Overexpression of truncated ERG from TMPRSS2-ERG fusion and prostate cancer development

Melanie Leong1*
Wen-feng Shi1*
Jun Tian2
Ellen Cho1
Abbas Raza1
Saquib A Siddiqi1
Abdulhafez Selim3
Han-chun Chen4
Dianzheng Zhang1,4

1Department of Biochemistry and Molecular Biology and Center for Chronic Disorders of Aging, Philadelphia College of Osteopathic Medicine, Philadelphia, PA, USA;
2Department of Renal Transplantation, Qilu Hospital of Shandong University, Jinan, Shandong, People’s Republic of China; 3Osteotech Inc, Eatontown, NJ, USA; 4Department of Biochemistry, School of Biological Science and Technology, Central South University, Changsha, Hunan, People’s Republic of China; *These authors contributed equally to this work

Abstract: In men, prostate cancer is one of the most common cancers worldwide and the second leading cause of death among all cancer types in Europe and North America, with the numbers of those affected continuing to increase. Recent studies have identified a recurrent fusion of TMPRSS2 with members of the ETS family of transcription factors in about 80% of prostate cancer tissues. Among them, the TMPRSS2-ERG fusion accounts for approximately 50% of these cases. TMPRSS2 is highly regulated by androgen receptor and the chromosomal rearrangement abnormally induces ERG production by androgen. To investigate the effects of ERG overexpression on its target genes expression and prostate cancer development, plasmids were first constructed by inserting the truncated ERG into an expression vector in the forward or reverse directions. A predicted three-dimensional model of the protein structure of the truncated ERG, along with immunofluorescence assays, suggest that the minor deletion on the N-terminus does not appear to affect the structure or function of ERG. Results from ERG target gene expression profile indicate that TMPRSS2-ERG fusion-induced aberrant ERG overexpression is likely involved in prostate cancer development by enhancing tumor angiogenesis.

Keywords: prostate cancer, androgen receptor, TMPRSS2, ERG, chromosomal translocation

Introduction

Prostate cancer is one of the most prevalent threats to men’s health in the Western world and it accounts for the second largest number of male cancer deaths in the United States.1-5 It is understood to be a disease of complex and multifactorial origins. It appears that race, family history, and age are major factors related to its incidence. Men who have first-degree relatives with prostate cancer are at increased risk for the disease, with the risk of developing prostate cancer increasing with the number of relatives diagnosed.6 Males under the age of 50 have low rates of occurrence. However, rates increase rapidly after the age of 50 years, with 73 years being the average age at time of diagnosis.5,7 Because of improvements in the standards of modern living and advances in medicine, life expectancy has progressively increased and as a result, prostate cancer incidences are expected to increase proportionally. Hormone therapy benefits about 80% of prostate cancer patients by retarding the progression of the disease.3 However, prostate cancer in the majority of the patients who initially benefited from hormonal therapy will eventually develop into an aggressive, hormone-independent form, with little hope for further intervention.3 Despite the high prevalence of prostate cancer, knowledge about the molecular and cellular mechanisms that underlie the cause of the disease is limited, and no cure yet exists for late stage prostate cancer.
The recently identified high incidence recurrent fusion of TMPRSS2 with members of the E26 transformation-specific (ETS) family of transcription factors, including ERG, ETV1, ETV4, and ETV5, in prostate cancer samples has opened a new field in prostate cancer research. This is significant not only for understanding the underlying molecular mechanisms of prostate cancer development, but also for designing strategies for diagnosis, prognosis, and the differential treatment of prostate cancer. About 80% of prostate tumors contain genomic fusions of TMPRSS2 and members of the ETS family of transcription factors. Of these, ~50% contains TMPRSS2-ERG fusions. Furthermore, it has been reported that TMPRSS2-ERG-positive prostate cancers have more aggressive phenotypes than those of TMPRSS2-ERG-negative cancers. Consequently, both short-term and long-term survival rates of patients with such fusions are significantly affected. Screening of prostate cancer samples found that among the 27 ETS factors, ERG is the most frequently overexpressed. The TMPRSS2 gene is located on chromosome 21 at 21q22.3 (Figure 1A) and encodes for a type II transmembrane serine protease. Normal prostate cells express little to no detectable TMPRSS2 protein, but high levels of TMPRSS2 protein have been found in a large percentage of prostate cancer cells. A recent study has demonstrated that the full-length TMPRSS2 protein is expressed in LNCaP cells, an androgen-responsive prostate cancer cell line, but is not detected in the androgen-unresponsive prostate cancer cells PC-3 and DU-145. The expression of TMPRSS2 quickly increases in LNCaP cells after exposure to androgenic hormones indicating that it is likely a primary androgen receptor (AR) target gene. In fact, the TMPRSS2 promoter region has a 15-bp sequence that

![Figure 1](https://www.dovepress.com/)

**Figure 1** Chromosomal deletion and TMPRSS2-ERG fusion. **A** Ideogram of chromosomal 21. **B** Intersitial deletion and formation of the TMPRSS2-ERG fusion. **C** Schematic illustration of the exons and gene sequences in the TMPRSS2-ERG gene fusion. The gray box represents the open reading frame of TMPRSS2 while the striped box represents the open reading frame of ERG. Each exon is indicated by numbers in the boxes and the numbers above the boxes are the last base of each exon. Sequences at the fusion junction are indicated and the sequence from TMPRSS2 is underlined.
contains a consensus AR-responsive element (ARE) which is likely involved in AR-ARE interaction and in the regulation of TMPRSS2 expression.\textsuperscript{13} Previous experiments have also suggested that the androgen-responsive promoter elements of TMPRSS2 are implicated in prostate cancer through the overexpression of members of the E26 transformation-specific (ETS) family.\textsuperscript{19}

One effect of the interstitial 3 Mb deletion from chromosomal region 21q22 is the generation of the TMPRSS2-ERG fusion gene, which contains the promoter/enhancer region of the TMPRSS2 gene and a part of the open reading frame of the ERG gene (Figure 1B). There are at least 17 distinct TMPRSS2-ERG transcripts generated from fusions between exons 1, 2, and 3 of TMPRSS2 and exons 2, 3, 4, 5, or 6 of ERG which have been identified,\textsuperscript{24} with the most commonly identified fusion in prostate cancer containing the promoter/enhancer region and first exon of TMPRSS2 gene and exon 4 of ERG gene (Figure 1C). ERG is known to be a pivotal transcription factor in cell differentiation and growth control pathways, and there are lines of evidence to suggest that overexpression of ERG may be involved in different stage of prostate cancer development.\textsuperscript{13,14,16,25} However, the molecular mechanisms underlying the chromosomal translocation and prostate cancer development are not known.\textsuperscript{20} We hypothesize that overexpression of the ETS factors resulting from TMPRSS2-ETS fusion plays an essential role in prostate cancer development, and the data from our research suggest that overexpression of the truncated ERG may affect prostate cancer development, most likely through enhancing angiogenesis.

Materials and methods

Construction of overexpression vector, PSG5-ERG

Total RNA was purified from the VCaP prostate cancer cell line using an RNAqueous-4PCR kit (Ambion, Inc., Austin, TX). For cDNA synthesis, 5 µg of total RNA was reverse transcribed using random primers (Stratagene, La Jolla, CA) and SuperScript II reverse transcriptase (Invitrogen, Carlsbad, CA). The truncated open reading frame (ORF) of ERG was amplified by PCR using the following primers: Upper primer, 5'-ATGACCCGCTCTCTCAGCCGA-3'; Lower primer, 5'-TTAGTATGAGTGCCAGATGAG-3'. Xhol and Nhel restriction enzyme sites were added to the upper and lower primer ends, respectively. The full length of the ORF in the fusion gene was approximately 1.2 kb, and the amplified PCR product was purified with a Qiagen Gel Extraction kit (Qiagen, Inc., Valencia, CA). The PCR product was cloned into a TOPO-TA vector (Invitrogen). ERG was then sub-cloned into the PSG5 expression vector (J. Wong lab, Baylor College of Medicine, Houston, TX) on the XhoI site using T4 DNA ligase (Promega, Madison, WI). A negative control was constructed with the insert oriented in the opposite direction. Orientations were determined by digesting with Nhel (Promega). DNA sequence and cloning junction were verified by sequencing (University of Pennsylvania DNA Sequencing Facility, Philadelphia, PA).

Cell culture and transient transfection

HeLa cells were maintained in Dulbecco’s modified Eagle’s medium (DMEM) (Cellgro, Mediatech, Inc., Herndon, VA) containing 10% fetal bovine serum (FBS) (Atlanta Biologicals, Lawrenceville, GA) and 1% antibiotics/antimycotics (Abx/Amx) (Gibco, Carlsbad, CA). NIH 3T3 cells were maintained in DMEM (Cellgro, Mediatech, Inc.) containing 10% bovine calf serum (BCS) (Atlanta Biologicals, Lawrenceville, GA) and 1% Abx/Amx (Gibco). Cells were incubated at 37 °C and 4.7% CO\textsubscript{2}. HeLa and NIH 3T3 cells were transfected with plasmids containing either the forward insert or the reverse insert, using lipofectamine and PLUS reagent (Invitrogen). Cells were incubated in reduced serum opti-modified Eagle’s medium (Opti-MEM) (Gibco) for six hours. After this six-hour incubation, DMEM containing 1% Abx/Amx and either 10% FBS or 10% BCS was added to the HeLa and NIH 3T3 cells, respectively. Cells were incubated for an additional 36 hours at 37 °C and 4.7% CO\textsubscript{2}.

Lysis and western blotting

Cells were collected and lysed with EBC buffer containing 20 mM Tris HCl (pH 8.0), 125 mM EDTA, and 0.5% NP-40. Protein samples were separated by SDS-PAGE electrophoresis and transferred to a nitrocellulose membrane. All membranes were stained with Ponceau S Solution to confirm successful transfer of the proteins (Sigma-Aldrich, St. Louis, MO). They were then blocked with 5% milk in TBST (Tris-base, NaCl, and 1% Tween 20, pH 7.6). Proteins were incubated in 2.5% milk in TBST with ERG-1/2/3 rabbit polyclonal IgG (0.2 µg/mL) (Santa Cruz Biotechnology, Santa Cruz, CA) as a primary antibody and donkey-anti-rabbit IgG (0.08 µg/mL) (Pierce Biotechnology, Rockford, IL) as a secondary antibody. The blot was stripped and reprobed by anti-actin (1 µg/mL) (Calbiochem, San Diego, CA) and peroxidase goat anti-mouse IgM (0.01 µg/mL) (Calbiochem). The blots were detected using the Super Signal West Pico Chemiluminescent Substrate detection system (Pierce Biotechnology).
**Immunofluorescence**

After transfection and incubation on glass coverslides for 24 hours at 37 °C and 4.7% CO₂, HeLa cells were washed once with 1X phosphate-buffered solution (PBS), fixed with Cytofix/Cytoperm (BD Biosciences, San Diego, CA), and blocked for 30 minutes with 1X Perm/Wash (BD Biosciences). Cells were incubated with primary and secondary antibodies in 1X Perm/Wash at 37 °C for one hour each. ERG-1/2/3 rabbit polyclonal IgG (Santa Cruz Biotechnology) was used as a primary antibody (0.2 µg/mL), Alexa Fluor 488 donkey anti-rabbit IgG was used as a secondary antibody (0.2 µg/mL; Molecular Probes, Eugene, OR), and bis-benzamide (1 µg/mL, Sigma-Aldrich) was used to stain the nucleus. Coverslips were mounted on glass slides using aqueous mounting media containing anti-fading reagent (Biomeda Corp., Foster City, CA). Slides were examined with the use of a Nikon Eclipse E800 microscope. Z-stack images were obtained using an RT Slider Spot camera (Diagnostic Instruments, Inc., Sterling Heights, MI) and analyzed with Image-Pro Plus, Version 6.2.0.424 (Media Cybernetics, Bethesda, MD) and Autoquant X, Version X2.0.0 (Media Cybernetics) softwares.

**Bioinformatic prediction of three-dimensional structures**

Three-dimensional structures corresponding to sequences of the full-length ERG, the truncated ERG, and the functional domains were predicted by using bioinformatic softwares. The MUSTER database was used to generate the three-dimensional models of the proteins, 26 Jmol was used for visualization (www.jmol.org), and PFAM was used for the domain search. 27

**Analysis of ERG target genes**

Total RNA was purified from transiently transfected HeLa or NIH 3T3 cells using an RNAsqueous-4PCR kit (Ambion, Austin, TX). For cDNA synthesis, 5 µg of total RNA was reverse transcribed using random primers (Stratagene, La Jolla, CA) and Superscript II reverse transcriptase (Invitrogen). ERG target genes were detected by PCR using the specifically-designed primers (Invitrogen). The following PCR conditions were used routinely in a Genius thermal cycler (Techne Inc., Burlington, NJ): one cycle, 95 °C, two minutes; 35 cycles, 95 °C, one minute, 60 °C, one minute, 72 °C, one minute; one cycle, 72 °C, five minutes. PCR products were analyzed on a 1% agarose gel and visualized with ultraviolet light. The amplified bands were quantified using Image-J software.

**Results and discussion**

**Cloning and overexpression of the truncated ERG**

The TMPRSS2-ERG gene fusion is heterogeneously expressed due to a variety of interstitial deletions, creating at least 17 distinct transcripts depending on the location of the fusion. Alternative splicing can cause fusions that occur between exons 1, 2, or 3 of TMPRSS2 and exons 2, 3, 4, 5, or 6 of ERG. The most common TMPRSS2-ERG fusion occurs between exon 1 of TMPRSS2 and exon 4 of ERG (Figure 1C). In this fusion gene, the native ATG is lost and translation is assumed to start from the first ATG. The truncated open reading frame (ORF) was amplified from VCaP prostate cancer cells and was bidirectionally cloned into expression vector PSG5 as described in Materials and methods (Figure 2A). As expected, for successful ligations between ERG and PSG5, XhoI digests of the clones show two distinct bands, corresponding to the 4.5 kb vector and the 1.5 kb insert (Figure 2B, lanes 1 and 2). Upon separate digestion with NheI, plasmids with the insert oriented in the forward direction have one large fragment of approximately 6 kb and a small undetectable fragment of less than 20 bp, which is not detectable on the agarose gel (Figure 2B, lane 3). In contrast, plasmids with the insert cloned in the reverse direction should contain fragments of 4.5 kb and 1.5 kb, which are of similar size to digestion with XhoI (Figure 2B, lane 4). These DNA sequences and orientations were verified by sequencing. Because of their size, plasmids containing the reverse insert served as an ideal negative control in future experiments. The overexpression of the truncated ERG was shown by western blotting (Figure 2C). The molecular weight of the full-length ERG has been estimated to be 49 kD. With the deletion of 31 amino acids from the N-terminus, the truncated ERG is expected to be approximately 45 kD. Figure 2C shows that after 36 hours of transient transfection with PSG5-ERG, both NIH3T3 and HeLa cell lines overexpress a protein of approximately 45 kD. In contrast, cells transfected with no DNA or with the vector containing a reverse ERG insert exhibit no bands. This indicates that the truncated ERG was successfully cloned and overexpressed in both cells. During the preparation of this manuscript, a published paper verified that the truncated ERG from the TMPRSS2-ERG fusion is indeed translated with the first ATG as an alternative translation initiation site. 15

**Predicted three-dimensional structure of the truncated ERG**

The ERG used in these experiments is modeled after the most common TMPRSS2-ERG fusion which occurs between
Truncated ERG and prostate cancer

Dové Press submit your manuscript | www.dovepress.com

Exon 1 of TMPRSS2 and exon 4 of ERG. The use of an alternate ATG start codon generates a truncated ERG, which deletes 31 amino acids from the N-terminus of the protein. However, this truncation does not affect the DNA binding domain (ETS) or the integrity of the pointed domain (PTN, Figure 3A), which is particularly important for both dimerization and recruitment of other transcriptional factors.\textsuperscript{26}

To determine if the truncation of ERG affects the structure of the protein, we analyzed sequences of both the full-length and truncated ERG with bioinformatic softwares. The predicted three-dimensional structure of the truncated ERG appears to be very similar to that of the full-length protein. Both visibly contain five alpha helices (colored in pink) and multiple beta sheets (shown in yellow, Figure 3B). Although the software

Figure 2 Cloning and overexpression of truncated ERG. A) Schematic strategy in cloning of the truncated ERG into the expression vector PSGS. B) Verification of the insertion by enzyme digestion restriction. Lanes 1 and 3: Forward clone; and lanes 2 and 4: Reverse clone. C) Overexpression of truncated ERG. C: transfection with no DNA; F: transfected with Forward clone; and R: transfected with Reverse clone. tERG: truncated ERG. The actin serves as internal control.

Figure 3 Schematic drawing of the truncated ERG protein A) and modeled three-dimensional protein structure B) of the truncated ERG. The activation domains, ADT1 and ADT2, are located in the N- and C-termini, respectively. PNT represents the pointed domain and ETS represents the ETS-binding site. The gray box depicts the 31 amino acid deletion from the N-terminus. Predicted three-dimensional models of the protein structures of the full-length ERG and truncated ERG show that the overall structure of the truncated ERG appears to be similar to that of the full length. The alpha helices are shown in pink and the beta sheets were in yellow.
has limited capabilities and could not specifically identify the location of the deletion, the overall structures of the native ERG and the truncated ERG are similar. This suggests that the small truncation in the ORF of ERG does not affect the overall protein structure significantly.

**Overexpressed truncated ERG protein localizes to the nucleus**

Since ERG is a transcription factor that binds to the promoter regions of target genes, its function depends on its ability to locate to the nucleus. The nuclear translocation signal sequence has not been identified for ERG. We cannot completely exclude the possibility that it is located in the deleted N-terminus. Therefore, it is important to verify that the nucleus translocation is not affected by the deletion. Immunofluorescent microscopy was performed on HeLa cells transiently transfected with no ERG or with plasmids containing either the forward or reverse ERG insert. Three-dimensional images show that cells transfected with the forward insert exhibit intense green staining intermingled with blue nuclear staining, signifying that the protein is not simply residing in cytosolic areas around the nucleus (Figure 4, middle panel). This indicates that the truncated ERG is located in the nucleus. Control cells transfected with no ERG (Figure 4, left panel) or with the reverse insert (Figure 4, right panel) do not express ERG in their nuclei, as evidenced by the lack of green staining. Since cells transfected with PSG5-ERG demonstrate no signs of impaired ERG nuclear localization, this suggests that the small truncation in the ERG does not have a significant effect on subcellular location.

**Overexpression of truncated ERG and regulation of target gene expression**

Although the cellular functions of ERG have been established, the genes as well as the genetic pathways regulated by the TPRS2-ERG gene fusion product have not been characterized. However, because ERG is a pivotal transcription factor, it is reasonable to speculate that its actions are mediated through the regulation of its target genes. It is expected that the TPRS2-ERG gene fusion will first induce ERG overexpression and the overexpressed ERG will

![Figure 4](image)

*Figure 4* ERG is expressed in nuclei of cells transfected with PSG5-ERG. HeLa cells transfected with plasmids containing the forward insert, the reverse insert, or no DNA were fixed, permeabilized, and stained for ERG (Green) using ERG-1/2/3 rabbit polyclonal IgG as the primary antibody and Alexa Fluor 488 donkey anti-rabbit IgG as the secondary antibody. Bis-benzamide (Blue) was used for nuclear staining. Images are taken with a 60X objective. The top and bottom panels are top- and side-views, respectively.
then affect the expression of its target genes either through activation or repression. Ultimately, it is the combinations of the target gene products that will lead to prostate cancer development. Previous literature has found that there is a wide variety of target genes that are up- or downregulated by ERG. Through comparison of the ERG-positive and ERG-negative prostate cancers, Iljin and colleagues identified a wide array of genes up- or downregulated by ERG. Based on the microarray data, a broad screening of seven ERG target genes was first conducted by reverse transcriptase-polymerase chain reaction (RT-PCR) using specifically designed primers (Table 1). It was initially determined that three genes, AGT, GPR116, KHDRBS3, appear to be upregulated, whilst the RAMP1 is downregulated in certain degrees by ERG overexpression in both HeLa and NIH3T3 cells. Our results are consistent with the previous findings. However, the AGT appears to be affected the most. Figure 5 showed that the mRNA level of AGT is significantly upregulated in cells transfected with the forward insert. In contrast, cells transfected with the reverse insert do not exhibit this effect and express levels comparable to those of control cells transfected with no DNA. GAPDH, a housekeeping protein, was used as a loading control.

We speculate that aberrant expression of the truncated ERG could enhance prostate cancer angiogenesis through upregulating AGT expression. The human AGT gene, located on chromosome 1, consists of 5 exons and 4 introns and encodes for angiotensinogen. Angiotensinogen (AGT) is a 452-amino acid glycoprotein that is synthesized and secreted mainly by the liver, but also by the brain. So far, it is the only known precursor of angiotensin I (Agt I), which is produced when AGT is cleaved by renin, an enzyme secreted by the kidneys. Agt I is a decapeptide which remains inactive, with no known function except to be converted by angiotensin-converting enzyme (ACE) into its active form, angiotensin II (Agt II). Agt II plays a major role in the renin-angiotensin system (RAS) and acts as a hormone through its interactions with its specific angiotensin receptors, AT1 and AT2. Renin, AGT, ACE, Agt II, and the angiotensin receptors have been found in the human prostate, prostate cancer cell lines, and rat prostatic cells. The presence of Agt II has been observed in basal cells of the normal prostate, as well as in proliferating neoplastic cells in high grade prostatic intraepithelial neoplasia and malignant prostate cancer cells. Other research has provided evidence that Agt II has been shown to increase the rate of proliferation of prostate cancer cells and it may act as an early marker of malignancy. It has been suggested that Agt II is a cell proliferative factor due to the fact that blocking its receptors has been shown to have anti-proliferative activity, as well as roles in decreasing prostate-specific antigen and improving the statuses of patients with hormone-refractory prostate cancer. Chronic inflammation has long been associated with the development of cancer, by releasing reactive oxygen species that cause permanent DNA damage and genomic alterations, particularly in proliferating epithelium. 

**Table 1** Primers used in reverse transcriptase-polymerase chain reaction

<table>
<thead>
<tr>
<th>Name of the genes</th>
<th>Abbreviations</th>
<th>Primer sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angiotensinogen</td>
<td>AGT</td>
<td>Upper: 5’-CTCTCTGGACCTTCACAGAAGCTGGA-3’</td>
</tr>
<tr>
<td>Ca²⁺-dependent secretion activator</td>
<td>CADPS</td>
<td>Lower: 5’-GGAGGAGTTTCGGAAGGAGACAC-3’</td>
</tr>
<tr>
<td>Ca²⁺-dependent activator protein for secretion 2</td>
<td>CADPS2</td>
<td>Upper: 5’-TACATCTGGAGAAGGTCTGCTG-3’</td>
</tr>
<tr>
<td>G protein-coupled receptor 116</td>
<td>GPR116</td>
<td>Lower: 5’-TTCACTCTACTGATCTGCTGCTGCTC-3’</td>
</tr>
<tr>
<td>KH domain containing, binding, signal transduction associated 3</td>
<td>KHDRBS3</td>
<td>Upper: 5’-CATCCATTTCGGCGCTACATC-3’</td>
</tr>
<tr>
<td>Phospholipase A2, group VII (platelet-activating factor acetylhydrolase, plasma)</td>
<td>2G7</td>
<td>Lower: 5’-GGAGAGTTGGAAGGACATGCTG-3’</td>
</tr>
<tr>
<td>Receptor activity modifying protein 1</td>
<td>RAMP1</td>
<td>Upper: 5’-ATCACCTCTTCATAGCACCAGTGC-3’</td>
</tr>
<tr>
<td>Glyceraldehyde-3-phosphate dehydrogenase (control)</td>
<td>GAPDH</td>
<td>Lower: 5’-TACACATGCAACGCATTGCTAAGTC-3’</td>
</tr>
</tbody>
</table>
of Health Statistics and Research; partially by Center for Chronic Disorders of Aging, Philadelphia College of Osteopathic Medicine, Philadelphia, PA, USA, and partially by the Christian R. Mary F. Lindback Foundation, Philadelphia, PA, USA. The authors report no conflicts of interest in this work.

References


