lung adenocarcinomas, which could influence precision medicine strategies when considering different communities.

It has been suggested that there is a genetic factor associated with Hispanic or Native American ancestry that influences the prognosis of these patients [19,20]. In a germline variants analysis conducted in our cohort of Mexican patients, we observed a pathogenic variant in TP53, also affected in the TCGA-PanCancer adenocarcinoma cohort [15]. TP53 is associated with Li-Fraumeni syndrome, which has been described as a significant risk factor for many types of cancer, particularly in young patients. Germline variants in RAD50 have been reported in patients with lung adenocarcinoma, but they have not been linked to an increased risk of LUAD [21]. Loss of function in *CDKN2A* has been associated with resistance to immunotherapy in patients with NSCLC [22]. Similarly, germline variants in *MUTYH* have been linked to an increased risk of lung cancer, particularly in non-smokers [23]. EGFR is the gene most strongly associated with susceptibility to lung cancer, yet we do not find such variants in our cohort [24–27]. On the other hand, SERPINA1, which encodes for alpha-1 antitrypsin, is affected in three of the patients in our cohort. Deficiency of this protein is strongly associated with the development of lung emphysema and chronic obstructive pulmonary disease. As a result, patients carrying pathogenic variants in this gene have an increased risk of lung cancer, especially adenocarcinoma subtype [24,28–31]. It has been proposed that one mechanism associated with increased LUAD risk is that alpha-1 antitrypsin deficiency increases the contact with agents in airways, leading to greater exposure to environmental carcinogens [32]. There is evidence for a link between this variant and tobacco or exposure to wood smoke in Mexicans with respiratory diseases, although this has yet to be studied in LUAD patients [33]. However, there are reports indicating that variants in this gene lead to a poorer response to chemotherapy [32]. Although this variant has been reported to be more prevalent in Latin American populations, further investigation is required to assess its potential clinical significance.

Patients in our cohort exhibit a higher frequency of pathogenic or likely pathogenic germline variants in cancer susceptibility genes (32% vs. 6.4%), although pathogenic variants have been reported in up to 18.5% of patients of European ancestry [34]. This could be due to the lack of clinical information to better discriminate each variant. However, this difference could also be influenced by technical factors, different methodologies, differences in variant classification criteria, and the fact that global reference databases are predominantly composed of individuals of European ancestry. As a result, variants that are common in Latin American populations may be erroneously classified. Further studies and

a more comprehensive characterization of germline variants in Latin American populations are needed to explore the relevance and potential involvement of variants in SERPINA1 and other genes in LUAD among Mexican patients.

Smoking is the primary risk factor for lung adenocarcinoma globally. However, in developing countries, additional risk factors such as exposure to wood smoke and biofuels gain significance [35–37]. A total of 40% of the patients in our cohort had a history of smoking, which is lower than the reported proportion of patients of European ancestry (approximately 85%) [38]. Smoking status has also been associated with mutations in KRAS. Of the four patients with KRAS mutations, three have a history of smoking, and the smoking status of the remaining patient is unknown. Exposure to wood smoke is reported in 30–40% of lung adenocarcinoma cases, and this percentage increases in rural areas. In our cohort, we observed it in 22% (4 out of 18) of the patients. This risk factor has been linked to mutations in EGFR, adenocarcinoma histology, non-smokers, and better prognosis. No specific pattern was observed in the analysis of mutational signatures that could be related to wood smoke exposure.

Interestingly, none of the smoking patients from our cohort carried the SBS4 signature which is known for its robust association with this etiology [39,40]. Two independent studies on Asian patients with LUAD have demonstrated that the impact of this signature in smokers is less pronounced compared to individuals of European ancestry [41,42]. In a large Chinese cohort, the correlation with smoking status was not observed [42]. These findings support the hypothesis that carcinogenic processes are influenced by population-specific biases. Ancestry-related genetic variants in genes involved in carcinogen metabolism or DNA repair mechanisms may affect the processing of carcinogens or alter DNA topology, resulting in distinct mutagenic mechanisms and biases in mutational patterns. Another important consideration is the potential influence of germline variants on individual susceptibility to carcinogens, as well as the presence of other factors, such as exposure to uncharacterized environmental agents, that may mask the traditional effects of tobacco.

We did not observe well-known lung cancer signatures, such as APOBEC (SBS13), or they appeared with very low contributions (SBS2) despite being related with non- or light-smokers. A total of 66% of patients with EGFR mutations have a contribution from the SBS2 signature [39,40]. The mutational signatures with the most statistically significant differences in our cohort are SBS32, SBS85, SBS12, SBS19, and SBS95. The SBS32 signature has been associated with immunosuppressive treatment with Azathioprine. Although there are reports of its use in the treatment of interstitial lung disease, we lack information regarding exposure to this compound among our patients [43]. The SBS85 signature is associated with the indirect effect of induced activation of cytidine deaminase (AID), an enzyme homologous to APOBEC implicated in the development and progression of cancer. This signature is associated with mutations in TP53, which we observed in 2 out of 3 patients with this signature [44]. Somatic mutation-related features have been associated with response to immunotherapy, with tumor mutational burden (TMB) being the most common [45–47]. Additionally, hypermutation phenomena such as SBS32 and SBS85 signatures have also been considered. Moreover, some studies suggest that differences in the immune microenvironment may exist between populations [48,49]. Signature SBS32 has been associated with improved overall OS in ROS1-positive patients undergoing immunotherapy, although the underlying mechanism remains unclear [50]. In the case of SBS85, it has been hypothesized that the better prognosis observed in patients treated with immune checkpoint inhibitors (ICIs) may result from somatic hypermutation driven by AID activity which generates neoepitopes that enhance immunogenicity and leads to T-cell exhaustion—factors that could increase sensitivity to ICIs [50,51]. The SBS95 signature was originally described as an artifact; however, recent reports have indicated

an association with the deamination process, similar to the SBS85 signature [52]. The SBS12 and SBS19 signatures are classified as of unknown etiology that show strain biases, suggesting that they are the product of the effect of exogenous agents. Associations with particular chemicals have been suggested [18,53,54]. The mutational signatures SBS19 and SBS5 have been linked to carcinogenic substances from environmental pollution, with SBS5 recently associated not only with aging but also with exposure to PM2.5 particles [55]. Additionally, the SBS5 and SBS12 signatures, along with SBS4, exhibit population-specific biases in Asian patients. These biases in mutational signature analysis highlight potential differences in the mechanisms of LUAD initiation and progression driven by environmental carcinogens. Such findings are particularly important as they can inform targeted health policies and diagnostic strategies for this specific population.

The genes with the highest potential for positive selection that are not shared with the Pan-cancer cohort were found in the top decile; there was an enrichment in protein–protein interactions greater than expected by chance, suggesting functional interactions among these genes, which may be likely involved in the carcinogenic process specific to our cohort. Among the top 25 genes most highly ranked as potential drivers, the necroptosis pathway was significantly enriched. This pathway is not enriched when evaluating the reference cohort or even the gene sets from cohorts with other ancestries. This cellular process is a poorly understood regulated cell death mechanism that, unlike apoptosis, involves an inflammatory response. Some studies in NSCLC patients have suggested that this pathway plays an important role in both tumor progression and antitumor processes through the regulation of the immune system. It can activate the adaptive immune response and trigger inflammation, or alternatively, contribute to the formation of an immunosuppressive microenvironment [56]. Targeted therapies have also been proposed for this pathway as they amplify anti-tumor immunity when combined with other treatments. Several genes within this pathway, including JAK2, IFNA6, and MAPK8, have been identified as drug targets. In addition, it has been suggested that high expression of key genes in this pathway is associated with improved progression-free survival, disease-free survival, and better response to chemotherapy and radiotherapy. However, there are also reports linking overexpression of these genes with a worse prognosis, which could be explained by excessive necroptosis-mediated damage. This pathway could be used to predict prognosis after treatment, and necroptosis-based anti-tumor therapies could be promising therapeutic strategies [57–59]. However, further work is needed to properly ascertain their clinical application.

The genes with the most significant differences in frequencies, when compared to the reference cohort, are SLC36A4, AP1S1, and TP53. The SLC36A4 gene belongs to a family of amino acid transporters that have been poorly studied in cancer; however, it is well known that amino acid transport is essential for the metabolism of cancer cells. Its critical role in migration is independent of its involvement in migration. It has been shown to play a role in cell growth regulation as well as in migration, generation of a immunosuppressive microenvironment and autophagy [60,61]. However, AP1S1 is a key component of cellular trafficking and receptor regulation, controlling the availability of surface molecules like growth factor receptors and extracellular matrix elements, reducing immunogenicity in cancer cells. Also, it has been associated with increased susceptibility to TKIs in patients with EGFR mutations due to EGFR lysosomal degradation [62]. Its fundamental role in copper homeostasis, which is important for carcinogenesis, has also been recognized [63]. On the other hand, TP53 plays a fundamental role in lung adenocarcinoma and ethnic differences have even been reported and proposed as prognostic markers for immunotherapy [64–66]. EGFR and KRAS are clinically relevant genes due to the availability of targeted therapies. In this study, we confirm the frequencies of somatic mutations in these genes in the Mexican population, with 24% for EGFR and 16% for KRAS, compared to 20% and 10%, respectively

in the Carrot-Zhang recent study that included Mexican patients with lung adenocarcinoma [19]. The Hispanic paradox posits that Hispanic patients have better prognosis, at all stages, than European ancestry patients even in the presence of more adverse conditions, such as lower socioeconomic status [67]. A recent study suggests an association of EGFR mutations with Native American ancestry [19]; however, in our study these patients do not have a significantly higher contribution of Mexican indigenous ancestry. This discrepancy may be due to differences in Native American ancestry markers used in the cited study, or because the Carrot-Zhang study was conducted using tumor-only data.

Among the genes with high mutation frequencies in our cohort, we identified several potential biomarkers or targets of clinical interest (Table S2) [68–92]. One notable example is TET1, which encodes an enzyme involved in DNA demethylation and is generally regarded as a tumor suppressor. However, under hypoxic conditions, TET1 may assume an oncogenic role. Mutations in this gene have been proposed as biomarkers for predicting the efficacy of immunotherapy and tyrosine kinase inhibitors (TKIs), such as Erlotinib, a commonly used targeted therapy for lung adenocarcinoma [93,94]. Furthermore, the co-occurrence of mutations in KRAS and TET1 has been associated with poorer survival outcomes, underscoring their clinical significance [95].

We also identified other genes, such as RBMX and PDE4D, which were present in our cohort but absent in the reference cohort and have previously been suggested as biomarkers. RBMX is a multifunctional RNA-binding protein involved in various processes, including RNA splicing, cell cycle progression, DNA repair, and metastasis. Notably, RBMX exhibited the highest number of mutations in our cohort. This gene has been linked to poor survival in bladder cancer, and tobacco-induced mutations in RBMX are thought to predispose individuals to lung cancer [96]. Additionally, both RBMX and PDE4D have been proposed as potential predictors of response to immunotherapy in patients with lung adenocarcinoma [88,97,98]. Moreover, some of these genes have been considered therapeutic targets, either alone or due to their synergistic effects when blocked in combination with existing therapies. In the case of RBMX, it has been proposed as a potential therapeutic target in lung cancer due to its association with telomere instability, and its depletion has been shown to sensitize tumor cells to cytotoxic agents. Similarly, the inhibition of AP1S1 has shown promising results in cells resistant to anti-EGFR TKIs [62]. Furthermore, PDE4D has been suggested as a potential therapeutic target in lung cancer as its silencing affects both proliferation and metastasis in tumor cells [99,100].

This initial study aims to identify genetic variations relevant to Mexican patients, though determining clinical relevance at the gene level remains challenging. One of the limitations of the sample size is that we could miss rare mutations in the somatic variants analysis. This issue could be further investigated in the next phase of the study, using a larger cohort. Despite this, we observe biases that suggest significant differences across various levels (germline, somatic signatures, and potential drivers) which warrant further exploration. Confirming these findings in a larger cohort and conducting functional assays are essential for a more thorough characterization of these differences. Our results have potential implications for precision medicine strategies tailored to Latin American populations. Addressing the current limitations and continuing this approach will bring us closer to developing personalized medicine that accounts for the unique genetic characteristics of populations like ours.

4. Materials and Methods

4.1. Mexican Lung Adenocarcinoma Cohort

Mexican patients treated at the Instituto Nacional de Enfermedades Respiratorias (INER) in Mexico City were included. The inclusion criteria were as follows: 1—confirmed

lung adenocarcinoma by histopathological report, 2—smokers and non-smokers, 3—any age, 4—any stage, 5—signed informed consent. The project was approved by Instituto Nacional de Medicina Genomica's Ethical and Research committees (C1_12/2010).

4.2. Sample Processing and Sequencing

DNA extraction was performed from frozen or paraffin-embedded lung tumor tissue and, for the ones available, peripheral blood or adjacent lung tissue as a paired sample. DNA extraction from these samples was carried out using Qiagen kits: QIAamp DNA Mini Kit QIAmp® (Qiagen, Hilden, Germany, Cat. No: 51304), DNA FFPE Tissue® (Qiagen, Hilden, Germany, Cat. No: 56404), and y QIAamp DNA Blood® (Qiagen, Hilden, Germany, Cat. No: 51104), according to the manufacturer's instructions. Library preparation and whole-exome sequencing were conducted through three external services: BGI Americas (Cambridge, MA, USA), INMEGEN, and Broad Institute, using Illumina platforms (San Diego, CA, USA), with paired-end sequencing.

4.3. Preprocessing

The quality of the raw reads was assessed using FastQC (v.0.11.9) and aligned to the hg19 reference genome (build GRCh37) obtained from the Broad Institute bioinformatics resource base using BWA (0.7.12-r1039) [101]. Preprocessing of the raw files was performed according to GATK best practices [102].

4.4. Ancestry Analysis

The ancestry inference was performed using ADMIXTURE (v.1.3.0) [103], which included 145,815 SNP's and assumed three clusters (K = 3) for the calculation of the proportion of African, European and Indigenous Mexican. The references were derived from African (Yoruba), northern European (Utah), and indigenous Mexican from individuals included in the García-Ortiz's study [12].

4.5. Germline Variants Analysis

The HaplotypeCaller (GATK 4.0.5.1) algorithm was run in joint calling mode according to best practices. We used GenotypeGVCFs, VariantRecalibrator, and ApplyVQSR using reference resources from: HapMap (3.3), 1000 Genomes (phase 1 and omni 2.5), and (dbsnp138). We filtered snps using: "QD < 2.0", "FS > 60.0", "MQ < 30.0", "MQRankSum and "ReadPosRankSum < −8.0". We focused on variants that: 1. were reported as "Pathogenic" or "Likely pathogenic"; 2. met ACMG and AMP recommendations [16]; 3. had an allele frequency of less than 0.05 in any population according to gnomAD and/or 1000 Genomes Phase 3; 4. belonged to the set of 152 cancer susceptibility genes [15], and 5. passed manual review using IGV (v.2.14.1) [104]. The variants of "Uncertain Significance" were selected after passing additionally two of three criteria: CADD (v.1.6) phred score higher than 20, classified as "Probably damaging" or "Possibly damaging" by Polyphen (v.2.2.2) or "Deleterious" by SIFT (v.5.2.2). Predictions were run through the Variant Effect Predictor (v110) [105].

4.6. Somatic Variants

Somatic variant identification was performed using Mutect2 (GATK 4.0.5.1), employing a combined panel of normal samples composed of non-tumor tissue samples from this cohort, a set of normal samples from the Mexican population used in a previous study by our group [106], and the panel of normal samples provided by the Broad Institute bioinformatics resource base. Variant annotation was performed using VEP (v110), followed by a manual review with IGV (v.2.14.1). The TMB was calculated on the basis

of non-synonymous mutations (splice site, missense, nonsense, and nonstop mutations) divided by 37.5 Mb of the exome library used.

4.6.1. Mutational Signatures Analysis

The extraction of mutational signatures was performed using SigProfilerMatrixGenerator (v.1.2.17) [107] and SigProfilerAssignment (v.0.1.0) [108] on the entire set of somatic variants after quality filtering. Subsequently, for cohort comparisons, we aim to mitigate biases related to TMB by calculating the mean individual contribution (as a proportion) to determine the contribution of each signature to the cohort.

4.6.2. Positive Selection Analysis

To identify genes under positive selection in our cohort, we used dNdScv (v.0.0.1.0) [109], MutsigCV (v.1.3.5) [110], and Oncodrivefml (v.2.2.0) [111]. Due to the lack of statistical significance resulting from sample size, we designed a strategy with numerical tests (using bootstrapping) to find out if the rankings are randomly ordered by taking deciles of results from each tool and using the ranking of significant genes (syn1661643) as a control. Subsequently, we continued the analysis by considering the top decile and searching for intersections between tools and cohorts: ours and TCGA-PanCancer (syn1759475) run under the same parameters. We keep the genes that were in the top decile in two out of the three tools used. Overrepresentation analyses were performed using WebGestalt (v.151) [112]. Protein–protein interaction analyses were performed using STRING (v.12) [113] focusing on physical protein–protein interactions.

4.7. Statistical Analyses

Statistical analyses were conducted using R software (v.4.2.1). All comparisons and p-value calculations were performed using numerical tests, Wilcoxon–Mann–Whitney or Fisher tests with significance thresholds of $p < 0.05$ or lower, as specified in each case. Adjusted p-values in the analysis of mutational signatures were corrected for multiple testing using the Bonferroni method.

More detailed information can be found in the GitHub repository: https://github.com/bgrueda/LUAD_MX_WES (accessed on 14 May 2025).

References

1. Arrieta, O.; Zatarain-Barrón, Z.L.; Aldaco, F.; Barrón, F.; Báez-Saldaña, R.; Campos-Gómez, S.; Trejo, R.; De la Garza, J. Lung Cancer in Mexico. *J. Thorac. Oncol.* **2019**, *14*, 1695–1700. [CrossRef] [PubMed]

2. Gridelli, C.; Rossi, A.; Carbone, D.P.; Guarize, J.; Karachaliou, N.; Mok, T.; Petrella, F.; Spaggiari, L.; Rosell, R. Non-small-cell lung cancer. *Nat. Rev. Dis. Prim.* **2015**, *1*, 15009. [CrossRef] [PubMed]

3. Chen, J.; Yang, H.; Teo, A.S.M.; Amer, L.B.; Sherbaf, F.G.; Tan, C.Q.; Alvarez, J.J.S.; Lu, B.; Lim, J.Q.; Takano, A.; et al. Genomic landscape of lung adenocarcinoma in East Asians. *Nat. Genet.* **2020**, *52*, 177–186. [CrossRef]

4. Midha, A.; Dearden, S.; McCormack, R. EGFR mutation incidence in non-small-cell lung cancer of adenocarcinoma histology: A systematic review and global map by ethnicity (mutMapII). *Am. J. Cancer Res.* **2015**, *5*, 2892–2911.

5. Arrieta, O.; Cardona, A.F.; Martín, C.; Más-López, L.; Corrales-Rodríguez, L.; Bramuglia, G.; Castillo-Fernandez, O.; Meyerson, M.; Amieva-Rivera, E.; Campos-Parra, A.D.; 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.* **2015**, *10*, 838–843. [CrossRef]

6. Zhou, W.; Christiani, D.C. East meets West: Ethnic differences in epidemiology and clinical behaviors of lung cancer between East Asians and Caucasians. *Chin. J. Cancer* **2011**, *30*, 287–292. [CrossRef] [PubMed]

7. De Ver Dye, T.; Tavarez, Z.Q.; Pérez Ramos, J.G.; Fernandez, I.D.; Vega, C.V.; Vega Ocasio, D.M.; Avendaño, E.; Cardona Cordero, N.R.; Hering, C.D.; Dozier, A.M.; et al. Participation in genetic research among Latinx populations by Latin America birth-residency concordance: A global study. *J. Community Genet.* **2021**, *12*, 603–615. [CrossRef] [PubMed]

8. Molina-Aguilar, C.; Robles-Espinoza, C.D. Tackling the lack of diversity in cancer research. *Dis. Models Mech.* **2023**, *16*, dmm050275. [CrossRef]

9. Park, S.L.; Cheng, I.; Haiman, C.A. Genome-Wide Association Studies of Cancer in Diverse Populations. *Cancer Epidemiol. Biomark. Prev.* **2018**, *27*, 405–417. [CrossRef]

10. Arrieta, O.; Cardona, A.F.; Bramuglia, G.; Cruz-Rico, G.; Corrales, L.; Martín, C.; Imaz-Olguín, V.; Castillo, O.; Cuello, M.; Rojas-Bilbao, É.; et al. Molecular Epidemiology of ALK Rearrangements in Advanced Lung Adenocarcinoma in Latin America. *Oncology* **2019**, *96*, 207–216. [CrossRef]

11. Hernández-Pedro, N.; Soca-Chafre, G.; Alaez-Versón, C.; Carrillo-Sánchez, K.; Avilés-Salas, A.; Vergara, E.; Arrieta, O. Mutational profile by targeted next generation sequencing of non-small cell lung cancer in the Mexican population. *Salud Publica Mex.* **2019**, *61*, 308. [CrossRef]

12. García-Ortiz, H.; Barajas-Olmos, F.; Contreras-Cubas, C.; Cid-Soto, M.Á.; Córdova, E.J.; Centeno-Cruz, F.; Mendoza-Caamal, E.; Cicerón-Arellano, I.; Flores-Huacuja, M.; Baca, P.; et al. The genomic landscape of Mexican Indigenous populations brings insights into the peopling of the Americas. *Nat. Commun.* **2021**, *12*, 5942. [CrossRef] [PubMed]

13. Ordoñez, G.; Romero, S.; Orozco, L.; Pineda, B.; Jiménez-Morales, S.; Nieto, A.; García-Ortiz, H.; Sotelo, J. Genomewide admixture study in Mexican Mestizos with multiple sclerosis. *Clin. Neurol. Neurosurg.* **2015**, *130*, 55–60. [CrossRef]

14. Carrot-Zhang, J.; Chambwe, N.; Damrauer, J.S.; Knijnenburg, T.A.; Robertson, A.G.; Yau, C.; Zhou, W.; Berger, A.C.; Huang, K.L.; Newberg, J.Y.; et al. Comprehensive Analysis of Genetic Ancestry and Its Molecular Correlates in Cancer. *Cancer Cell* **2020**, *37*, 639–654.e6. [CrossRef] [PubMed]

15. Huang, K.; Mashl, R.J.; Wu, Y.; Ritter, D.I.; Wang, J.; Oh, C.; Paczkowska, M.; Reynolds, S.; Wyczalkowski, M.A.; Oak, N.; et al. Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell* **2018**, *173*, 355–370.e14. [CrossRef]

16. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W.W.; Hegde, M.; Lyon, E.; Spector, E.; et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **2015**, *17*, 405–424. [CrossRef]

17. Dong, F.; Davies, K.D. Mutational Signatures in Cancer. *J. Mol. Diagn.* **2023**, *25*, 790–795. [CrossRef] [PubMed]

18. Alexandrov, L.B.; Kim, J.; Haradhvala, N.J.; Huang, M.N.; Tian Ng, A.W.; Wu, Y.; Boot, A.; Covington, K.R.; Gordenin, D.A.; Bergstrom, E.N.; et al. The repertoire of mutational signatures in human cancer. *Nature* **2020**, *578*, 94–101. [CrossRef]

19. Carrot-Zhang, J.; Soca-Chafre, G.; Patterson, N.; Thorner, A.R.; Nag, A.; Watson, J.; Genovese, G.; Rodriguez, J.; Gelbard, M.K.; Corrales-Rodriguez, L.; et al. Genetic Ancestry Contributes to Somatic Mutations in Lung Cancers from Admixed Latin American Populations. *Cancer Discov.* **2021**, *11*, 591–598. [CrossRef]

20. Patel, M.I.; Schupp, C.W.; Gomez, S.L.; Chang, E.T.; Wakelee, H.A. How Do Social Factors Explain Outcomes in Non–Small-Cell Lung Cancer Among Hispanics in California? Explaining the Hispanic Paradox. *J. Clin. Oncol.* **2013**, *31*, 3572–3578. [CrossRef]

21. Donner, I.; Katainen, R.; Sipilä, L.J.; Aavikko, M.; Pukkala, E.; Aaltonen, L.A. Germline mutations in young non-smoking women with lung adenocarcinoma. *Lung Cancer* **2018**, *122*, 76–82. [CrossRef] [PubMed]

22. Gutiontov, S.I.; Turchan, W.T.; Spurr, L.F.; Rouhani, S.J.; Chervin, C.S.; Steinhardt, G.; Lager, A.M.; Wanjari, P.; Malik, R.; Connell, P.P.; et al. CDKN2A loss-of-function predicts immunotherapy resistance in non-small cell lung cancer. *Sci. Rep.* **2021**, *11*, 20059. [CrossRef]

23. Trendowski, M.R.; Lusk, C.M.; Wenzlaff, A.S.; Neslund-Dudas, C.; Purrington, K.S.; Beebe-Dimmer, J.L.; Schwartz, A.G. Association of Germline Pathogenic Variants in *MUTYH* and Other DNA Damage Response Genes with Lung Cancer Risk Among Non-Hispanic Whites and African Americans. *JCO Precis. Oncol.* **2025**, *9*, e2400558. [CrossRef]

24. Benusiglio, P.R.; Fallet, V.; Sanchis-Borja, M.; Coulet, F.; Cadranel, J. Lung cancer is also a hereditary disease. *Eur. Respir. Rev.* **2021**, *30*, 210045. [CrossRef]

25. Liu, M.; Liu, X.; Suo, P.; Gong, Y.; Qu, B.; Peng, X.; Xiao, W.; Li, Y.; Chen, Y.; Zeng, Z.; et al. The contribution of hereditary cancer-related germline mutations to lung cancer susceptibility. *Transl. Lung Cancer Res.* **2020**, *9*, 646–658. [CrossRef]

26. Peng, W.; Li, B.; Li, J.; Chang, L.; Bai, J.; Yi, Y.; Chen, R.; Zhang, Y.; Chen, C.; Pu, X.; et al. Clinical and genomic features of Chinese lung cancer patients with germline mutations. *Nat. Commun.* **2022**, *13*, 1268. [CrossRef]

27. Yang, T.; Liu, H.; Li, M.; Zhang, Y.; Zhang, Y.; Liang, X.; Lou, F.; Cao, S.; Wang, H. Germline mutations among patients with non-small cell lung cancer. *J. Clin. Oncol.* **2021**, *39* (Suppl. S15), e20527. [CrossRef]

28. Strnad, P.; McElvaney, N.G.; Lomas, D.A. Alpha1-Antitrypsin Deficiency. *N. Engl. J. Med.* **2020**, *382*, 1443–1455. [CrossRef] [PubMed]

29. Torres-Durán, M.; Ruano-Ravina, A.; Parente-Lamelas, I.; Abal-Arca, J.; Leiro-Fernández, V.; Montero-Martínez, C.; Pena, C.; Castro-Añón, O.; Golpe-Gómez, A.; González-Barcala, F.J.; et al. Alpha-1 Antitrypsin Deficiency and Lung Cancer Risk. *J. Thorac. Oncol.* **2015**, *10*, 1279–1284. [CrossRef]

30. Yang, P. Alpha1-Antitrypsin Deficiency Carriers, Tobacco Smoke, Chronic Obstructive Pulmonary Disease, and Lung Cancer Risk. *Arch. Intern. Med.* **2008**, *168*, 1097. [CrossRef]

31. Ercetin, E.; Richtmann, S.; Delgado, B.M.; Gomez-Mariano, G.; Wrenger, S.; Korenbaum, E.; Liu, B.; DeLuca, D.; Kühnel, M.P.; Jonigk, D.; et al. Clinical Significance of *SERPINA1* Gene and Its Encoded Alpha1-antitrypsin Protein in NSCLC. *Cancers* **2019**, *11*, 1306. [CrossRef] [PubMed]

32. Tubío-Pérez, R.A.; Torres-Durán, M.; Fernández-Villar, A.; Ruano-Raviña, A. Alpha-1 antitrypsin deficiency and risk of lung cancer: A systematic review. *Transl. Oncol.* **2021**, *14*, 100914. [CrossRef]

33. Pérez-Rubio, G.; Ambrocio-Ortiz, E.; López-Flores, L.A.; Juárez-Martín, A.I.; Jiménez-Valverde, L.O.; Zoreque-Cabrera, S.; Galicia-Negrete, G.; Ramírez-Díaz, M.E.; Cruz-Vicente, F.; Castillejos-López, M.d.J.; et al. Heterozygous Genotype rs17580 AT (PiS) in *SERPINA1* is Associated with COPD Secondary to Biomass-Burning and Tobacco Smoking: A Case–Control and Populational Study. *Int. J. Chron. Obs. Pulmon. Dis.* **2020**, *15*, 1181–1190. [CrossRef]

34. Sorscher, S.; LoPiccolo, J.; Heald, B.; Chen, E.; Bristow, S.L.; Michalski, S.T.; Nielsen, S.M.; Lacoste, A.; Keyder, E.; Lee, H.; et al. Rate of Pathogenic Germline Variants in Patients with Lung Cancer. *JCO Precis. Oncol.* **2023**, *7*, e2300190. [CrossRef]

35. Arrieta, O.; Campos-Parra, A.D.; Zuloaga, C.; Avilés, A.; Sánchez-Reyes, R.; Manríquez, M.E.V.; Covián-Molina, E.; Martínez-Barrera, L.; Meneses, A.; Cardona, A.; et al. Clinical and Pathological Characteristics, Outcome and Mutational Profiles Regarding Non–Small-Cell Lung Cancer Related to Wood-Smoke Exposure. *J. Thorac. Oncol.* **2012**, *7*, 1228–1234. [CrossRef] [PubMed]

36. Corrales, L.; Rosell, R.; Cardona, A.F.; Martín, C.; Zatarain-Barrón, Z.L.; Arrieta, O. Lung cancer in never smokers: The role of different risk factors other than tobacco smoking. *Crit. Rev. Oncol. Hematol.* **2020**, *148*, 102895. [CrossRef]

37. Báez-Saldaña, R.; Canseco-Raymundo, A.; Ixcot-Mejía, B.; Juárez-Verdugo, I.; Escobar-Rojas, A.; Rumbo-Nava, U.; Castillo-González, P.; León-Dueñas, S.; Arrieta, O. Case–control study about magnitude of exposure to wood smoke and risk of developing lung cancer. *Eur. J. Cancer Prev.* **2021**, *30*, 462–468. [CrossRef] [PubMed]

38. Molina, J.R.; Yang, P.; Cassivi, S.D.; Schild, S.E.; Adjei, A.A. Non-Small Cell Lung Cancer: Epidemiology, Risk Factors, Treatment, and Survivorship. *Mayo Clin. Proc.* **2008**, *83*, 584–594. [CrossRef] [PubMed]

39. Ernst, S.M.; Mankor, J.M.; van Riet, J.; von der Thüsen, J.H.; Dubbink, H.J.; Aerts, J.G.J.V.; de Langen, A.J.; Smit, E.F.; Dingemans, A.M.C.; Monkhorst, K. Tobacco Smoking-Related Mutational Signatures in Classifying Smoking-Associated and Nonsmoking-Associated NSCLC. *J. Thorac. Oncol.* **2023**, *18*, 487–498. [CrossRef]

40. van den Heuvel, G.R.M.; Kroeze, L.I.; Ligtenberg, M.J.L.; Grünberg, K.; Jansen, E.A.M.; von Rhein, D.; de Voer, R.M.; van den Heuvel, M.M. Mutational signature analysis in non-small cell lung cancer patients with a high tumor mutational burden. *Respir. Res.* **2021**, *22*, 302. [CrossRef]

41. Wang, J.; Guo, C.; Wang, J.; Zhang, X.; Qi, J.; Huang, X.; Hu, Z.; Wang, H.; Hong, B. Tumor Mutation Signature Reveals the Risk Factors of Lung Adenocarcinoma with *EGFR* or *KRAS* Mutation. *Cancer Control* **2025**, *32*, 10732748241307363. [CrossRef] [PubMed]

42. Chevalier, A.; Guo, T.; Gurevich, N.Q.; Xu, J.; Yajima, M.; Campbell, J.D. Characterization of highly active mutational signatures in tumors from a large Chinese population. *medRxiv* **2023**. [CrossRef]

43. van den Bosch, L.; Luppi, F.; Ferrara, G.; Mura, M. Immunomodulatory treatment of interstitial lung disease. *Ther. Adv. Respir. Dis.* **2022**, *16*, 175346662211170. [CrossRef] [PubMed]

44. Shinmura, K.; Igarashi, H.; Goto, M.; Tao, H.; Yamada, H.; Matsuura, S.; Tajima, M.; Matsuda, T.; Yamane, A.; Funai, K.; et al. Aberrant Expression and Mutation-Inducing Activity of AID in Human Lung Cancer. *Ann. Surg. Oncol.* **2011**, *18*, 2084–2092. [CrossRef]

45. Shao, M.; Xu, Y.; Zhang, J.; Mao, M.; Wang, M. Tumor mutational burden as a predictive biomarker for non-small cell lung cancer treated with immune checkpoint inhibitors of PD-1/PD-L1. *Clin. Transl. Oncol.* **2024**, *26*, 1446–1458. [CrossRef]

46. Greillier, L.; Tomasini, P.; Barlesi, F. The clinical utility of tumor mutational burden in non-small cell lung cancer. *Transl. Lung Cancer Res.* **2018**, *7*, 639–646. [CrossRef]

47. Vryza, P.; Fischer, T.; Mistakidi, E.; Zaravinos, A. Tumor mutation burden in the prognosis and response of lung cancer patients to immune-checkpoint inhibition therapies. *Transl. Oncol.* **2023**, *38*, 101788. [CrossRef]

48. Trendowski, M.R.; Watza, D.; Lusk, C.M.; Lonardo, F.; Ratliff, V.; Wenzlaff, A.S.; Mamdani, H.; Neslund-Dudas, C.; Boerner, J.L.; Schwartz, A.G.; et al. Evaluation of the Immune Response within the Tumor Microenvironment in African American and Non-Hispanic White Patients with Non–Small Cell Lung Cancer. *Cancer Epidemiol. Biomark. Prev.* **2024**, *33*, 1220–1228. [CrossRef]

49. Xu, Y.; Zhang, L.; Thaiparambil, J.; Mai, S.; Perera, D.N.; Zhang, J.; Pan, P.Y.; Coarfa, C.; Ramos, K.; Chen, S.H.; et al. Patients with Lung Cancer of Different Racial Backgrounds Harbor Distinct Immune Cell Profiles. *Cancer Res. Commun.* **2022**, *2*, 884–893. [CrossRef]

50. Li, Y.; Zhao, H.; Huang, J.; Yan, H.; Zhao, B. The association of ROS1 mutation with cancer immunity and its impact on the efficacy of pan-cancer immunotherapy. *J. Transl. Med.* **2024**, *22*, 403. [CrossRef]

51. Hernández-Verdin, I.; Akdemir, K.C.; Ramazzotti, D.; Caravagna, G.; Labreche, K.; Mokhtari, K.; Hoang-Xuan, K.; Peyre, M.; Bielle, F.; Touat, M.; et al. Pan-cancer landscape of AID-related mutations, composite mutations, and their potential role in the ICI response. *npj Precis. Oncol.* **2022**, *6*, 89. [CrossRef]

52. Degasperi, A.; Zou, X.; Dias Amarante, T.; Martinez-Martinez, A.; Koh, G.C.C.; Dias, J.M.L.; Heskin, L.; Chmelova, L.; Rinaldi, G.; Wang, V.Y.W.; et al. Substitution mutational signatures in whole-genome–sequenced cancers in the UK population. *Science* **2022**, *376*, abl9283. [CrossRef] [PubMed]

53. Hu, X.; Biswas, A.; Sharma, A.; Sarkodie, H.; Tran, I.; Pal, I.; De, S. Mutational signatures associated with exposure to carcinogenic microplastic compounds bisphenol A and styrene oxide. *NAR Cancer* **2021**, *3*, zcab004. [CrossRef]

54. Riva, L.; Pandiri, A.R.; Li, Y.R.; Droop, A.; Hewinson, J.; Quail, M.A.; Iyer, V.; Shepherd, R.; Herbert, R.A.; Campbell, P.J.; et al. The mutational signature profile of known and suspected human carcinogens in mice. *Nat. Genet.* **2020**, *52*, 1189–1197. [CrossRef] [PubMed]

55. Díaz-Gay, M.; Zhang, T.; Hoang, P.H.; Khandekar, A.; Zhao, W.; Steele, C.D.; Otlu, B.; Nandi, S.P.; Vangara, R.; Bergstrom, E.N.; et al. The mutagenic forces shaping the genomic landscape of lung cancer in never smokers. *medRxiv* **2024**. [CrossRef]

56. Wang, G.; Liu, X.; Liu, H.; Zhang, X.; Shao, Y.; Jia, X. A novel necroptosis related gene signature and regulatory network for overall survival prediction in lung adenocarcinoma. *Sci. Rep.* **2023**, *13*, 15345. [CrossRef] [PubMed]

57. Gong, Y.; Fan, Z.; Luo, G.; Yang, C.; Huang, Q.; Fan, K.; Cheng, H.; Jin, K.; Ni, Q.; Yu, X.; et al. The role of necroptosis in cancer biology and therapy. *Mol. Cancer* **2019**, *18*, 100. [CrossRef]

58. Lee, J.H.; Park, J.E.; Hong, M.J.; Choi, J.E.; Kang, H.; Do, S.K.; Lee, S.; Jeong, J.Y.; Shin, K.M.; Do, Y.W.; et al. Genetic variants in key necroptosis regulators predict prognosis of non-small cell lung cancer after surgical resection. *Thorac. Cancer* **2023**, *14*, 2678–2686. [CrossRef]

59. Wang, L.; Zhou, L.; Zhou, Y.; Liu, L.; Jiang, W.; Zhang, H.; Liu, H. Necroptosis in Pulmonary Diseases: A New Therapeutic Target. *Front. Pharmacol.* **2021**, *12*, 737129. [CrossRef]

60. Bröer, S. Amino Acid Transporters as Targets for Cancer Therapy: Why, Where, When, and How. *Int. J. Mol. Sci.* **2020**, *21*, 6156. [CrossRef]

61. Zhang, N.; Yang, X.; Yuan, F.; Zhang, L.; Wang, Y.; Wang, L.; Mao, Z.; Luo, J.; Zhang, H.; Zhu, W.G.; et al. Increased Amino Acid Uptake Supports Autophagy-Deficient Cell Survival upon Glutamine Deprivation. *Cell Rep.* **2018**, *23*, 3006–3020. [CrossRef] [PubMed]

62. Jeong, J.; Hwang, Y.E.; Lee, M.; Keum, S.; Song, S.; Kim, J.; Choi, J.; Rhee, S. Downregulation of AP1S1 causes the lysosomal degradation of EGFR in non-small cell lung cancer. *J. Cell. Physiol.* **2023**, *238*, 2335–2347. [CrossRef]

63. Su, F.; Fang, Y.; Yu, J.; Jiang, T.; Lin, S.; Zhang, S.; Lv, L.; Long, T.; Pan, H.; Qi, J.; et al. The Single Nucleotide Polymorphisms of AP1S1 are Associated with Risk of Esophageal Squamous Cell Carcinoma in Chinese Population. *Pharmgenom. Pers. Med.* **2022**, *15*, 235–247. [CrossRef] [PubMed]

64. Shi, H.; Seegobin, K.; Heng, F.; Zhou, K.; Chen, R.; Qin, H.; Manochakian, R.; Zhao, Y.; Lou, Y. Genomic landscape of lung adenocarcinomas in different races. *Front. Oncol.* **2022**, *12*, 946625. [CrossRef] [PubMed]

65. Li, H.; Yang, L.; Wang, Y.; Wang, L.; Chen, G.; Zhang, L.; Wang, D. Integrative analysis of TP53 mutations in lung adenocarcinoma for immunotherapies and prognosis. *BMC Bioinform.* **2023**, *24*, 155. [CrossRef]

66. Lin, X.; Wang, L.; Xie, X.; Qin, Y.; Xie, Z.; Ouyang, M.; Zhou, C. Prognostic Biomarker TP53 Mutations for Immune Checkpoint Blockade Therapy and Its Association with Tumor Microenvironment of Lung Adenocarcinoma. *Front. Mol. Biosci.* **2020**, *7*, 602328. [CrossRef]

67. Kumar, R.; Castillero, F.; Bhandari, S.; Malapati, S.; Kloecker, G. The Hispanic paradox in non-small cell lung cancer. *Hematol. Oncol. Stem Cell Ther.* 2021, *in press*. [CrossRef] [PubMed]

68. Bai, C.; Chen, D. NRCAM acts as a prognostic biomarker and promotes the tumor progression in gastric cancer via EMT pathway. *Tissue Cell* **2022**, *77*, 101859. [CrossRef]

69. Zhou, L.; He, L.; Liu, C.-H.; Qiu, H.; Zheng, L.; Sample, K.M.; Wu, Q.; Li, J.; Xie, K.; Ampuero, J.; et al. Liver cancer stem cell dissemination and metastasis: Uncovering the role of NRCAM in hepatocellular carcinoma. *J. Exp. Clin. Cancer Res.* **2023**, *42*, 311. [CrossRef]

70. Liu, F.; Ma, J.; Wang, K.; Li, Z.; Jiang, Q.; Chen, H.; Li, W.; Xia, J. High expression of PDE4D correlates with poor prognosis and clinical progression in pancreaticductal adenocarcinoma. *J. Cancer* **2019**, *10*, 6252–6260. [CrossRef]

71. Xu, Y.; Jin, J.; Xu, J.; Shao, Y.W.; Fan, Y. JAK2 variations and functions in lung adenocarcinoma. *Tumor Biol.* **2017**, *39*, 101042831771114. [CrossRef]

72. Zhang, Y.; Gu, X.; Zhu, F.; Li, Y.; Huang, Y.; Ju, S. High expression of GPR176 predicts poor prognosis of gastric cancer patients and promotes the proliferation, migration, and invasion of gastric cancer cells. *Sci. Rep.* **2023**, *13*, 9360. [CrossRef] [PubMed]

73. Ni, L.; Chen, S.; Liu, J.; Li, H.; Zhao, H.; Zheng, C.; Zhang, Y.; Huang, H.; Huang, J.; Wang, B.; et al. GPR176 Is a Biomarker for Predicting Prognosis and Immune Infiltration in Stomach Adenocarcinoma. *Mediat. Inflamm.* **2023**, *2023*, 7123568. [CrossRef]

74. Liu, X.; Rothe, K.; Yen, R.; Fruhstorfer, C.; Maetzig, T.; Chen, M.; Forrest, D.L.; Humphries, R.K.; Jiang, X. A novel AHI-1–BCR-ABL–DNM2 complex regulates leukemic properties of primitive CML cells through enhanced cellular endocytosis and ROS-mediated autophagy. *Leukemia* **2017**, *31*, 2376–2387. [CrossRef] [PubMed]

75. Chen, L.; Wang, Y.; Hu, Q.; Liu, Y.; Qi, X.; Tang, Z.; Hu, H.; Lin, N.; Zeng, S.; Yu, L. Unveiling tumor immune evasion mechanisms: Abnormal expression of transporters on immune cells in the tumor microenvironment. *Front. Immunol.* **2023**, *14*, 1225948. [CrossRef] [PubMed]

76. Fan, S.-J.; Snell, C.; Turley, H.; Li, J.-L.; McCormick, R.; Perera, S.M.W.; Heublein, S.; Kazi, S.; Azad, A.; Wilson, C.; et al. PAT4 levels control amino-acid sensitivity of rapamycin-resistant mTORC1 from the Golgi and affect clinical outcome in colorectal cancer. *Oncogene* **2016**, *35*, 3004–3015. [CrossRef]

77. Goberdhan, D.C.I. Intracellular amino acid sensing and mTORC1-regulated growth: New ways to block an old target? *Curr. Opin. Investig. Drugs* **2010**, *11*, 1360–1367.

78. Yang, Q.; Dang, H.; Liu, J.; Wang, X.; Wang, J.; Lan, X.; Ji, M.; Xing, M.; Hou, P. Hypoxia switches TET1 from being tumor-suppressive to oncogenic. *Oncogene* **2023**, *42*, 1634–1648. [CrossRef]

79. Tang, J.; Peng, W.; Ji, J.; Peng, C.; Wang, T.; Yang, P.; Gu, J.; Feng, Y.; Jin, K.; Wang, X.; et al. GPR176 Promotes Cancer Progression by Interacting with G Protein GNAS to Restrain Cell Mitophagy in Colorectal Cancer. *Adv. Sci.* **2023**, *10*, 2205627. [CrossRef]

80. Lucchetta, M.; da Piedade, I.; Mounir, M.; Vabistsevits, M.; Terkelsen, T.; Papaleo, E. Distinct signatures of lung cancer types: Aberrant mucin O-glycosylation and compromised immune response. *BMC Cancer* **2019**, *19*, 824. [CrossRef]

81. Uhlen, M.; Zhang, C.; Lee, S.; Sjöstedt, E.; Fagerberg, L.; Bidkhori, G.; Benfeitas, R.; Arif, M.; Liu, Z.; Edfors, F.; et al. A pathology atlas of the human cancer transcriptome. *Science* **2017**, *357*, eaan2507. [CrossRef] [PubMed]

82. Ansari, J.; Shackelford, R.E.; El-Osta, H. Epigenetics in non-small cell lung cancer: From basics to therapeutics. *Transl. Lung Cancer Res.* **2016**, *5*, 155–171. [CrossRef] [PubMed]

83. Ghazimoradi, M.H.; Pakravan, K.; Khalafizadeh, A.; Babashah, S. TET1 regulates stem cell properties and cell cycle of Cancer stem cells in triple-negative breast cancer via DNA demethylation. *Biochem. Pharmacol.* **2024**, *219*, 115913. [CrossRef]

84. Filipczak, P.T.; Leng, S.; Tellez, C.S.; Do, K.C.; Grimes, M.J.; Thomas, C.L.; Walton-Filipczak, S.R.; Picchi, M.A.; Belinsky, S.A. p53-Suppressed Oncogene TET1 Prevents Cellular Aging in Lung Cancer. *Cancer Res.* **2019**, *79*, 1758–1768. [CrossRef]

85. Alrehaili, A.A.; Gharib, A.F.; Alghamdi, S.A.; Alhazmi, A.; Al-Shehri, S.S.; Hagag, H.M.; Alsaeedi, F.A.; Alhuthali, H.M.; Raafat, N.; Etewa, R.L.; et al. Evaluation of TET Family Gene Expression and 5-Hydroxymethylcytosine as Potential Epigenetic Markers in Non-small Cell Lung Cancer. *In Vivo* **2023**, *37*, 445–453. [CrossRef]

86. Tuersun, H.; Liu, L.; Zhang, J.; Maimaitizunong, R.; Tang, X.; Li, H. m6A reading protein RBMX as a biomarker for prognosis and tumor progression in esophageal cancer. *Transl. Cancer Res.* **2023**, *12*, 2319–2335. [CrossRef] [PubMed]

87. Liu, J.; Zheng, T.; Chen, D.; Huang, J.; Zhao, Y.; Ma, W.; Liu, H. RBMX involves in telomere stability maintenance by regulating TERRA expression. *PLoS Genet.* **2023**, *19*, e1010937. [CrossRef]

88. Sheng, Y.; Lei, K.; Sun, C.; Liu, J.; Tu, Z.; Zhu, X.; Huang, K. Aberrant RBMX expression is relevant for cancer prognosis and immunotherapy response. *Aging* **2024**, *16*, 226. [CrossRef]

89. Wang, Z.; Xing, Y.; Li, B.; Li, X.; Liu, B.; Wang, Y. Molecular pathways, resistance mechanisms and targeted interventions in non-small-cell lung cancer. *Mol. Biomed.* **2022**, *3*, 42. [CrossRef]

90. Ramalingam, S.S.; Yang, J.C.-H.; Lee, C.K.; Kurata, T.; Kim, D.-W.; John, T.; Nogami, N.; Ohe, Y.; Mann, H.; Rukazenkov, Y.; et al. Osimertinib As First-Line Treatment of *EGFR* Mutation–Positive Advanced Non–Small-Cell Lung Cancer. *J. Clin. Oncol.* **2018**, *36*, 841–849. [CrossRef]

91. Cannon, M.; Stevenson, J.; Stahl, K.; Basu, R.; Coffman, A.; Kiwala, S.; McMichael, J.F.; Kuzma, K.; Morrissey, D.; Cotto, K.; et al. DGIdb 5.0: Rebuilding the drug–gene interaction database for precision medicine and drug discovery platforms. *Nucleic Acids Res.* **2024**, *52*, D1227–D1235. [CrossRef]

92. Zhou, Y.; Zhang, Y.; Zhao, D.; Yu, X.; Shen, X.; Zhou, Y.; Wang, S.; Qiu, Y.; Chen, Y.; Zhu, F. TTD: *Therapeutic Target Database* describing target druggability information. *Nucleic Acids Res.* **2024**, *52*, D1465–D1477. [CrossRef] [PubMed]

93. Forloni, M.; Gupta, R.; Nagarajan, A.; Sun, L.-S.; Dong, Y.; Pirazzoli, V.; Toki, M.; Wurtz, A.; Melnick, M.A.; Kobayashi, S.; et al. Oncogenic EGFR Represses the TET1 DNA Demethylase to Induce Silencing of Tumor Suppressors in Cancer Cells. *Cell Rep.* **2016**, *16*, 457–471. [CrossRef]

94. Wu, H.-X.; Chen, Y.-X.; Wang, Z.-X.; Zhao, Q.; He, M.-M.; Wang, Y.-N.; Wang, F.; Xu, R.H. Alteration in TET1 as potential biomarker for immune checkpoint blockade in multiple cancers. *J. Immunother. Cancer* **2019**, *7*, 264. [CrossRef] [PubMed]

95. Xu, Q.; Wang, C.; Zhou, J.-X.; Xu, Z.-M.; Gao, J.; Sui, P.; Walsh, C.P.; Ji, H.; Xu, G.L. Loss of TET reprograms Wnt signaling through impaired demethylation to promote lung cancer development. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2107599119. [CrossRef]

96. Yan, Q.; Zeng, P.; Zhou, X.; Zhao, X.; Chen, R.; Qiao, J.; Feng, L.; Zhu, Z.; Zhang, G.; Chen, C. RBMX suppresses tumorigenicity and progression of bladder cancer by interacting with the hnRNP A1 protein to regulate PKM alternative splicing. *Oncogene* **2021**, *40*, 2635–2650. [CrossRef] [PubMed]

97. Renieri, A.; Mencarelli, M.A.; Cetta, F.; Baldassarri, M.; Mari, F.; Furini, S.; Piu, P.; Ariani, F.; Dragani, T.A.; Frullanti, E. Oligogenic germline mutations identified in early non-smokers lung adenocarcinoma patients. *Lung Cancer* **2014**, *85*, 168–174. [CrossRef]

98. Feng, B.; Pan, B.; Huang, J.; Du, Y.; Wang, X.; Wu, J.; Ma, R.; Shen, B.; Huang, G.; Feng, J. PDE4D/cAMP/IL-23 axis determines the immunotherapy efficacy of lung adenocarcinoma via activating the IL-9 autocrine loop of cytotoxic T lymphocytes. *Cancer Lett.* **2023**, *565*, 216224. [CrossRef]

99. Pullamsetti, S.S.; Banat, G.A.; Schmall, A.; Szibor, M.; Pomagruk, D.; Hänze, J.; Kolosionek, E.; Wilhelm, J.; Braun, T.; Grimminger, F.; et al. Phosphodiesterase-4 promotes proliferation and angiogenesis of lung cancer by crosstalk with HIF. *Oncogene* **2013**, *32*, 1121–1134. [CrossRef]

100. Hsien Lai, S.; Zervoudakis, G.; Chou, J.; Gurney, M.E.; Quesnelle, K.M. PDE4 subtypes in cancer. *Oncogene* **2020**, *39*, 3791–3802. [CrossRef]

101. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef]

102. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef] [PubMed]

103. Alexander, D.H.; Novembre, J.; Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **2009**, *19*, 1655–1664. [CrossRef] [PubMed]

104. Robinson, J.T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative genomics viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. [CrossRef] [PubMed]

105. McLaren, W.; Gil, L.; Hunt, S.E.; Riat, H.S.; Ritchie, G.R.S.; Thormann, A.; Flicek, P.; Cunningham, F. The Ensembl Variant Effect Predictor. *Genome Biol.* **2016**, *17*, 122. [CrossRef]

106. Sanchez, P.; Espinosa, M.; Maldonado, V.; Barquera, R.; Belem-Gabiño, N.; Torres, J.; Cravioto, A.; Melendez-Zajgla, J. Pancreatic ductal adenocarcinomas from Mexican patients present a distinct genomic mutational pattern. *Mol. Biol. Rep.* **2020**, *47*, 5175–5184. [CrossRef]

107. Bergstrom, E.N.; Huang, M.N.; Mahto, U.; Barnes, M.; Stratton, M.R.; Rozen, S.G.; Alexandrov, L.B. SigProfilerMatrixGenerator: A tool for visualizing and exploring patterns of small mutational events. *BMC Genomics* **2019**, *20*, 685. [CrossRef]

108. Díaz-Gay, M.; Vangara, R.; Barnes, M.; Wang, X.; Islam, S.M.A.; Vermes, I.; Duke, S.; Narasimman, N.B.; Yang, T.; Jiang, Z.; et al. Assigning mutational signatures to individual samples and individual somatic mutations with SigProfilerAssignment. *Bioinformatics* **2023**, *39*, btad756. [CrossRef]

109. Martincorena, I.; Raine, K.M.; Gerstung, M.; Dawson, K.J.; Haase, K.; Van Loo, P.; Davies, H.; Stratton, M.R.; Campbell, P.J. Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **2017**, *171*, 1029–1041.e21. [CrossRef]

110. Lawrence, M.S.; Stojanov, P.; Polak, P.; Kryukov, G.V.; Cibulskis, K.; Sivachenko, A.; Carter, S.L.; Stewart, C.; Mermel, C.H.; Roberts, S.A.; et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **2013**, *499*, 214–218. [CrossRef]

111. Mularoni, L.; Sabarinathan, R.; Deu-Pons, J.; Gonzalez-Perez, A.; López-Bigas, N. OncodriveFML: A general framework to identify coding and non-coding regions with cancer driver mutations. *Genome Biol.* **2016**, *17*, 128. [CrossRef] [PubMed]

112. Wang, J.; Vasaikar, S.; Shi, Z.; Greer, M.; Zhang, B. WebGestalt 2017: A more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* **2017**, *45*, W130–W137. [CrossRef] [PubMed]

113. Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; Nastou, K.; Mehryary, F.; Hachilif, R.; Gable, A.L.; Fang, T.; Doncheva, N.T.; Pyysalo, S.; et al. The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **2023**, *51*, D638–D646. [CrossRef] [PubMed]