

Determination of ERG(+), EZH2, NKX3.1, and SPINK-1 Subtypes to Evaluate Their Association With Clonal Origin and Disease Progression in Multifocal Prostate Cancer

Yenifer Yamile SEGURA-MORENO

Instituto Nacional de Cancerología

María Carolina SANABRIA-SALAS

Instituto Nacional de Cancerología

Jorge Andrés Mesa-López de Mesa

Instituto Nacional de Cancerología

Rodolfo VARELA-RAMIREZ

Instituto Nacional de Cancerología

Natalia Lizeth ACOSTA-VEGA

Instituto Nacional de Cancerología

Martha Lucía Serrano (✉ mlserranol@unal.edu.co)

Instituto Nacional de Cancerología

Research Article

Keywords: Prostate cancer, genetic heterogeneity, clonal evolution, TMPRSS2-ERG fusion, SPOP gene, ERG gene, EZH2 gene, NKX3.1 gene, SPINK-1 gene

Posted Date: January 17th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1193535/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

The prognostic relevance of prostate cancer (PCa) molecular subtypes remains controversial, given the presence of multiple foci, with the possibility of different subtypes, in the same PCa patient. To help elucidate the clonal origin of heterogeneity in PCa and its association with disease progression, SPOP, ERG(+), EZH2, NKX3.1, and SPINK-1 subtypes were analyzed.

Methods

Eighty-three samples of different foci (n=50) and positive lymph nodes (LN) (n=13) from 20 PCa patients were used to analyze the presence of the TMPRSS2-ERG fusion and expression levels of ERG, EZH2, NKX3.1, and SPINK-1, using RT-PCR. Mutations in exons 6 and 7 of the SPOP gene were searched by sequencing. The presence of subtypes and molecular patterns were established, combining the presence of all subtypes analyzed. To establish the clonal origin of multifocal PCa, the molecular concordance between different foci of the same patient was determined. Association of these subtypes with histopathological groups and time to biochemical recurrence (BCR) was assessed.

Results

No mutation in SPOP was found in any sample. The ERG(+) subtype was the most frequent. The molecular pattern with the four PCa subtypes was detected only in 3 samples (4%), all LN, but it was the most frequent (40%) in patients. Molecular discordance was the predominant status (55%) when all the analyzed molecular characteristics were considered. All the subtypes could be found starting from preneoplastic lesion, and all but one LN molecular subtype were ERG(+) and NKX3.1 subtypes. Only the expression of the NKX3.1 gene was significantly different among the histopathological groups. No association was found between BCR time in patients, and molecular subtypes or molecular concordance, or between clinicopathological characteristics and molecular subtype with ERG, EZH2, and SPINK-1 subtypes.

Conclusion

The predominance of molecular discordance in prostatic foci per patient, which reflects the multifocal origin of PCa foci, highlights the importance of analyzing multiple samples, to establish the prognostic and therapeutic relevance of molecular subtypes in a patient. All the subtypes analyzed here are of early onset, starting from preneoplastic lesions. NKX3.1 gene expression is the only molecular characteristic that shows a progression pattern by sample.

Background

According to the report of the International Agency for Research on Cancer (IARC), prostate cancer (PCa) is the second type of cancer with the highest incidence worldwide, and the fifth cause of death by cancer worldwide for both sexes and all ages (1). PCa starts as a preneoplastic lesion known as high-grade prostatic intraepithelial neoplasia (HGPIN), it progresses to cancer, and can cause regional and distant metastasis. PCa in a large percentage of cases behaves as an indolent disease, but in other cases it can behave in an aggressive manner and

cause death. The criteria to establish PCa prognosis include PSA level, Gleason score, and clinical TNM stage, but these criteria still do not establish the prognosis with sufficient certainty (2).

Many studies have established different classifications of PCa in molecular subtypes that can contribute to determining the prognosis, among them ETS(+) subtypes, characterized by chromosomal fusions, which cause the overexpression of ETS transcription factors. The most prevalent ETS fusion in PCa is TMPRSS2-ERG, which leads to ERG overexpression (3-5). ERG overexpression leads to an increase in the expression of EZH2, which acts as a histone methyltransferase enzyme, as well as to a decrease in the expression of the tumor suppressor gene NKX3.1, which is an androgen-regulated prostate-specific homeobox gene and a transcription factor. NKX3.1 also binds to TMPRSS2 upstream sequences, and in TMPRSS2-ERG cases, it negatively regulates ERG expression and constitutes a negative feedback control since ERG in turn directly represses NKX3.1 (6,7). EZH2 and NKX3.1 had also been described as PCa subtypes. Another subtype reported in PCa is characterized by SPINK-1 overexpression, which is specifically overexpressed in a subset of ETS(-) and is associated with another subtype, the SPOP subtype, characterized by a mutation in this gene (3). SPINK-1 protein is a trypsin inhibitor, which can function as an autocrine growth factor (8,9), and SPOP is an E3 ubiquitin ligase substrate-binding subunit of the proteasome complex that mediates the ubiquitination of target proteins, leading more frequently to their proteasomal degradation (10).

The prognostic relevance of PCa molecular subtypes remains controversial (11-20). The presence of multiple foci—with the possibility of different subtypes—in the same patient with PCa could contribute to the difficulty of predicting the clinical behavior of a specific PCa subtype (21-23). It is not well known whether these foci have a monoclonal origin that expands through the prostate, giving rise to new foci, and share some molecular characteristics by origin, which begins to acquire changes in subclonal events; or whether each focus appears independently, so they have a multiclonal origin and they are expected to be very heterogeneous at the molecular level (22,24,25).

To help elucidate the origin of heterogeneity in PCa and its association with disease progression, ERG(+), SPOP, EZH2, SPINK-1, and NKX3.1 subtypes were analyzed in 83 samples from prostatic foci and regional metastasis to lymph nodes (LN) from 20 PCa patients with poor prognosis. We also established molecular patterns in samples and patients, combining the presence of all subtypes analyzed.

Methods

Samples

Samples from formalin-fixed paraffin-embedded (FFPE) tissues obtained from radical prostatectomy (RP) of 20 PCa patients with poor prognosis (regional metastases and/or biochemical recurrence (BCR)) and 5-years of follow-up were included. Cases were selected from the project “Exploration of potential predictor biomarkers in patients diagnosed with prostate cancer,” approved by the Ethics Committee of the Instituto Nacional de Cancerología, (Bogotá, Colombia), for the development of this study. Tissues were obtained as part of routine care and was not obtained specifically for this study and data used in this study was anonymised before. The Ethics Committee of the Instituto Nacional de Cancerología, (Bogotá, Colombia) waived the need of informed consent for this study.

Four to five FFPE samples per patient were selected to find foci with different degrees of differentiation; a total of 83 samples were analyzed. These samples corresponded to: HGPIN (n=17), PCa foci with different Gleason scores (n= 53; GL3, n=23; GL4, n=23; and GL5, n=7), and LN (n=13). One sample of adjacent non-tumor prostate tissue of each patient was used in this study as reference. We also used two frozen PCa tissue samples as a positive control of the TMPRSS2-ERG fusion, as well as the PC3 cell line as a negative control for gene fusions. All methods were carried out in accordance with relevant guidelines and regulations at national and international level.

Cell culture

The PC3 cell line was derived from androgen-independent PCa. This line was kindly provided by Dr. Fabio Aristizábal, from the Universidad Nacional de Colombia. The culture conditions were 5% CO₂ at 37°C, in DMEM medium (Dulbecco/Vogt modified Eagle's minimal essential medium), supplemented with 10% fetal bovine serum and a cocktail composed of penicillin, streptomycin, and 1% amphotericin B.

Extraction and quantification of DNA/RNA from FFPE tissue samples

The DNA/RNA was isolated using four punch biopsies from FFPE tissue samples of each selected foci or LN, by automated extraction in the QIAcube (QIAGEN, Hilden, Germany), as well as using a modified protocol from the commercial kit AllPrep DNA/RNA FFPE (QIAGEN, Hilden, Germany). Tissue cores were incubated with xylol at 56°C for 1.5 hour, followed by two steps of washing with ethanol to deparaffinize the tissues. Subsequently, the sample was prepared for automated extraction, following the steps set out in the standard protocol (<https://www.qiagen.com/us/qiacube/standard/protocolview.aspx?StandardProtocolID=975>). The modification consisted in the incubation of the sample before starting the automated protocol with proteinase K at 56°C for 3 hours, instead of 15 minutes. The RNA was quantified by using the Qubit® 2.0 fluorometer for HS RNA (Invitrogen, Waltham, Massachusetts, USA). The RNA was stored at -70°C until its use. Good-quality DNA was obtained with this same kit, which was also quantified by the Qubit® 2.0 fluorometer for HS DNA and stored at -20°C.

Analysis of the presence or absence of the TMPRSS2-ERG fusion

RT-PCR was performed to determine the presence of the TMPRSS2-ERG fusion (exon 1 of TMPRSS2: exon 4 of ERG), using the KAPA SYBR FAST One-Step kit to carry out qRT-PCR (Kapa biosystems, Wilmington, Massachusetts, USA), following the recommendations of the manufacturer. Briefly, 4 µl of RNA sample (10 ng/µl) were combined with the KAPA SYBR FAST qPCR Master Mix (2X) Universal, 0.4 µl of 50X KAPA RT Mix, 2 µl of each pair of primers, and 1.6 µl of nuclease-free water for RT-PCR. The primers used are presented in Table 1 (5). The protocol begins with reverse transcription at 42°C for 5 minutes during one cycle, activation of the enzyme at 95°C for 3 minutes during a cycle, then 40 cycles of denaturation at 95°C for 10 seconds, annealing temperature for 20 seconds for each transcript presented in Table 1, and extension at 72°C for 1 second. It was run at PTC-200 Thermal Cycler (PCR System) (MJ Research, Reno, Nevada, USA). The presence or absence of the TMPRSS2-ERG fusion was determined by visualizing the corresponding amplicons in 2% agarose gels with SYBR Safe.

Analysis of the presence or absence of SPOP mutations

The identification of SPOP mutations was done by sequencing exons 6 and 7. First, we amplified exons 6 and 7 through PCR with the Platinum High Fidelity PCR kit (Invitrogen, Waltham, Massachusetts, USA), using the SPOP-6 and SPOP-7 primers (26) (Table 1), with the following cycling conditions: incubation at 94°C for 2 minutes,

denaturation at 94°C for 30 seconds, primer annealing at 55°C for 30 seconds, and extension at 68°C for 1 minute; 35 cycles were run. Sample processing was carried out according to the manufacturer's recommendations. Subsequently, for sequencing the two exons in each sample, the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystem, Foster City, California, USA) was used with the same PCR primers. The reaction conditions were incubation at 96°C for 1 minute, denaturation at 96°C for 10 seconds, annealing of primers at 50°C for 5 seconds, and extension at 60°C for 4 minutes; 25 cycles were performed. The sequences were purified according to the BigDye XTerminator KIT protocol (Applied Biosystem, Foster City, California, USA), and obtained using the 3500 Genetic Analyzer (Applied Biosystem, Foster City, California, USA).

RT-PCR for ERG, EZH2, NKX3.1, SPINK-1, and UBC gene expression

RT-PCR was carried out in a single step using the qScript 1-Step SYBR Green qRT-PCR kit (Quanta Biosciences, Cummings Center, Beverly, Massachusetts, USA), following the manufacturer's instructions. In sum, 10 µl of the One-Step SYBR Green Master Mix (2X), 0.4 µl of qScript One-Step RT, 2 µl of each pair of primers, 4 µl of RNA sample (10 ng/µl), and 1.6 µl of nuclease-free water were used for qRT-PCR. The primers used for each gene and respective amplicon length, concentration, and annealing temperature are described in Table 1 (8,27-30). The protocol starts with reverse transcription at 50°C for 10 minutes during one cycle; then inactivation of the reverse transcriptase at 95°C for 5 minutes during a cycle; subsequently, 40 cycles of denaturation at 95°C for 10 seconds and primer annealing for 20 seconds; and, finally, an extension at 72°C for 1 minute. A LightCycler® 480 thermal cycler (Roche Life Science, Penzberg, Upper Bavaria, Germany) was used for RT-qPCR. To obtain the melting curves, the PCR products were heated at 95°C for 5 seconds, at 65°C for 1 minute, at 97°C continuously with 5-10 adq/°C, and then cooled at 40°C for 10 seconds. UBC was used as a reference expression gene. All tests were carried out in duplicate. All 103 study samples showed similar values of UBC gene expression: cross points (CPs) median=29.61, mean= 29.68.

Elaboration of the standard curves of ERG, EZH2, NKX3.1, SPINK-1, and UBC gene expression

The standard curves were prepared with PCR products obtained from a FFPE sample of adjacent non-tumor prostate tissue, purified from agarose gel using the Illustra GFX PCR DNA and Gel Band Purification commercial kit (GE Healthcare, Little Chalfont, United Kingdom), and quantified using the Qubit® 2.0 fluorometer for HS DNA (Invitrogen, Waltham, Massachusetts, USA). The number of copies of the cDNA was calculated using the URI Genomics & Sequencing Center's online calculator (2004) (<http://cels.uri.edu/gsc/cndna.html>). Seven serial dilutions were made in duplicate, between 1×10^1 and 1×10^7 copy number. The efficiency of the amplification (E) was calculated as: $E = 10^{-1/\text{slope}}$. E values were: 1.903 for UBC, 1.873 for EZH2, 2.003 for NKX3.1, and 1.904 for SPINK-1; standard deviations were <0.2 with respect to the duplicates. The validity of the points, efficiency calculations, and errors in the equipment were verified by the method of the second derivative.

Relative expression of ERG, EZH2, NKX3.1, and SPINK-1

With the E values found in the standard curves, the relative expression of each sample was calculated using the method of Pfaffl (31), which uses the following equation:

$$\text{Relative expression} = \left[\frac{E_{ref}^{CP_{ref} \text{ Sample}}}{E_{targ}^{CP_{targ} \text{ Sample}}} \right] \div \frac{E_{ref}^{CP_{ref} \text{ Calibrator}}}{E_{targ}^{CP_{targ} \text{ Calibrator}}}, \text{ where } E_{target}$$

is the E of target gene; E_{ref} is the E of a reference gene (UBC); CP_{target} is the cross point of the target gene; CP_{ref} is the CP of the reference gene (UBC).

Calibrator is the adjacent non-tumor prostate tissue for each patient. For the analysis of the relative expression of ERG, EZH2, NKX3.1, and SPINK-1, the values obtained were converted into three categories: highly expressed (>1.20), normally expressed (between 0.80-1.20), and lowly expressed (<0.80).

Determination of PCa molecular subtypes

PCa molecular subtypes were assigned as follows: ERG(+) subtype, when TMPRSS2-ERG fusion is present and/or ERG mRNA is highly expressed; SPOP subtype, in the presence of mutations in the SPOP gene; EZH2 subtype, when EZH2 mRNA is highly expressed; NKX3.1 subtype, when NKX3.1 mRNA is lowly expressed; and SPINK-1 subtype, when SPINK-1 mRNA is highly expressed.

These subtypes were assigned by sample and by patient; to consider that a patient is positive for a subtype, at least one sample must have this subtype. A molecular pattern is the combination of the presence of all subtypes analyzed in samples and patient.

Concordance patterns of the presence/absence of the TMPRSS2-ERG fusion, SPOP mutations, and gene expression of ERG, EZH2, NKX3.1, and SPINK-1 in prostatic foci of patients

Concordance for the presence/absence of the TMPRSS2-ERG fusion, SPOP mutations, and gene expression of ERG, EZH2, NKX3.1, and SPINK-1 were assessed separately and together, aiming to establish the clonal origin of all prostatic foci (HGPIN and PCa) analyzed by patient. LN were not used for concordance analyses due to being extraprostatic lesions. Figure 1A-E graphically describe the five concordance patterns found: two concordant patterns (patterns A and B), two partially concordant patterns (patterns C and D), and one discordant pattern (pattern E).

Data analysis

Clinicopathological characteristics were analyzed as continuous variables: age of diagnosis (years), prediagnosis PSA (ng/ml), tumor percentage in RP, index diameter (cm), and time to BCR (months), as well as categorical variables: Gleason score, Gleason grade group at RP, margin status in RP, perineural invasion, lymphovascular invasion, primary tumor size and/or amount of spread into nearby structures (pT), lymph node compromise (pN), and 5-years BCR. Molecular characteristics were analyzed as categorical variables: detection of the TMPRSS2-ERG fusion, SPOP mutations, and relative expression of ERG, EZH2, NKX3.1, and SPINK-1. The values obtained were converted into three categories: highly expressed (>1.20), normally expressed (between 0.80-1.20), and lowly expressed (<0.80). The distribution of the relative expression of ERG, EZH2, NKX3.1, and SPINK-1 was also analyzed.

The assumption of normality in the quantitative variables was validated using the Shapiro-Wilk test. For the variables that fulfilled the assumption, the mean (standard deviation) was used as a descriptive measure, and for those that did not, the median (interquartile range). For comparison between groups, the t-Student test for independent samples was used in the case of variables with normal distribution, and the Wilcoxon signed-rank test otherwise. In the case of categorical variables, absolute and relative frequencies were used as descriptive measures. To determine whether there was a statistically significant association between the categorical variables, the Fisher exact test was used, given the sample size.

The Shapiro-Wilk test showed that the data of relative gene expression did not comply with the criteria of normality in any histopathological group. Although the data complied with normality when using log₂ transformation, they did not fulfill the assumption of homogeneity of variances according to Levene's test. The significant differences were verified by groups with the molecular alterations evaluated in the study, using the Kruskal-Wallis nonparametric test for categorical variables and the Fisher test for quantitative variables. According to the results obtained with the Kruskal-Wallis test, a paired comparison was subsequently applied between histopathological groups and clinical-histopathological characteristics, and the results were considered significant with a $p < 0.05$, using the Mann-Whitney U-test. The results were visualized according to the distribution of median values and data dispersion, plotted on a logarithmic scale log₂ with a box-and-whisker diagram. All analyses were carried out with the statistical software SPSS v18.

Results

Clinicopathological characteristics of the cohort

One hundred and three samples from 20 PCa patients were analyzed: foci of adjacent non-tumor prostate tissue (n=20), HGPIN (n=17), GL3 (n=23), GL4 (n=23), GL5 (n=7), and LN (n=13). The main clinicopathological characteristics of the 20 patients are summarized in Table 2.

Determination of molecular subtypes and molecular patterns by sample and by patient

The molecular subtypes analyzed in this study were: ERG(+), EZH2, NKX3.1, and SPINK-1. The ERG(+) subtype includes presence of the TMPRSS2-ERG fusion and/or highly expressed ERG mRNA. Since no mutations were found in the SPOP gene, this was not considered for analysis. Eighty-three samples (70 prostatic foci and 13 LN) obtained from 20 patients were analyzed. All the results detailed by patient and sample are shown in Table 1S.

Figure 2A shows the presence of different molecular subtypes by samples and divided by the presence/absence of the ERG(+) subtype, which is considered the most frequent and most studied finding in PCa. The presence of each subtype in the samples analyzed range from 41 to 60%. Fifty samples (60%) were considered ERG(+), 45 (54%) were EZH2, 38 (46%) were NKX3.1, and 34 (41%) were SPINK-1. In 7 samples, a nonmolecular subtype was identified. The presence of each subtype in the samples was predominantly in combination with ERG(+) ranging from 60 to 74%: 27 (60%) of EZH2 samples, 27 (71%) of NKX3.1 samples, and 25 (74%) of SPINK-1 samples. ERG(+) alone was only found in one sample (Figure 2A).

To establish patient subtypes, a patient is considered positive for a subtype when they have at least one sample where the subtype is present. The presence of each subtype in the samples analyzed ranges from 75 to 85%; 17 patients (85%) were ERG(+), 17 (85%) were EZH2, 15 (75%) were NKX3.1, and 15 (75%) were SPINK-1. The presence of each subtype in patients was predominantly in combination with ERG(+), ranging from 82 to 93%; ERG(+) alone was not found in any patient (Figure 2B).

When considering the four subtypes simultaneously, we identified 15 molecular patterns in the samples, with the frequency of these patterns ranging from 1 to 13%. Eight molecular patterns were ERG(+) and they were found in 50 samples (60%), while 7 were ERG(-) in 33 samples (40%) (Figures 2A,C). The simultaneous presence of ERG(+)/EZH2/NKX3.1, which are related among them, were found in 11 samples (13%): 8 samples (9,6%) with this pattern and 3 samples (3,6%) also including SPINK-1 (ERG(+)/EZH2/NKX3.1/SPINK-1). The most frequent pattern was EZH2 alone (13%), followed by ERG(+)/EZH2/SPINK-1 and ERG(+)/NKX3.1 (11% each).

We found 8 different molecular patterns in patients, with the frequency of these patterns ranging from 5 to 40%. Five molecular patterns were ERG(+) and they were found in 17 patients (85%), while 3 patterns were ERG(-) in 3 patients (15%) (Figures 2B,C). The simultaneous presence of ERG(+)/EZH2/NKX3.1, which are related among them, were found in 11 patients (55%): 3 samples (15%) with this pattern and 8 (40%) including also SPINK-1 (ERG(+)/EZH2/NKX3.1/SPINK-1); the latter was the most frequent pattern in patients. The 3 molecular patterns with ERG(-) are present in only one patient each (5%), but all have EZH2 subtype (Figures 2B,C).

Determination of concordance for each molecular alteration evaluated in the prostatic foci

Aiming to establish the clonal origin of the different prostatic foci analyzed in each of the 20 patients, the concordance of molecular characteristics was assessed with respect to the presence or absence of the TMPRSS2-ERG fusion and the expression of ERG, EZH2, NKX3.1, and SPINK-1 genes in 70 prostatic foci (17 HGPIN and 53 PCa) (Figure 1, Table 1S). LN samples were excluded from these analyses.

Figures 1A-E show the five possible concordance patterns found: two concordant (Figures A,B), two partially concordant (Figures C,D), and one discordant (Figure E). The concordant and partially concordant patterns are compatible with monoclonal origin because discordances related with lesion progression findings in the partially concordant pattern are taken as subclonal events; one discordant pattern is compatible with multiclonal origin. Figure 1F shows concordance patterns with respect to each molecular alteration evaluated by patient. The concordant pattern was the predominant state for all evaluated alterations (40 to 55%), except for SPINK-1 expression, where the partially concordant state was the predominant pattern in 45%. The partially concordant pattern ranged between 20-45%, and total molecular concordance, which includes concordant and partially concordant cases, was the most common pattern for all molecular characteristics analyzed, ranging between 70-90% (90% for ERG, 85% for fusion status, NKX3.1 and SPINK-1 expression, and 70% for EZH2 expression), which is compatible with monoclonal origin. Discordance was the less common pattern (10-30%), which would support a multiclonal origin (Figure 1F).

Among the cases with concordant pattern, the most common was overexpression of ERG (73%), EZH2 (70%), and NKX3.1 (56%) (pattern A), as well as the absence of fusion (82%) and underexpression of SPINK-1 (75%) (pattern B). The most common pattern for partially concordant cases was pattern C: acquisition of the fusion (100%) and progressive increase in gene expression of ERG (57%), EZH2 (75%), NKX3.1 (63%), and SPINK-1 (67%) (Figure 1F).

Determination of concordance in prostatic foci for all molecular alterations evaluated in each patient

After analyzing concordance for each molecular characteristic separately, we examine concordance for all molecular characteristics analyzed in this study simultaneously in prostatic foci (n=70) (Table 1S). Three out of 20 patients were concordant (15%), 6 patients were partially concordant (30%), and 11 patients were discordant (55%), which shows that heterogeneity and the possibility of multiclonal origin increases from 15 to 30% by molecular characteristic analyzed separately to 55% when the five molecular characteristics are analyzed simultaneously.

Intratumoral concordance was also analyzed in the same GL score foci in 6 patients who had samples with more than one focus with the same GL. Although the foci of the same GL in patients 1, 2, and 9 showed concordance, there were differences in patients 7, 10, and 8 in the presence of fusion and expression of ERG, NKX3.1, and SPINK-1, which evidences intratumoral heterogeneity (Table 1S).

Histopathological and clinical significance of the presence of ERG, EZH2, SPINK-1, and NKX3.1 subtypes

Figure 3A shows the frequency of molecular subtypes by histopathological group. All analyzed subtypes were detected starting from the preneoplastic lesion (HGPIN) in a range of 47-53%, but the SPINK-1 subtype was in a low percentage (18%). In LN, the range was between 54-92%; the presence of each subtype increased compared to HGPIN, where the most frequent subtypes were ERG and NKX3.1 (92% each). It is very clear that ERG overexpression increases in frequency from HGPIN to LN (53 to 92%).

Figure 3B shows the frequency of 15 molecular patterns by histopathological group; a very interesting finding is that pattern heterogeneity decreases with lesion progression; while 7 molecular patterns were found in HGPIN, 9 in GL3, and 11 in GL4, only 5 were found in GL5 and 6 in LN. All LN and GL5 have a molecular pattern, but in 7 foci (4 HGPIN (24%), 1 GL3 (4%), and 2 GL4 (9%)) no pattern was found, which probably reflects different patterns not identified with the molecular subtypes analyzed here and increased heterogeneity in these histopathological groups. The molecular pattern with the 4 PCa subtypes (ERG+)/EZH2/NKX3.1/SPINK-1) was detected only in LN, in 23% of them. The constant in all but one LN molecular pattern was the NKX3.1 subtype.

We also looked for association between relative gene expression of ERG, EZH2, SPINK-1, and NKX3.1, and histopathological group. The distribution of the relative gene expression of ERG, EZH2, and SPINK-1 showed no significant differences between histopathological groups, but the relative gene expression of NKX3.1 showed a statistically significant difference between groups ($p < 0.001$). A lower expression in LN samples with respect to all the other groups (LN vs. HGPIN, $p = 0.0015$; LN vs. GL3, $p < 0.001$, LN vs. GL4, $p < 0.001$; LN vs. GL5, $p = 0.0236$), and between GL5 compared to GL4 samples ($p = 0.0071$), indicates that there is an inversely proportional association between PCa progression (HGPIN, GL3, GL4 vs. GL5, LN) and the expression levels of NKX3.1 (Figure 4).

An analysis was made by comparing clinicopathological characteristics and molecular subtypes and concordance patterns in our 20 patients by BCR time, categorized as ≤ 3 months and > 3 months (Table 2). A statistically significant association was found between BCR time with tumoral percentage ($p = 0.040$), pT ($p = 0.049$), pN ($p = 0.029$), but no association was found between BCR time and molecular subtypes or molecular concordance. A comparison of the clinicopathological characteristics by molecular subtype with ERG, EZH2, NKX3.1, and SPINK-1 subtypes did not find any statistically significant association (Supplementary Table 2S-5S).

Discussion

Although PCa is a multifocal disease, it is not clear to what extent molecular heterogeneity between these foci is a determining factor in the natural history of the disease or its role to establish a prognostic molecular biomarker of this disease. In this study, five molecular subtypes associated with PCa were analyzed: ERG(+), EZH2, NKX3.1, SPINK-1, and SPOP mutation, to establish their association with different degrees of lesion and regional metastasis in LN in 20 PCa patients with poor prognosis, as well as the clonal origin of the foci in these patients. The ERG(+) subtype includes TMPRSS2-ERG fusion and ERG overexpression. Since no mutations were found in the SPOP gene, which may be due to sample size, this was not considered for analysis.

Initially, when determining the presence of molecular subtypes, the ERG(+) subtype was found in 60% of the samples and in 85% of the patients. The frequency of this subtype reported before for PCa in Caucasian men ranges between 40-66% (3, 32-35), similar to what we found in the samples, but our frequency is higher in patients, probably because our study includes HGPIN and LN, and various GL foci that the other studies did not include.

We identified the reported EZH2 subtype in 54% of the samples and 85% of the patients, which was more frequently associated with the ERG(+) subtype (60% in samples and 82% in patients). All patients with ERG(-) (18%; n=3) had EZH2 subtype. In the literature, the EZH2/ERG-independent subtype has been reported in 30% of the cases by immunohistochemistry (IHC) (36), which is similar to what was found in our samples (40%, and a decrease to 18% in patients), but it is important to remember that our study analyzes many foci per patients, which, given the heterogeneity of PCa, increases the probability to find a focus ERG(+).

With respect to the NKX3.1 subtype, 46% of the foci and 75% of the patients have this subtype, and the presence of NKX3.1 subtype is more frequent in combination with ERG(+), found in 71% of the foci and 93% of the patients. We have not found a frequency of this PCa subtype in the literature to compare our results. Given the relationship among the expressions of ERG, EZH2, and NKX3.1, the simultaneous presence of these three subtypes (ERG+)/EZH2/NKX3.1) was found in 11 samples (13%) and in 11 patients (55%). It is important to highlight that these subtypes are also found separately, which indicates that the expression of these genes is more complex than only the reciprocal relationship among them (6).

The SPINK-1 subtype was found in 41% of the samples and 75% of the patients. Although most studies reported that the SPINK-1 subtype is found in ERG(-) PCa, in approximately 10-15% of the cases (9,15,16), in this study, the presence of SPINK-1 was predominantly in combination with ERG(+) (74% of SPINK-1 samples and 87% of the patients). Although SPOP mutation has been linked to SPINK-1 overexpression in various studies, mutations were not found in any focus.

Considering the simultaneous presence/absence of the four subtypes, we identified more molecular patterns (15 patterns) in the samples than in patients (8 patterns). It is very interesting that the most frequent pattern for samples was an ERG(-) pattern, EZH2/ERG-independent, which was found in 11 samples (13%), although this number is not too high and there are 3 more patterns that range between 10-11%. In patients, the most common pattern (40%) was the simultaneous presence of the four subtypes (ERG+)/EZH2/NKX3.1/SPINK-1), which is a noteworthy finding.

Similarly, a comparative analysis was carried out between clinicopathological characteristics and molecular subtypes in the 20 study patients by BCR time. No association was found between BCR time in patients, categorized as ≤ 3 months and > 3 months, and molecular subtypes or molecular concordance. We also compared clinicopathological characteristics and the molecular subtype with ERG, EZH2, SPINK-1 subtypes, which did not show any statistically significant association.

It is not surprising that the frequency of the analyzed subtypes increases when analyzed by patient (75-85%) with respect to foci (41-60%), while the presence of molecular patterns decreases. This is because, contrary to most of the literature that only take into account one focus (index tumor) (23) to establish subtypes, we consider that a subtype is present when it is detected at least in one sample of a patient, and we analyzed a range of 3-5 samples (HGPIIN, PCa foci, LN) from each patient. Molecular patterns decrease in patients because patterns with presence of more subtypes increase when we sum the presence of the subtypes in all samples analyzed by case.

The presence of the molecular alteration of a subtype in at least one sample is important because the clone that has a specific subtype can be the most aggressive one and determine prognosis and response to therapies. For this reason, taking into account more adaptative tumoral cells, what we did in LN to establish the presence of a subtype is not common, but very interesting, and increases the presence of subtypes found in patients. In 7

prostatic foci (8%), the analyzed nonmolecular subtype was identified, but there are other molecular subtypes, such as mutations in the IDH1 and FOXA1 genes, described in the TCGA study (3), that were not evaluated in this work. Interestingly, although the TCGA made a complete molecular analysis of PCa, 26% of the cases did not belong to any of the 7 subtypes described by them.

In relation to the concordance analysis of prostatic foci per patient, the concordance of the 5 molecular characteristics (TMPRSS2-ERG fusion, expression of ERG, EZH2, NKX3.1, and SPINK-1) was used to inquire about the clonal origin of PCa. When each molecular characteristic was evaluated separately, we found that total concordance was the predominant state for most of them. Previous studies have found that between 50-60% of multifocal PCa tumors are concordant in terms of the presence or absence of the TMPRSS2-ERG fusion, which is similar to the 55% found in this study (32,37). There is no record of concordance among foci regarding the expression of ERG, NKX3.1, or EZH2.

The second most frequent pattern for these four alterations was partially concordant cases (20-45%). Among these, the molecular characteristics vary in relation to disease progression, as observed in subclonal alterations acquired during tumor cell evolution. Four of the 5 molecular characteristics were present as alterations characterized as PCa subtypes that appear with disease progression (e.g., increase in ERG expression that characterizes the ERG(+) subtype of late appearance). The progressive decrease in gene expression related to PCa was less frequent, only in NKX3.1. These results are not comparable with other studies in the literature, since these studies of several foci only analyzed 2 or 3 tumor samples, without considering HGPIN, and did not present results with respect to Gleason score per focus.

Total molecular concordance, which includes concordant and partially concordant cases, was high (70-90%) for each alteration and support a monoclonal origin with subclonal events; between 15-30% of the cases show discordance, which could be associated with a multiclonal origin. When all alterations were analyzed simultaneously by patient, the degree of discordance (55%) was higher than what we found when alterations were analyzed separately (10-30%), as expected. Other studies have reported a decrease in molecular concordance as the number of analyzed molecular alterations increases, especially the whole genome or transcriptome show that the foci vary widely in the same patient (38-40). However, based on the results of our study, we cannot discard that some multifocal PCa may have a monoclonal origin, and others a multiclonal origin; that is, there is no single origin for all PCa.

It is noteworthy that there were intratumoral differences found in foci with the same GL in 3 out of 6 patients who had samples with more than one focus with the same GL, but concordance in the other 3 patients support the probability of a monoclonal origin. Wei et al. (22) showed that the correlation between different PCa foci in an individual is low. These findings highlight the importance of evaluating the role of multiclonality, which presents different alterations and subtypes in the same patient, complicating thus its application in translational oncology, since alterations other than the index tumor may be determining metastases (23,41).

LN were not considered for concordance analysis, but it is important to highlight that from the 13 LN analyzed, none was the same as the other foci analyzed, which underlines that malignant cells need other subclonal events that can form metastatic foci in LN. This is in accordance with what was said above about subclonal alterations, which must provide additional advantages for such sophisticated processes as metastasis.

Regarding the association between subtypes and histopathological group, it was found that all of them could be observed starting from HGPIN lesions, which leads to the conclusion that all the alterations evaluated here are early events in the natural history of PCa. It is worth noting that the constant in all but one LN molecular subtype were the ERG(+) and NKX3.1 subtypes (92%). The molecular pattern with the four PCa subtypes (ERG+)/EZH2/NKX3.1/SPINK-1) was detected only in LN (23%). The 13 analyzed LN had more similarity among them than with the prostatic foci of the same patient. Findings in molecular patterns are in accordance with less heterogeneity in more advanced lesions (GL5 and LN). The less intermetastatic heterogeneity found was expected if mutations in the same genes are needed to confer the same properties to a metastatic microenvironment and at a distant site.

Only the expression of the NKX3.1 gene was significantly different among the histopathological groups. Its expression decreased in more advanced lesions; therefore, the lowest levels were found in the LN group that represents regional metastasis, and this finding is consistent with other reports where NKX3.1 has been found to be downregulated in many PCa tissues; even in the most dysplastic and metastatic PCa tumors, it has been reported to be lost or completely eliminated (11,42-44).

No association was found between BCR time or clinicopathological characteristics and molecular subtypes or molecular concordance.

Conclusions

The results found in this study support the notion that, given the molecular heterogeneity of PCa, the molecular evaluation of a single focus does not provide a complete molecular picture of the disease in each patient. Although molecular subtypes can be assigned in this way, this type of analysis hinders the association of molecular subtype with prognosis and the choice of appropriate treatment. All the subtypes analyzed here are of early onset, starting from preneoplastic lesions. NKX3.1 expression level is the only molecular characteristic analyzed that shows a progression pattern.

Abbreviations

DNA: Deoxyribonucleic acid. Dx: Histopathological diagnosis. ERG(-): Low expression of ERG and absence of the TMPRSS2-ERG fusion gene. ERG(+): High expression of ERG and/or presence of the TMPRSS2-ERG fusion gene. ERG: ETS-related gene. ETS: Erythroblast transformation-specific family. EZH2: Enhancer of zeste 2. F: Fusion; (+) indicates the cases where the fusion is present and (-) where it is not. GL: Gleason pattern. LN: Lymph node. HGPIN: High-Grade Prostatic Intraepithelial Neoplasia. HGPIN: High-grade prostatic intraepithelial neoplasia. MYC: v-myc avian myelocytomatosis viral oncogene homolog. NKX3-1: NK3 Homeobox 1. P: Patient. PCa: Prostate cancer. RNA: Ribonucleic acid. RT-PCR: Reverse transcription-polymerase chain reaction. RT-qPCR: Real-time quantitative polymerase chain reaction. UBC: Ubiquitin C.

Declarations

Ethics approval and consent to participate.

The present research was approved by the Ethics Committee of the Instituto Nacional de Cancerología, Bogotá, Colombia. . Tissues were obtained as part of routine care and was not obtained specifically for this study and data

used in this study was anonymised before. The Ethics Committee of the Instituto Nacional de Cancerología, (Bogotá, Colombia) waived the need of informed consent for this study. All methods were carried out in accordance with relevant guidelines and regulations at national and international level.

Consent for publication: N/A.

Availability of data and materials: All data generated or analysed during this study are included in this published article.

Competing interests: Authors declare no conflicts of interest in this study.

Acknowledgments and funding: We thank the Instituto Nacional de Cancerología (Bogotá, Colombia) for its financial support, and the Universidad Nacional de Colombia for the academic contribution, as well as Ricardo Sánchez Pedraza and Marcela Núñez for collaboration with statistical analysis.

Author contributions: The authors have participated in the research that resulted in this work, as well as in the writing, reviewing, and final approval of this manuscript. Yenifer Yamile Segura-Moreno and Martha Lucía Serrano wrote the main manuscript text and prepared the figures and tables.

References

1. International Agency for Research on Cancer-World Health Organization. Cancer Today [Internet]. Lyon: IARC; 2020 [cited 2021 Aug 31]. Available from: <https://gco.iarc.fr/today/home>.
2. D'Amico AV. Risk-based management of prostate cancer. *N Engl J Med*. 2011;365(2):169-71. doi:10.1056/NEJMe1103829.
3. Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. *Cell*. 2015;163(4):1011-25. doi:10.1016/j.cell.2015.10.025.
4. Tomlins SA, Alshalalfa M, Davicioni E, Erho N, Yousefi K, Zhao S, et al. Characterization of 1577 primary prostate cancers reveals novel biological and clinicopathologic insights into molecular subtypes. *Eur Urol*. 2015;68(4):555-67. doi:10.1016/j.eururo.2015.04.033.
5. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science*. 2005;310(5748):644-8. doi:10.1126/science.1117679.
6. Kunderfranco P, Mello-Grand M, Cangemi R, Pellini S, Mensah A, Albertini V, et al. ETS transcription factors control transcription of EZH2 and epigenetic silencing of the tumor suppressor gene NKX3.1 in prostate cancer. *PLoS One*. 2010;5(5):e10547. doi:10.1371/journal.pone.0010547.
7. Thangapazham R, Saenz F, Katta S, Mohamed AA, Tan SH, Petrovics G, et al. Loss of the NKX3.1 tumorsuppressor promotes the TMPRSS2-ERG fusion gene expression in prostate cancer. *BMC Cancer*. 2014;14:16. doi:10.1186/1471-2407-14-16.
8. Paju A, Hotakainen K, Cao Y, Laurila T, Gadaleanu V, Hemminki A, et al. Increased expression of tumor-associated trypsin inhibitor, TATI, in prostate cancer and in androgen-independent 22Rv1 cells. *Eur Urol*.

2007;52(6):1670-9. doi:10.1016/j.eururo.2007.01.096.

9. Tomlins SA, Rhodes DR, Yu J, Varambally S, Mehra R, Perner S, et al. The role of SPINK1 in ETS rearrangement-negative prostate cancers. *Cancer Cell*. 2008;13(6):519-28. doi:10.1016/j.ccr.2008.04.016.
10. Duan S, Pagano M. SPOP Mutations or ERG Rearrangements Result in Enhanced Levels of ERG to Promote Cell Invasion in Prostate Cancer. *Mol Cell*. 2015;59(6):883-4. doi:10.1016/j.molcel.2015.09.003.
11. Lapointe J, Li C, Giacomini CP, Salari K, Huang S, Wang P, et al. Genomic profiling reveals alternative genetic pathways of prostate tumorigenesis. *Cancer Res*. 2007;67(18):8504-10. doi:10.1158/0008-5472.CAN-07-0673.
12. Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature*. 2002;419(6907):624-9. doi:10.1038/nature01075.
13. Bowen C, Bubendorf L, Voeller HJ, Slack R, Willi N, Sauter G, et al. Loss of NKX3.1 expression in human prostate cancers correlates with tumor progression. *Cancer Res*. 2000;60(21):6111-5. PMID: 11085535.
14. Taris M, Irani J, Blanchet P, Multigner L, Cathelineau X, Fromont G. ERG expression in prostate cancer: the prognostic paradox. *Prostate*. 2014;74(15):1481-7. doi:10.1002/pros.22863.
15. Grupp K, Diebel F, Sirma H, Simon R, Breitmeyer K, Steurer S, et al. SPINK1 expression is tightly linked to 6q15- and 5q21-deleted ERG-fusion negative prostate cancers but unrelated to PSA recurrence. *Prostate*. 2013;73(15):1690-8. doi:10.1002/pros.22707.
16. Flavin R, Pettersson A, Hendrickson WK, Fiorentino M, Finn S, Kunz L, et al. SPINK1 protein expression and prostate cancer progression. *Clin Cancer Res*. 2014;20(18):4904-11. doi:10.1158/1078-0432.CCR-13-1341.
17. García-Flores M, Casanova-Salas I, Rubio-Briones J, Calatrava A, Domínguez-Escrig J, Rubio L, et al. Clinico-pathological significance of the molecular alterations of the SPOP gene in prostate cancer. *Eur J Cancer*. 2014;50(17):2994-3002. doi:10.1016/j.ejca.2014.08.009.
18. Liu D, Takhar M, Alshalalfa M, Erho N, Shoag J, Jenkins RB, et al. Impact of the SPOP Mutant Subtype on the Interpretation of Clinical Parameters in Prostate Cancer. *JCO Precis Oncol*. 2018;2018:PO.18.00036. doi:10.1200/PO.18.00036.
19. Shoag J, Liu D, Ma X, Oromendia C, Christos P, Ballman K, et al. Prognostic value of the SPOP mutant genomic subclass in prostate cancer. *Urol Oncol*. 2020;38(5):418-22. doi:10.1016/j.urolonc.2020.02.011.
20. Lippolis G, Edsjö A, Stenman UH, Bjartell A. A high-density tissue microarray from patients with clinically localized prostate cancer reveals ERG and TATI exclusivity in tumor cells. *Prostate Cancer Prostatic Dis*. 2013;16(2):145-50. doi:10.1038/pcan.2013.7.
21. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA, et al. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. *Am J Surg Pathol*. 2016;40(2):244-52. doi:10.1097/PAS.0000000000000530.

22. Wei L, Wang J, Lampert E, Schlanger S, DePriest AD, Hu Q, et al. Intratumoral and Intertumoral Genomic Heterogeneity of Multifocal Localized Prostate Cancer Impacts Molecular Classifications and Genomic Prognosticators. *Eur Urol.* 2017;71(2):183-92. doi:10.1016/j.eururo.2016.07.008.
23. Huang CC, Deng FM, Kong MX, Ren Q, Melamed J, Zhou M. Re-evaluating the concept of “dominant/index tumor nodule” in multifocal prostate cancer. *Virchows Arch.* 2014;464(5):589-94. doi:10.1007/s00428-014-1557-y.
24. Wise AM, Stamey TA, McNeal JE, Clayton JL. Morphologic and clinical significance of multifocal prostate cancers in radical prostatectomy specimens. *Urology.* 2002;60(2):264-9. doi:10.1016/s0090-4295(02)01728-4.
25. Cheng L, Song SY, Pretlow TG, Abdul-Karim FW, Kung HJ, Dawson DV, et al. Evidence of independent origin of multiple tumors from patients with prostate cancer. *J Natl Cancer Inst.* 1998;90(3):233-7. doi:10.1093/jnci/90.3.233.
26. Zhang P, Wang D, Zhao Y, Ren S, Gao K, Ye Z, et al. Intrinsic BET inhibitor resistance in SPOP-mutated prostate cancer is mediated by BET protein stabilization and AKT-mTORC1 activation. *Nat Med.* 2017;23(9):1055-62. doi:10.1038/nm.4379.
27. Saramäki OR, Harjula AE, Martikainen PM, Vessella RL, Tammela TL, Visakorpi T. TMPRSS2:ERG fusion identifies a subgroup of prostate cancers with a favorable prognosis. *Clin Cancer Res.* 2008;14(11):3395-400. doi:10.1158/1078-0432.CCR-07-2051.
28. Börno ST, Fischer A, Kerick M, Fälth M, Laible M, Brase JC, et al. Genome-wide DNA methylation events in TMPRSS2-ERG fusion-negative prostate cancers implicate an EZH2-dependent mechanism with miR-26a hypermethylation. *Cancer Discov.* 2012;2(11):1024-35. doi:10.1158/2159-8290.CD-12-0041.
29. Richter E, Masuda K, Cook C, Ehrich M, Tadese AY, Li H, et al. A role for DNA methylation in regulating the growth suppressor PMEPA1 gene in prostate cancer. *Epigenetics.* 2007;2(2):100-9. doi:10.4161/epi.2.2.4611.
30. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 2002;3(7):RESEARCH0034. doi:10.1186/gb-2002-3-7-research0034.
31. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 2001;29(9):e45. doi:10.1093/nar/29.9.e45.
32. Barry M, Perner S, Demichelis F, Rubin MA. TMPRSS2-ERG fusion heterogeneity in multifocal prostate cancer: clinical and biologic implications. *Urology.* 2007;70(4):630-3. doi:10.1016/j.urology.2007.08.032.
33. Furusato B, Tan SH, Young D, Dobi A, Sun C, Mohamed AA, et al. ERG oncoprotein expression in prostate cancer: clonal progression of ERG-positive tumor cells and potential for ERG-based stratification. *Prostate Cancer Prostatic Dis.* 2010;13(3):228-37. doi:10.1038/pcan.2010.23.
34. Furusato B, Gao CL, Ravindranath L, Chen Y, Cullen J, McLeod DG, et al. Mapping of TMPRSS2-ERG fusions in the context of multi-focal prostate cancer. *Mod Pathol.* 2008;21(2):67-75. doi:10.1038/modpathol.3800981.
35. Fontugne J, Davis K, Palanisamy N, Udager A, Mehra R, McDaniel AS, et al. Clonal evaluation of prostate cancer foci in biopsies with discontinuous tumor involvement by dual ERG/SPINK1 immunohistochemistry. *Mod*

Pathol. 2016;29(2):157-65. doi:10.1038/modpathol.2015.148.

36. Melling N, Thomsen E, Tsourlakis MC, Kluth M, Hube-Magg C, Minner S, et al. Overexpression of enhancer of zeste homolog 2 (EZH2) characterizes an aggressive subset of prostate cancers and predicts patient prognosis independently from pre- and postoperatively assessed clinicopathological parameters. *Carcinogenesis*. 2015;36(11):1333-40. doi:10.1093/carcin/bgv137.
37. Mehra R, Han B, Tomlins SA, Wang L, Menon A, Wasco MJ, et al. Heterogeneity of TMPRSS2 gene rearrangements in multifocal prostate adenocarcinoma: molecular evidence for an independent group of diseases. *Cancer Res*. 2007;67(17):7991-5. doi:10.1158/0008-5472.CAN-07-2043.
38. Wei Y, Wu J, Gu W, Wang J, Lin G, Qin X, et al. Prognostic Value of Germline DNA Repair Gene Mutations in De Novo Metastatic and Castration-Sensitive Prostate Cancer. *Oncologist*. 2020;25(7):e1042-e50. doi:10.1634/theoncologist.2019-0495.
39. Cooper CS, Eeles R, Wedge DC, Van Loo P, Gundem G, Alexandrov LB, et al. Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat Genet*. 2015;47(4):367-72. doi:10.1038/ng.3221.
40. Boutros PC, Fraser M, Harding NJ, de Borja R, Trudel D, Lalonde E, et al. Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat Genet*. 2015;47(7):736-45. doi:10.1038/ng.3315.
41. Løvlf M, Zhao S, Axcróna U, Johannessen B, Bakken AC, Carm KT, et al. Multifocal Primary Prostate Cancer Exhibits High Degree of Genomic Heterogeneity. *Eur Urol*. 2019;75(3):498-505. doi:10.1016/j.eururo.2018.08.009.
42. Gurel B, Ali TZ, Montgomery EA, Begum S, Hicks J, Goggins M, et al. NKX3.1 as a marker of prostatic origin in metastatic tumors. *Am J Surg Pathol*. 2010;34(8):1097-105. doi:10.1097/PAS.0b013e3181e6cbf3.
43. Akamatsu S, Takata R, Ashikawa K, Hosono N, Kamatani N, Fujioka T, et al. A functional variant in NKX3.1 associated with prostate cancer susceptibility down-regulates NKX3.1 expression. *Hum Mol Genet*. 2010;19(21):4265-72. doi:10.1093/hmg/ddq350.
44. Bethel CR, Faith D, Li X, Guan B, Hicks JL, Lan F, et al. Decreased NKX3.1 protein expression in focal prostatic atrophy, prostatic intraepithelial neoplasia, and adenocarcinoma: association with gleason score and chromosome 8p deletion. *Cancer Res*. 2006;66(22):10683-90. doi:10.1158/0008-5472.CAN-06-0963.

Tables

Table 1. Primers used in the study for RT-PCR, qRT-PCR, and PCR assays with their respective standardized optimal conditions

Gene	Primers	Amplicon length (bp)	Annealing temperature (°C)	Primer concentration (nM)	Reference
TMPRSS2-ERG (exon 1: exon 4)	F: TAGGCGCGAGCTAAGCAGGAG R: GTAGGCACACTCAAACAACGACTGG	125	61	100	(5)
SPOP-6	F: ACCCATAGCTTTGGTTTCTTCTCCC R: TATCTGTTTTGGACAGGTGTTTGCG	170	55	200	(26)
SPOP-7	F: ACTCATCAGATCTGGGAACTGC R: AGTTGTGGCTTTGATCTGGTT	240	55	200	(26)
ERG	F: TCTTGGACCAACAAGTAGCC R: GTCGGGATCCGTCATCTTG	151	63	150	(27)
EZH2	F: TGTGGATACTCCTCCAAGGAA R: GAGGAGCCGTCCTTTTTCA	90	55	250	(28)
NKX3.1	F: CCGAGACGCTGGCAGAGA R: GAAGGGCGCCTGAAGTGTT	81	55	200	(29)
SPINK-1	F: CAAAAATCTGGGCCTTGCTGAGAAC R: AGGCCTCGCGGTGACCTGAT	62	55	200	(8)
UBC	F: ATTTGGGTCGCGGTTCTTG R: TGCCTTGACATTCTCGATGGT	133	55	150	(30)

F: Forward; R: Reverse.

Table 2. Clinicopathological and molecular characteristics of the full cohort, classified by BCR time, categorized as ≤ 3 months and > 3 months

Characteristic	Full cohort (n=20)	BCR time*		p-value
		≤ 3 months (n=9)	> 3 months (n=8)	
Age of diagnosis (years)				
Median [IQR]	66.0 [12.5]	65.0 [12.0]	63.0 [21.5]	1.00
Prediagnosis_PSA				
Median [IQR]	14.3 [12.7]	14.3 [9.75]	11.5 [11.8]	0.681
Gleason score, n (%)				
7	19	8 (50.0)	8 (50.0)	1.00
8	1	1 (100)	0 (0.00)	
Gleason grade group at RP (%)				
G2	5	2 (40.0)	3 (60.0)	1.00
G3	14	6 (54.5)	5 (45.5)	
G4	1	1 (100)	0 (0.00)	
Tumor percentage (%)				
Mean ± SD	39.3 ± 19.2	49.4 ± 21.6	29.9 ± 12.5	0.040**
Index diameter				
Mean ± SD	2.00 ± 0.48	2.19 ± 0.55	1.79 ± 0.39	0.110
Margin status in RP, n (%)				
Negative	4	1 (33.3)	2 (66.4)	0.576
Positive	16	8 (57.2)	6 (42.8)	
Perineural invasion, n (%)				
Intraprostatic	8	3 (50.0)	3 (50.0)	1.00
Intraprostatic + extraprostatic	12	6 (54.5)	5 (45.5)	
Lymphovascular invasión, n (%)				
No	10	2 (25.0)	6 (75.0)	0.131
Yes	9	6 (75.0)	2 (25.0)	
Unknown	1	1 (100)	0 (0.00)	
Pathologic stage.pT, n (%)				
2c-3a	7	1 (16.7)	5 (83.3)	0.049**
3b	13	8 (72.7)	3 (27.3)	

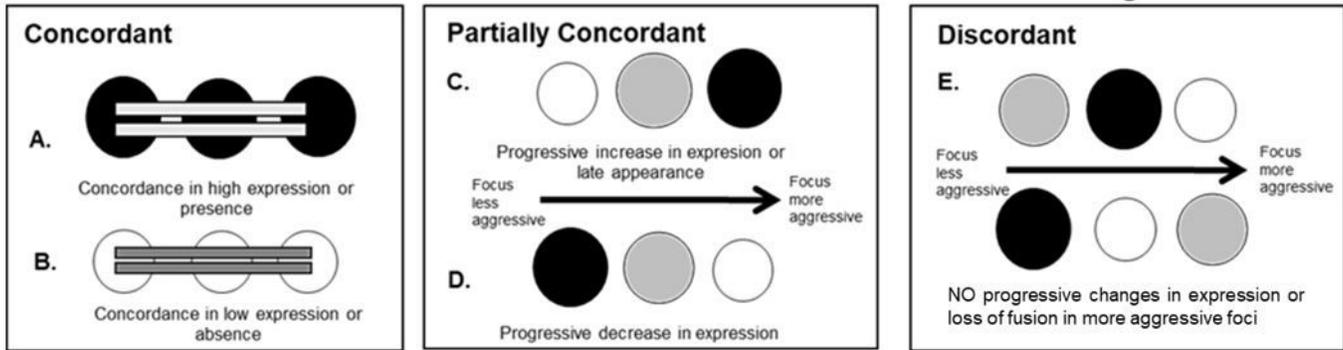
Pathologic stage.pN, n (%)				
0	4	0 (0.00)	4 (100)	0.029**
1	16	9 (69.2)	4 (30.8)	
Fusion, n (%)				
Positive	10	4 (50.0)	4 (50.0)	1.00
Negative	10	5 (55.6)	4 (44.4)	
ERG, n (%)				
Positive	14	6 (54.6)	5 (45.4)	1.00
Negative	6	3 (50.0)	3 (50.0)	
EZH2, n (%)				
Positive	15	5 (41.7)	7 (58.3)	0.294
Negative	5	4 (80.0)	1 (20.0)	
NKX3.1, n (%)				
Positive	9	4 (57.2)	3 (42.8)	1.00
Negative	11	5 (50.0)	5 (50.0)	
SPINK-1, n (%)				
Positive	12	4 (36.4)	7 (63.6)	0.131
Negative	8	5 (83.3)	1 (16.7)	
Concordance pattern, n (%)				
Concordant	3	2 (66.7)	1 (33.3)	1.00
Partially concordant	6	2 (40.0)	3 (60.0)	
Discordant	11	5 (55.6)	4 (44.4)	

IQR: Interquartile range; SD: Standard deviation; RP: Radical prostatectomy. *: In 3 cases out of 20, we could not establish BCR. **: < 0.05.

Figures

Compatible with monoclonal origin

Compatible with multiclonal origin



F.

Concordance patterns	Molecular characteristics				
	TMPRSS2-ERG fusion	Gene expression			
		ERG	EZH2	NKX3.1	SPINK-1
Concordant	11 (55%)	11 (55%)	10 (50%)	9 (45%)	8 (40%)
Pattern A	2 (18%)	8 (73%)	7 (70%)	5 (56%)	2 (25%)
Pattern B	9 (82%)	3 (27%)	3 (30%)	4 (44%)	6 (75%)
Partially concordant	6 (30%)	7 (35%)	4 (20%)	8 (40%)	9 (45%)
Pattern C	6 (100%)	4 (57%)	3 (75%)	5 (63%)	6 (67%)
Pattern D	NA	3 (43%)	1 (25%)	3 (37%)	3 (33%)
Discordant	3 (15%)	2 (10%)	6 (30%)	3 (15%)	3 (15%)
Pattern E					

Figure 1

Concordance patterns found for each molecular characteristic analyzed among prostatic foci (HGPIIN and PCa: GL3, GL4, GL5) of each patient. Color pattern of gene expression or fusion: black: highly expressed (>1.20) or presence of fusion; medium gray: normally expressed (between 0.80-1.20); and white: lowly expressed: (<0.80). There are five concordance patterns (Figure 1A-E). Two concordant patterns (A,B). A. Homogeneous agreement of high expression or presence of fusion (pattern A); B. Homogeneous agreement of low expression or absence of fusion (pattern B). Two partially concordant patterns (C,D): C. Progressive increase in gene expression or late appearance of fusion related to tumor progression (pattern C); D. Progressive decrement in expression, related to tumor progression (pattern D). One discordant pattern (E): E. Molecular mismatch among different foci independently of lesion progression or disappearance of fusion with tumor progression, since once the fusion appears it is expected not to disappear (pattern E). Concordant and partially concordant patterns are compatible with monoclonal origin. Discordant pattern is compatible with multiclonal origin. F. Concordance patterns with respect to each molecular alteration evaluated by patient.

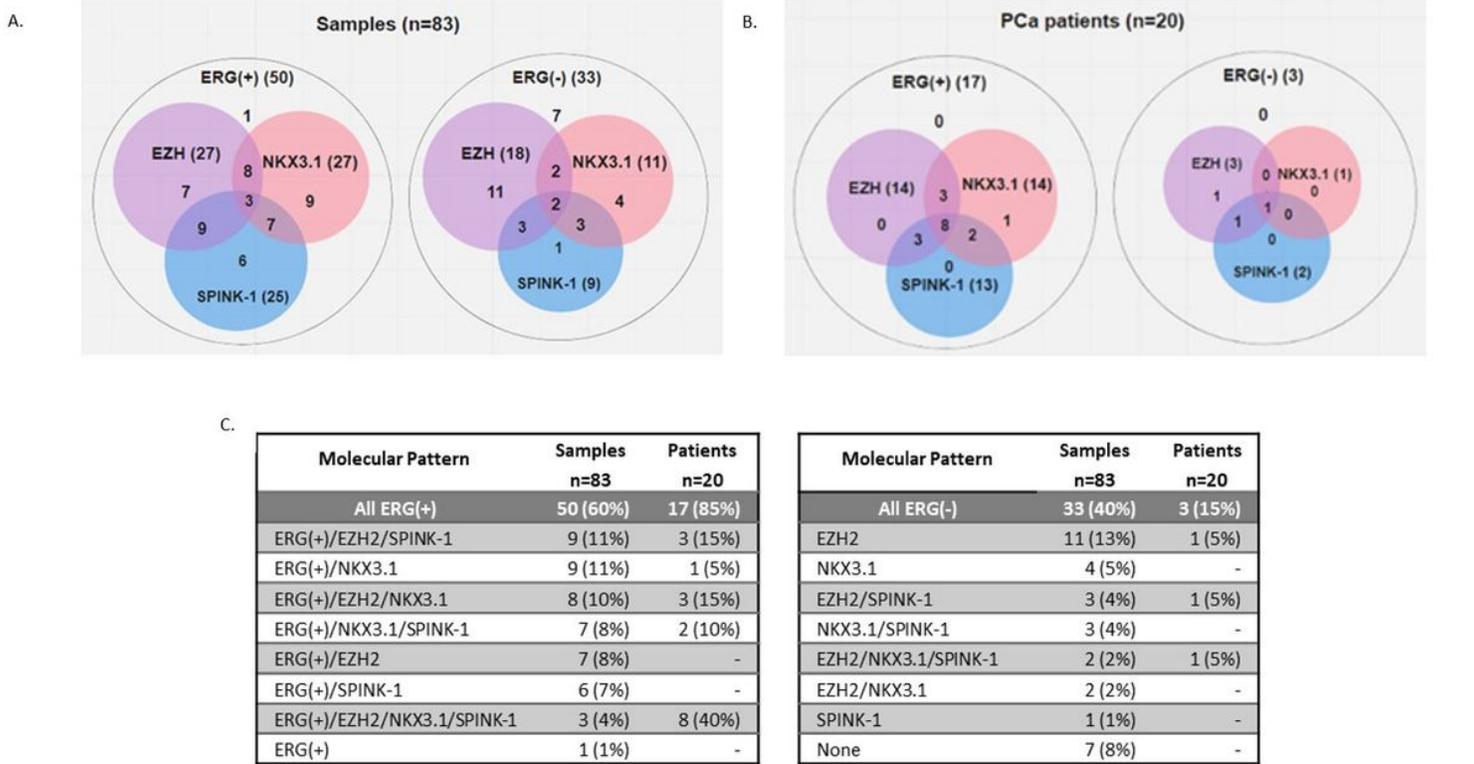
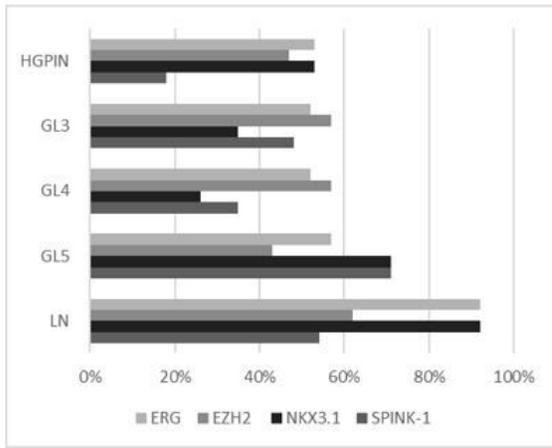


Figure 2

Molecular subtypes and molecular patterns of samples and patients. A. Venn diagrams showing the presence of different molecular patterns by sample. Numbers between parentheses correspond to the number of samples with a subtype, and the number without parentheses is for each molecular pattern. B. Venn diagrams showing the presence of different subtypes and molecular patterns by patients. Numbers between parentheses corresponds to the number of patients with a specific subtype, and the number without parentheses is for each patient's molecular pattern. C. Frequency table of molecular patterns by sample and by patient.

A.



B.

Molecular Pattern*	HGPIN n=17	GL3 n=23	GL4 n=23	GL5 n=7	LN n=13
All ERG(+)	9 (53%)	12 (52%)	13 (57%)	4 (57%)	12 (92%)
All ERG(-)	8 (47%)	11 (48%)	10 (43%)	3 (43%)	1 (8%)
ERG(+)/NKX3.1/SPINK-1	-	2 (9%)	-	1 (14%)	4 (31%)
ERG(+)/EZH2/NKX3.1	2 (12%)	-	2 (9%)	1 (14%)	3 (23%)
ERG(+)/EZH2/NKX3.1/SPINK-1	-	-	-	-	3 (23%)
ERG(+)/NKX3.1	4 (24%)	2 (9%)	2 (9%)	-	1 (8%)
ERG(+)/EZH2	2 (12%)	1 (4%)	3 (13%)	-	1 (8%)
EZH2/NKX3.1	1 (6%)	-	-	-	1 (8%)
ERG(+)/EZH2/SPINK-1	-	5 (22%)	2 (9%)	2 (29%)	-
NKX3.1/SPINK-1	-	-	1 (4%)	2 (29%)	-
NKX3.1	-	3 (13%)	-	1 (14%)	-
EZH2	1 (6%)	5 (22%)	5 (22%)	-	-
ERG(+)/SPINK-1	1 (6%)	2 (9%)	3 (13%)	-	-
EZH2/NKX3.1/SPINK-1	-	1 (4%)	1 (4%)	-	-
ERG(+)	-	-	1 (4%)	-	-
SPINK-1	-	-	1 (4%)	-	-
EZH2/SPINK-1	2 (12%)	1 (4%)	-	-	-
None	4 (24%)	1 (4%)	2 (9%)	-	-

Figure 3

Molecular subtypes and molecular patterns analyzed by histopathological group. A. Frequency of molecular subtypes of PCa. B. Frequency of molecular patterns by histopathological group. *: The molecular patterns were organized so that they allow showing less heterogeneity in more advanced groups (GL5 and LN).

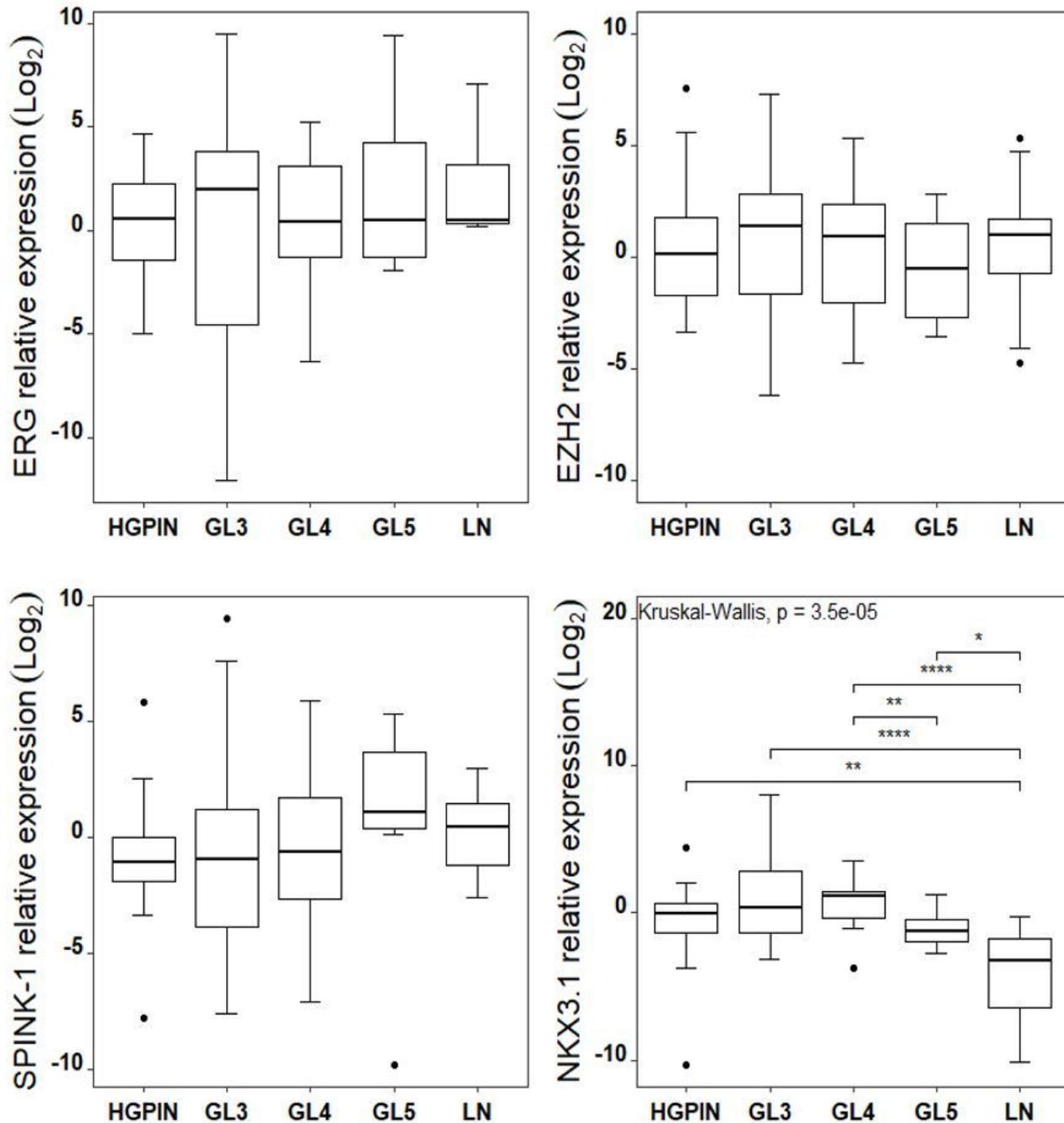


Figure 4

Distribution of the relative expression of ERG, EZH2, SPINK-1, and NKX3.1 by histopathological group. Data were graphed with a logarithmic scale (log₂). The points are outliers. The Kruskal-Wallis nonparametric test for the log₂(NKX3.1) for the four groups shows a statistically significant difference (p=3.5e-05). The braces with asterisks represent significant differences (*: p < 0.05; **: p < 0.01; ****: p < 0.001) between the groups, compared with the Mann-Whitney U test statistic.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTablesSubtypesPCa6JAN22.docx](#)