Delineation of TMPRSS2-ERG Splice Variants in Prostate Cancer

Ying Hu,1 Albert Dobi,1,4 Taduru Sreenath,1 Christopher Cook,1 Atekelt Y. Tadase,1 Lakshmi Ravindranath,1 Jennifer Cullen,1 Bungo Furusato,1,2 Yongmei Chen,1 Rajesh L. Thangapazham,1 Ahmed Mohamed,1 Chen Sun,1 Isabell A. Sesterhenn,2 David G. McLeod,1,3,4 Gyorgy Petrovics,1 and Shiv Srivastava1,4

Abstract

Purpose: The expression of the ETS-related gene (ERG) is low or undetectable in benign prostate epithelial cells. High prevalence of ERG overexpression in prostate cancer cells due to TMPRSS2-ERG fusions suggest for causal roles of ERG protein in the neoplastic process. TMPRSS2-ERG fusion junctions have been extensively studied in prostate cancer. However, virtually nothing is known about the nature of full-length transcripts and encoded proteins. This study focuses on qualitative and quantitative features of full-length TMPRSS2-ERG transcripts in prostate cancer.

Experimental Design: Full-length TMPRSS2-ERG transcripts were cloned and sequenced from a cDNA library generated from pooled RNA of six TMPRSS2-ERG fusion–positive prostate tumors. The encoded ERG proteins were analyzed in HEK293 cells. Copy numbers of TMPRSS2-ERG splice variants were determined by quantitative reverse transcription-PCR in laser capture microdissected prostate cancer cells.

Results: Two types of TMPRSS2-ERG cDNAs were identified: type I, which encodes full-length prototypical ERG protein (ERG1, ERG2, ERG3), and type II, encoding truncated ERG proteins lacking the ETS domain (ERG8 and a new variant, TEPC). In microdissected prostate tumor cells from 122 patients, relative abundance of these variants was in the following order: ERG8 > TEPC > ERG3 > ERG1/2 with combined overexpression rate of 62.3% in prostate cancer. Increased ratio of type I over type II splice forms showed a trend of correlation with less favorable pathology and outcome.

Conclusions: Qualitative and quantitative features of specific ERG splice variants defined here promise to enhance the utility of ERG as a biomarker and therapeutic target in prostate cancer.

Molecular genetic evaluations of prostate cancer are defining mutational and expression alterations of critical oncogenes involved in disease onset and/or progression (reviewed in refs. 1–3). Discovery of prevalent chromosomal rearrangements/ translocations leading to the activation of ETS transcription factors (predominantly ERG) through the androgen receptor–regulated TMPRSS2 gene promoter underscore the critical roles of ERG-encoded protein in prostate cancer (4–7). Because ERG represents the majority of TMPRSS2-ETS factor alterations described thus far (6, 7), we have focused on the expression and regulation of TMPRSS2-ERG in prostate cancer. Oncogenic functions of ETS factors, including ERG, have also been implicated in diverse cancers (8).

Structure and function of ERG-encoded proteins remain to be defined in prostate cancer. ERG consists of 17 exons spanning about 300 kb and generates at least nine alternate splice forms, seven of them coding for protein products of varying sizes (9). These ERG splice variants have been primarily described in nonprostate tissues. Despite the large body of data on the TMPRSS2-ERG fusion junctions in prostate cancer (reviewed in refs. 6, 7), virtually nothing is known about the full-length TMPRSS2-ERG transcripts in prostate cancer, including the existence and relative abundance of specific splice variants. In this context, it is important to note that the cancer-associated splice variants of numerous genes, e.g., androgen receptor, fibroblast growth factor receptor, survivin, and MDM2, have functional implications (10, 11). Thus, characterization of full-length TMPRSS2-ERG transcripts is essential to better understand ERG function(s) in prostate cancer and to further enhance its utility as biomarker and therapeutic target.
In this study, we have cloned and sequenced full-length cDNAs from TMPRSS2-ERG fusion–positive prostate tumors, and from the VCaP cell line. We have identified two types of TMPRSS2-ERG cDNAs, one (type I) encoding full-length prototypical ERG protein (ERG1, ERG2, ERG3) and the other (type II) encoding a shorter version lacking the ETS domain (ERG8 and a new variant, TEPC). We have further quantified and validated the expression of these ERG splice forms in a large cohort of prostate cancer specimens. The ERG exons at the TMPRSS2-ERG fusion junction have been the subject of a number of studies (fusion junction variants; refs. 5–7, 12–19). However, these exons are present in all ERG splice forms and do not identify specific splice variants. In recent in vivo models assessing the role of ERG in prostate cancer, only a type I splice variant, specifically NH₂-terminally truncated ERG3, was tested (20, 21). Intriguingly, the data presented here shows a more abundant expression of type II splice variants in prostate cancer cells. Our new findings on ERG splice variants in prostate cancer have promise in improving the understanding of ERG functions and its therapeutic targeting in prostate cancer, as well as in enhancing the detection of ERG alterations in clinical specimens.

Materials and Methods

Tissue specimens, laser capture microdissection, and quantitative gene expression analysis. The prostate tissue specimens used in this study...
were obtained from radical prostatectomy procedures under an Institutional Review Board–approved protocol at Walter Reed Army Medical Center. Laser capture microdissection (LCM) of tumor and benign epithelial cells from optimum cutting temperature–embedded frozen tissues obtained from the radical prostatectomy specimens, RNA isolation from the LCM samples, and real-time quantitative reverse transcription-PCR (TaqMan) were essentially done as described previously (4, 22). The differentiation status of microdissected cells was recorded independently from the overall pathologic Gleason grade of the prostate, which was determined from whole-mounted, formalin-fixed, paraffin-embedded prostate specimens of each patient. The small amounts of tissue specimens (~5 mm³) were obtained for optimum cutting temperature embedding from radical prostatectomy specimen before whole-mount prostate processing. Selection of specimens for LCM was primarily driven by the presence of sufficient amount of tumor cells for the LCM. The predominant tumor cell type (by differentiation) present in a frozen section was microdissected. Most of the time, but not always, the predominant differentiation grade in frozen tissue section represented the prevalent differentiation grade of the tumor cells in the prostate. Overall, 88.4% of LCM samples were collected from the primary Gleason pattern of the index tumor. Because of this, we have compared the quantitative gene expression of TMPRSS2-ERG splice variants to the differentiation grade of the microdissected cells, as well as to the overall differentiation grade of the tumor cells in the prostate. Overall conclusions were similar by two-way comparisons but an increased statistically significant relationship was noted when the gene expression ratios of ERG I/II in LCM-RNAs was correlated with the differentiation grade of the LCM-dissected cells. TaqMan primers and probes are listed in the Supplementary data.

Detection of the TMPRSS2-ERG and TMPRSS2-ETV fusion transcripts was done essentially as described (5). The different TMPRSS2-ERG fusion junction types (A, B, and C) are described in Fig. 1A in a schematic diagram. All three fusion types have been previously described (5, 19). The expression of GAPDH was simultaneously analyzed as endogenous control, and the target gene expression in each sample (in duplicates) was normalized to GAPDH. RNA samples without reverse transcription were included as the negative control in each assay.

**Generation and screening of cDNA library from prostate tumors.** For the generation of the cDNA library, frozen tumor tissues from index tumors of six patients were selected based on available tissue size (over 30 mg), highest tumor cell content (over 70%), and the presence of TMPRSS2-ERG fusion transcripts by reverse transcription-PCR. Polyadenylated RNA was isolated from the optimum cutting temperature–embedded frozen tumor tissues. A cDNA library was generated from the pooled RNA (Loefstrand Laboratories) and cloned into the XhoI–EcoRI sites of lambdaZAP Express vector (Stratagene). Screening of the expression library was carried out according to the protocol described by the manufacturer (Stratagene). The primary library of about 400,000 plaques were screened by ERG2 probe (NM_004449; cDNA obtained from Dr. Dennis Watson, Medical University of South Carolina, Charleston, SC) and found 84 hybridized with different intensities. The positive plaques were further screened for the presence of TMPRSS2 fusions by fusion-specific PCR (5). A total of 12 plaques showed

---

**Fig. 2.** ERG splice forms in prostate tumors and their expression in the VCaP prostate cancer cell line. **A,** schematic representation of full-length type I and type II ERG transcripts expressed in prostate cancer cells with TMPRSS2-ERG fusion. Numbered boxes, ERG exons (B); boxes with * and **, unique regions of TEPC and ERG8, respectively. Solid lines above the exons, TaqMan primers and probes used for the detection of the ERG splice variants. **B,** type I transcripts code for both transactivation (SAM Pointed) and DNA-binding (ETS) domains. The relative positions coding for the two major functional domains of ERG protein are shown in type I and type II splice variants. **C,** columns, copy numbers of the ERG splice forms in VCaP cells determined by TaqMan quantitative reverse transcription-PCR. The median of three experiments using triplicates are shown. **D,** protein products expressed from TMPRSS2-ERG2, TMPRSS2-ERG3, and TMPRSS2-ERG8 clones transiently transfected into HEK293 cells are shown by Western blot analysis. Anti-Flag antibody was used for the detection of Flag-tagged ERG8 protein.
amplification. Detailed cDNA sequence analysis revealed the presence of two types of TMPRSS2-ERG fusion transcripts. Within the positively identified plaques, three represented type I (with both SAM domain and DNA-binding ETS domain) and five type II (without ETS domain). Fusion-positive type I– and type II–containing phages were amplified with T3 and T7 primers, subcloned into TOPO vector (Invitrogen), and verified by DNA sequencing.

Cell culture and Western blot. The prostate cancer cell line VCaP, which has type A TMPRSS2-ERG fusion (5), and human embryonic kidney HEK 293 cells were obtained from the American Type Culture Collection. Cells were cultured according to the provider’s instructions5 and harvested upon confluence of 70%. RNA was isolated by RNAzol B method (Tel-Test, Inc.). TMPRSS2-ERG2, TMPRSS2-ERG3, and TMPRSS2-ERG8 (Flag-tagged) constructs were cloned from prostate cancer cDNA library into pIRES-EGFP plasmid vector (Clontech) and were verified by DNA sequencing. HEK293 cells transfected with the constructs were lysed in M-PER mammalian protein extraction reagent (Pierce) supplemented with protease and phosphatase inhibitor cocktails (Sigma). ERG2 and ERG3 proteins were detected by Western blot (NuPAGE Bis-Tris gel, Invitrogen) using immunoaffinity-purified anti-ERG peptide polyclonal antibody prepared in our laboratory (DFHGI AQALQ HPPE SLYK YPSDL PYMGS YHAHP QKMNF VAPH PAL). The tagged ERG8 protein was detected by Flag-tag antibody (Sigma).

Statistical analyses of clinical and gene expression data. Measures of central tendency (median) and dispersion (range) are used to describe continuously measured patient characteristics, whereas frequencies and percentages are used to describe categorical patient characteristics. χ² and Fisher’s exact tests were conducted to compare TMPRSS2-ERG splice variant transcript expression across patient clinical and demographic characteristics. P values <0.05 are considered statistically significant.

Results

Quantitative analysis of TMPRSS2-ERG expression in prostate tumors. Quantitative analyses of the transcript levels of various TMPRSS2-ETS fusion genes were done in LCM matched benign and tumor epithelium of prostate cancer specimens (122 patients; 244 specimens; Fig. 1A). The demographic, clinical, and pathologic variables of the patient cohort are summarized in Supplementary Table S1. The workflow of LCM and quantitative reverse transcription-PCR analysis is summarized in Fig. 1B. The most frequently observed TMPRSS2-ERG fusion transcript junctions (6, 7) were detected in 57% of the patients, and among these 95% expressed TMPRSS2-ERG fusion type A (Fig. 1A). Fusions with other ETS family members, such as TMPRSS2-ETV1 or TMPRSS2-ETV4, were not detected in this cohort. No fusions were detected in matched benign prostate epithelial cells dissected from the same prostate.

Identification of full-length TMPRSS2-ERG transcripts in prostate tumors. To investigate the nature of TMPRSS2-ERG – encoded proteins in prostate cancer, a cDNA library was generated from RNA pooled from six prostate tumors with
TMPRSS2-ERG fusion. Screening of the library (see flow chart in Supplementary Fig. S1) by both ERG and TMPRSS2 probes resulted in the identification of the following ERG splice variants: ERG1 (M21535), ERG2 (NM004449), ERG3 (NM182918), ERG8 (AY204742), and TEPC, a novel splice variant (EU1432099; Fig. 2A). ERG1, ERG2, and ERG3 contain both SAM (pointed) and ETS (DNA-binding) domain (type I); however, ERG8 and TEPC lack the ETS domain (type II; Fig. 2B). Among the positively identified cDNA library clones, 30% were type I and 70% were type II. Both types of ERG transcripts are expressed in VCaP cells, a human prostate cancer cell line derived from vertebral metastasis that harbors TMPRSS2-ERG fusion, with the type II transcripts being more abundant (Fig. 2C).

HEK293 cells were transiently transfected with the TMPRSS2-ERG2, TMPRSS2-ERG3, and TMPRSS2-ERG8 constructs and the expressed ERG proteins were detected by Western blot showing the expected molecular weight of type I and type II proteins (Fig. 2D). For the detection of type I splice forms (ERG2 and ERG3), an anti-peptide polyclonal ERG antibody was used, which was developed in our laboratory. ERG8, a type II splice form, was Flag-tagged and detected by anti-Flag antibody, because ERG8 lacks part of our ERG peptide epitope.

Relative abundance of type II ERG splice forms in tumor cells of prostate cancer patients. Quantitative expression of the ERG splice variants were determined in microdissected tumor cells of 122 prostate cancer patients: 66 with TMPRSS2-ERG fusion and 56 with no detectable fusion A transcript (Fig. 3A). At least two or more ERG splice variants were detectable in all TMPRSS2-ERG fusion A–positive prostate cancer patients. ERG8 and TEPC represented the most abundant ERG splice forms analyzed (Fig. 3B) and were detected in 65 of 66 TMPRSS2-ERG fusion A–positive patients (Fig. 3A). Expression of at least two of the ERG splice forms was detected in 10 of 56 fusion A–expression–negative cases. Three of these tumors were positive for TMPRSS2-ERG fusion types B or C. It is likely that other such tumors may harbor other TMPRSS2-ERG fusion junctions. Thus, quantitative analysis of ERG splice variants, especially ERG8 and TEPC, provide a reliable surrogate for TMPRSS2-ERG fusion in prostate cancer, and in addition it detects ERG overexpression even if the fusion junction type is unknown.

The order of median abundance (copies/ng total RNA) of ERG splice forms in prostate cancer cells of 76 patients with detectable ERG expression was ERG8 (~3,200) > TEPC (~1,800) > ERG3 (~1,500) > ERG1&2 (~800; Fig. 3B). Overall, the type II splice variants (with no ETS domain) were present in higher copy numbers in prostate cancer cells than the type I splice forms (Fig. 3B and C), and 77% of ERG-positive prostate cancer patients tested have more copies of type II than type I splice forms (Fig. 3A). We conclude that quantitative detection of ERG splice variants, especially ERG8 and TEPC, may provide increased sensitivity in assessing overall frequency of TMPRSS2-ERG fusions in prostate cancer cells.

Expression of ERG splice forms in relation to clinicopathologic variables of prostate cancer patients. In comparison with

<table>
<thead>
<tr>
<th>Clinicopathologic characteristics</th>
<th>ERG splice variant expression</th>
<th>Type I/ type II ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No (46)</td>
<td>Yes (76)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>29 (33%)</td>
<td>59 (67%)</td>
</tr>
<tr>
<td>African American</td>
<td>15 (56%)</td>
<td>12 (44%)</td>
</tr>
<tr>
<td>Family history</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>28 (38%)</td>
<td>45 (62%)</td>
</tr>
<tr>
<td>Yes</td>
<td>10 (33%)</td>
<td>20 (67%)</td>
</tr>
<tr>
<td>Pathologic T stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pT2</td>
<td>14 (36%)</td>
<td>25 (64%)</td>
</tr>
<tr>
<td>pT3</td>
<td>29 (42%)</td>
<td>40 (58%)</td>
</tr>
<tr>
<td>Pathologic Gleason sum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-6</td>
<td>9 (24%)</td>
<td>29 (76%)</td>
</tr>
<tr>
<td>7</td>
<td>20 (36%)</td>
<td>36 (64%)</td>
</tr>
<tr>
<td>8-10</td>
<td>14 (70%)</td>
<td>6 (30%)</td>
</tr>
<tr>
<td>LCM differentiation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well</td>
<td>31 (32%)</td>
<td>66 (68%)</td>
</tr>
<tr>
<td>Poorly</td>
<td>15 (62%)</td>
<td>9 (38%)</td>
</tr>
<tr>
<td>Margin status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>28 (38%)</td>
<td>45 (62%)</td>
</tr>
<tr>
<td>Positive</td>
<td>16 (39%)</td>
<td>25 (61%)</td>
</tr>
<tr>
<td>PSA recurrence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>33 (37%)</td>
<td>56 (63%)</td>
</tr>
<tr>
<td>Yes</td>
<td>11 (41%)</td>
<td>16 (59%)</td>
</tr>
</tbody>
</table>

NOTE: ERG splice variant expression (N = 122): In comparison with patients with no detectable expression of ERG in their prostate cancer cells (n = 46), the ERG-positive patient cohort (n = 76) has a decreased proportion of patients with high Gleason grade, poor prostate cancer cell differentiation, and African American ethnicity. Type I/type II ratio (n = 76): The ratio of ERG type I/type II splice variants in prostate cancer cells is increased in patients with poor tumor cell differentiation and with prostate-specific antigen recurrence.

Abbreviation: PSA, prostate-specific antigen.

* Two-sided test, P < 0.05 was considered statistically significant.

† One-sided test, P < 0.05 was considered statistically significant.
prostate cancer patients with no detectable ERG expression \((n = 46)\), the ERG expression–positive patient cohort \((n = 76)\) has a smaller proportion of patients with high pathologic Gleason grade \((8–10)\), poor prostate cancer cell differentiation, or African American ethnicity (Table 1). The levels of type I or type II ERG splice forms in the cohort of prostate cancer patients with ERG expression \((n = 76)\) did not show significant correlations with clinicopathologic variables. However, there was a trend of correlation of higher copy number ratio of type I over type II splice forms with poor differentiation of prostate cancer cells, higher pathologic Gleason sum, positive margin, and biochemical recurrence (Table 1).

**Discussion**

ERG overexpression as a result of TMPRSS2-ERG fusion represents a highly prevalent oncogenic alteration in prostate cancer. Remarkable progress has been made in just over 2 years in establishing the diagnostic and prognostic features of TMPRSS2-ERG fusion in prostate cancer \((6, 7)\). Despite the large body of data on the TMPRSS2-ERG fusion junctions, virtually nothing is known about the full-length TMPRSS2-ERG transcripts, including the existence and relative abundance of specific splice variants in human prostate tumors. However, splice variants of numerous genes, e.g., androgen receptor, fibroblast growth factor receptor, survivin, and MDM2, are known to play critical roles in various human cancers \((10, 11)\).

This study establishes the nature of full-length TMPRSS2-ERG transcripts and encoded proteins in prostate cancer cells. In addition to expected full-length TMPRSS2-ERG transcripts, we have identified relatively abundant ERG splice forms with unique 3’ sequences that lack a conserved region coding for the DNA binding ETS domain. Parallel quantitative analyses of ERG splice variants in precisely microdissected cells from well-defined histologic features of the tumor provided accurate data with respect to the presence, abundance, and distribution of various ERG splice forms in prostate cancer in relation with clinicopathologic status.

Monitoring the expression of ERG splice variants, we detected more prostate cancer cases than by monitoring the fusion transcript junctions, likely because unknown or undetected fusions are present in a subset of cases. Furthermore, the number of various fusion junctions in prostate cancer is far more than the number of ERG splice variants.

Recent reports revealed that specific junction types of TMPRSS2-ERG fusion transcripts, genomic deletions, or the presence of TMPRSS2-ERG fusion are associated with poor prognosis (reviewed in refs. 6, 7). However, others reported that fusion-positive tumors were associated with lower Gleason grade and/or better disease outcome \((12)\). In this study, we found that compared with patients with no detectable expression of ERG in their prostate cancer cells, the ERG expression–positive patient cohort has a decreased proportion of patients with high Gleason grade, poor prostate cancer cell differentiation, and African American ethnicity \((N = 122)\). This is in agreement with our previous study on ERG expression in prostate cancer \((4)\). Lower or no ERG expression in a subset of aggressive tumors with TMPRSS2-ERG fusion may reflect attenuation of androgen signaling pathway during prostate cancer progression \((23)\). The levels of type I or type II ERG splice forms did not show significant correlations with clinicopathologic variables. It will be useful to combine multiple approaches, including quantitative assessment of TMPRSS2-ERG expression levels, evaluation of genomic rearrangements, and different types of transcripts in multicenter cohort to confirm prognostic values of qualitative and quantitative aspects of ERG alterations in prostate cancer.

The diversity of TMPRSS2-ERG fusion transcripts has recently been emphasized focusing on the fusion junction region of the transcripts \((13, 14)\) and by using exon arrays \((15)\) that did not allow for the discovery of the type II splice variants described here. Our results highlight the importance of understanding the expression and distribution of full-length splice forms of ERG, including variants with no DNA binding domain, in the tumor cells. Our data show a trend of correlation of relatively more type I over type II splice forms, with less favorable pathology and outcome that need to be confirmed in a larger patient cohort. The heterogeneity of TMPRSS2-ERG rearrangements in multifocal prostate cancer reported by our group and others \((18, 19)\) further adds to the complexity of understanding the roles of ERG in prostate cancer.

In conclusion, this study establishes two major types of full-length transcripts from the TMPRSS2-ERG locus in prostate cancer. Further, we establish the protein products translated from type I and type II transcripts. The presence of these specific ERG splice forms, especially the more abundant type II splice forms, may provide new opportunities in as prostate cancer biomarker. Finally, overall status of the type I and II forms in prostate cancer cells, such as the ratio of their expression levels, has potential to enhance our understanding of the biology of prostate tumors with TMPRSS2-ERG fusion.

**Disclosure of Potential Conflicts of Interest**

None of the authors have competing financial interests.

**Acknowledgments**

We thank Drs. Vasantha Srikantan and Govindan Vaidyanathan for their contributions to this study; Drs. Dennis K. Watson (Medical University of South Carolina) and Charles J. Bieberich (University of Maryland, Baltimore Campus) for their valuable advice and comments during the progress of this study; and Amina Ali and Yinghui Shi and for their superb technical support.

**References**


Clinical Cancer Research

Delineation of TMPRSS2-ERG Splice Variants in Prostate Cancer

Ying Hu, Albert Dobi, Taduru Sreenath, et al.


Updated version
Access the most recent version of this article at:
http://clincancerres.aacrjournals.org/content/14/15/4719

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2008/08/25/14.15.4719.DC1

Cited articles
This article cites 23 articles, 8 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/14/15/4719.full#ref-list-1

Citing articles
This article has been cited by 8 HighWire-hosted articles. Access the articles at:
http://clincancerres.aacrjournals.org/content/14/15/4719.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.