TMPRSS2 Fusions with Oncogenic ETS Factors in Prostate Cancer Involve Unbalanced Genomic Rearrangements and Are Associated with HDAC1 and Epigenetic Reprogramming

Kristiina Iljin,1 Majia Wolf,1 Henrik Edgren,1 Santosh Gupta,1 Sami Kilpinen,1 Rolf I. Skotheim,1,2 Mari Peltola,1 Frank Smit,3 Gerald Verhaegh,3 Jack Schalken,1 Matthias Nees,1 and Olli Kallioniemi1

1Medical Biotechnology, VTT Technical Research Centre of Finland and University of Turku, Turku, Finland; 2Department of Cancer Prevention, Institute for Cancer Research, Rikshospitalet-Radiumhospitalet Medical Center, Oslo, Norway; and 3Department of Urology, Radboud University Nijmegen Medical Center, the Netherlands

Abstract

Translocations fusing the strong androgen-responsive gene, TMPRSS2, with ERG or other oncogenic ETS factors may facilitate prostate cancer development. Here, we studied 18 advanced prostate cancers for ETS factor alterations, using reverse transcription-PCR and DNA and RNA array technologies, and identified putative ERG downstream gene targets from the microarray data of 410 prostate samples. Out of the 27 ETS factors, ERG was most frequently overexpressed. Seven cases showed TMPRSS2:ERG gene fusions, whereas the TMPRSS2:ETV4 fusion was seen in one case. In five out of six tumors with high ERG expression, array-CGH analysis revealed interstitial 2.8 Mb deletions between the TMPRSS2 and ERG loci, or smaller, unbalanced rearrangements. In silico analysis of the ERG gene coexpression patterns revealed an association with high expression of the histone deacetylase 1 gene, and low expression of its target genes. Furthermore, we observed increased expression of WNT-associated pathways and down-regulation of tumor necrosis factor and cell death pathways. In summary, our data indicate that the TMPRSS2:ERG translocation is common in advanced prostate cancer and occurs by virtue of unbalanced genomic rearrangements. Activation of ERG by fusion with TMPRSS2 may lead to epigenetic reprogramming, WNT signaling, and down-regulation of cell death pathways, implicating ERG in several hallmarks of cancer with potential therapeutic importance. (Cancer Res 2006; 66(21): 10242-6)

Introduction

Fusion transcripts arising from translocations of a prostate-specific TMPRSS2 gene with oncogenic ETS factors, ERG, ETV1, or ETV4, were recently identified in up to half of all human prostate cancers (1, 2). Although both translocation partners have been previously studied in prostate cancer (3, 4), the identification of a translocation between the two genes provides new insights into the mechanisms of androgen-dependent prostate tumorigenesis. TMPRSS2 encodes for an androgen-dependent transmembrane serine protease which is strongly expressed in both normal and prostate cancer tissues. ETS transcription factors are characterized by a conserved DNA binding domain and they have been previously implicated in oncogenic translocations in Ewing’s sarcomas, leukemias, lymphomas, fibrosarcomas, and secretory breast carcinomas (5–7). ETS factors regulate the expression of genes with important cancer-relevant biological processes, such as cell growth, differentiation, and transformation (5). However, their specific roles in prostate cancer development and progression have remain unclear, and the molecular consequences of the TMPRSS2:ERG fusion gene formation are unknown. The aims of our study were to explore the role of ETS factor alterations in advanced hormone-refractory prostate cancers, to study the genetic mechanisms leading to oncogenic ETS gene fusions, and to identify potential downstream pathways associated with the ETS activation.

Materials and Methods

Patient data and prostate cancer cell lines. Nineteen advanced prostate cancer samples were obtained from 18 patients. The tissue samples included into this analysis were leftovers after the clinical diagnosis was completed, and were used according to the contemporary guidelines and informed consent of the patients was obtained. The frozen tissue blocks were sectioned, and 20-μm sections were collected for DNA and RNA extractions. Clinical data of the patients are summarized in the Supplementary Table S1. In addition, five prostate cancer cell lines; VCaP, 22Rv1, DU145, LNCaP, and PC-3 were included in the analyses.

Sample preparation, gene copy number, and expression data. DNA was extracted after overnight protease K treatment using standard protocols, whereas total RNA was extracted using either TRIzol (Invitrogen, Carlsbad, CA) or LICI/urea isolation-based methods. Genome-wide data for DNA copy number and matching gene expression data from the same samples will be described in detail elsewhere. The data were merged to analyze the corresponding copy number and expression levels for the ERG, ETV4, ETV1, and TMPRSS2 loci. Briefly, array-based CGH was done using human genome CGH 44A and 44B oligo microarrays according to the version 2 protocol provided by Agilent Technologies (Palo Alto, CA), with minor modifications. Male genomic DNA (Promega, Madison, WI) was used as a reference in all hybridizations. Three micrograms of digested tumor DNA and reference DNA were labeled using Cy5-dUTP and Cy3-dUTP incorporation (Perkin-Elmer, Wellesey, MA) and the Bioprime Array CGH genomic labeling module (Invitrogen). A laser confocal scanner (Agilent Technologies) was used to obtain signal intensities from targets, and Agilent Feature Extraction software (version 8.1.1.1) was applied using the manufacturer’s recommended settings. To analyze the aCGH data, we used the CGH Analytics software (version 3.2.32; Agilent Technologies).

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Requests for reprints: Olli Kallioniemi, Medical Biotechnology Centre, VTT Technical Research Centre of Finland and University of Turku, Turku, Finland. Phone: 358-2478-8602, Fax: 358-2478-8601; E-mail: Olli.Kallioniemi@vtt.fi.

©2006 American Association for Cancer Research. doi:10.1158/0008-5472.CAN-06-1986

4 M. Wolf et al., in preparation.
Gene expression levels were measured using Affymetrix GeneChip Human Genome U133 Plus 2.0 arrays (Affymetrix, Santa Clara, CA). Sample processing and labeling were done according to the protocol provided by Affymetrix. Three micrograms of total RNA from each sample was used for the initial one-cycle cDNA synthesis. Arrays were scanned immediately after staining using a GeneChip scanner (Affymetrix).

Reverse transcription-PCR to identify TMPRSS2:ERG, TMPRSS2-ETV1, and TMPRSS2-ETV4 fusion transcripts. Reverse transcription used 50 ng of total RNA, Sensiscript reverse transcription kit (Qiagen, Valencia, CA) and oligo dT-primers. The fusion transcripts were PCR-amplified, using gene-specific primers for TMPRSS2 (5’-CAGAGCTTGCATAACAGGAGGCGGAGC-3’), ERG (5’-CATAGTTAGTGACCCAGGAGGCGG-3’), ETV1 (5’-TGTGGTGGGAAAGGGATGTCT-3’), and ETV4 (5’-CAAGGTCCGCTGCTTGCCTCTG-3’) cDNAs. The PCR was done with Phusion High-Fidelity DNA polymerase (Finzymes, Espoo, Finland). All PCR experiments included reverse transcription (RT)-negative controls and a blank with no template. PCR products were isolated from agarose gels, treated with Taq polymerase to generate polyadenylc acid overhangs, and cloned into pCRII-TOPO cloning vector (Invitrogen). Sequencing reactions using the same primers as for amplification were prepared by using the ABI BigDye Terminator V3.1 cycle sequencing kit, according to the manufacturer’s instructions and analyzed on the ABI 3100 genetic Analyzer (Applied Biosystems, Foster City, CA).

Data filtering and normalization, clustering, and statistical analysis. Affymetrix U133 Plus 2.0 arrays were normalized using It (8) and the RMA (9) implementation in Bioconductor package affy. Multiple probe sets mapping to the same genes were combined using mean values. Both genes and samples were clustered hierarchically using Euclidean distance and complete linkage analyses.

In silico analysis of potential ERG target genes. We assembled expression data from 410 human prostate tissue samples, consisting of 178 normal samples and 232 tumors and metastases (Supplementary Table S2). We analyzed the patterns of ERG coexpressed genes by four independent methods. First, significance analysis of microarrays (SAM; ref. 10) was done to identify genes that correlate with ERG in prostate tissues. Prostate tissue samples from the HG-195A platform were divided into ERG-positive samples (group 2, n = 16), expressing ERG at higher levels than any of the normal samples, and into ERG-negative samples including both normal and tumor tissues (group 1, n = 93). Normalized gene expression data for samples used in SAM are provided in Supplementary Table S3. Another SAM analysis was done with ERG-positive (n = 16) versus ERG-negative (n = 48) tumor samples only. The false discovery rate was set to zero in both analyses. Second, hierarchical clustering analysis was done to identify the genes coexpressed with ERG in prostate samples in an unsupervised manner. Third, we identified genes whose expression was most closely associated with that of ERG’s by using a Perl implementation of Pearson’s correlation (correlation factors >0.5 or ≤0.5). Finally, Gene Ontology analysis using the DAVID GO analysis tool (http://www.niaid.abcc.ncifcrf.gov) and gene set enrichment analysis (Broad Institute of MIT and Harvard) were done using the same expression data as in the SAM analyses. The data from the three different patient cohorts and the four different analysis methods were overlaid to define the most consistent alterations associated with ERG gene expression.

Results and Discussion

ETS factor gene expression in hormone-refractory prostate cancers. Affymetrix gene expression data for 27 ETS family members were determined for the 14 advanced prostate cancer samples with informative CGH profiles (Fig. 1). ERG was found to be the most frequently up-regulated ETS factor in advanced prostate tumors, including both hormone-refractory (four of nine, nos. 3, 4, 7, and 13) and untreated clinically advanced prostate cancers (two of five, nos. 14 and 16). None of the ETS factors were consistently associated with hormone-refractory tumors. Androgen receptor expression was highly increased in seven of nine hormone-refractory prostate cancers (Supplementary Table S4). However, the androgen receptor expression levels were not associated with ERG activation. In androgen receptor–overexpressing cancers, the ERG expression was 102 ± 110 (mean ± SD) ranging from 16 to 264, whereas in other advanced prostate cancers, ERG expression was 179 ± 208 (range, 12–481). Therefore, our results show that the ERG overexpression is a frequent event in prostate cancer, but that this alteration is not associated with the hormone-refractory nature of the tumors, nor with the androgen receptor expression levels.

TMPRSS2-ERG fusion genes in prostate cancer cell lines and tumors. Nineteen tumor samples and five prostate cancer cell lines were screened for the presence of the three previously identified TMPRSS2:ETS transcription factor fusions by RT-PCR. A fragment of the expected size was amplified with TMPRSS2:ERG-specific primers from the VCaP cell line (1), whereas the other prostate cancer cell lines analyzed were negative. The TMPRSS2:ERG fusion transcript was also detected in 7 of the 19 prostate cancer samples (Fig. 1; Supplementary Table S4), indicating that the fusion transcript is expressed in ∼40% of advanced prostate cancers. Six of these samples (nos. 3, 4, 7, 13, 14, and 16) were strongly positive, whereas in one advanced prostate cancer sample (no. 10), the fusion transcript was expressed at very low levels (data not shown). Sequence analysis of the RT-PCR products indicate that the most common fusion transcript was a fusion of exon 1 of TMPRSS2 with the beginning of exon 4 of ERG (TMPRSS2:ERGa, ref. 1). In one tumor (no. 10), exon 1 of TMPRSS2 was fused to the beginning of exon 2 of ERG (TMPRSS2:ERGb, ref. 1). In sample no. 16, exon 3 of TMPRSS2 was fused to the beginning of exon 4 of ERG. In tumor no. 10, there was also a more complicated fusion
consisting of exon 2 of TMPRSS2 fused to 95 nucleotides identical to a segment of ERG splice form 6 (ERG6 mRNA, AY204740.1; bp, 225-286), followed by the entire exon 4 of ERG. The schematic presentation of the novel TMPRSS2:ERG fusion transcript is presented in Supplementary Fig. S1. In summary, the sequence analysis of TMPRSS2:ERG fusion transcripts indicated significant variation in fusion breakpoints.

**Genomic rearrangements at the ERG locus by array-CGH analysis.** Genomic rearrangements, either deletions at the ERG locus or interstitial deletions between the TMPRSS2 and ERG loci, were identified by array CGH in five out of the six samples displaying TMPRSS2:ERG gene fusions with high ERG expression (nos. 3, 4, 13, 14, and 16; Fig. 2). This indicates that the ERG activation is not caused by simple balanced translocation, but by a variety of unbalanced genetic rearrangements that bring together these two adjacent loci. The proximity of the two genes (with a genomic distance of only 2.8 Mb), and location in the same DNA strand, may facilitate the fusion gene formation and allow a simple intergenic deletion to activate the ERG gene by the fusion to TMPRSS2. This was the most prevalent genetic alteration by array-CGH. In one of the tumors (no. 4), there were two microdeletions both at the TMPRSS2 and ERG loci (measuring 911 and 159 kb, respectively), suggesting that complex genetic rearrangements led to fusion gene formation.

**TMPRSS2-ETV1 and ETV4 fusion genes.** The occurrence of TMPRSS2:ETV1 fusion (1, 11) could not be detected in any of our prostate cancer samples. Recently, another rearrangement fusing an 8 kb upstream region of TMPRSS2 and a third member of the ETS transcription factor family, ETV4, was described (2). In two of our tumor samples (nos. 8 and 18), ETV4 overexpression was detected. The results from the RT-PCR analyses indicate that tumor no. 8 contains a TMPRSS2:ETV4 fusion transcript, but all other samples were negative.

**In silico analysis of ERG coexpressed genes in prostate tumors.** The ERG gene is the most consistently overexpressed oncogene in malignant epithelial cells of the prostate (4). However, its functional role in prostate cancer development and progression has not yet been clearly determined. To identify potential ERG

---

**Figure 2.** Array-based CGH data showing deletions between ERG and TMPRSS2 in advanced prostate cancers and metastases. Chromosome 21 copy number profiles at ERG/TMPRSS2 region at q22.2-q22.3 from specimens with a reduction of copy number. In sample no. 10, only a modest indication of copy number reduction was detected, whereas in sample no. 18, the actual breakpoint occurred proximally from ERG. Examples from cases with interstitial deletion (sample no. 14) and smaller microdeletions affecting ERG and TMPRSS2 loci (sample no. 4) are also shown.
target genes and deregulated biological processes in vivo, we analyzed gene coexpression data from three different prostate data sets, altogether consisting of 410 human prostate tissue samples. The largest data set \((n = 284)\) consisted of nine previously published Affymetrix gene expression studies. These data were normalized to render them directly comparable.5 Expression data were analyzed by multiple statistical methods to characterize the most consistent ERG-associated genes. First, the results from SAM analysis indicated that 136 genes were significantly differentially expressed between ERG-positive and ERG-negative prostate samples; 92 genes were positively correlated (>1.5-fold change, positive hits), and 44 genes were negatively correlated with ERG (Supplementary Table S5). SAM was also done with prostate cancers only (Fig. 3). Second, the ERG clusters generated by K means and hierarchical clustering showed a 46% and 59% overlap, respectively, with the positive SAM hit list. Third, ERG correlation analyses across all prostate samples \((n = 284)\) or all prostate cancer samples \((n = 147)\) were done. For the top 200 genes correlating with ERG in prostate cancers, there was an 85% overlap with the SAM-positive hits using linear correlation and a 75% overlap for log-transformed correlation.

For validation, we used an independent large prostate data set \((n = 112)\) based on cDNA microarray analysis (12). Half of the top 200 positively ERG-correlating genes in the second data set were represented on Affymetrix microarray platforms, and of these, 46% overlapped with the positive SAM hits.

The third data set \((n = 14)\) consisted of the advanced prostate cancers analyzed in this study. Although the number of samples was small and many samples were derived from hormone-refractory prostate cancers, three genes among the top ERG-correlating genes were also among the top 10 positive SAM hit gene list. The correlation results from the three different data sets overlaid with the SAM results are presented in Supplementary Table S5.

Histone deacetylase 1 (HDAC1) was the only gene among the top ERG coexpressed genes in all three data sets and therefore was the most consistent feature of ERG-overexpressing prostate cancers. By RT-PCR validation of the HDAC1 expression levels, all ERG-positive

Figure 3. Identification of ERG-associated genes in prostate cancer. A, scatter plot of the observed relative difference score (X-axis) versus the expected relative difference score (Y-axis) from the comparison of ERG-positive prostate cancers with ERG-negative prostate cancers by SAM. ERG and HDAC1 are indicated. The gene with the second highest observed d-score is PEX10, which was not found among the top correlating genes with ERG in the other data sets analyzed and is therefore not further discussed. B, coexpression plot of ERG and HDAC1 in prostate cancer samples used in SAM. Red, samples with high ERG expression; black, samples with low ERG expression. C, heat-map of the SAM-positive hits. The most highly correlated genes with ERG are indicated.

5 Autio et al., in preparation.
prostate cancers were strongly HDAC1-positive, whereas the ERG-negative tumors showed more variable, but significantly lower ($P = 0.0001$) expression levels (Supplementary Fig. S2). HDAC1 catalyzes the deacetylation of lysine residues of the core histones and other proteins, leading to epigenetic silencing of target genes. HDAC1 has been previously shown to be strongly expressed in hormone-refractory prostate cancers (13). Currently, it remains unclear if HDAC1 is a direct transcriptional target of ERG or whether HDAC1 up-regulation results from other changes occurring in ERG-overexpressing tumors. ERG has been shown to interact indirectly with HDAC1 via SETDB1 methyltransferase (14, 15).

Results from gene ontology analyses indicated that the “organogenesis” as well as “cell growth and maintenance” were the most significantly ($P = 0.001$ and 0.02) overrepresented gene ontology terms in ERG-positive tumors. Gene set enrichment analysis results, presented in Fig. 4, indicated that the WNT and P1T2X2 pathways were among the most highly enriched pathways in ERG-overexpressing tumors (16, 17). The WNT pathway controls organogenesis by inducing, e.g., P1T2X2 transcription factor, which serves as an important modulator of growth control genes (17). HDAC1 itself is linked to these two pathways. The down-regulated gene sets in ERG-overexpressing tumors included CCR5, cell death, and tumor necrosis factor/FAS. Interestingly, the HDAC1 pathway with known HDAC target gene regulators was also highlighted by this analysis, suggesting that the up-regulation of HDAC1 in ERG-positive tumors led, as could be expected, to the down-regulation of HDAC target genes (18, 19). The genes showing core enrichment in the identified target pathways are presented in Supplementary Table S6.

In conclusion, we have taken the first steps to shed new light into the mechanism of disease in ERG-positive prostate cancers by discovering that ERG may contribute, directly or indirectly, to several different phenotypic hallmarks of cancer. Of significant potential therapeutic interest are epigenetic mechanisms, involving HDAC1 up-regulation and target gene silencing. HDAC inhibitors have been tested in animal models of prostate cancer and have shown promising antitumor activity (20). Consistent with this hypothesis, ERG-positive prostate cancer cell lines, VCaP and DuCaP, showed strong responses with anti-HDAC drug trichostatin A (Supplementary Fig. S3). Based on these early stage in vitro functional studies and the extensive in vivo correlations, we suggest that patients with ERG-positive prostate cancer could benefit from epigenetic therapy with HDAC inhibitors and other epigenetic drugs, perhaps in combination with traditional treatments, such as androgen-deprivation therapy. This study also illustrates how the linking of a pathogenetic event—the ERG gene activation, with its downstream phenotypic consequences of therapeutic potential, such as sensitivity to epigenetic compounds—could pave the way towards future individualized therapies for human prostate cancer.

Acknowledgments

Received 5/31/2006; revised 9/5/2006; accepted 9/21/2006.

Grant support: Marie Curie Canceromics (MEXT-CT-2003-2728), EU-PRIMA project (contract no. LSHC-CT-2004-504876), Academy of Finland, Sigrid Juselius Foundation, and the Cancer Organizations of Finland.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

References


Figure 4. Gene sets showing significant enrichment among ERG-positive prostate cancers. Both the up-regulated (nominal $P < 0.03$, enrichment score >0.8) and down-regulated (nominal $P < 0.03$, enrichment score, <0.5) gene sets in ERG-positive prostate cancers identified by gene set enrichment analysis. Enrichment score is plotted in the Y-axis. Stars, the significance of the enrichment score (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).
# TMPRSS2 Fusions with Oncogenic ETS Factors in Prostate Cancer Involve Unbalanced Genomic Rearrangements and Are Associated with HDAC1 and Epigenetic Reprogramming

Kristiina Iljin, Maija Wolf, Henrik Edgren, et al.


<table>
<thead>
<tr>
<th>Updated version</th>
<th>Access the most recent version of this article at: <a href="http://cancerres.aacrjournals.org/content/66/21/10242">http://cancerres.aacrjournals.org/content/66/21/10242</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary Material</td>
<td>Access the most recent supplemental material at: <a href="http://cancerres.aacrjournals.org/content/suppl/2006/11/02/66.21.10242.DC1">http://cancerres.aacrjournals.org/content/suppl/2006/11/02/66.21.10242.DC1</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cited articles</th>
<th>This article cites 19 articles, 9 of which you can access for free at: <a href="http://cancerres.aacrjournals.org/content/66/21/10242.full.html#ref-list-1">http://cancerres.aacrjournals.org/content/66/21/10242.full.html#ref-list-1</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Citing articles</td>
<td>This article has been cited by 31 HighWire-hosted articles. Access the articles at: <a href="http://cancerres.aacrjournals.org/content/66/21/10242.full.html#related-urls">http://cancerres.aacrjournals.org/content/66/21/10242.full.html#related-urls</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E-mail alerts</th>
<th>Sign up to receive free email-alerts related to this article or journal.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reprints and Subscriptions</td>
<td>To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at <a href="mailto:pubs@aacr.org">pubs@aacr.org</a>.</td>
</tr>
<tr>
<td>Permissions</td>
<td>To request permission to re-use all or part of this article, contact the AACR Publications Department at <a href="mailto:permissions@aacr.org">permissions@aacr.org</a>.</td>
</tr>
</tbody>
</table>