Article

Mutational Landscape of Bladder Cancer in Mexican Patients: *KMT2D* Mutations and Chr11q15.5 Amplifications are Associated with Muscle Invasion

María D. Pérez-Montiel[1], Dennis Cerrato-Izaguirre[2,3], Yesennia Sánchez-Pérez[3], Jose Diaz-Chavez[3], Carlo Cortés-González[3], Jairo A. Rubio[3], Miguel A. Jiménez-Ríos[4], Luis A. Herrera[5], Anna Scavuzzo[4], Abelardo Meneses-García[7], Ricardo Hernández-Martínez[3], Felipe Vaca-Paniagua[8], Andrea Ramírez[9], Alicia Orozco[9], David Cantú-de-León[6], Diddier Prada[3,10]*

[1] Departamento de Patología, Instituto Nacional de Cancerología (INCan), Ciudad de México, México
[2] Departamento de Farmacología, Centro de Investigación y de Estudios Avanzados del I.P.N. (CINVESTAV), Avenida Instituto Politécnico Nacional No. 2508, Ciudad de México CP. 07360, México
[3] Subdirección de Investigación Básica, Instituto Nacional de Cancerología (INCan), San Fernando No. 22, Tlalpan, Ciudad de México CP. 14080, México
[4] Departamento de Urología, Instituto Nacional de Cancerología (INCan), Ciudad de México, México
[5] Dirección General, Instituto Nacional de Medicina Genómica, Ciudad de México, México
[6] Dirección de Investigación, Instituto Nacional de Cancerología (INCan), Ciudad de México, México
[7] Dirección GENERAL, Instituto Nacional de Cancerología (INCan), Ciudad de México, México
[8] Laboratorio Nacional en Salud, Diagnóstico Molecular y Efecto Ambiental en Enfermedades Crónico-Degenerativas, Facultad de Estudios Superiores Iztacala, Tlalnepantla 54090, México.
[9] Unidad de Apoyo Molecular a la Investigación Clínica, Instituto Nacional de Cancerología (INCan), San Fernando No. 22, Tlalpan, Ciudad de México CP. 14080, México
[10] Department of Environmental Health Sciences, Mailman School of Public Health, New York, NY, USA
* Correspondence: dpradao@incan.edu.mx; Phone.: +52 553693 5200 Ext 241.

Abstract:

Bladder cancer (BC) is the most common neoplasm of the urinary tract, which originates in the epithelium that covers the inner surface of the bladder. The molecular BC profile has led to the development of different classifications of non-muscle invasive bladder cancer (NMIBC) and muscle-invasive (MIBC). However, the genomic BC landscape profile of the Mexican population, including NMIBC and MIBC, is unknown. In this study, we aimed to identify somatic single nucleotide variants (SNVs) and copy number variations (CNVs) in Mexican patients with BC and their associations with clinical and pathological characteristics. We retrospectively evaluated 37 patients treated between 2012 and 2021 at the National Cancer Institute - Mexico (INCan). DNA samples were obtained from paraffin-embedded tumor tissues and exome sequenced. Strelka2 and Lancet packages were used to identify SNVs and insertions or deletions. FACETS was used to determine CNVs. We found a high frequency of mutations in *TP53* and *KMT2D*, gains in 11q15.5 and 19p13.11-q12, and losses in 7q11.23. *STAG2* mutations and 1q11.23 deletions were also associated with NMIBC and low histologic grade.

Keywords: Bladder cancer; Hispanics; Mexicans population; Mutations; Cancer genomics; non-muscle invasive bladder cancer; muscle-invasive bladder cancer

1. Introduction

Bladder cancer (BC) is the most common neoplasm of the urinary tract. It generally originates from the epithelium that covers the inner surface of the bladder [1]. BC is the tenth most common cancer type, with an estimated 570,000 new cases worldwide in 2020 [2]. Smoking is the leading risk factor associated with BC incidence [3]. Within the total incidence rate, men tend to be diagnosed with bladder cancer 3 to 4 times more frequently than women. In addition, Non-Hispanic Whites patients are the most frequently affected, followed by Non-Hispanic Blacks and Hispanics [4].

BC arises in two different pathways depending on the origin site. Non-muscular invading bladder cancer (NMIBC) or superficial BC represents about 60% of all BC cases. NMIBC is characterized by the loss of heterozygosity in chromosome 9 and mutations in *FGFR3*. In 10% to 20% of the cases, NMIBC can invade the muscular layer and evolve as muscular invasive bladder cancer (MIBC); this evolution is characterized by the loss of the tumor suppressor genes *TP53* and *RB1* [5]. MIBC is the most aggressive BC type, and it is characterized by genomic instability and a high mutational rate. Patients with MIBC tend to present a 5-year survival rate of 60% when having a localized tumor but less than 10% when distant metastases are present [6].

The molecular analysis of BC has led to the development of different classifications of both NMIBC and MIBC [7-10]. After an international consensus, MIBC was classified into six different molecular subtypes: Basal/Squamous, Neuroendocrine-like, Stroma-rich, Luminal-papillary, Luminal-unstable, and Luminal-non-specified [6]. The improvements in molecular biology and our understanding of tumorigenesis open the window to the personalized medicine era for patients with BC [1]. However, the molecular and genomic data found for BC has been obtained principally from European and European descendant patients; other ethnic groups, like Hispanics, are underrepresented in these classifications. Here we aim to identify the somatic single nucleotide variants (SNVs) and copy number variations (CNVs) present in Mexican patients with bladder cancer, exploring their association with clinical-pathological characteristics.

2. Results

2.1 Clinical-pathological characteristics of the BC patients

We included 37 patients treated at the National Cancer Institute - Mexico (INCan) between 2012 and 2021. The mean age of patients was 62.46 years (Standard deviation [SD]: 11.41 years) with a mean time to follow-up of 39.91 months (SD: 19.44). Men accounted for 70.73% of the patients included in this study. Of the total patients, 59.46% had an educational level of high school or less, and 48.56% had a history of smoking. Nearly half of the patients had a MIBC phenotype (45.95%), whereas 59.46 were found with subepithelial infiltration. Details of the characteristics of the patients are shown in Table 1.

Table 1. Demographic characteristics of patients with bladder cancer attended at the National Cancer Institute - Mexico between 2012 and 2021 (N=37).

Variable	Mean / n	SD / %
Age (Years)	62.49	11.41
BMI (kg/m²)	26.75	3.99
Time to follow-up (Months)	39.91	19.44
Sex		
Women	11	29.73%
Men	26	70.73%

Education level		
High school or less	22	59.46%
College or vocational school	5	13.51%
Grad school or higher	10	27.03%
Smoking	18	48.65%
Symptoms at diagnosis		
Asymptomatic	1	2.70%
Symptomatic[1]	2	5.41%
Hematuria	34	91.89%
Muscle invasion disease		0.00%
Yes	17	45.95%
Subepithelial infiltration		
Yes	22	59.46%
Histological grade		
Low	10	27.03%
High	27	72.97%
Recurrence	11	29.73%

SD=standard deviation. [1]Symptomatic= Irritative symptoms during urination (dysuria or burning sensation). BMI: Body Mass Index.

When evaluating the overall survival of the patients, a median value for survival in this cohort was not reached (Figure 1a). When grouping patients according to their muscle invasion phenotype, the patients with MIBC showed poor overall survival compared to patients with NMIBC (*p-value* = 0.001), with a median survival of 25 months (Figure 1b). We did not observe any differences related to overall survival in the Kaplan-Meier curves related to histological grade, smoking, or sex (Supplementary Figure S1).

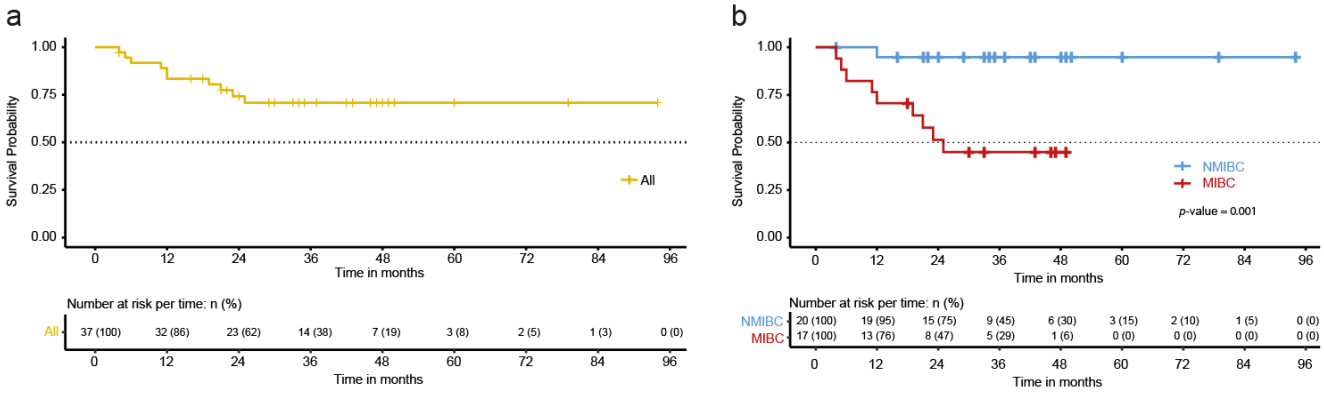

Figure 1. Overall survival of the patients with bladder cancer treated at the INCan (N= 37) between 2012 and 2021. a) Kaplan-Meier curve depicting the overall survival of all the patients included in the study. b) Kaplan-Meier curve showing the overall survival differences between the patients with non-muscle invasive bladder (NMIBC) cancer and muscle-invasive bladder cancer (MIBC). *p-value* was calculated using Log Rank test.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | "Posted: 31 October 2022" doi:10.20944/preprints202210.0463.v1

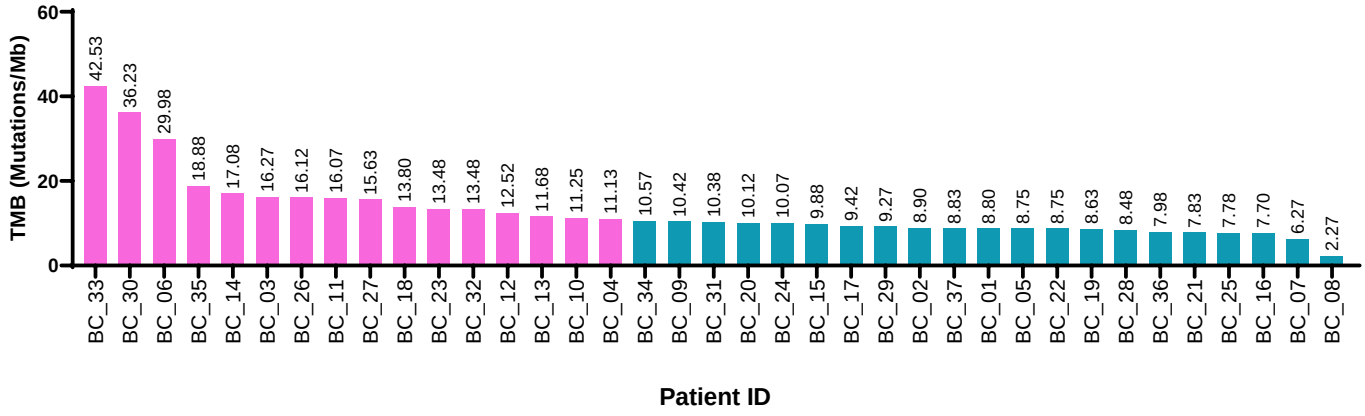

Figure 2. Tumor mutational burden (TMB) patients with bladder cancer treated at the NCI-Mx (INCan, N= 37) between 2012 and 2021. Bar plot depicts TMB distribution among all the patients. Pink bars represents patients with high TMB, and blue bars repre-

2.1 Tumor mutational burden and clinical-pathological characteristics

Whole exome sequencing (WES) was achieved successfully with an average sequencing depth on the target region of 183.90x and an average coverage of the target region of 99.65%. Samples had a mean of 94.83% Q30 quality in the sequenced bases. FASTQ files were obtained, and further bioinformatics analysis was performed to detect somatic and structural variations. Samples showed a median of 623 somatic variants (ranging from 136 to 2552 mutations per case), including SVNs and InDels (Insertion-or-Deletion). The total number of somatic variants, including driver and passenger

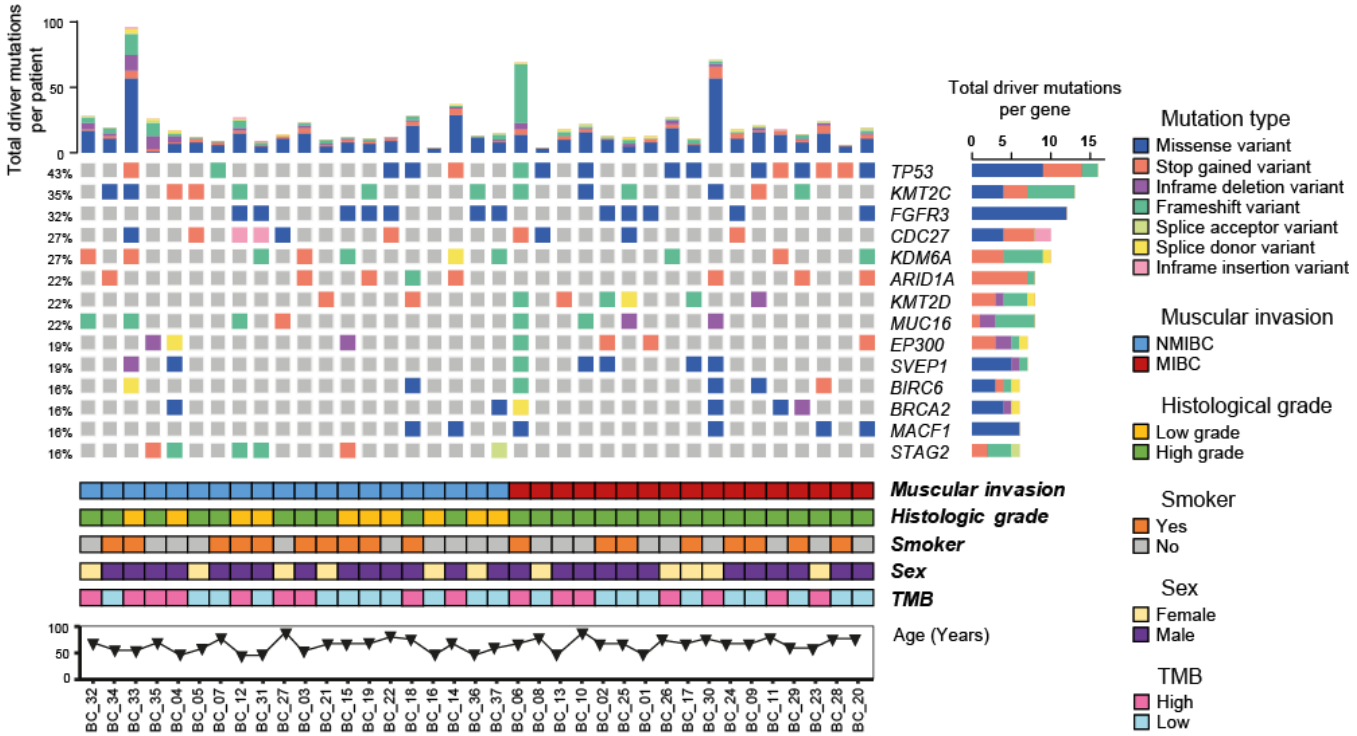

Figure 3. Driver somatic variants and clinical characteristics of patients with bladder cancer treated at INCan between 2012 and 2021 (n = 37). Oncoprint, sorted by muscular invasion, depicting the genes affected by driver somatic variants (single nucleotide variations, small insertions, and small deletions) in 16% or more of the samples. The variants are represented according to the mutation type, each described on the right color panel. The upper bar plot represents the number of driver somatic variants per patient, and the right bar plot represents the number of driver somatic variants per gene. The lower paneer represents the muscular invasion, histologic grade, smoking history, and sex of the patients.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | "Posted: 31 October 2022" doi:10.20944/preprints202210.0463.v1

mutations, present in the exome region, was used to identify the tumor mutational burden (TMB) of each patient. The TMB values ranged from 2.27 to 42.53 mutations/Mb. A total of 43% of the patients had TMB values greater than 10 mutations / Mb and were classified as having high TMB status (Figure 2). The relationship between TMB and clinical features is shown in Supplementary Figure S2.

2.2 Somatic variants and clinical-pathological characteristics of the patients

High-confidence somatic variants were classified as driver and passenger mutations. All the samples had driver somatic variants. The tumor suppressor gene *TP53* was the most frequently affected in our study, which was mutated in 43% of the patients. Other genes frequently mutated included *KMT2C* (35%), *FGRFR3* (32%), *CDC27* (27%), *KDM6A* (27%), *ARID1A* (27%), and *KTM2D* (27%). Missense SNVs were the most common type of somatic variant (Figure 3).

When grouping the patients according to their muscle invasion phenotype, we observed that *TP53* (*p*-value = 0.014, False Discovery Rate [FDR]: 0.54) and *KMT2D* (*p*-value = 0.060, FDR = 0.540) tend to have a high frequency of mutations in patients with MIBC (Figure 4a). All patients with MIBC had a high histologic grade, but not all patients with NMIBC had a low histologic grade. Mutations in *STAG2* were exclusive of patients with NMIBC phenotype and tended to be associated with a low histologic grade (*p*-value = 0.003, FDR = 0.72). Mutations in *FGFR3* also tend to be associated to patients with low histologic grade (p-value = 0.005, FDR = 0.729) (Figure 4b). Patients with positive smoking

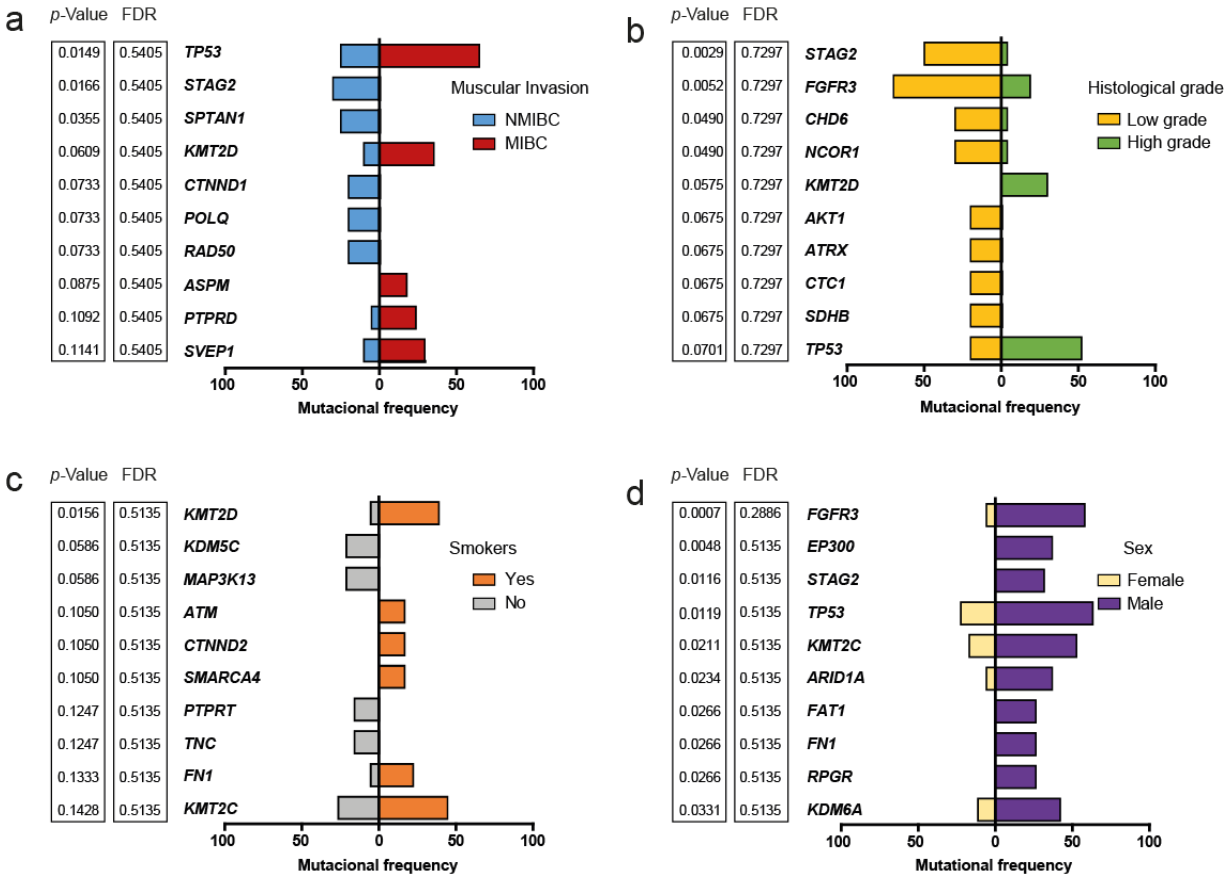

Figure 4. Muscle invasion, histologic grade, smoking status, sex, and frequency of somatic variants. The 10 most affected genes were ordered in ascending order according to smallest *p*-value on the top, obtained with the Fisher's exact test. a) Comparison of somatic variant frequency according to the muscular invasion phenotype. b) Comparison of somatic variant frequency according to the histologic grade c) Comparison of somatic variant frequency according to the smoking status. d) Comparison of somatic variant frequency according to the sex of the patients.

history were mainly associated with a high frequency of mutations of *KMT2D* (p-value = 0.015, FDR = 0.514) (Figure 4c). On the other hand, male patients tend to present a high frequency of mutations in *FGFR3* (p-value < 0.001, FDR = 0.288), *EP300* (p-value = 0.004, FDR=0.513), and *STAG2* (p-value = 0.011, FDR =0.513) (Figure 4d).

2.3 Structural variants and clinical-pathological characteristics of the patients

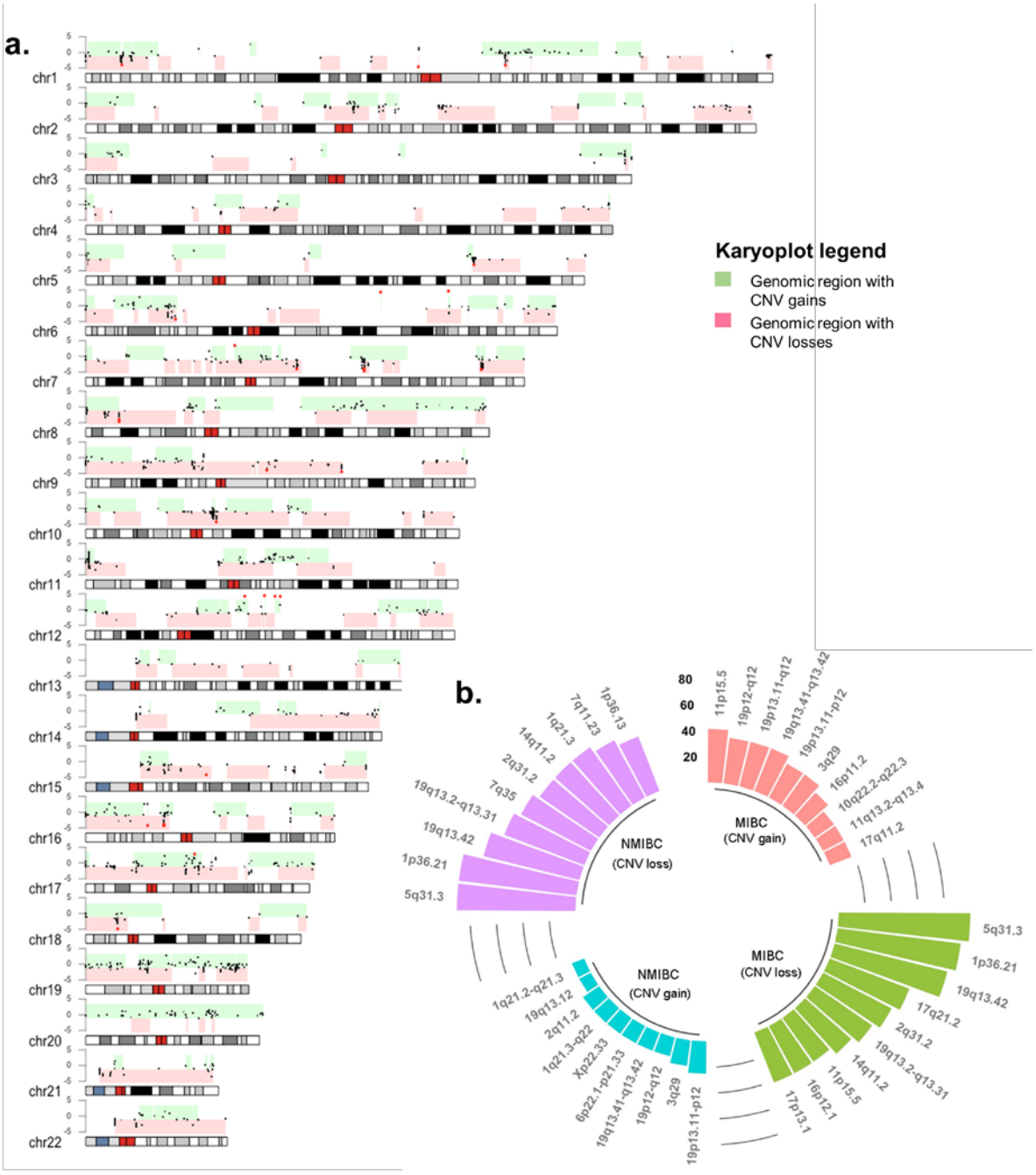

Figure 5. Structural variants present in Mexican patients with bladder cancer attended the NCI-Mx (INCan, n= 37). a) Karyoplot depicting the regions on each chromosome affected by copy number variations (CNV). The red bar represents losses, and the green bar indicates gains of cytoband copy number on each chromosome. The black dots indicate the start position of the CNV on each patient. The red dot indicates CNV with a log2 greater than 3 or lower than -3 on each patient. b) Bar plot presenting the cytobands more frequently affected by CNV gains and losses in patients with non-muscle invasive bladder cancer (NMIBC) and muscle invasive bladder cancer (MIBC).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | "Posted: 31 October 2022" doi:10.20944/preprints202210.0463.v1

The structural variants were identified for the patients as CNV gains and CNV losses. As shown in figure 5a, chromosomes 8, 19 and were affected almost entirely by CNV

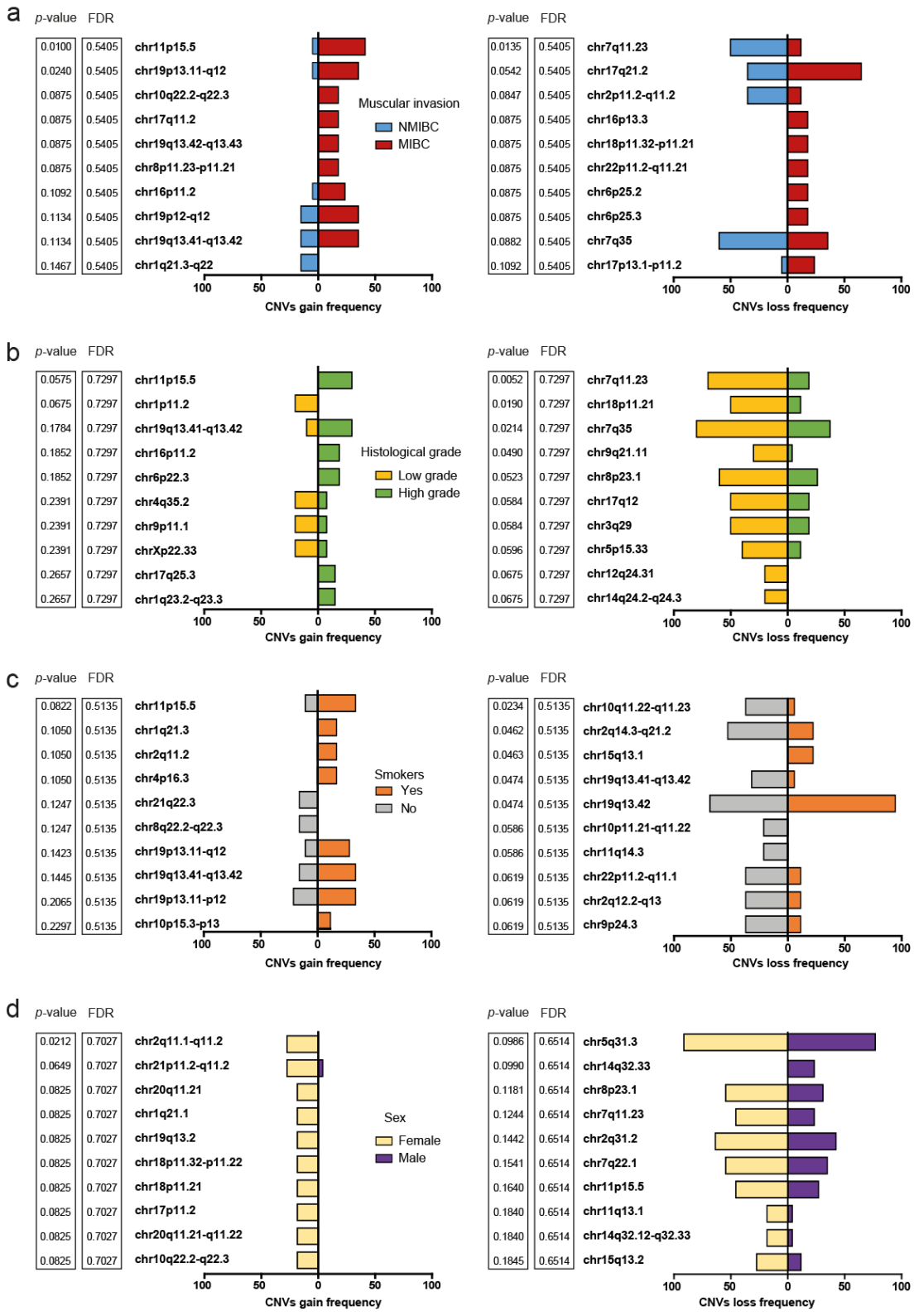

Figure 6. Muscle invasion, histologic grade, smoking, gender, and frequency of cytobands affected by CNV. The 10 most affected cytobands are ordered in ascending order from the smallest *p*-value according to Fisher's exact test on top a) Comparison of the frequency of gains and losses in copy number according to the muscle invasion phenotype. b) Comparison of the frequency of copy number gains and losses according to histological grade c) Comparison of the frequency of copy number gains and losses according to smoking. d) Comparison of the frequency of gains and losses in the number of copies according to the sex of the patients.

 doi:10.20944/preprints202210.0463.v1

gains, and chromosomes 9, 10, 17, 21, and 22 were affected mainly by CNV losses. Even though large chromosomic regions were affected by structural variants, our cohort also showed cytobands with a higher frequency of CNVs (Figure 5a). Cytoband chr5q31.3 was the most frequently affected, showing CNV losses in 98% of the samples. Therefore, the protocadherin gamma cluster genes (*PCDHG*), encoded in chr5q31.3, were the most frequently affected genes, with CNV losses in 95% of the patients. *ZMAT2* was also present in this cytoband and was found to be affected in 73% of the patients. Chr1p36.21 was affected completely by CNV losses in 95% of the cases and chr19q13.42 was affected in 82% of the cases by CNV gains and losses. Within chr1p36.21 and ch19q13.42, the genes frequently affected by CNV included: *RPS9* (78%), *LILRB3* (73%), *PRAMEF12* (54%) and *PRAMEF1* (38%). Figure 5b shows an arrangement of the cytobands frequently affected by CNV gains and losses among the patients evaluated and according to their muscle invasive phenotype.

When grouping according to the muscle invasive phenotype of the patients (Figure 6a), those with MIBC tend to show a high frequency of CNV gains in cytobands chr11p15.5 (p-value = 0.010, FDR = 0.540) and chr19p13.11-q12 (p-value = 0.024, FDR = 0.540). *MUC2*

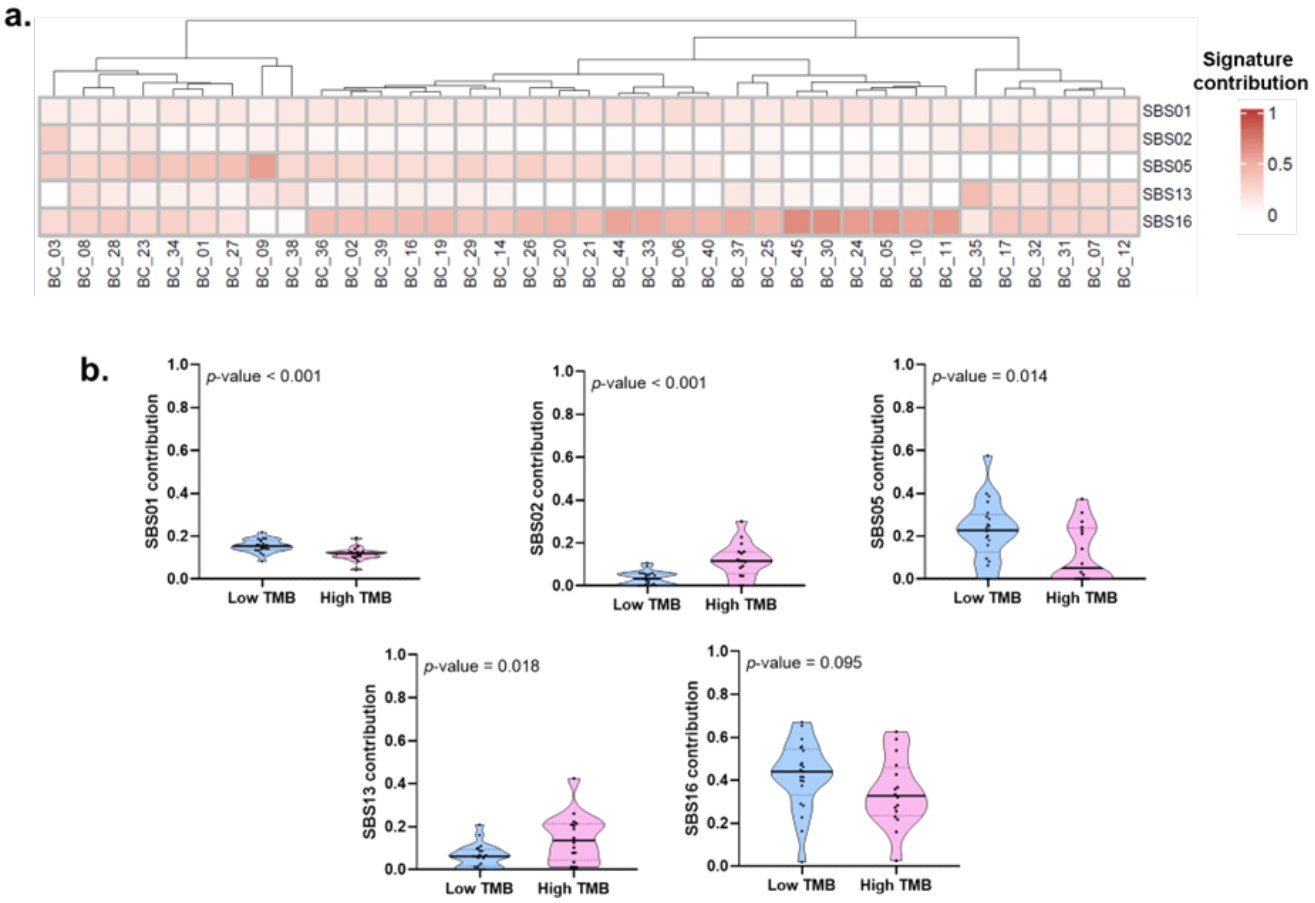

Figure 7. Mutational processes and their association with TMB status of patients with bladder cancer attended at the INCan (n= 37) between 2012 and 2021. a) Heatmap showing the contribution of the mutational signatures to the carcinogenic process. Only single-base substitution signatures (SBS) with a contribution more significant than cero were depicted. The gradient red color indicates the grade of the contribution of each signature. b) Violin plot depicting the association between the tumoral mutational burden (TMB) status of the patients and SBS01, SBS02, SBS05, SBS13, and SBS16. We did not find any association between TMB and clinical features (Supplementary Figure S1).

encoded in chr11p15.5 was found in all the patients affected by CNV gains, with log2 values ranging from 0.22 to 2.19. In the same direction, *ZNF429* harbored in chr19p13.11-q12 was found in all the patients affected by CNV gains. CNV gains in chr11p15.5 also tended to be frequent in patients with a high histologic grade, but did not reach the threshold of statistical significant for the raw *p*-value (*p*-value = 0.057, FDR 0.729) whereas CNV losses in chr7q11.23 (*p*-value = 0.005, FDR =0.729), chr18p11.21 (*p*-value = 0.019, FDR =0.729), and chr7q35 (*p*-value = 0.021, FDR = 0.729) were frequently affected in patients with low histologic grade (Figure 6b). Regarding the smoking status of the patients, CNV deletions in cytoband chr15q13.1 (*p*-value = 0.046, FDR =0.513) were found as significant in patients with a smoking history (Figure 6c). No significant associations between the sex of the patients and the affected cytobands were found (Figure 6d).

2.4 Mutational processes and clinical-pathological characteristics

The mutational processes implicated in bladder cancer were evaluated through mutational signature analysis, identifying the contribution of single-base substitution signatures (SBS). Signature SBS16, related to liver cancer and defective nucleotide excision repair pathway, was the most frequent mutational process, with a contribution greater than 0.2 and found in 89% of the patients, followed by SBS05 present in 51% and SBS13 in 19% of the patients. (Figure 7a) When evaluating the association of the mutational processes, we observed that patients with smoking history were associated with greater contribution values of SBS05 (Supplementary Figure S3). High TMB status was also associated with SBS02 and SBS13, whereas low TMB status was associated with SBS01 and SBS05 (Figure 7b).

3. Discussion

To our best knowledge, this is the first report of the somatic mutational landscape of bladder cancer in Mexican patients. Using WES, we identified a differential mutational pattern between patients with MIBC and NMIBC. Patients with MIBC were associated with a higher frequency of mutations in *TP53* and *KMT2D* and CNV gains in chr11q15.5 and chr19p13.11-q12. In contrast, mutations in *STAG2* and CNV deletions in chr7q11.23 were associated with patients with NMIBC and a low histologic grade.

This group of patients evaluated included a similar proportion of patients with MIBC and NMIBC, with a greater proportion of cases with high histologic grades. We identified *KMT2D* as a gene frequently mutated in patients with MIBC and high histologic grade. *KMT2D* codifies for the lysine methyltransferase 2D; the loss of this gene has been associated with abnormal epigenetic reprogramming of different molecular pathways [11]. *KMT2D* is mutated in different types of cancer, including bladder cancer [12]. In bladder cancer cell lines, *KMT2D* overexpression was associated with tumor suppressor effects as it promotes the expression of *PTEN* and *TP53*, as well as the repression of *STAG2* [13]. In lung cancer, it has been observed that *KMT2D* mutations increase the glycolysis of tumor cells. For this reason, it has been proposed to study the pharmacological inhibition of glycolysis in tumors deficient in *KMT2D* [14]. The expression of *KMT2D* has also been proposed as a prognostic biomarker for bladder cancer in European patients, observing a trend towards greater survival in patients with higher expression of this protein [15]. However, it is unknown whether this effect was also present in Mexican patients.

We also observed that patients with MIBC frequently showed copy number gains of the chr11p15.5 cytoband. *MUC2* was seen to be affected in all patients with these gains. This protein has been seen to contribute to the development of colorectal cancer [16]. Similarly, there is evidence suggesting that the expression of *MUC2* is associated with invasion and metastasis in various malignant tumors, including gastric cancer, prostate cancer, and colorectal cancer [17]. In the case of bladder cancer, the presence of *MUC2* has been associated with the non-invasive proliferation of tumors or with a favorable outcome for patients [18]. It has also been seen that the expression of *MUC2* is observed more

frequently in low-grade urothelial carcinomas and correlates with a low pathological stage [18].

All the genomic studies have limitations due to the complexity of the tumor, microenvironment, and changes in molecular and epigenetics of these. One particular in this study has some limitations. First, a reduced number of cases in our study could hide the frequency of genomic variants identified. Also, this could be a real trouble to replicate and find the same variants with other small cohorts. The second is the advance and chances in the bioinformatics analyses because analytical methods are continually being improved, and we may be omitting genetic variants of low expression and great relevance. Nonetheless, this study provides new information about the clinical and pathological characteristics of the Hispanic population regarding bladder cancer, which could guide us in the design of further studies that feature a larger cohort to test some of the identified variants as biomarkers of prognosis for patients treated.

4. Material and Methods

4.1. Population

We included patients treated between 2012 and 2021 at the National Cancer Institute – Mexico (INCan), located in Mexico City. Patients in this study were older than 18, had a positive diagnosis of urothelial bladder cancer, and had a transurethral resection of the bladder tumor (TURBT) before sample selection. Patients with a history of other cancer or any type of treatment prior to TURBT were not included in this study. Tumor samples were selected from formalin-fixed paraffin-embedded tissue (FFPE) blocks for an experienced pathologist. Selected tumor samples had at least 70% of tumor cellularity. This project was approved by the institutional ethics board (017/034/IBI) and the institutional ethics in research committee (CEI/1175/17) of the INCan. Due to the minimal risk and the nature of this study for the patients and following the guidelines of the institutional ethics board, the informed consent was waived.

4.2. Clinical data collection

Clinical data, including age at diagnosis, body mass index (BMI), sex, education level, smoking status, symptoms at diagnosis, muscular invasion, subepithelial invasion, and histological grade, were collected from the electronic clinical records of each patient. The follow-up time was calculated from the clinical records as the period between the date of diagnosis and the date of death or loss of follow-up.

4.3. DNA extraction and quality control

Samples were homogenized using QIAshredder (QIAGEN, 79654), and DNA was extracted using the QIAamp® DNA FFPE Tissue Kit (QIAGEN, 56404), following the protocol recommendations. Purity of DNA was evaluated with a Thermo Fisher Scientific NanoDrop 2000. Quantity and DNA fragmentation status were evaluated with the Agilent 2200 TapeStation System using the genomic DNA ScreenTape assay (Agilent, 5067-5365) to obtain a DNA integrity number [22]. Samples with DIN from 6 to 10 and a minimum concentration of 20 ng/μL were selected for WES.

4.4. Library preparation, hybridation capture, and WES

Library preparation, hybridization capture, and WES procedures were performed by the New York Genome Center. TruSeq DNA PCR-Free libraries were prepared from FFPE tissues using 1μg of input DNA according to the manufacturer's instructions (Ilumina, San Diego, CA). Sequencing was performed on HiSeq2500 (Ilumina, San Diego, CA).

4.5 Bioinformatics pipeline

Sequencing reads for the tumor samples were first trimmed for adapters using TrimGalore (v0.4.0). The trimmed reads were then aligned to the reference genome using BWA-MEM (v0.7.15) [23], GATK (v4.1.0) [24] FixMateInformation was run to verify and fix mate-pair information, followed by Novosort (v1.03.01) markDuplicates to merge individual lane BAM files into a single BAM file per sample. Duplicates were then sorted

and marked, and GATK's base quality score recalibration (BQSR) was performed to obtain a coordinated sorted BAM file for each sample.

A matched normal sample was not available. In its place, we used the HapMap sample NA12878, which was prepared and sequenced using the same protocol as the tumor sample. This normal sample was used to remove some of the false positives that were due to library preparation and sequencing (that would manifest in the same way in the tumor and NA12878), as well as some germline variants that are common to the tumor sample and NA12878.

The tumor and normal BAM files were processed using (GATK v4.0.5.1) [24], Strelka2 (v2.9.3) [25], and Lancet (v1.0.7) [26] for calling SNVs and short Indels. SvABA (v0.2.1) [27] for calling Indels. FACETS (v0.5.5) [28] for calling CNVs. High-confidence variants identified by at least two variant callers and variants with a variant allele frequency equal to or in between 0.1 and 0.45 were selected for subsequent analysis.

SNVs and Indels were annotated with Ensembl as well as databases such as COSMIC (v86) [29], 1000Genomes (Phase3) [30], ClinVar (201706) [31], PolyPhen (v2.2.2) [32], SIFT (v5.2.2) [33], FATHMM (v2.1) [34], gnomAD (r2.0.1) [35] and dbSNP (v150) [36] using Variant Effect Predictor (v93.2) [37]. Synonymous mutations and mutations annotated in non-coding regions were filtered out.

For CNVs, segments with log2 > 0.2 were categorized as amplifications, and segments with log2 < -0.235 were categorized as deletions (corresponding to a single copy change at 30% purity in a diploid genome, or a 15% Variant Allele Fraction). CNVs of size less than 20Mb are denoted as focal and the rest are considered large-scale. Only focal CNV were selected for subsequent analysis. We use Bedtools [38] for annotating CNVs. All predicted CNVs were annotated with germline variants by overlapping with known variants in 1000 Genomes and Database of Genomic Variants (DGV) [39].

4.5 Mutational signature analysis

High Confidence somatic SNVs within autosomes were used to estimate the contribution of known COSMIC mutational signatures (v2 - March 2015) 5 in each tumor sample using deconstructSigs (v1.8.0) [40]. The SNV mutation count data was normalized to reflect the absolute frequency of each trinucleotide context as it would occur across the whole genome. The "tri.counts.method" parameter in deconstructSigs was set to "exome2genome", and a custom exome trinucleotide counts file based on the target interval was provided.

4.6 Statistical analysis

Associations between clinical characteristics such as muscular invasion, histologic grade smoking history, and sex were assessed with the presence of somatic variations and CNVs. Patients were grouped according to their clinical characteristics, then the frequency of the variants between the groups was compared using Fisher's exact test. p-values were adjusted using the Benjamini–Hochberg procedure. The Kaplan-Meier curve was analyses by Log rank test. The mutational signatures contribution was compared with the clinical characteristics of the patients using the Mann-Whitney U test. Statistical significance was set at p-value < 0.05. All statistical analyses were performed using R software (https://cran.r-project.org).

5. Conclusions

This is the first work in a Mexican population in which the mutational panorama of BC is characterized for both MIBC and NMIBC. Patients with MIBC showed a higher frequency of mutations in *TP53* and *KMT2D*, gains in chr11q15.5 and chr19p13.11-q12, and losses in chr7q11.23. *STAG2* mutations and CNV deletions at chr1q11.23 were frequently found in patients with NMIBC and low histologic grade. These genomic changes may open new research lines toward their specific detection at diagnosis in prospective studies.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Figure S1: Relationship between TMB value and clinical features. a) Comparison of TMB value with the muscular invasion; Figure S2: Association between mutational signature contribution and clinical characteristics of patients with bladder cancer. Figure S3: Overall survival of the patients with bladder cancer with other clinical features.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, M.D.P.M. and D.P.; methodology, D.C.I., Y.S.P.; software, D.C.I.; formal analysis, D.C.I., Y.S.P., J.D.C, C.C.C., J.A.R., M.J.R., L.A.H., A.S., D.C.L., A.M.G., R.H.M., F.V.P., A.R., A.O.M., D.P.; investigation, R.H.M., A.R., A.O.M., D.P.; resources, D.P.; data curation, D.C.I.; writing—original draft preparation, D.C.I., R.H.M., D.P.; writing—review and editing, D.C.I., R.H.M., D.P.; visualization, D.C.I.; supervision, D.P.; project administration, A.O.M.; funding acquisition, D.P. All authors have read and agreed to the published version of the manuscript." Please turn to the CRediT taxonomy for the term explanation.

Funding: This research was funded by CONACYT, grant numbers FOSISS-2017-290412 and FOSISS-2017-289503.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of the National Cancer Institute – Mexico (protocol code CEI/1193/17).

Informed Consent Statement: Because of the retrospective design of this study, informed consent was not obtained from all subjects involved in the study.

Data Availability Statement: The datasets used and/or analyzed during the current study are available on reasonable request.

Acknowledgments: We thank Clementina Castro for the critical revision of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Supplementary Fig 1.
Supplementary Fig 2.
Supplementary Fig 3.

References

1. Dobruch, J.; Oszczudłowski, M., Bladder Cancer: Current Challenges and Future Directions. *Medicina (Kaunas)* **2021,** 57, (8).
2. Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F., Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians* **2021,** 71, (3), 209-249.
3. Freedman, N. D.; Silverman, D. T.; Hollenbeck, A. R.; Schatzkin, A.; Abnet, C. C., Association between smoking and risk of bladder cancer among men and women. *JAMA* **2011,** 306, (7), 737-45.
4. Kirkali, Z.; Chan, T.; Manoharan, M.; Algaba, F.; Busch, C.; Cheng, L.; Kiemeney, L.; Kriegmair, M.; Montironi, R.; Murphy, W. M.; Sesterhenn, I. A.; Tachibana, M.; Weider, J., Bladder cancer: epidemiology, staging and grading, and diagnosis. *Urology* **2005,** 66, (6 Suppl 1), 4-34.
5. Siracusano, S.; Rizzetto, R.; Porcaro, A. B., Bladder cancer genomics. *Urologia* **2020,** 87, (2), 49-56.
6. Kamoun, A.; de Reyniès, A.; Allory, Y.; Sjödahl, G.; Robertson, A. G.; Seiler, R.; Hoadley, K. A.; Groeneveld, C. S.; Al-Ahmadie, H.; Choi, W.; Castro, M. A. A.; Fontugne, J.; Eriksson, P.; Mo, Q.; Kardos, J.; Zlotta, A.; Hartmann, A.; Dinney, C. P.; Bellmunt, J.; Powles, T.; Malats, N.; Chan, K. S.; Kim, W. Y.; McConkey, D. J.; Black, P. C.; Dyrskjøt, L.; Höglund, M.; Lerner, S. P.; Real, F. X.; Radvanyi, F.; Group, B. C. M. T., A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. *Eur Urol* **2020,** 77, (4), 420-433.

7. Hedegaard, J.; Lamy, P.; Nordentoft, I.; Algaba, F.; Høyer, S.; Ulhøi, B. P.; Vang, S.; Reinert, T.; Hermann, G. G.; Mogensen, K.; Thomsen, M. B. H.; Nielsen, M. M.; Marquez, M.; Segersten, U.; Aine, M.; Höglund, M.; Birkenkamp-Demtröder, K.; Fristrup, N.; Borre, M.; Hartmann, A.; Stöhr, R.; Wach, S.; Keck, B.; Seitz, A. K.; Nawroth, R.; Maurer, T.; Tulic, C.; Simic, T.; Junker, K.; Horstmann, M.; Harving, N.; Petersen, A. C.; Calle, M. L.; Steyerberg, E. W.; Beukers, W.; van Kessel, K. E. M.; Jensen, J. B.; Pedersen, J. S.; Malmström, P. U.; Malats, N.; Real, F. X.; Zwarthoff, E. C.; Ørntoft, T. F.; Dyrskjøt, L., Comprehensive Transcriptional Analysis of Early-Stage Urothelial Carcinoma. *Cancer Cell* **2016,** 30, (1), 27-42.
8. Sjödahl, G.; Lauss, M.; Lövgren, K.; Chebil, G.; Gudjonsson, S.; Veerla, S.; Patschan, O.; Aine, M.; Fernö, M.; Ringnér, M.; Månsson, W.; Liedberg, F.; Lindgren, D.; Höglund, M., A molecular taxonomy for urothelial carcinoma. *Clin Cancer Res* **2012,** 18, (12), 3377-86.
9. Network, C. G. A. R., Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **2014,** 507, (7492), 315-22.
10. Robertson, A. G.; Kim, J.; Al-Ahmadie, H.; Bellmunt, J.; Guo, G.; Cherniack, A. D.; Hinoue, T.; Laird, P. W.; Hoadley, K. A.; Akbani, R.; Castro, M. A. A.; Gibb, E. A.; Kanchi, R. S.; Gordenin, D. A.; Shukla, S. A.; Sanchez-Vega, F.; Hansel, D. E.; Czerniak, B. A.; Reuter, V. E.; Su, X.; de Sa Carvalho, B.; Chagas, V. S.; Mungall, K. L.; Sadeghi, S.; Pedamallu, C. S.; Lu, Y.; Klimczak, L. J.; Zhang, J.; Choo, C.; Ojesina, A. I.; Bullman, S.; Leraas, K. M.; Lichtenberg, T. M.; Wu, C. J.; Schultz, N.; Getz, G.; Meyerson, M.; Mills, G. B.; McConkey, D. J.; Weinstein, J. N.; Kwiatkowski, D. J.; Lerner, S. P.; Network, T. R., Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. *Cell* **2018,** 174, (4), 1033.
11. Dhar, S. S.; Lee, M. G., Cancer-epigenetic function of the histone methyltransferase KMT2D and therapeutic opportunities for the treatment of KMT2D-deficient tumors. *Oncotarget* **2021,** 12, (13), 1296-1308.
12. Ding, B.; Yan, L.; Zhang, Y.; Wang, Z.; Xia, D.; Ye, Z.; Xu, H., Analysis of the role of mutations in the KMT2D histone lysine methyltransferase in bladder cancer. *FEBS Open Bio* **2019,** 9, (4), 693-706.
13. Sun, P.; Wu, T.; Sun, X.; Cui, Z.; Zhang, H.; Xia, Q.; Zhang, D., KMT2D inhibits the growth and metastasis of bladder Cancer cells by maintaining the tumor suppressor genes. *Biomed Pharmacother* **2019,** 115, 108924.
14. Alam, H.; Tang, M.; Maitituoheti, M.; Dhar, S. S.; Kumar, M.; Han, C. Y.; Ambati, C. R.; Amin, S. B.; Gu, B.; Chen, T.-Y.; Lin, Y.-H.; Chen, J.; Muller, F. L.; Putluri, N.; Flores, E. R.; DeMayo, F. J.; Baseler, L.; Rai, K.; Lee, M. G., KMT2D Deficiency Impairs Super-Enhancers to Confer a Glycolytic Vulnerability in Lung Cancer. *Cancer Cell* **2020,** 37, (4), 599-617.e7.
15. Laukhtina, E.; Lemberger, U.; Bruchbacher, A.; Ilijazi, D.; Korn, S.; Berndl, F.; D'Andrea, D.; Susani, M.; Enikeev, D.; Compérat, E.; Shariat, S. F.; Hassler, M. R., Expression Analysis and Mutational Status of Histone Methyltransferase. *J Pers Med* **2021,** 11, (11).
16. Betge, J.; Schneider, N. I.; Harbaum, L.; Pollheimer, M. J.; Lindtner, R. A.; Kornprat, P.; Ebert, M. P.; Langner, C., MUC1, MUC2, MUC5AC, and MUC6 in colorectal cancer: expression profiles and clinical significance. *Virchows Arch* **2016,** 469, (3), 255-65.
17. Li, C.; Zuo, D.; Yin, L.; Lin, Y.; Liu, T.; Wang, L., Prognostic Value of MUC2 Expression in Colorectal Cancer: A Systematic Review and Meta-Analysis. *Gastroenterol Res Pract* **2018,** 2018, 6986870.
18. Gonul, I. I.; Cakir, A.; Sozen, S., Immunohistochemical expression profiles of MUC1 and MUC2 mucins in urothelial tumors of bladder. *Indian J Pathol Microbiol* **2018,** 61, (3), 350-355.
19. McGrail, D. J.; Pilié, P. G.; Rashid, N. U.; Voorwerk, L.; Slagter, M.; Kok, M.; Jonasch, E.; Khasraw, M.; Heimberger, A. B.; Lim, B.; Ueno, N. T.; Litton, J. K.; Ferrarotto, R.; Chang, J. T.; Moulder, S. L.; Lin, S. Y., High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. *Ann Oncol* **2021,** 32, (5), 661-672.
20. Lv, J.; Zhu, Y.; Ji, A.; Zhang, Q.; Liao, G., Mining TCGA database for tumor mutation burden and their clinical significance in bladder cancer. *Biosci Rep* **2020,** 40, (4).

21. Natesan, D.; Zhang, L.; Martell, H. J.; Jindal, T.; Devine, P.; Stohr, B.; Espinosa-Mendez, C.; Grenert, J.; Van Ziffle, J.; Joseph, N.; Umetsu, S.; Onodera, C.; Turski, M.; Chan, E.; Desai, A.; Aggarwal, R.; Wong, A.; Porten, S.; Chou, J.; Friedlander, T.; Fong, L.; Small, E. J.; Sweet-Cordero, A.; Koshkin, V. S., APOBEC Mutational Signature and Tumor Mutational Burden as Predictors of Clinical Outcomes and Treatment Response in Patients With Advanced Urothelial Cancer. *Front Oncol* **2022,** 12, 816706.
22. Pena-Llopis, S.; Brugarolas, J., Simultaneous isolation of high-quality DNA, RNA, miRNA and proteins from tissues for genomic applications. *Nat Protoc* **2013,** 8, (11), 2240-55.
23. Li, H.; Durbin, R., Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **2010,** 26, (5), 589-95.
24. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; DePristo, M. A., The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **2010,** 20, (9), 1297-303.
25. Saunders, C. T.; Wong, W. S.; Swamy, S.; Becq, J.; Murray, L. J.; Cheetham, R. K., Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **2012,** 28, (14), 1811-7.
26. Narzisi, G.; Corvelo, A.; Arora, K.; Bergmann, E. A.; Shah, M.; Musunuri, R.; Emde, A. K.; Robine, N.; Vacic, V.; Zody, M. C., Genome-wide somatic variant calling using localized colored de Bruijn graphs. *Commun Biol* **2018,** 1, 20.
27. Wala, J. A.; Bandopadhayay, P.; Greenwald, N. F.; O'Rourke, R.; Sharpe, T.; Stewart, C.; Schumacher, S.; Li, Y.; Weischenfeldt, J.; Yao, X.; Nusbaum, C.; Campbell, P.; Getz, G.; Meyerson, M.; Zhang, C. Z.; Imielinski, M.; Beroukhim, R., SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome Res* **2018,** 28, (4), 581-591.
28. Shen, R.; Seshan, V. E., FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res* **2016,** 44, (16), e131.
29. Tate, J. G.; Bamford, S.; Jubb, H. C.; Sondka, Z.; Beare, D. M.; Bindal, N.; Boutselakis, H.; Cole, C. G.; Creatore, C.; Dawson, E.; Fish, P.; Harsha, B.; Hathaway, C.; Jupe, S. C.; Kok, C. Y.; Noble, K.; Ponting, L.; Ramshaw, C. C.; Rye, C. E.; Speedy, H. E.; Stefancsik, R.; Thompson, S. L.; Wang, S.; Ward, S.; Campbell, P. J.; Forbes, S. A., COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **2019,** 47, (D1), D941-D947.
30. Auton, A.; Brooks, L. D.; Durbin, R. M.; Garrison, E. P.; Kang, H. M.; Korbel, J. O.; Marchini, J. L.; McCarthy, S.; McVean, G. A.; Abecasis, G. R.; Consortium, G. P., A global reference for human genetic variation. *Nature* **2015,** 526, (7571), 68-74.
31. Landrum, M. J.; Lee, J. M.; Riley, G. R.; Jang, W.; Rubinstein, W. S.; Church, D. M.; Maglott, D. R., ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **2014,** 42, (Database issue), D980-5.
32. Adzhubei, I.; Jordan, D. M.; Sunyaev, S. R., Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* **2013,** Chapter 7, Unit7.20.
33. Vaser, R.; Adusumalli, S.; Leng, S. N.; Sikic, M.; Ng, P. C., SIFT missense predictions for genomes. *Nat Protoc* **2016,** 11, (1), 1-9.
34. Shihab, H. A.; Gough, J.; Mort, M.; Cooper, D. N.; Day, I. N.; Gaunt, T. R., Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. *Hum Genomics* **2014,** 8, 11.
35. Lek, M.; Karczewski, K. J.; Minikel, E. V.; Samocha, K. E.; Banks, E.; Fennell, T.; O'Donnell-Luria, A. H.; Ware, J. S.; Hill, A. J.; Cummings, B. B.; Tukiainen, T.; Birnbaum, D. P.; Kosmicki, J. A.; Duncan, L. E.; Estrada, K.; Zhao, F.; Zou, J.; Pierce-Hoffman, E.; Berghout, J.; Cooper, D. N.; Deflaux, N.; DePristo, M.; Do, R.; Flannick, J.; Fromer, M.; Gauthier, L.; Goldstein, J.; Gupta, N.; Howrigan, D.; Kiezun, A.; Kurki, M. I.; Moonshine, A. L.; Natarajan, P.; Orozco, L.; Peloso, G. M.; Poplin, R.; Rivas, M. A.; Ruano-Rubio, V.; Rose, S. A.; Ruderfer, D. M.; Shakir, K.; Stenson, P. D.; Stevens, C.; Thomas, B. P.; Tiao, G.; Tusie-Luna, M. T.; Weisburd, B.; Won, H. H.; Yu, D.; Altshuler, D. M.; Ardissino, D.; Boehnke, M.; Danesh, J.; Donnelly, S.;

Elosua, R.; Florez, J. C.; Gabriel, S. B.; Getz, G.; Glatt, S. J.; Hultman, C. M.; Kathiresan, S.; Laakso, M.; McCarroll, S.; McCarthy, M. I.; McGovern, D.; McPherson, R.; Neale, B. M.; Palotie, A.; Purcell, S. M.; Saleheen, D.; Scharf, J. M.; Sklar, P.; Sullivan, P. F.; Tuomilehto, J.; Tsuang, M. T.; Watkins, H. C.; Wilson, J. G.; Daly, M. J.; MacArthur, D. G.; Consortium, E. A., Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **2016,** 536, (7616), 285-91.

36. Wang, K.; Yuen, S. T.; Xu, J.; Lee, S. P.; Yan, H. H.; Shi, S. T.; Siu, H. C.; Deng, S.; Chu, K. M.; Law, S.; Chan, K. H.; Chan, A. S.; Tsui, W. Y.; Ho, S. L.; Chan, A. K.; Man, J. L.; Foglizzo, V.; Ng, M. K.; Ching, Y. P.; Cheng, G. H.; Xie, T.; Fernandez, J.; Li, V. S.; Clevers, H.; Rejto, P. A.; Mao, M.; Leung, S. Y., Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat Genet* **2014,** 46, (6), 573-82.
37. McLaren, W.; Gil, L.; Hunt, S. E.; Riat, H. S.; Ritchie, G. R.; Thormann, A.; Flicek, P.; Cunningham, F., The Ensembl Variant Effect Predictor. *Genome Biol* **2016,** 17, (1), 122.
38. Quinlan, A. R.; Hall, I. M., BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010,** 26, (6), 841-2.
39. MacDonald, J. R.; Ziman, R.; Yuen, R. K.; Feuk, L.; Scherer, S. W., The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res* **2014,** 42, (Database issue), D986-92.
40. Rosenthal, R.; McGranahan, N.; Herrero, J.; Taylor, B. S.; Swanton, C., DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol* **2016,** 17, 31.