NEOPLASIA
www.neoplasia.com

Volume 21 Number 10 October 2019 pp. 989–1002 989

Pseudogene Associated Recurrent Gene Fusion in Prostate Cancer

Balabhadrapatruni VSK Chakravarthi[*,1], Pavithra Dedigama-Arachchige[†,1], Shannon Carskadon[†], Shanker Kalyana Sundaram[‡], Jia Li[§], Kuan-Han Hank Wu[§], Darshan Shimoga Chandrashekar[*], James O Peabody[†], Hans Stricker[†], Clara Hwang[**], Dhananjay A Chitale[¶], Sean R Williamson[¶], Nilesh S Gupta[¶], Nora M Navone[††], Craig Rogers[†], Mani Menon[†], Sooryanarayana Varambally[*,#,2] and Nallasivam Palanisamy[†,*,2]

[*]Department of Pathology, University of Alabama at Birmingham, Birmingham, AL; [†]Vattikuti Urology Institute, Department of Urology, Henry Ford Health System, Detroit, MI; [‡]Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, MI; [§]Department of Public Health Sciences, Henry Ford Health System, Detroit, MI; [¶]Department of Pathology, Henry Ford Health System, Detroit, MI; [#]O'Neal Comprehensive Cancer Center, University of Alabama at Birmingham, Birmingham, AL; [**]Department of Hematology and Oncology, Henry Ford Health System, Detroit, MI; [††]Department of Genitourinary Medical Oncology-The University of Texas, M D, Anderson Cancer Center, TX.

Abstract

We present the functional characterization of a pseudogene associated recurrent gene fusion in prostate cancer. The fusion gene *KLK4-KLKP1* is formed by the fusion of the protein coding gene *KLK4* with the noncoding pseudogene *KLKP1*. Screening of a cohort of 659 patients (380 Caucasian American; 250 African American, and 29 patients from other races) revealed that the *KLK4-KLKP1* is expressed in about 32% of prostate cancer patients. Correlative analysis with other ETS gene fusions and *SPINK1* revealed a concomitant expression pattern of *KLK4-KLKP1* with *ERG* and a mutually exclusive expression pattern with *SPINK1, ETV1, ETV4, and ETV5.* Development of an antibody specific to KLK4-KLKP1 fusion protein confirmed the expression of the full-length KLK4-KLKP1 protein in prostate tissues. The *in vitro* and *in vivo* functional assays to study the oncogenic properties of KLK4-KLKP1 confirmed its role in cell proliferation, cell invasion, intravasation, and tumor formation. Presence of strong ERG and AR binding sites located at the fusion junction in *KLK4-KLKP1* suggests that the fusion gene is regulated by *ERG* and *AR*. Correlative analysis of clinical data showed an association of *KLK4-KLKP1* with lower preoperative PSA values and in young men (<50 years) with prostate cancer. Screening of patient urine samples showed that *KLK4-KLKP1* can be detected noninvasively in urine. Taken together, we present *KLK4-KLKP1* as a class of pseudogene associated fusion transcript in cancer with potential applications as a biomarker for routine screening of prostate cancer.

Neoplasia (2019) 21, 989–1002

Address all correspondence to: Nallasivam Palanisamy, PhD, Department of Urology, Vattikuti Urology Institute, Henry Ford Health System, Detroit, MI, USA. E-mail:
Funding: National Cancer Institute, grant number: R21CA176330; Department of Defense -CDMRP W81XWH-16-1-0544 and Dykstra Foundation to N. P.
Conflicts of interest: none.

[1]Equal contribution.

[2]Sharing senior authorship.

Received 10 June 2019; Revised 22 July 2019; Accepted 23 July 2019

https://doi.org/10.1016/j.neo.2019.07.010

Introduction

Prostate cancer is the most common cancer among men in the United States. Advances in diagnosis, treatment, and management have resulted in increased survival rate, yet prostate cancer still remains the second leading cause of cancer-related deaths among American men [1,2]. One of the major barriers to achieving successful prostate cancer control is the underlying molecular complexity of the disease itself [3]. Morphologically, prostate cancer is well known to be a diverse disease with patients developing tumors with varying pathological characteristics [4,5]. Many studies have also indicated that prostate cancer is highly heterogeneous with distinct molecular aberrations observed in patient subgroups [6–8]. For example, roughly 50%-60% of prostate cancer patients are known to carry E26 transformation-specific (ETS) family rearrangements, where *ERG, ETV1, ETV4,* or *ETV5* genes are fused with androgen regulated 5′ partner genes [9]. Additionally, the overexpression of *SPINK1* has been observed in about 5%-10% of prostate cancer patients [10]. Furthermore, 1%-2% of the cases are known to carry *RAF* kinase (*BRAF, RAF1*) gene fusions [11], while the genetic underpinnings in the remaining 30%-40% of the prostate cancer cases are not known [6]. Importantly, distinct molecular changes have been linked with unique disease outcomes [10,12,13], indicating complex heterogeneity among patients with respect to disease progression. Therefore, discovery of new molecular markers for further patient stratification is an urgent unmet clinical need to facilitate targeted therapy and effective prostate cancer management.

Currently, prostate cancer diagnosis is primarily based on prostate-specific antigen (PSA) levels and Gleason Grade Group, a scoring system based on the morphology of the prostate tissue [14]. Following the detection of elevated PSA or pro-PSA levels, prostate cancer is identified by the presence of Gleason Graded cancer on needle biopsies. The decision to pursue immediate treatment or continue active surveillance is determined using the Gleason Grade Group. However, the rise in PSA is not prostate cancer specific and is multifactorial [15]. Therefore, PSA has been an inadequate diagnostic marker, in some cases leading to overdiagnosis and unnecessary treatment. On the other hand, though high–Gleason Grade Group tumors are known to be clinically aggressive, whether low–Gleason Grade Group tumors require treatment has been debated [16]. While intervention in low–Gleason Grade Group cancers may result in overtreatment, watchful waiting may also pose an unnecessary risk and additional burden of repeat biopsies. Given these limitations of the existing markers and the recognition of prostate cancer as a heterogeneous disease, molecular markers specific to distinct patient subgroups are required as alternatives for both initial cancer diagnosis and distinguishing aggressive cancer from indolent disease.

Although several recurrent molecular alterations have been identified in a subset of prostate cancer cases, the genetic aberrations in prostate cancer patients negative for all the known molecular makers remain to be studied. Moreover, most prostate cancer molecular studies have been carried out on Caucasian American patients with little representation of the African American population [17]. Despite the unique ancestral background of African Americans and the aggressive nature of prostate cancer in African American patients, the genetic underpinnings behind the racial disparity of prostate cancer markers are not well studied. Therefore, the study of additional molecular aberrations using large cohorts of a racially diverse population is a pressing need in prostate cancer research. In addition to identifying subtype-specific prostate cancer diagnostic and prognostic markers, such studies may also facilitate the development of novel therapeutic approaches by uncovering molecular alterations, which may be pharmacologically targeted in distinct patient subgroups.

Given the need for identifying novel molecular markers in prostate cancer patients, we investigated the expression patterns of pseudogenes in 89 prostate cancer patient samples using a paired-end next-generation sequencing approach [18]. Often considered as dysfunctional relatives of known protein-coding genes, pseudogenes have recently been implicated in cancer with roles in gene regulation [19]. While we observed distinct expression changes in several pseudogenes in prostate cancer compared to normal prostate tissue, we also noted the rare occurrence of a chimeric transcript formed through the fusion of the androgen regulated gene *KLK4* (Kallikrein Related Peptidase 4) with the adjacent pseudogene *KLKP*1 (Kallikrein Pseudogene 1). Importantly, the fusion converts the *KLKP1* pseudogene to a protein-coding gene with a predicted chimeric protein of 164 amino acids, of which 55 amino acids are derived from the pseudogene part due to a shift in the open reading frame [18]. Although a few pseudogenes have been previously reported to be expressed as proteins [20,21], *KLK4-KLKP1* is a rare example where gene fusion leads to the conversion of a noncoding pseudogene to a protein-coding gene. Further studies showed that *KLK4-KLKP1* fusion is both prostate tissue and cancer specific, suggesting a role in prostate cancer formation [18]. Both the prostate cancer specific expression and the intriguing nature of the *KLK4-KLKP1* fusion warrant further functional studies to understand the role of *KLK4-KLKP1* in prostate cancer development. Therefore, in this study, we explored the prevalence, the expression pattern, noninvasive detection, and the oncogenic properties of *KLK4-KLKP1* to investigate the potential of *KLK4-KLKP1* fusion gene as a novel molecular marker in prostate cancer.

Materials and Methods

Tissue Microarray Construction

Prostatectomy samples collected from 659 patients who underwent radical prostatectomy at Henry Ford Health Systems were reviewed, and tissue cores containing benign and tumors from different regions of the radical prostatectomy tissues were isolated to construct formalin--fixed and paraffin-embedded tissue microarrays. In most cases, a total of three tissue cores from different regions were obtained from each whole mount radical prostatectomy sample. In all cases, appropriate informed consent and Institutional Review Board approval were obtained. The Gleason Grade Group of each tissue core was reviewed by the study pathologists (N.G., D.C., and S.W.). Clinical and pathological information of patients such as age, race, family history of prostate cancer, preoperative PSA, prostatectomy date, Gleason Grade Group, tumor stage, cancer status of the lymph nodes, tumor volume, perineural invasion, presence of lymph vascular invasion, last PSA, and presence of biochemical recurrence was also recorded.

KLK4-KLKP1 RNA In Situ Hybridization (RNA-ISH)

RNA-ISH was performed as described previously using RNAscope 2.5 HD Reagent Kit (ACDBio, catalog #322350) according to the manufacturer's instructions [1]. Briefly, after baking, deparaffinization, and target retrieval per manufacturer's instructions, tissue microarray (TMA) slides were incubated with target probes for KLK4-KLKP1 (ACDBio, catalog #405501, NM_001136154, region

2933–3913) for 2 hours at 40°C in a humidity chamber. After detection and color development, slides were washed twice in deionized water and then counterstained in hematoxylin (Agilent DAKO, catalog #K800821-2) for 5 minutes. Slides were washed several times in tap water, then dried, dipped in xylene, and mounted in EcoMount (Fisher, catalog #50-828-32). Next, the slides were scanned using a digital imaging system (Aperio Scanner, Leica). The images were reviewed, and the RNA-ISH signal on the TMAs was scored. A staining pattern of distinct punctuate cytoplasmic dots was considered as a positive RNA-ISH signal for KLK4-KLKP1 expression. Depending on the intensity of the RNA-ISH staining, a score ranging from +1 to +4 was given to tissue cores with positive RNA-ISH signal, with +1 assigned to the weakest RNA-ISH staining and +4 given to the cores showing the most intense RNA-ISH staining. A score of 0 was assigned to tissue cores with no visible RNA-ISH staining. The highest score observed among the tissue cores was then assigned to each patient case. If all tissue cores of a patient were 0, the case was recorded as negative.

Cell Culture

HEK-293 cells and prostate benign epithelial cells (RWPE-1, #CRL-11609) were purchased from American Type Culture Collection (Manassas, VA). Primary prostate epithelial cells (PrEC) were purchased from Lonza (Walkersville, MD). HEK-293 cells were cultured in MEM media (Thermo Fisher Scientific, catalog #11095080,) supplemented with 10% FBS (fetal bovine serum, Thermo Fisher Scientific, catalog number #10082147). RWPE-1 cells were cultured in keratinocyte serum free medium (K-SFM, Gibco, Thermo Fisher Scientific, catalog #17005-042, Carlsbad, CA) supplemented with Bovine Pituitary Extract (BPE, 0.05 mg/ml, Thermo Fisher Scientific, catalog #17005-042), human recombinant Epidermal Growth Factor 1-53 (EGF 1-53, 5 ng/ml, Thermo Fisher Scientific, catalog #17005-042), and 1% penicillin/streptomycin. PrEC cells were cultured in Prostate Epithelial Cell Basal Medium (PrEGM) supplemented with Prostate Epithelial Cell Growth Kit (Clonetics PrEGM, BulletKit, Lonza). All cell cultures were maintained at 37 °C in an incubator with a controlled humidified atmosphere composed of 95% air and 5% CO_2.

In Vitro Overexpression of KLK4-KLKP1

KLK4-KLKP1 cDNA was PCR amplified using a forward primer with DDK tag and a reverse primer from KLK4-KLKP1 template and was cloned into the Gateway expression system (Life Technologies). To generate lentiviral and adenoviral constructs, PCR8-KLK4-KLKP1 (DDK tagged) was recombined with pLenti6/V5-Dest (Life Technologies) or pAD/CMV/V5-Dest (Life Technologies), respectively, using LR Clonase II (Life Technologies). For transient overexpression in HEK-293, RWPE-1, and PrEC cells, adenoviruses carrying KLK4-KLKP1, EZH2, or lacZ were added to the culture media after cells reached 50%-70% confluency. At the same time, cells were treated with or without bortezomib (100 nM in ethanol, 10 µl, Cayman Chemical, catalog #10008822). After incubation for 48 hours at 370°C, cells were harvested by scraping. For stable overexpression, RWPE-1 cells were infected with lentiviruses expressing KLK4-KLKP1 or lacZ, and stable clones were selected with blasticidin (3.5 µg/ml, Sigma-Aldrich, St Louis, MO). Lenti- and adenoviruses were generated by the University of Michigan Vector Core (Ann Arbor, MI).

Western Blotting

Harvested cells were spun down (1000 rpm, 5 minutes, 4°C). For HEK-293 cells, the cell pellet was resuspended in RIPA lysis buffer (Thermo Fisher Scientific, catalog #PI89900) supplemented with protease inhibitor (1×, genDEPOT, catalog #50-101-5488). For RWPE-1 cells, NP-40 lysis buffer (Boston BioProducts, Ashland, MA) with protease inhibitor was used to lyse the cells. With xenograft tissues, frozen tissues were cut into small pieces and then sonicated on ice in RIPA lysis buffer. The debris from cells or tissues were removed by centrifugation (13.2 rpm, 10 minutes, 4°C). Protein concentration of the supernatant was determined using Micro BCA protein assay kit (Thermo Fisher Scientific, catalog #23235). The lysates were separated on a 12% SDS-PAGE or a NuPAGE 4%-12% Bis-Tris protein gel. After separation, proteins were transferred onto a PVDF membrane (Milipore Immobilon-P, Fisher, catalog #IPVH00010). Then the membranes were probed with specific antibodies: Flag (Sigma, catalog #F1804), KLK4/KLKP1 (Eurogentec custom synthesized antibody), and β-actin (Sigma, catalog #A2228). The membranes were visualized on an imaging system (ChemiDoc, BIO-RAD) using a chemiluminescence developing kit (Clarity Western ECL Blotting Substrates, BIO-RAD, catalog #1705060).

Measurement of Cell Proliferation

Cell proliferation was measured by cell counting. For this, stable RWPE-1 cells overexpressing KLK4-KLKP1 (DDK-tagged) or lacZ were used. The cells were seeded at a density of 10,000 cells per well in 24-well plates (n = 3). Next, the cells were trypsinized and counted at specified time points by a Z2 Coulter particle counter (Beckman Coulter, Brea, CA). LacZ cells served as controls. Each experiment were performed with three replicates per sample.

Matrigel Invasion Assay

Matrigel invasion assays were performed using BD BioCoat Matrigel matrix (Corning Life Sciences, Tewksbury, MA). The parental and transfected clones of RWPE-1 and PrEC cells were seeded at 1 × 105 cells in serum-free medium in the upper chamber of a 24-well culture plate. The lower chamber containing respective medium was supplemented with 10% serum as a chemoattractant. After 48 hours, the noninvading cells and Matrigel matrix from the upper side of the chamber were gently wiped with a cotton swab. Invasive cells located on the lower side of the chamber were stained with 0.2% crystal violet in methanol, air-dried, and photographed using an inverted microscope (4×). Invasion was quantified by colorimetric assay or by counting the number of cells. For colorimetric assays, the inserts were treated with 150 µl of 10% acetic acid and the absorbance measured at 560 nm.

Chicken Chorioallantoic Membrane Assay (CAM) assay

CAM assay was performed as described earlier [22]. Briefly, fertilized eggs were incubated in a rotary humidified incubator at 38°C for 10 days. CAM was dropped by making two holes, one through the eggshell into the air sac and a second hole near the allantoic vein that penetrates the eggshell membrane but not the CAM. Subsequently, a cutoff wheel (Dremel) was used to cut a 1-cm^2 window to expose the underlying CAM near the allantoic vein. After 3 days of implanting the $2*10^6$ cells in 50 µl medium on the top of each egg, lower CAM was harvested and analyzed for the presence of tumor cells by quantitative human Alu-specific PCR. Genomic DNA from lower CAM and livers was prepared using Puregene DNA

purification system (Qiagen USA), and quantification of human-Alu was performed as described earlier [22]. After 7 days of implantation, extraembryonic tumors were isolated and weighed. An average of eight eggs per group was used in each experiment.

Gene Expression Microarray Analysis. Two-channel microarray experiment was performed with two replicates using the Agilent Whole Human Genome Oligo Microarray (Agilent, catalog #G4851C Whole Human Genome Microarray 8×60K). Raw data from each replicate were independently processed using Bioconductor packages. "agilp" Bioconductor package [1] was used to apply loess normalization on raw expression values. Fold change for each probe was obtained by taking difference of loess-normalized, log-2–transformed signal intensity between sample with KLK4-KLKP1 gene fusion and control sample. Probes showing differential expression in both two-channel experiments were considered for functional analysis. In total, 1956 probes were upregulated (with Log2FC > =1) and 1918 probes were downregulated (with log2FC = −1) in KLK4-KLKP1 gene fusion sample. Heatmap of differentially expressed genes was created using heatmap.2 of "gplots" R package.

Gene Set Enrichment Analysis (GSEA). GSEA was performed using the curated gene sets [C2] (n = 1267) from Molecular Signature Database (MSigDB v5.0) provided by the Broad institute [2] Differentially expressed genes were ranked by average log2FC from two arrays and submitted to GSEAPreranked module in GSEA software.

KEGG Pathway Analysis. Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.8 [3] was used to identify enriched KEGG pathways in these differentially expressed genes. With default parameters (gene count of 2 and EASE of 0.1), functional annotation chart was obtained, and KEGG pathways with P value < .05 were considered to be enriched.

Screening of KLK4-KLKP1 in the Urine Samples of Prostate Cancer Patients. Random urine samples were collected with informed consent and Institutional Review Board approval from PCa patients visiting the Hematology Oncology clinic at Henry Ford Hospital in Detroit, MI. RNA was isolated using ZR urine RNA isolation kitTM (Zymo Research, catalog # R1038 & R1039) according to manufacturer's instructions. cDNA synthesis and qRT-PCR were performed as described earlier.

Statistical Analysis

Pearson's chi-square test was used to evaluate the association of KLK4-KLKP1 fusion with race, age, Gleason Grade Group, and other molecular markers. For association between KLK4-KLKP1 and preoperative PSA, two-sample t test was performed to evaluate the difference in log-transformed preoperative PSA between KLK4-KLKP1 positive and negative cases. Multivariable Cox regression was used to estimate the association between KLK4-KLKP1 and the risk of biochemical recurrence. Cox regression model was adjusted for patients' age group (<50; ≥50), Gleason Grade Group (1 or 2; 3 or 4+), and tumor stage (pT2; pT3 or pT4). For all analyses, a P value < .05 was considered statistically significant. All analyses were performed using the Statistical Analysis System statistical software package, version 9.1.3. For the rest of the experiments, Student's two-sample t test was used to determine significant differences between two groups. P values < .05 were considered significant.

Results

Both *KLK4* and *KLKP1* belong to the kallikrein family of serine proteases, a cluster of genes located on chromosome 19 (q13.33-q13.41). The gene cluster contains 15 members, including *KLK3*, which is commonly known as PSA [23]. The *KLK4-KLKP1* fusion is formed by a trans-splicing mechanism or an in-frame fusion due to a microdeletion of the region between the adjacent genes, *KLK4* and *KLKP1*, leading to the fusion of the first two exons of *KLK4* with exon 4 and 5 of *KLKP1* (Figure 1*A* **and Supplementary Figure 1,** GenBank ID 2227664). The resulting chimeric sequence predicts a 164–amino acid protein, of which 55 amino acids are derived from KLKP1 (Figure 1*B*). According to data on the GTEx portal, full-length KLKP1 is exclusively expressed in normal prostate tissue (**Supplementary Figure 2**). In contrast, quantitative PCR (qRT-PCR) analysis of prostate cancer samples, prostate cell lines, benign prostate tissues, and other solid cancers revealed that *KLK4-KLKP1* fusion transcript is prostate cancer specific and expressed in a subset of cases [18]. However, the study included only a limited number of prostate cancer samples (n = 36), and the occurrence of *KLK4-KLKP1* in a large, racially inclusive cohort must be explored to determine the prevalence of *KLK4-KLKP1* in the prostate cancer patient population. Therefore, we studied the expression of *KLK4-KLKP1* on a larger patient cohort by using RNA *in situ* hybridization (RNA-ISH). Specifically, we constructed TMAs using prostate cancer tissues obtained from 659 radical prostatectomy (RP) specimens at the Henry Ford Health Systems. The cohort was racially inclusive with 380 Caucasian Americans (58%), 250 African Americans (38%), and 29 patients (4%) belonging to other racial groups. Each TMA contained three cores obtained from different regions of the radical prostatectomy tissue from each patient (**Supplementary Figure 3**). The individual tissue cores in each patient were reviewed, and the highest tumor grade observed was assigned to each case. Thus, the TMAs included 612 patient cases with all cores carrying prostate cancer (Gleason Grade Group 1 [3 + 3 = 6]: 110, Gleason Grade Group 2 [3 + 4 = 7]: 247, Gleason Grade Group 3 [4 + 3 = 7]: 119, Gleason Grade Group 4 [4 + 4 = 8]: 94, and Gleason Grade Group 5 [4 + 5 = 9; 5 + 4 = 9 and 5 + 5 = 10]: 42). The rest of the cases consisted of 23 cases with benign, 21 cases with high-grade prostate intraepithelial neoplasia (HGPIN), 2 cases with stroma, and 1 case with atypical cores. RNA-ISH was carried out using an antisense RNA oligonucleotide probe specific to the *KLK4-KLKP1* fusion. The TMA slides were then reviewed for the intensity of the RNA-ISH signal. A score of expression ranging from 0 to 4+ was given according to the intensity of the RNA-ISH signal, where 0 indicated no detectable RNA-ISH signal, while 4+ was assigned to the highest level of RNA-ISH signal [24].

Of the 659 cases in the cohort, 209 (32%) were positive for *KLK4-KLKP1* fusion, indicating the recurrent nature of *KLK4-KLKP1* among prostate cancer patients. Most of the *KLK4-KLKP1*–positive cases showed RNA-ISH signal intensity of 1 + (130 cases; Figure 1*C*), while more intense RNA-ISH signal 2+ was observed in 66 cases, 3+ in 12 cases, and 4 + in 1 case, suggesting varying expression levels among patients. The remaining cases were "0" or negative. To further confirm that KLK4-KLKP1 is specific to prostate cancer, we then explored the association of *KLK4-KLKP1* RNA-ISH signal with Gleason Grade Group by using Pearson's chi-square test. The results showed that *KLK4-KLKP1* is exclusively expressed in prostate cancer tissues compared to benign, high-grade

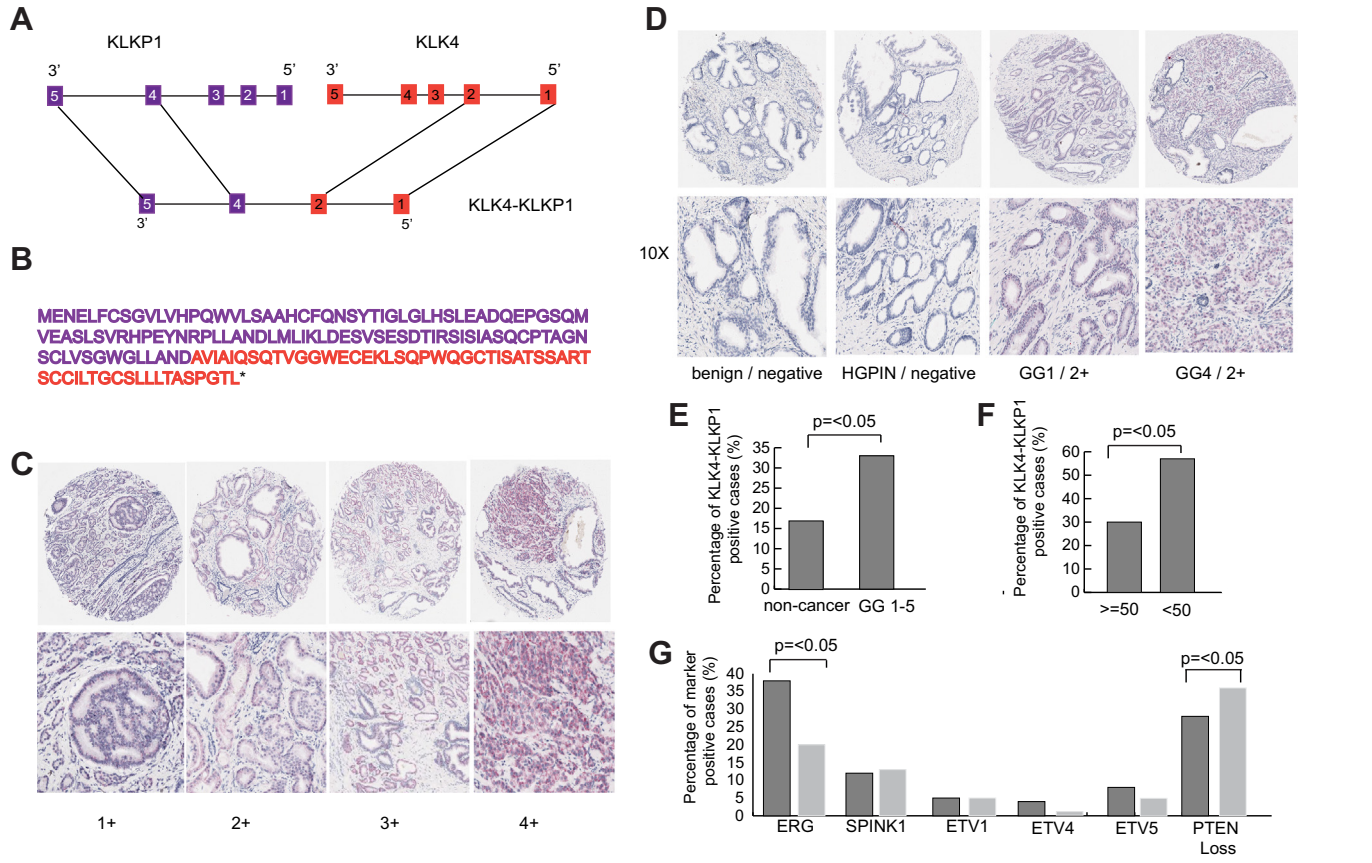

Figure 1. The structure of *KLK4-KLKP1* fusion and the RNA-ISH screening of *KLK4-KLKP1* in tissue micro arrays. (A) Schematic diagram showing the structure of *KLK4-KLKP1* fusion. *KLK4-KLKP1* is formed through the fusion of exon 1 and 2 of *KLK4* gene with exon 4 and 5 of *KLKP1*. (B) The predicted sequence of *KLK4-KLKP1* fusion protein. The sequence in purple is derived from KLK4, while the sequence in red is originating from KLKP1. (C) The expression of *KLK4-KLKP1* in prostate tissue cores detected by RNA-ISH. The bottom set of images shows an enlarged section of the corresponding tissue core in the top set of images. Values 1+ to 4+ indicate the intensity of *KLK4-KLKP1* RNA-ISH staining. (D) Prostate cancer specific expression of *KLK4-KLKP1*. *KLK4-KLKP1* RNA-ISH staining in benign, HGPIN, and prostate cancer tumor cores is shown. The bottom set of images contains a magnified area of the images on the top. Values 1+ to 4+ refer to the intensity of the *KLK4-KLKP1* RNA-ISH staining. (E) *KLK4-KLKP1* is expressed more in the prostate cancer patients (Gleason Grade Group 1-5) compared to noncancer (benign, HGPIN, atypical, and stroma) cases. The percentage of cases showing a positive *KLK4-KLKP1* RNA-ISH signal among noncancer and Gleason Grade Group1-5 is shown. *P* value was calculated based on Pearson's chi-square test. (F) *KLK4-KLKP1* is expressed more in young prostate cancer patients. The percentages of cases with positive *KLK4-KLKP1* RNA-ISH signal in the young patient (age lower than 50 years) and old patient groups (age equal to or higher than 50 years) are shown. *P* value was calculated based on Pearson's chi-square test. (G) *KLK4-KLKP1* expression is associated with *ERG* overexpression. *SPINK1*, *ETV1*, *ETV4,* and *ETV5* overexpression is mutual from *KLK4-KLKP1* expression. *PTEN* loss is significantly lower in cases with *KLK4-KLKP1* expression. The percentages of cases showing positive signal for *ERG*, *SPINK1*, *ETV1*, *ETV4*, *ETV5,* or *PTEN* loss among *KLK4-KLKP1* RNA-ISH positive cases (dark gray bars) and *KLK4-KLKP1* RNA-ISH negative cases (light gray bars) are shown. *P* value was calculated based on Pearson's chi-square test. Abbreviations: GG, Gleason Grade Group; HGPIN, high-grade prostate intraepithelial neoplasia.

prostate intraepithelial neoplasia and atypical prostate tissues (Figure 1*D*, *E* and Table 1), confirming that *KLK4-KLKP1* expression is prostate cancer specific. Additionally, we also analyzed if KLK4-KLKP1 expression is associated with Gleason Grade Group. No associations were observed between KLK4-KLKP1 RNA-ISH signal and distinct Gleason Grade Groups (Table 2).

Next, we investigated if *KLK4-KLKP1* fusion displays racial disparity in the incidence. The 209 positive cases included 128 Caucasian Americans (34%), 69 African Americans (28%), and 12

Table 1. The Comparison of KLK4-KLKP1 RNA-ISH Status Between Noncancer (Benign, HGPIN, Atypical, Stroma) and GG 1-5 Cases

Tumor Grade	No. of Cases with KLK4-KLKP1 RNA-ISH Negative	No. of Cases with KLK4-KLKP1 RNA-ISH Positive	*P* Value
Noncancer (benign, HGPIN, atypical, stroma)	39 (83%) 8	8 (17%)	
GG 1, 2, 3, 4, 5	409 (67%)	201 (33%)	.02

GG refers to Gleason grade. The number of cases in each tumor grade with KLK4-KLKP1 RNA-ISH signal positive (1+ to 4+) and KLK4-KLKP1 negative is shown. The *P* value was calculated using Pearson's chi-square test. $P < .05$ was considered statistically significant.

patients of other races (41.4%). The prevalence of *KLK4-KLKP1* was higher in Caucasian Americans compared with African American patients. However, when analyzed by Pearson's chi-square test, the difference in *KLK4-KLKP1* expression between Caucasian American and African American patients was not statistically significant (Table 3 and Supplementary Figure 4). Then, we also explored if *KLK4-KLKP1* expression is related to patient age. We categorized the patients into two groups as young (age ranging from 40 to 49 years) and old (age ranging from 50 to 83 years). Pearson's chi-square test showed significantly higher expression of *KLK4-KLKP1* in young age group compared to the older age group (Figure 1*F* and Table 4).

The other common prostate cancer–specific mutations such as ETS gene fusions and *SPINK1* overexpression are known to occur in a mutually exclusive manner. Therefore, we also analyzed the association of *KLK4-KLKP1* fusion expression with ETS gene fusions and *SPINK1* expression. We screened the same set of TMAs by using dual immunohistochemistry (IHC) for *ERG* and *SPINK1* and dual RNA-ISH for *ETV1, ETV4,* and *ETV5*. By using Pearson's chi-square test, we observed that *KLK4-KLKP1* expression is associated with *ERG*+ cases (Figure 1*G*, Supplementary Figure 5, and Table 5). However, no such association was observed with *SPINK1*, *ETV1*, *ETV4,* and *ETV5* (Figure 1*G* and Table 5), suggesting the concurrent expression of *KLK4-KLKP1* with distinct ETS gene fusion positive cases. Next, we investigated if *KLK4-KLKP1* is related with *PTEN* loss, another common prostate cancer mutation that is associated ERG+ and aggressive disease [25–27]. We carried out IHC for *PTEN* on the same set of TMAs and found that *PTEN* deletion was significantly lower in *KLK4-KLKP1*–positive cases compared to *KLK4-KLKP1*–negative cases (Figure 1*G* and Table 5). Given that *ERG* is known to co-occur with *PTEN* loss [28], we further analyzed if there is any significant difference in *PTEN* loss in cases showing both *ERG* fusion and *KLK4-KLKP1* compared to the rest of the cases. No significant difference in *PTEN* status was observed in cases with *ERG* fusion and *KLK4-KLKP1* expression, suggesting that *KLK4-KLKP1* may represent a distinct subtype of prostate cancer.

Having thus confirmed the recurrent and the prostate cancer specific occurrence of *KLK4-KLKP1* fusion, we then studied the expression of *KLK4-KLKP1* fusion protein. Based on the sequence, the *KLK4-KLKP1* fusion gene is predicted to generate a full-length protein of 164 amino acids of which 55 are derived from the KLKP1 pseudogene (Figure 1*B*). To validate the *KLK4-KLKP1* expression as a full-length protein, we generated adenoviral constructs carrying the N-FLAG-tagged *KLK4-KLKP1* fusion gene and transfected to HEK293 cells. To stabilize the protein levels of *KLK4-KLKP1*, the cells were treated with bortezomib, a proteasome inhibitor. As a control, bortezomib-treated cells transfected with vector DNA alone were used. Expression of the fusion transcript was confirmed by qRT-PCR using fusion specific primers (Figure 2*A*). Cell lysates were analyzed by Western blotting using an anti-N-FLAG antibody. Importantly, we observed a FLAG-specific protein band around 17kDA (Figure 2, *B*, *C*), confirming the expression of *KLK4-KLKP1* as a full-length protein. For additional validation, we also checked the expression of N-FLAG-tagged *KLK4-KLKP1* using the anti-FLAG antibody in the normal prostate cell line RWPE-1, transfected with and without N-FLAG-tagged *KLK4-KLKP1* adenovirus construct. Notably, we detected anti-FLAG specific protein band only in the transfected RWPE1 cells (Supplementary Figure 6). Furthermore, we also developed a *KLK4-KLKP1* specific polyclonal antibody (Eurogentech, Seraing, Belgium) using the antigenic peptide "CTISATS-SARTS" derived from the *KLKP1* pseudogene region of the fusion protein (Figure 1*B*). After cell lysis and SDS-PAGE, we probed HEK293 lysates transfected with and without N-FLAG-tagged *KLK4-KLKP1* adenovirus construct with the *KLK4-KLKP1* specific antibody using Western blot. A protein band around 17 kDa was observed (Figure 2, *B*, *C*), further confirming the expression of the chimeric *KLK4-KLKP1* protein and the specificity of the *KLK4-KLKP1* antibody to the fusion protein.

As an additional validation experiment, we then analyzed the expression of *KLK4-KLKP1* in prostate cancer patient derived xenografts (PDX) [29]. We first screened the expression of *KLK4-KLKP1* using qRT-PCR and identified 17 out of 31 PDX models positive for endogenous expression of *KLK4-KLKP1* (Supplementary Figure 7; Table 6). Then, we selected one of the PDX tissues (MDA PCa 153-7) expressing high levels of *KLK4-KLKP1* and one with no detectable levels of *KLK4-KLKP1* (MDA PCA 144-13). After protein isolation and separation on SDS-PAGE, the lysates were probed with the *KLK4-KLKP1* specific antibody using Western blot. Importantly, we observed a protein band around 17 kDa only in the *KLK4-KLKP1*–positive PDX (Figure 2, *D*, *E*), indicating the endogenous expression of *KLK4-KLKP1* fusion protein in prostate cancer patients. Additionally, we also screened the expression of *KLK4-KLKP1* in xenograft tissues using IHC with the *KLK4-KLKP1* specific antibody. While *KLK4-KLKP1* expression was observed in qRT-PCR positive PDX tissues, minimal or no *KLK4-KLKP1* IHC signal was seen in qRT-PCR negative xenografts (Figure 2*F*), further suggesting the presence of *KLK4-KLKP1* protein in a subset of prostate cancer patients. Then, we compared the detection of KLK4-KLKP1 by IHC using the KLK4-KLKP1 antibody to that of RNA-ISH with the fusion specific RNA probe. We carried out IHC using the KLK4-KLKP1 antibody on TMAs that were already analyzed with RNA ISH for *KLK4-KLKP1* (Figure 1). Then we compared the staining with IHC with the *KLK4-KLKP1* RNA ISH signal we observed previously. The comparison showed positive IHC signal predominantly in the RNA ISH positive tissue cores compared to RNA ISH negative tumor cores (Figure 2*G*).

Given the exclusive expression of *KLK4-KLKP1* in prostate cancer, next we explored the functions of *KLK4-KLKP1* by studying the oncogenic properties of the fusion gene. Specifically, we established RWPE-1 cells with stable expression of *KLK4-KLKP1* by transfection with lentiviral constructs carrying FLAG-tagged *KLK4-KLKP1*. As controls, cells stably transfected with a LACZ control (LACZ) and un-transfected RWPE-1 cells were used. We first confirmed the

Table 2. KLK4-KLKP1 RNA-ISH Status Compared Between Different Gleason Grade Groups

Tumor Grade	No. of Cases with KLK4-KLKP1 RNA-ISH Negative	No. of Cases with KLK4-KLKP1 RNA-ISH Positive	*P* Value
GG 1	70	39	>.05
GG 2	165	82	>.05
GG 3	79	39	>.05
GG 4	64	30	>.05
GG 5	31	11	>.05

GG refers to Gleason Grade Group. The number of KLK4-KLKP1 RNA-ISH signal positive (1+ to 4+) cases in each Gleason Grade Group is shown. The association of KLK1-KLKP1 RNA-ISH signal with Gleason Grade Group was analyzed using Pearson's chi-square test. $P < .05$ was considered statistically significant.

Table 3. KLK4-KLKP1 RNA-ISH Status Compared Between AA and CA Patients

Race	No. of cases with KLK4-KLKP1 RNA ISH Positive	No. of cases with KLK4-KLKP1 RNA-ISH Negative	*P* Value
African American (AA)	69 (28%)	181 (72%)	
Caucasian American (CA)	128 (34%)	250(66%)	.098

The number of cases with KLK4-KLKP1 RNA-ISH signal positive (1+ to 4+) and KLK4-KLKP1 negative in each race group is shown. The *P* value was calculated using Pearson's chi-square test. $P < .05$ was considered statistically significant.

expression of *KLK4-KLKP1* by qRT-PCR. The results showed significant expression of *KLK4-KLKP1* in transfected cells compared to both the untransfected cells and the LACZ control (Figure 3*A*). Then we investigated the effect of *KLK4-KLKP1* on cell proliferation by measuring the number of cells using a Coulter particle counter. Compared to the un-transfected cells and the LACZ control, a notable increase in the cell number was seen over time in *KLK4-KLKP1* transfected cells (Figure 3*B*), indicating a role of *KLK4-KLKP1* on cell proliferation. Next, we studied the effect of *KLK4-KLKP1* in cell invasion using the Matrigel invasion assay. Importantly, a significant increase in the number of invading cells was observed with *KLK4-KLKP1* transfected cells compared to both the untransfected and the LACZ control (Figure 3*C*). For additional validation, we also transiently transfected PrEC, another normal prostate cell line with *KLK4-KLKP1*. As controls, untransfected cells and cells transfected with a LACZ control were used. Additionally, we also used cells transfected with *EZH2*, which has been shown to increase invasion of prostate cancer and other cancer cells [30,31], as a positive control. The invasion of cells was then examined by the Matrigel invasion assay. Like RWPE-1, PrEC cells also showed a significant increase in the number of invading cells in *KLK4-KLKP1* transfected cells compared to both the untransfected and the LACZ control (Figure 3*D*). As expected, cells transfected with EZH2 also demonstrated increased invasion compared to the LACZ control and the untransfected cells (Figure 3*D*). In all, our studies indicate that *KLK4-KLKP1* promote both cell proliferation and invasion of prostate cells, suggesting an oncogenic role for *KLK4-KLKP1* fusion.

In order to further understand the oncogenic properties of *KLK4-KLKP1*, we also studied the effects of *KLK4-KLKP1* fusion on intravasation and tumor formation using the chicken chorioallantoic membrane (CAM) *in vivo* assay [22,32]. We implanted eggs with RWPE-1 cells stably expressing *KLK4-KLKP1* and then checked for the presence of intravasated cells in the lower CAM by using quantitative human Alu-specific PCR. As controls, eggs implanted with either untransfected cells or cells stably transfected with a LACZ control were used. Notably, we observed a marked intravasation by *KLK4-KLKP1* transfected cells in the lower CAM compared to both untransfected cells and LACZ control (Figure 3*E*). Additionally, we also isolated and weighed the extraembryonic tumors from eggs implanted with either *KLK4-KLKP1* transfected cells or controls. The tumors isolated from eggs implanted with cells expressing *KLK4-KLKP1* showed significantly higher weight than the tumors isolated from eggs treated with the untransfected cells and the LACZ control (Figure 3*F*). Overall, the results establish that *KLK4-KLKP1* drives intravasation and tumor formation in prostate cells, indicating a potential role in prostate cancer development.

Further, we investigated the molecular mechanisms underlying the oncogenic functions of *KLK4-KLKP1* fusion. We conducted a gene expression microarray analysis using RWPE-1 cells stably transfected with *KLK4-KLKP1*. As the control, cells transfected with LACZ control were used. After RNA isolation and microarray analysis, we observed a significant number of genes expressed differently between the RWPE-1 cells transfected with *KLK4-KLKP1* and the LACZ control. We selected the genes showing a fold change value of more than one in two independent replicates and generated a heat map with the top 100 genes differentially expressed (Figure 4*A*). We noted genes both up- and

Table 4. KLK4-KLKP1 RNA-ISH Status Compared Between Young (Age Lower Than 50 Years) and Old (Age Equal to or Higher than 50 Years) Patients

Age Group	No. of Cases with KLK4-KLKP1 RNA-ISH Positive	No. of Cases with KLK4-KLKP1 RNA-ISH Negative	*P* Value
Young age (less than 50 years)	26 (57%)	20 (43%)	
Old age (equal to 50 years or higher)	183 (30%)	427 (70%)	.0002

The number of cases with KLK4-KLKP1 RNA-ISH signal positive (1+ to 4+) and KLK4-KLKP1 negative in each age group is shown. The *P* value was calculated using Pearson's chi-square test. $P < .05$ was considered statistically significant.

Table 5. ERG, SPINK1, ETV1, ETV4, and ETV5 Marker Status Compared with KLK4-KLKP1 Status

Molecular Marker	No. of Cases with the Corresponding Marker Status	No. of Cases with KLK4-KLKP1 RNA-ISH Positive	No. of Cases with KLK4-KLKP1 RNA-ISH Negative	*P* Value
ERG	No. of cases with ERG negative	130 (63%)	355 (80%)	
	No. of cases with ERG positive	78 (38%)	90 (20%)	<.001
SPINK1	No. of cases with SPINK1 negative	184 (88%)	387 (87%)	
	No. of cases with SPINK1 positive	24 (12%)	57 (13%)	.703
ETV1	No. of cases with ETV1 negative	197 (95%)	423 (95%)	
	No. of cases with ETV1 positive	11 (5%)	22 (5%)	.849
ETV4	No. of cases with ETV4 negative	200 (96%)	439 (99%)	
	No. of cases with ETV4 positive	8 (4%)	6 (1%)	.077
ETV5	No. of cases with ETV5 negative	193 (92%)	425 (95%)	.217
	No. of cases with ETV5 positive	16 (8%)	23 (5%)	
PTEN	No. of cases with PTEN loss	57 (28%)	159 (36%)	.032
	No. of cases without PTEN loss	150 (72%)	281 (64%)	

The number of cases with KLK4-KLKP1 RNA-ISH signal positive (1+ to 4+) and KLK4-KLKP1 negative in each marker status is shown. The *P* value was calculated using Pearson's chi-square test. $P < .05$ was considered statistically significant.

downregulated in cells expressing the *KLK4-KLKP1* fusion, suggesting a role for *KLK4-KLKP1* in gene expression regulation. Further, we also carried out a gene set enrichment analysis [33] to explore any overlap between the differentially expressed genes observed with *KLK4-KLKP1* transfection and other curated gene sets. Importantly, we noted enrichment of two curated gene sets, one involving genes upregulated in endometroid endometrial metastatic tumor and the other containing genes overexpressed in melanoma metastatic cancer (Figure 4*B*), indicating that the genes affected by *KLK4-KLKP1* are associated with metastatic cancer. As a further step, we also carried out a KEGG pathway analysis using the DAVID tool [34]. The genes differentially affected by *KLK4-KLKP1* were shown to be associated with several cancer-related pathways (Figure 4*C*), further implying that *KLK4-KLKP1* may regulate the expression of genes involved in cancer and metastasis.

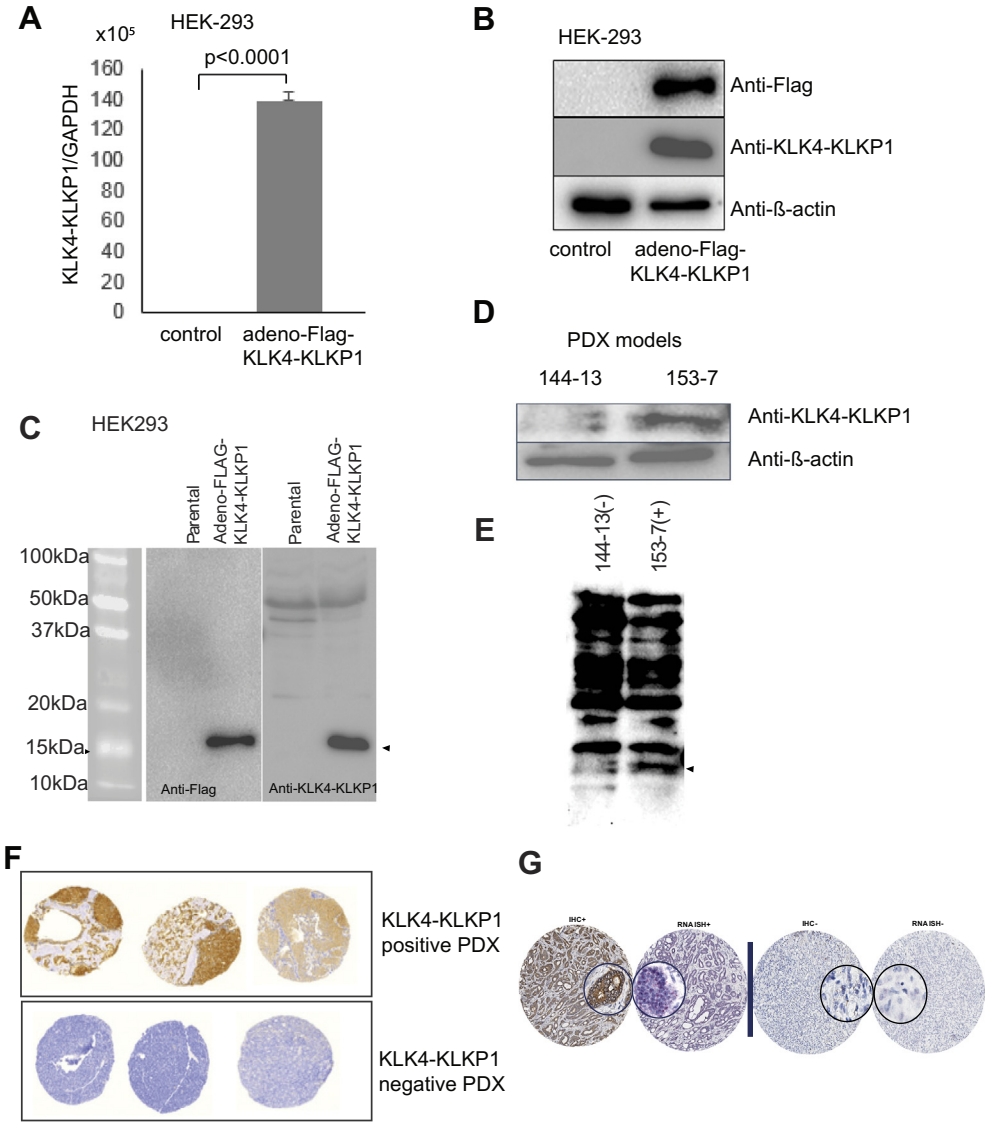

Figure 2. Validation of the expression of *KLK4-KLKP1* protein in HEK-293 cells and PDX tissues. (A) The qRT-PCR analysis HEK-293 cells transfected with and without FLAG tagged-*KLK4-KLKP1*. HEK-293 cells were transfected with adenoviral vectors carrying FLAG tagged-*KLK4-KLKP1* (adeno-FLAG-*KLK4-KLKP1*). As a control, untransfected cells treated with bortezomib were used. The expression of *KLK4-KLKP1* was confirmed by qRT-PCR. (B) Western blot analysis of HEK-293 cells transfected with FLAG-tagged *KLK4-KLKP1* using anti-FLAG, anti–*KLK4-KLKP1,* and anti–β-actin antibody. (C) Full images of Western blot analysis of Figure 2B. The full anti-Flag blot (left side) and the full anti–KLK4-KLKP1 blot (right side) are shown. The molecular weight ladder is also shown on the left. The arrowhead indicates KLK4-KLKP1 protein observed at the expected molecular weight. (D) Western blot analysis of *KLK4-KLKP1* qRT-PCR negative (MDA PCa144-13) and qRT-PCR positive (MDA PCa 153-7) PDX tissues using anti–*KLK4-KLKP1* and anti–β-actin antibody. (E) Full image of Western blot analysis of Figure 2D. The arrowhead shows a band observed around the expected molecular weight only in the *KLK4-KLKP1* qRT-PCR positive tissue sample (MDA PCa 153-7). (F) IHC staining of KLK4-KLKP1 qRT-PCR positive and qRT-PCR negative PDX models. (G) Representative images from a TMA stained with KLK4-KLKP1 RNA ISH probe (left) and KLK4-KLKP1 antibody (right). Comparison between IHC staining with anti-*KLK4-KLKP1* antibody and RNA-ISH on two representative TMA tissue cores. The IHC staining with anti–*KLK4-KLKP1* on a *KLK4-KLKP1* RNA ISH positive tumor core (left side) and a *KLK4-KLKP1* RNA ISH negative tumor core is shown. The inner circles show the enlarged images of a tumor area of the TMA tissue core.

Table 6. List of PDX Models Used to Screen KLK4-KLKP1 Fusion Gene

MDA PCa PDXs and a Cell Line	Source	Treatment
MDA PCa 173-13	Testis	Therapy naïve
MDA PCa 149-1	Bladder, local extension of prostate cancer	CRPC
MDA PCa 153-7	Thyroid	CRPC
MDA PCa 2b	Bone	CRPC
MDA PCa 183-A	Bone	Therapy naïve
MDA PCa 203-A	Bone	CRPC
MDA PCa 101	Liver	CRPC
MDA PCa 152-1	Brain	CRPC
MDA PCa 175-10	Testis	CRPC
MDA PCa 180-30	Prostate	CRPC
MDA PCa 153-14	Thyroid	CRPC
MDA PCa 175-2	Testis	CRPC
MDA PCa 230-A	Chest wall	CRPC
MDA PCa 146-12	Bladder, local extension of prostate cancer	CRPC
MDA PCa 188-2	Bladder, local extension of prostate cancer	CRPC
MDA PCa 133-4	Bone	CRPC
MDA PCa 180-11	Bladder, local extension of prostate cancer	CRPC
MDA PCa 273-A	Retroperitoneal LN	CRPC
MDA PCa 94	Pleural effusion	CRPC
MDA PCa 118a	Bone	CRPC
MDA PCa 144-4	Prostate	CRPC
MDA PCa 144-13	Bladder, local extension of prostate cancer	CRPC
MDA PCa 146-10	Bladder, local extension of prostate cancer	CRPC
MDA PCa 146-20	Bladder, local extension of prostate cancer	CRPC
MDA PCa 150-1	Bone	CRPC
MDA PCa 155-12	Bladder, local extension of prostate cancer	CRPC
MDA PCa 155-16	Prostate	CRPC
MDA PCa 166-8	Bladder, local extension of prostate cancer	CRPC
MDA PCa 177-B	Prostate	CRPC
MDA PCa 118b	Bone	CRPC

Given the well-established role of androgen receptor (*AR*) in gene expression in prostate cancer [35], we also explored if *AR* is driving the expression of *KLK4-KLKP1* in prostate cancer. Additionally, since we observed concurrent expression of ERG with *KLK4-KLKP1* (Figure 1*G*), we also studied if *ERG* is involved in the expression of *KLK4-KLKP1*. Therefore, to identify any AR or ERG binding sequences on *KLK4* or *KLKP1*, we examined data from a previous study where a chromatin immunoprecipitation assay was carried out using antibodies specific to *AR* and *ERG* [36]. Notably, we observed both *AR* and *ERG* binding sites at the fusion junction of *KLKP1* (Supplementary Figure 8), suggesting that both *AR* and *ERG* may modulate the expression of *KLK4-KLKP1* during prostate cancer formation.

For further characterization of the functional role of *KLK4-KLKP1*, we also studied the cellular localization of *KLK4-KLKP1*. We carried out immunofluorescence studies on RWPE-1 cells transfected with adeno-FLAG tagged-*KLK4-KLKP1* using fluorescent anti-FLAG antibody. As a control, cells transfected with adeno-LacZ were used. While cells transfected with adeno-Lacz showed minimal immunofluorescence as expected, notably, we observed colocalization of *KLK4-KLKP1* immunofluorescence signal with DAPI (4, 6-diamidino-2-phenylindole, Supplementary Figure 9), indicating that KLK4-KLKP1 is localized in the nucleus of the cells. Further, the nuclear localization signal predicting program, cNLS mapper [37], predicted a nuclear localization signal "**RPLLANDLKLIKLDESVSESDTIRSISIASQCPTA"** in KLK4-KLKP1 sequence, validating the immunofluorescence results. IHC analysis using KLK4-KLKP1 specific polyclonal antibody and RNA ISH analysis recognized samples positive for KLK4-KLKP1 expression; however, the nucleus specific expression is not ruled out (Figure 2, *F*, *G*).

The prostate cancer specific expression of *KLK4-KLKP1* in a subset of patients indicates the possible use of *KLK4-KLKP1* as a biomarker for prostate cancer. Therefore, to further explore the potential utility of *KLK4-KLKP1* as a prostate cancer marker, we investigated the association between *KLK4-KLKP1* expression and preoperative PSA of the 659 patients in our cohort. Specifically, we performed a *t* test to evaluate the difference in log-transformed preoperative PSA between cases with and without *KLK4-KLKP1* expression. Interestingly, patients with *KLK4-KLKP1* expression showed slightly lower preoperative PSA values compared to patients without *KLK4-KLKP1* expression (Supplementary Figure 10). As a further step, we also analyzed the association between *KLK4-KLKP1* and the time to biochemical recurrence using multivariable Cox regression model. Patients with *KLK4-KLKP1* showed a lower risk of biochemical recurrence (HR = 0.58; Supplementary Figure 11) after adjusting for age, Gleason Grade Group, and tumor stage. However, the difference in recurrence was not statistically significant (P = .12) due to the small number of patients showing recurrence (n = 49). Additionally, we also analyzed the association of *KLK4-KLKP1* with other clinical and pathological parameters such as family history, tumor stage, tumor volume, metastasis to lymph nodes, perineural invasion, and the presence of lymph vascular invasion using Pearson's chi-square test. No statistically significant association was observed between *KLK4-KLKP1* and the clinicopathological variables. Lastly, we explored the feasibility of noninvasively detecting *KLK4-KLKP1* in urine samples of prostate cancer patients like TMPRSS2-ERG gene fusions [38]. We collected urine samples from 90 unselected prostate cancer patients. All patients had confirmed prostate cancer, with most having metastatic or biochemically recurrent disease. Then we screened for *KLK4-KLKP1* transcript using qRT-PCR. As a positive control, RWPE-1 cells stably expressing *KLK4-KLKP1* were used. Importantly, *KLK4-KLKP1* expression was detected in 15 out of 90 (17%) patient samples (Supplementary Figure 12), suggesting the potential for noninvasive detection in patient urine samples. Overall, our study establishes *KLK4-KLKP1* as a recurrent chimeric transcript exclusively expressed in prostate cancer tissues with implications on disease progression and feasibility of being noninvasively detected in patient urine samples.

Discussion

Given the complex, heterogeneous nature of prostate cancer, the identification of distinct patient subgroups based on molecular markers is a necessary step towards targeted disease management. Therefore, in this study, we further explored and characterized a pseudogene associated gene fusion *KLK4-KLKP1*. We established that *KLK4-KLKP1* is a recurrent, prostate cancer specific fusion transcript that occurs at a significant incidence rate (32%) among prostate cancer patients. Similar to other distinct molecular aberrations such as ETS rearrangements [9] and *SPINK1* mutation [10], *KLK4-KLKP1* was observed only in a subset of prostate cancer patients. However, unlike the mutually exclusive pattern of expression of ETS rearrangements and *SPINK1*, *KLK4-KLKP1* showed concomitant expression with *ERG*, indicating possible cross talk with ERG. Notably, *KLK4-KLKP1* expression was associated with intact *PTEN* status, suggesting these fusion positive tumors are distinct molecular subtypes from ERG+/PTEN– tumors. Interestingly, full-length normal *KLKP1* transcript showed normal prostate specific expression (GTEX portal) and not in prostate cancer.

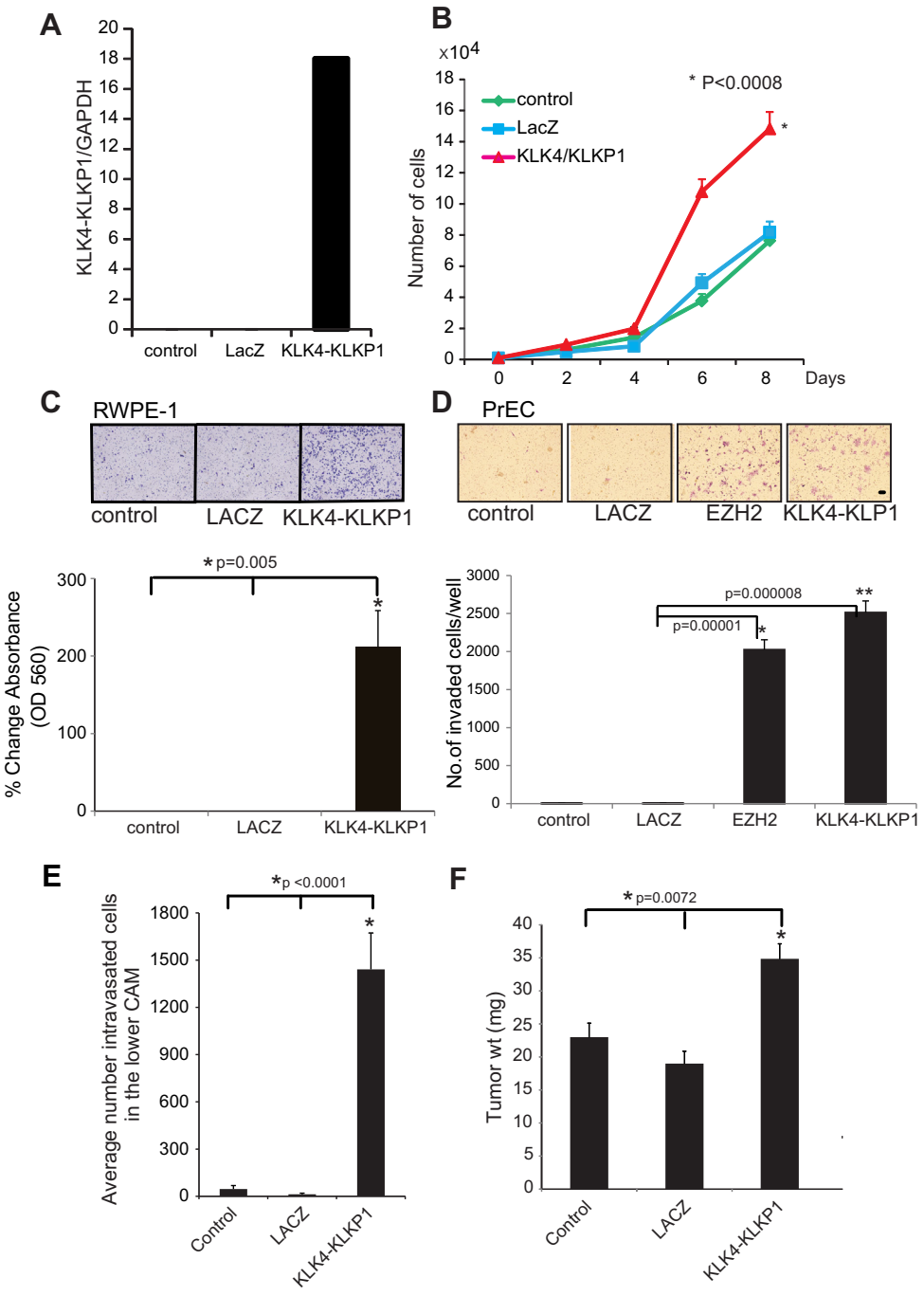

Figure 3. Functional characterization of *KLK4-KLKP1*. (A) qRT-PCR validation of *KLK4-KLKP1* expression in RWPE-1 cells after stable transfection with FLAG tagged-KLK4-KLKP1. As controls, untransfected cells (control) and cells transfected with LacZ were used. (B) Analysis of cellular proliferation in RWPE-1 cells stably expressing FLAG tagged *KLK4-KLKP1*. Cells were plated in 96-well plates. The number of cells was measured on days 2, 4, 6, and 8 using a Coulter particle counter. Cells untransfected and transfected with LACZ were used as controls. (C) Analysis of cell invasion in RWPE-1 cells. The invasion of RWPE-1 cells stably transfected with either FLAG tagged-*KLK4-KLKP1* or LacZ was studied using the Boyden chamber assay. Untransfected cells were also used as a control. After invasion of cells into the invasion chamber, cells were fixed and visualized using crystal violet. Additionally, the invasion chamber membranes carrying the fixed cells were dipped in glacial acetic acid, and the absorbance at 560 nm was also measured. Representative images of the crystal violet–stained cells that underwent invasion in each case and the absorbance at 560 nm are shown. (D) Analysis of cell invasion in PrEC cells. The cellular invasion in PrEC cells transfected with FLAG tagged-*KLK4-KLKP1* was performed as described in panel C. The number of invaded cells was counted and plotted. In addition to LACZ and untransfected cells, PrEC cells transfected with *EZH2* were also used a control. (E) Intravasation of RWPE-1 cells measured using CAM assay. RWPE-1 cells stably transfected with FLAG tagged-*KLK4-KLKP1* were implanted on eggs. The presence of intravasated cells in the lower CAM was assessed by quantitative human Alu-specific PCR. Untransfected cells and cells transfected with LACZ were used as controls. (F) Analysis of weight of extraembryonic tumors isolated from eggs implanted with RWPE-1 cells stably expressing FLAG-tagged *KLK4-KLKP1*. Cells transfected with LACZ and untransfected cells were used as controls. Abbreviations: CAM, chicken chorioallantoic membrane.

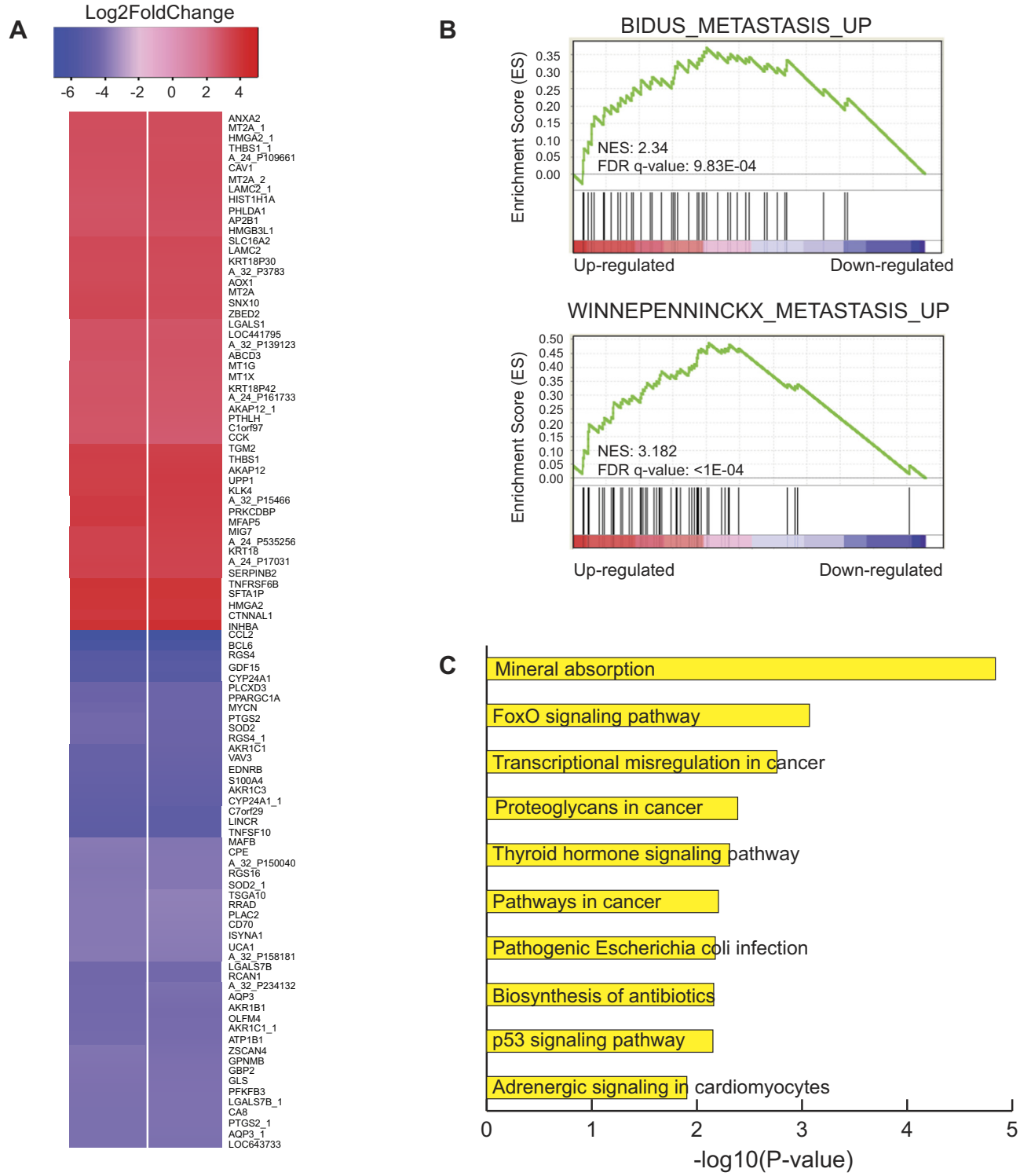

Figure 4. Gene expression analysis of *KLK4-KLKP1*. (A) Heat map showing the top 100 genes differentially expressed in RWPE-1 cells stably transfected with *KLK4-KLKP1* compared to cells transfected with LACZ. The results from two independent trials are shown. (B) Gene set enrichment analysis of differentially expressed genes. The genes were enriched in two curated gene sets, one involving genes upregulated in endometroid endometrial metastatic tumor "BIDUS_METASTASIS_UP" (top image) and the other including genes overexpressed in melanoma metastatic cancer "WINNEPENNINCKX_METASTASIS_UP" (bottom image). (C) Top 10 KEGG pathways enriched in differentially expressed genes obtained using DAVID tool.

However, despite *KLKP1* being categorized as a pseudogene, using Western blot, we showed that *KLK4-KLKP1* is expressed as a full-length protein in prostate cancer in a rare phenomenon where gene fusion leads to the inclusion of a pseudogene segment in an expressed protein. Importantly, *KLK4-KLKP1* promoted proliferation, invasion, intravasation, and tumor formation, suggesting

functional implications on prostate cancer development. Moreover, gene expression studies revealed considerable transcriptional changes in cancer-related genes in cells transfected with *KLK4-KLKP1*, which may indicate that *KLK4-KLKP1* may play a role in transcription during prostate cancer formation. In agreement with a role in transcriptional regulation, *KLK4-KLKP1* was also seen to be localized in the nucleus; however, further studies are needed to understand the nuclear expression and its effects. Furthermore, both *ERG* and *AR* were found to have strong binding sites on *KLKP1*, indicating that *KLK4-KLKP1* expression is modulated by *ERG* and *AR*. Finally, we showed that *KLK4-KLKP1* can be easily detected in patient urine samples, suggesting the feasibility to use as a biomarker for noninvasive detection of prostate cancer. Altogether, our study establishes *KLK4-KLKP1* as a novel player in a subset of prostate cancer cases with likely roles in tumor formation.

Long thought to be junk or nonfunctional units of the human genome, pseudogenes have been recently acknowledged to have key cellular roles, particularly in diseases such as cancer [39]. While some pseudogenes are known to be transcribed into noncoding RNA [39], a few pseudogenes have been shown to be even expressed as proteins [20]. Studies have revealed that several different variants of *KLKP1* pseudogene are transcribed exclusively in prostate tissues (Supplementary Figure S1) in an androgen-regulated manner [21,40]. Of the different variants, at least one *KLKP1* variant has been shown to be expressed as a protein in a transfected cell, although not *in vivo* [21]. Even though the variant chimeric transcripts of KLK*4-KLKP1* have been previously described [41,42], it has not been reported to be expressed as a protein and the functional characteristics have not been explored. Importantly, we verified that *KLK4-KLKP1* is expressed as a full-length protein in both transfected cells and endogenously in castration-resistant prostate cancer (PDX), suggesting the occurrence in prostate cancer tissues. In contrast to *KLK4*, which is overexpressed in prostate cancer with roles in cell proliferation, migration, and cancer metastasis [43–45], all KLKP1 variants are known to be expressed more in normal prostate tissues compared to prostate cancer [21,40]. However, *KLK4-KLKP1* is exclusively expressed in prostate cancer with co-occurrence with ERG+ tumors. Thus, our results indicate novel complexity in the *KLK4* and *KLKP1* locus and hint at the differential expression of the loci in prostate cancer cells compared to normal prostate cells. Given the presence of *AR* and *ERG* binding sites on *KLKP1* and the previous reports demonstrating *AR* regulation of *KLKP1* expression [21,40], it is likely that prostate cancer specific expression of *KLK4-KLKP1* is modulated by *AR* and *ERG*. Furthermore, additional variants of *KLK4-KLKP1*, which are different from the *KLK4-KLKP1* transcript observed in prostate cancer, have also been reported in renal cell cancer [42]. While the alternative *KLK4-KLKP1* transcripts were found to occur in a considerable subset of renal cell cancer cases (27%), none of the variants were shown to be expressed as proteins. Thus, *KLK4-KLKP1* may be spliced and expressed differently in a tissue specific manner in distinct cancers. Taken together, our results suggest that *KLK4* and *KLKP1* may be a diverse locus that undergoes differential splicing and transcription with functional implications in cancer. Consequently, our work highlights unprecedented roles of pseudogenes and complex molecular events involved in cancer.

In agreement with previous reports indicating significant molecular heterogeneity among prostate cancer cases [7], *KLK4-KLKP1* was expressed only in a subset of prostate cancer patients (32%). Additionally, *KLK4-KLKP1* expression was significantly higher in younger patients compared to older prostate cancer patients. Given the oncogenic properties and the transcriptional changes observed with *KLK4-KLKP1*, our results suggest that distinct molecular changes may dictate unique prostate cancer clinical outcomes among patients. Thus, our study further emphasizes the need for specific molecular markers for patient stratification in prostate cancer control.

In addition to enhancing cell proliferation, invasion, and tumor formation, *KLK4-KLKP1* also caused marked changes in gene expression. Notably, genes affected by *KLK4-KLKP1* were cancer-related and were involved in the metastasis of other cancers, implicating a functional role for *KLK4-KLKP1* in prostate cancer. Additionally, ERG was found to have a binding site on *KLK4-KLKP1*. Given that ERG expression was associated with *KLK4-KLKP1*, *ERG* may bind to the *KLKP1* locus and may promote the expression of *KLK4-KLKP1* in a subset of prostate cancer patients. Thus, it would be interesting to investigate the role of *ERG* on the expression and the oncogenic functions of *KLK4-KLKP1*.

Even though *KLK4-KLKP1* was implicated in metastatic prostate cancer, the association of *KLK4-KLKP1* with intact *PTEN* status and lower preoperative PSA values also hints indolent disease in prostate cancer patients with *KLK4-KLKP1* expression. Nevertheless, larger studies exploring the association between *KLK4-KLKP1* expression and prostate cancer clinical outcomes are necessary to establish *KLK4-KLKP1* as a biomarker for prostate cancer. Furthermore, detailed studies are also necessary to fully understand the molecular mechanisms through which *KLK4-KLKP1* promotes prostate cancer formation. Consequently, such studies will explore the potential of *KLK4-KLKP1* as a biomarker and a therapeutic target in prostate cancer, eventually making significant contributions towards achieving effective prostate cancer control.

Conclusion

In conclusion, here we describe a novel, prostate cancer specific fusion transcript involving the protein coding gene *KLK4* and the pseudogene *KLKP1*. The unique feature of *KLK4-KLKP1* transcript is the conversion of a noncoding pseudogene into a protein coding gene with expression as a full-length protein. Given the prostate cancer specific expression, the oncogenic properties, and the noninvasive detection, *KLK4-KLKP1* may have potential applications as a therapeutic target and a biomarker for early detection of prostate cancer.

Acknowledgements

We would like to thank Mireya Diaz-Insua for her help with identifying the prostate patient cohort; Natalia Draga and Jingli Yang for constructing the prostate tissue microarray; and Arul Chinnaiyan at the Michigan Center for Translational Pathology, University of Michigan, for his support during the early phase of the project. We also thank the University of Michigan Sequencing Core and Vector Core facilities for their assistance in sequencing of the clones and construction of the adenoviral and lentiviral constructs used in the study.

Author Contributions

Conception and design: N. P.
Development and methodology: S. V., N. P.
Acquisition of data: B. V. S. K. C., P. D. A., S. C., N.P.
Analysis and interpretation of data: S. K., J. L., K. H. H. W., D. S. C., N. P., S. V., P. D. A., N. G., S. W., D. C., N. P.

Writing, review, and/or revision of the manuscript: P. D. A., N. P.

Administrative, technical, or material support: N. N., J. P., H. S., C. R., M. M.

Study supervision: S. V., N. P.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neo.2019.07.010.

References

[1] Cronin KA, Lake AJ, Scott S, Sherman RL, Noone AM, and Howlader N, et al (2018). Annual report to the nation on the status of cancer, part i: national cancer statistics. *Cancer* **124**(13), 2785–2800. doi:10.1002/cncr.31551.

[2] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, and Jemal A (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **68**(6), 394–424. doi:10.3322/caac.21492.

[3] Abate-Shen C and Shen MM (2000). Molecular genetics of prostate cancer. *Genes Dev* **14**(19), 2410–2434.

[4] Arora R, Koch MO, Eble JN, Ulbright TM, Li L, and Cheng L (2004). Heterogeneity of Gleason grade in multifocal adenocarcinoma of the prostate. *Cancer* **100**(11), 2362–2366. doi:10.1002/cncr.20243.

[5] Cheng L, MacLennan GT, Lopez-Beltran A, and Montironi R (2012). Anatomic, morphologic and genetic heterogeneity of prostate cancer: implications for clinical practice. *Expert Rev Anticancer Ther* **12**(11), 1371–1374. doi:10.1586/era.12.127.

[6] Cancer Genome Atlas Research N (2015). The molecular taxonomy of primary prostate cancer. *Cell* **163**(4), 1011–1025. doi:10.1016/j.cell.2015.10.025.

[7] Tomlins SA, Alshalalfa M, Davicioni E, Erho N, Yousefi K, and Zhao S, et al (2015). Characterization of 1577 primary prostate cancers reveals novel biological and clinicopathologic insights into molecular subtypes. *Eur Urol* **68**(4), 555–567. doi:10.1016/j.eururo.2015.04.033.

[8] Yang L, Wang S, Zhou M, Chen X, Jiang W, and Zuo Y, et al (2017). Molecular classification of prostate adenocarcinoma by the integrated somatic mutation profiles and molecular network. *Sci Rep* **7**(1), 738. doi:10.1038/s41598-017-00872-8.

[9] Tomlins SA, Bjartell A, Chinnaiyan AM, Jenster G, Nam RK, and Rubin MA, et al (2009). ETS gene fusions in prostate cancer: from discovery to daily clinical practice. *Eur Urol* **56**(2), 275–286. doi:10.1016/j.eururo.2009.04.036.

[10] Tomlins SA, Rhodes DR, Yu J, Varambally S, Mehra R, and Perner S, et al (2008). The role of SPINK1 in ETS rearrangement-negative prostate cancers. *Cancer Cell* **13**(6), 519–528. doi:10.1016/j.ccr.2008.04.016.

[11] Palanisamy N, Ateeq B, Kalyana-Sundaram S, Pflueger D, Ramnarayanan K, and Shankar S, et al (2010). Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. *Nat Med* **16**(7), 793–798. doi:10.1038/nm.2166.

[12] Baena E, Shao Z, Linn DE, Glass K, Hamblen MJ, and Fujiwara Y, et al (2013). ETV1 directs androgen metabolism and confers aggressive prostate cancer in targeted mice and patients. *Genes Dev* **27**(6), 683–698. doi:10.1101/gad.211011.112.

[13] Nam RK, Sugar L, Yang W, Srivastava S, Klotz LH, and Yang LY, et al (2007). Expression of the TMPRSS2:ERG fusion gene predicts cancer recurrence after surgery for localised prostate cancer. *Br J Cancer* **97**(12), 1690–1695. doi:10.1038/sj.bjc.6604054.

[14] Gleason DF (1992). Histologic grading of prostate cancer: a perspective. *Hum Pathol* **23**(3), 273–279.

[15] Adhyam M and Gupta AK (2012). A review on the clinical utility of PSA in cancer prostate. *Indian J Surg Oncol* **3**(2), 120–129. doi:10.1007/s13193-012-0142-6.

[16] Carter HB, Partin AW, Walsh PC, Trock BJ, Veltri RW, and Nelson WG, et al (2012). Gleason score 6 adenocarcinoma: should it be labeled as cancer? *J Clin Oncol* **30**(35), 4294–4296. doi:10.1200/JCO.2012.44.0586.

[17] Tan SH, Petrovics G, and Srivastava S (2018). Prostate cancer genomics: recent advances and the prevailing underrepresentation from racial and ethnic minorities. *Int J Mol Sci* **19**(4)E1255. doi:10.3390/ijms19041255.

[18] Kalyana-Sundaram S, Kumar-Sinha C, Shankar S, Robinson DR, Wu YM, and Cao X, et al (2012). Expressed pseudogenes in the transcriptional landscape of human cancers. *Cell* **149**(7), 1622–1634. doi:10.1016/j.cell.2012.04.041.

[19] Poliseno L. Pseudogenes: newly discovered players in human cancer. Sci Signal **2012**;5(242):re5 doi https://doi.org/10.1126/scisignal.2002858.

[20] Brosch M, Saunders GI, Frankish A, Collins MO, Yu L, and Wright J, et al (2011). Shotgun proteomics aids discovery of novel protein-coding genes, alternative splicing, and "resurrected" pseudogenes in the mouse genome. *Genome Res* **21**(5), 756–767. doi:10.1101/gr.114272.110.

[21] Kaushal A, Myers SA, Dong Y, Lai J, Tan OL, and Bui LT, et al (2008). A novel transcript from the KLKP1 gene is androgen regulated, down-regulated during prostate cancer progression and encodes the first non-serine protease identified from the human kallikrein gene locus. *Prostate* **68**(4), 381–399. doi:10.1002/pros.20685.

[22] Chakravarthi BV, Goswami MT, Pathi SS, Robinson AD, Cieslik M, and Chandrashekar DS, et al (2016). MicroRNA-101 regulated transcriptional modulator SUB1 plays a role in prostate cancer. *Oncogene* **35**(49), 6330–6340. doi:10.1038/onc.2016.164.

[23] Clements J, Hooper J, Dong Y, and Harvey T (2001). The expanded human kallikrein (KLK) gene family: genomic organisation, tissue-specific expression and potential functions. *Biol Chem* **382**(1), 5–14. doi:10.1515/BC.2001.002.

[24] Warrick JI, Tomlins SA, Carskadon SL, Young AM, Siddiqui J, and Wei JT, et al (2014). Evaluation of tissue PCA3 expression in prostate cancer by RNA in situ hybridization—a correlative study with urine PCA3 and TMPRSS2-ERG. *Mod Pathol* **27**(4), 609–620. doi:10.1038/modpathol.2013.169.

[25] Leinonen KA, Saramaki OR, Furusato B, Kimura T, Takahashi H, and Egawa S, et al (2013). Loss of PTEN is associated with aggressive behavior in ERG-positive prostate cancer. *Cancer Epidemiol Biomarkers Prev* **22**(12), 2333–2344. doi:10.1158/1055-9965.EPI-13-0333-T.

[26] Ahearn TU, Pettersson A, Ebot EM, Gerke T, Graff RE, Morais CL, *et al.* A prospective investigation of PTEN loss and ERG expression in lethal prostate cancer. J Natl Cancer Inst **2016**;108(2): djv346 doi https://doi.org/10.1093/jnci/djv346.

[27] Fontugne J, Lee D, Cantaloni C, Barbieri CE, Caffo O, and Hanspeter E, et al (2014). Recurrent prostate cancer genomic alterations predict response to brachytherapy treatment. *Cancer Epidemiol Biomarkers Prev* **23**(4), 594–600. doi:10.1158/1055-9965.EPI-13-1180.

[28] Bismar TA, Hegazy S, Feng Z, Yu D, Donnelly B, and Palanisamy N, et al (2018). Clinical utility of assessing PTEN and ERG protein expression in prostate cancer patients: a proposed method for risk stratification. *J Cancer Res Clin Oncol* **144**(11), 2117–2125. doi:10.1007/s00432-018-2730-5.

[29] Navone NM, van Weerden WM, Vessella RL, Williams ED, Wang Y, and Isaacs JT, et al (2018). Movember GAP1 PDX project: an international collection of serially transplantable prostate cancer patient-derived xenograft (PDX) models. *Prostate* **78**(16), 1262–1282. doi:10.1002/pros.23701.

[30] Ren G, Baritaki S, Marathe H, Feng J, Park S, and Beach S, et al (2012). Polycomb protein EZH2 regulates tumor invasion via the transcriptional repression of the metastasis suppressor RKIP in breast and prostate cancer. *Cancer Res* **72**(12), 3091–3104. doi:10.1158/0008-5472.CAN-11-3546.

[31] Bryant RJ, Cross NA, Eaton CL, Hamdy FC, and Cunliffe VT (2007). EZH2 promotes proliferation and invasiveness of prostate cancer cells. *Prostate* **67**(5), 547–556. doi:10.1002/pros.20550.

[32] Chakravarthi BV, Pathi SS, Goswami MT, Cieslik M, Zheng H, and Nallasivam S, et al (2014). The miR-124-prolyl hydroxylase P4HA1-MMP1 axis plays a critical role in prostate cancer progression. *Oncotarget* **5**(16), 6654–6669.

[33] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, and Gillette MA, et al (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**(43), 15545–15550. doi:10.1073/pnas.0506580102.

[34] Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, *et al.* DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res **2007**;35(Web Server issue):W169–75 doi https://doi.org/10.1093/nar/gkm415.

[35] Heinlein CA and Chang C (2004). Androgen receptor in prostate cancer. *Endocr Rev* **25**(2), 276–308. doi:10.1210/er.2002-0032.

[36] Yu J, Yu J, Mani RS, Cao Q, Brenner CJ, and Cao X, et al (2010). An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell* **17**(5), 443–454. doi:10.1016/j.ccr.2010.03.018.

[37] Kosugi S, Hasebe M, Tomita M, and Yanagawa H (2009). Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci U S A* **106**(25), 10171–10176. doi:10.1073/pnas.0900604106.

[38] Tomlins SA, Day JR, Lonigro RJ, Hovelson DH, Siddiqui J, and Kunju LP, et al (2016). Urine TMPRSS2:ERG plus PCA3 for individualized prostate cancer risk assessment. *Eur Urol* **70**(1), 45–53. doi:10.1016/j.eururo.2015.04.039.

[39] Pink RC, Wicks K, Caley DP, Punch EK, Jacobs L, and Carter DR (2011). Pseudogenes: pseudo-functional or key regulators in health and disease? *RNA* **17**(5), 792–798. doi:10.1261/rna.2658311.

[40] Lu W, Zhou D, Glusman G, Utleg AG, White JT, and Nelson PS, et al (2006). KLK31P is a novel androgen regulated and transcribed pseudogene of kallikreins that is expressed at lower levels in prostate cancer cells than in normal prostate cells. *Prostate* **66**(9), 936–944. doi:10.1002/pros.20382.

[41] Lai J, Lehman ML, Dinger ME, Hendy SC, Mercer TR, and Seim I, et al (2010). A variant of the KLK4 gene is expressed as a cis sense-antisense chimeric transcript in prostate cancer cells. *RNA* **16**(6), 1156–1166. doi:10.1261/rna.2019810.

[42] Pflueger D, Mittmann C, Dehler S, Rubin MA, Moch H, and Schraml P (2015). Functional characterization of BC039389-GATM and KLK4-KRSP1 chimeric read-through transcripts which are up-regulated in renal cell cancer. *BMC Genomics* **16**, 247. doi:10.1186/s12864-015-1446-z.

[43] Veveris-Lowe TL, Lawrence MG, Collard RL, Bui L, Herington AC, and Nicol DL, et al (2005). Kallikrein 4 (hK4) and prostate-specific antigen (PSA) are associated with the loss of E-cadherin and an epithelial-mesenchymal transition (EMT)-like effect in prostate cancer cells. *Endocr Relat Cancer* **12**(3), 631–643. doi:10.1677/erc.1.00958.

[44] Klokk TI, Kilander A, Xi Z, Waehre H, Risberg B, and Danielsen HE, et al (2007). Kallikrein 4 is a proliferative factor that is overexpressed in prostate cancer. *Cancer Res* **67**(11), 5221–5230. doi:10.1158/0008-5472.CAN-06-4728.

[45] Gao J, Collard RL, Bui L, Herington AC, Nicol DL, and Clements JA (2007). Kallikrein 4 is a potential mediator of cellular interactions between cancer cells and osteoblasts in metastatic prostate cancer. *Prostate* **67**(4), 348–360. doi:10.1002/pros.20465.