NEOPLASIA
www.neoplasia.com

Volume 11 Number 8 August 2009 pp. 804–811 804

N-myc Downstream Regulated Gene 1 (*NDRG1*) Is Fused to *ERG* in Prostate Cancer[1,2]

Dorothee Pflueger*,†,3, David S. Rickman*,3, Andrea Sboner‡,3, Sven Perner*, Christopher J. LaFargue*, Maria A. Svensson*, Benjamin J. Moss*, Naoki Kitabayashi*, Yihang Pan*, Alexandre de la Taille§,¶, Rainer Kuefer†, Ashutosh K. Tewari#, Francesca Demichelis*,, Mark S. Chee††, Mark B. Gerstein‡,‡‡,§§ and Mark A. Rubin***

*Department of Pathology & Laboratory Medicine, Weill Cornell Medical College, New York, NY, USA; †Department of Urology, University Hospital Ulm, Ulm, Germany; ‡Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA; §INSERM, Unité 955, Créteil, France; ¶Department of Urology, CHU Mondor, Créteil, France; #Department of Urology, Weill Cornell Medical College, New York, NY, USA; **Institute for Computational Biomedicine, Weill Cornell Medical College, New York, NY, USA; ††Prognosys Biosciences Inc., La Jolla, CA, USA; ‡‡Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA; §§Department of Computer Science, Yale University, New Haven, CT, USA

Abstract

A step toward the molecular classification of prostate cancer was the discovery of recurrent erythroblast transformation–specific rearrangements, most commonly fusing the androgen-regulated *TMPRSS2* promoter to *ERG*. The *TMPRSS2-ERG* fusion is observed in around 90% of tumors that overexpress the oncogene *ERG*. The goal of the current study was to complete the characterization of these ERG-overexpressing prostate cancers. Using fluorescence *in situ* hybridization and reverse transcription–polymerase chain reaction assays, we screened 101 prostate cancers, identifying 34 cases (34%) with the *TMPRSS2-ERG* fusion. Seven cases demonstrated *ERG* rearrangement by fluorescence *in situ* hybridization without the presence of TMPRSS2-ERG fusion messenger RNA transcripts. Screening for known 5′ partners, we determined that three cases harbored the *SLC45A3-ERG* fusion. To discover novel 5′ partners in these ERG-overexpressing and *ERG*-rearranged cases, we used paired-end RNA sequencing. We first confirmed the utility of this approach by identifying the *TMPRSS2-ERG* fusion in a known positive prostate cancer case and then discovered a novel fusion involving the androgen-inducible tumor suppressor, *NDRG1* (*N-myc downstream regulated gene 1*), and *ERG* in two cases. Unlike *TMPRSS2-ERG* and *SCL45A3-ERG* fusions, the *NDRG1-ERG* fusion is predicted to encode a chimeric protein. Like *TMPRSS2*, *SCL45A3* and *NDRG1* are inducible not only by androgen but also by estrogen. This study demonstrates that most ERG-overexpressing prostate cancers harbor hormonally regulated *TMPRSS2-ERG*, *SLC45A3-ERG*, or *NDRG1-ERG* fusions. Broader implications of this study support the use of RNA sequencing to discover novel cancer translocations.

Neoplasia (2009) 11, 804–811

Abbreviations: RT-PCR, reverse transcription–polymerase chain reaction; FISH, fluorescence *in situ* hybridization; RNA-seq, RNA sequencing; (b/a), break-apart

Address all correspondence to: Mark A. Rubin, MD, Pathology and Laboratory Medicine, Weill Cornell Medical College, 1300 York Ave, Room C 410-A, New York, NY 10021. E-mail: rubinma@med.cornell.edu

[1]This work was supported by National Institutes of Health (NIH)/National Cancer Institute grant R01 CA125612 (M.A.R. and F.D.), Heinrich Warner Foundation (D.P.), Department of Defense grant PC61474 (S.P.), and NIH/National Human Genome Research Institute (NHGRI) grant R44HG004237 (M.S.C.). The authors thank the support of the "Yale University Biomedical High Performance Computing Center" and NIH grant no. RR19895 that funded the computer cluster instrumentation.
[2]This article refers to supplementary materials, which are designated by Tables W1 to W4 and Figures W1 to W5 and are available online at www.neoplasia.com.
[3]These authors contributed equally to this work.
Received 3 April 2009; Revised 16 May 2009; Accepted 20 May 2009

DOI 10.1593/neo.09572

Introduction

Most prostate cancers detected through prostate-specific antigen (PSA) screening harbor an acquired recurrent chromosomal rearrangement [1]. The promoter region of the androgen-regulated transmembrane protease, serine 2 (*TMPRSS2*) gene, is most often fused to the coding region of members of the erythroblast transformation–specific (ETS) family of transcription factors, most commonly v-ets erythroblastosis virus E26 oncogene homolog (avian) (*ERG*). Other, less common, fusion events occur involving ETS family members (*ETV1*, *ETV4*, and *ETV5*) fused to *TMPRSS2* or other 5′ partners that differ in their prostate specificity and response to androgen (*SLC45A3*, *HERV-K*, *C15orf21*, *HNRPA2B1*, *FLJ35294*, *DDX5*, *CANT1*, and *KLK2*, reviewed by Kumar-Sinha et al. [2] and more recently, *ACSL3* [3]). Moreover, variations in the structure of the gene fusions in prostate cancer yielding different fusion transcript isoforms have been reported [4].

Emerging data suggest that *ETS*-rearranged prostate cancer, similar to other translocation tumors, represent a distinct molecular subclass of prostate cancer based on studies demonstrating characteristic morphologic features [5], natural history [6,7], and specific genomic [8] and expression profiles [9]. Herein, we report a comprehensive characterization for *ERG* gene rearrangements in prostate cancer including the identification of a novel hormone-regulated 5′ fusion partner using paired-end RNA sequencing (RNA-seq).

Materials and Methods

Patient Population

The study is composed of 101 men with localized and locally advanced prostate cancer who underwent radical prostatectomy as a monotherapy. All prostate cancer cases were collected as part of institutional review board–approved research protocols.

Sample Processing for RNA Analyses

Hematoxylin and eosin slides were prepared from formalin-fixed paraffin-embedded material and evaluated for cancer extent and tumor grade (Gleason score). Hematoxylin and eosin slides were prepared from the corresponding frozen tissue block and evaluated for the extent of cancer involvement. To ensure for a high concentration of cancer cells and minimized benign tissue, tumor isolation was performed by first selecting for high-density cancer foci (<10% stromal and other nontumor tissue contamination) and then taking 1.5-mm biopsy cores from the frozen tissue block for RNA extraction. Sections for fluorescence *in situ* hybridization (FISH) evaluation were taken from the frozen tissue block used for molecular analysis. The cancer foci selected for RNA extraction were well characterized by FISH to evaluate the *ERG* rearrangement status throughout the entire focus. We took special care to extract the RNA from a single cancer focus to exclude the problem of heterogeneity when looking for putative fusion transcripts. RNA was isolated from frozen tissue using TRIzol LS reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. After DNase treatment (Invitrogen), RNA concentration was measured using a NanoDrop 8000 spectrophotometer (Thermo Scientific, Wilmington, DE). Quality was assessed using the Bioanalyzer 2100 (Agilent Technologies, Inc, Santa Clara, CA). The qualitative detection of fusion transcripts in the cases was performed using conventional reverse transcription–polymerase chain reaction (RT-PCR), agarose gel fractionation/purification, and subsequent complementary DNA (cDNA) sequencing. For this, amplified DNA fragments corresponding to the expected sizes of fusion transcripts were gel-extracted using the MinElute Gel Extraction Kit (Qiagen, Valencia, CA) and sequenced at the Life Sciences Core Laboratories Center's DNA sequencing facility of Cornell University (Ithaca, NY). Quantitative ERG and TMPRSS2-ERG RT-PCR was performed using QuantiTect SYBR Green PCR Kit (Qiagen). Each sample was run in duplicate. The amount of each target gene relative to a control gene was determined using the comparative C_t method (ABI Bulletin 2; Applied Biosystems, Foster City, CA). C_t values for ERG were first normalized using the average C_t values obtained for SART3 and TCFL1/VPS72 and then calibrated using normalized C_t values obtained from benign prostate. Protocols and primers are shown in Table W2.

Assessment of ERG, TMPRSS2, SLC45A3, *and* NDRG1 *Rearrangements Using Two-color FISH Assays*

To assess for rearrangement of *ERG*, *TMPRSS2*, *SLC45A3*, and *NDRG1*, we used break-apart (b/a) FISH assays for each gene and fusion assays for *SLC45A3-ERG* or *NDRG1-ERG* on sections from the corresponding frozen tissue blocks. The centromeric probes for *ERG*, *TMPRSS2*, *SLC45A3*, and *NDRG1* were RP11-24A11, RP11-354C5, RP11-249H15, and RP11-185E14, respectively. The telomeric probes for *ERG*, *TMPRSS2*, *SLC45A3*, and *NDRG1* were RP11-372O17, RP11-891L10, RP11-131E5, and RP11-1145H17, respectively. We used probes RP11-131E5 (*SLC45A3*), RP11-1145H17 (*NDRG1*), and RP11-24A11 (*ERG*) for the *SLC45A3-ERG* and *NDRG1-ERG* fusion assays. Correct chromosomal probe localization was confirmed on normal lymphocyte metaphase preparations (see Figure W4 for metaphase results for bacterial artificial chromosomes (BACs) targeting *NDRG1* locus). For each sample, a minimum of 100 nuclei were analyzed.

RNA-seq and Data Analysis

We used the Illumina Genome Analyzer II for paired-end RNA-sequencing. This provided a pair of approximately 30 to 36 base reads, from each end of a transcript fragment of relatively well-defined length (approximately 330 nucleotides). The paired reads were aligned independently to the human genome (hg18 assembly in the UCSC genome browser[1]: Homo_Sapiens March 2006) using "eland," a short-read alignment tool included in the Genome Analyzer software suite. For each read, eland provides the coordinate(s) of the alignment to the reference genome, allowing for up to two mismatches in the sequence. We kept only the reads that are mapped uniquely to the genome, although they might have up to two mismatches. To search for novel translocations involving *ERG*, two strategies were applied. First, we selected for mapped paired reads that are more than 2 Mb apart. This allows us to identify translocations similar to TMPRSS2-ERG messenger RNA (mRNA). Indeed, the two genes are approximately 3 Mb apart. Second, paired reads mapping to different chromosomes were also selected as potential candidates. Because we focused on novel *ERG* partners, we selected for paired reads where one of the reads lies within *ERG*. This allowed us to identify several candidate fusion transcripts spanning all chromosomes. We finally selected the chromosome with the highest number of reads and checked if those reads lie within a gene. This approach yielded numerous leads, which, because of the low number of copies (from one to three reads), were considered background. We also identified numerous examples of putative fusions

[1] http://genome.ucsc.edu.

that ambiguously mapped to multiple sites along the reference genome probably because of the small size of the paired reads (between 30 and 36 bp). Sequences for NDRG1-ERG v1 and v2 have been submitted to GenBank (Accession Nos. FJ627786 and FJ627787, respectively).

Hormone Treatment of LNCaP Cells

The prostate cancer cell line LNCaP was obtained from ATCC (Manassas, VA; catalog No. CRL-1740) and maintained according to the supplier's instructions. For hormonal treatment, cells were plated (500,000 cells/10 cm^2) in the presence of complete growth medium supplemented with 1% penicillin/streptomycin. Cells were starved for 48 hours in charcoal-stripped medium (RPMI-1640 1×, 5% charcoal-stripped FBS, 1% penicillin/streptomycin) and then treated with R1881 (1 nM), 17β-estradiol (10 nM), diarylpropionitrile (DPN, 10 nM) or ethanol vehicle for 3, 12, and 24 hours. RNA was extracted using the TRIzol reagent (Invitrogen), subjected to DNase treatment (DNA-free Kit; Applied Biosystems) according to the manufacturer's instructions. To test for the specificity of androgen stimulation, cells were treated with 10 μM flutamide for 2 hours and then treated with R1881 as described previously. TaqMan assays (Table W3) were used to quantify relative levels of *SLC45A3*, *NDRG1*, *PSA* (*KLK3*), and *IGF1R*.

Results

TMPRSS2-ERG *and* SLC45A3-ERG *Account for 84% of ERG Overexpression in Prostate Cancer*

We screened prostate cancer cases from 101 men with localized and locally advanced prostate cancer who underwent radical prostatectomy for *ERG* gene rearrangement using a FISH b/a assay. In total, 44 cases were positive for *ERG* rearrangement. Given the heterogeneity of TMPRSS2-ERG mRNA expression level [4] in prostate cancer, we screened for TMPRSS2-ERG mRNA variant expression using conventional RT-PCR and cDNA sequencing. Of the 44, 34 (77%) expressed seven different variants of TMPRSS2-ERG mRNA described by Wang et al. [4]. To determine the level of ERG mRNA overexpression, we performed quantitative PCR using cDNA from 29 cases (19 that were TMPRSS2-ERG mRNA-positive and 10 TMPRSS2-ERG mRNA-negative), 15 cases that did not show *ERG* rearrangement and 6 benign prostate tissue samples (Figure 1*A*). ERG mRNA was overexpressed up to 75 times (median, 27) in *ERG*-rearranged cases compared with baseline levels in benign prostate tissue and cases negative for both *ERG* rearrangement and TMPRSS2-ERG mRNA. Contrary to findings by Wang et al., TMPRSS2-ERG mRNA isoform expression was not associated with ERG overexpression or with prostate cancer progression (Gleason score, pathologic stage, or surgical margin status; Table W1).

TMPRSS2-ERG mRNA was absent in 10 (23% of 44) *ERG*-rearranged cases, of which 7 expressed high ERG mRNA levels (5-38 times). To confirm the absence of *TMPRSS2* rearrangement in these cases, we performed FISH using a *TMPRSS2* b/a assay. We observed *TMPRSS2* rearrangement in 2 of 10 cases (60T and 51T) suggesting a novel *TMPRSS2-ERG* fusion that was not detected using standard RT-PCR approaches. Targeting the exon boundary of exons 1 and 2 in TMPRSS2, we detected a TMPRSS2-ERG fusion transcript in sample 60T that lacks the 5′ end of TMPRSS2 exon 1. This isoform (isoform VII) has been previously reported [4]. To screen for other possible fusion events with *ERG*, we performed RT-PCR analysis targeting known ETS family fusion partners (*SLC45A3*, *HERV-K*, *C15ORF21*, *HNRPA2B1*, *DDX5*, *CANT1*, *KLK2*, and *ACSL3*). This screening revealed that exon 4 of *ERG* was fused to exon 1 of *SLC45A3* in three ERG mRNA-overexpressed cases (34T, 150B_M, and 145C_M; Figure 1*B*). This is consistent with the recent report from Han et al. [10]. The predicted open reading frame is identical to what is encoded by the most common TMPRSS2 (exon 1)–ERG (exon 4) mRNA transcript. We confirmed this fusion *in situ* using an *SLC45A3* b/a assay and *SLC45A3-ERG* fusion assay (Figure 1*C*).

Massively Parallel RNA-seq Discovers NDRG1-ERG *Fusion Prostate Cancer*

Having characterized all but two ERG-overexpressing/*ERG*-rearranged cases (509B and 99T), we used paired-end RNA-seq to identify potential 5′ partners (Table W4). Initially, we explored for fusion transcripts by looking for paired reads where each pair mapped to regions that were either greater than 2 Mb and less than 5 Mb, greater than 5 Mb and less than 10 Mb, or greater than 10 Mb apart from each other on chromosome 21. We also explored for fusion transcripts between ERG and reads that mapped to different chromosomes. Spurious fusion candidates occur for a few reasons. The main reason is the misalignment of the short-read sequences against the reference genome. Another source of potential error is the creation of random chimeric fragments during sample preparation. Therefore, to reduce these false fusion reads, we required that the fusion candidate have several supporting paired reads. First, in prostate cancer cases known to harbor the *TMPRSS2-ERG* fusion (e.g., case 1701A), we detected numerous TMPRSS2-ERG transcripts as the only fusion transcript arising within chromosome 21 as demonstrated in Figure 2. Second, SLC45A3-ELK4 transcripts were also detected in case 1701A as we have previously observed [11]. There were a few potential ERG fusion transcripts with paired reads to different chromosomes with less than four supporting reads suggesting that they are artificial. Finally, in one case (99T) with ERG overexpression but no *SLC45A3* or *TMPRSS2* rearrangement as determined by RT-PCR and FISH, RNA-seq demonstrated 17 copies of a fusion transcript that mapped paired reads to *ERG* and to exons of *NDRG1* (Figure 3*A*). This was confirmed by conventional RT-PCR (Figure 3*B*). Screening other TMPRSS2-ERG, SLC45A3-ERG mRNA–negative cases revealed another, slightly different, NDRG1-ERG transcript variant (variant 2) in 509B. We confirmed this translocation at the genome level using *NDRG1* b/a and *NDRG1-ERG* fusion FISH assays (Figure 3, *C* and *D*).

TMPRSS2-, SLC45A3-, and NDRG1-ERG Are Regulated by Androgen and Estrogen

ERG mRNA expression in cases positive for *SLC45A3-ERG* or *NDRG1-ERG* is similar in magnitude to those measured for *TMPRSS2-ERG*-positive cases. *TMPRSS2* [12], *SLC45A3* [13], and *NDRG1* [14–16] are all known androgen-induced genes. This was confirmed by treating LNCaP with a synthetic androgen (R1881, 1 nM; Figure 4, *A* and *B*). Androgen regulation of *NDRG1* is supported by the observation of an AR binding site ~30 kb upstream of the start site (chr8:134407748-134408779) in LNCaP cells (Figure W3). The induction of gene expression was abrogated in the presence of flutamide. If we consider KLK3 (PSA) mRNA a surrogate read-out of androgen signaling, we might expect to find similar profiles between PSA and ERG mRNA levels in TMPRSS2-ERG, SLC45A3-ERG, or NDRG1-ERG mRNA-positive prostate cancer cases. PSA mRNA levels, however, did not mimic the pattern of ERG mRNA levels in

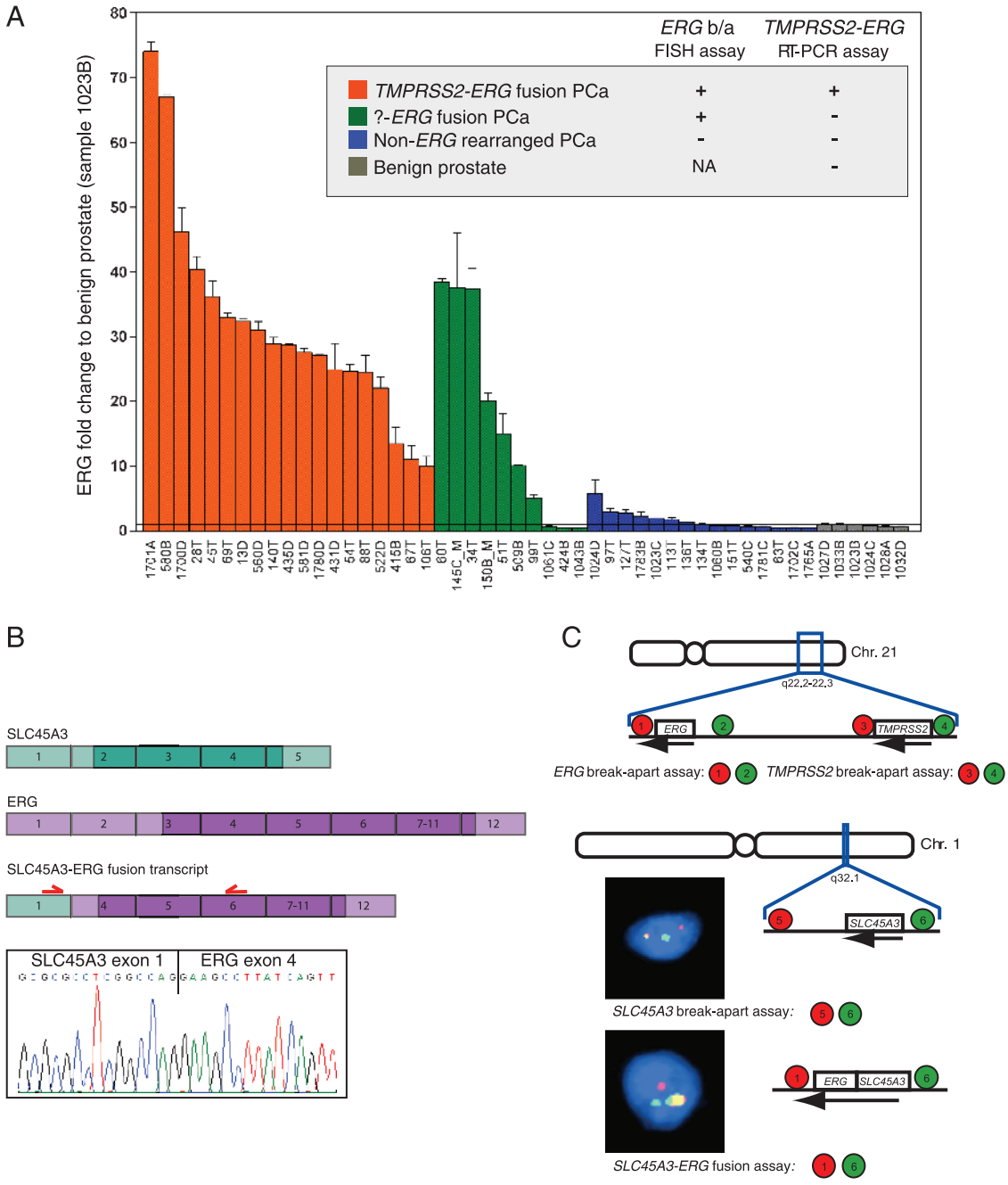

Figure 1. ERG mRNA expression in prostate cancer and benign tissue. (A) Quantitative RT-PCR of ERG expression in 29 *ERG* rearranged (including 19 TMPRSS2-ERG mRNA-positive (orange) and unknown mechanism–ERG (?–ERG, green)), 15 *ERG* nonrearranged (blue), and 6 benign prostate tissue samples (gray). (B) Exon composition and sequence covering the fusion junction of *SLC45A3-ERG* transcript. (C) FISH images of nuclei with *SLC45A3* rearrangement (upper) and SLC45A3-ERG fusion (lower) nucleus with yellow fusion signal.

TMPRSS2-ERG, SLC45A3-ERG, or NDRG1-ERG mRNA-positive cases, suggesting an additional mechanism for the regulation of the fusion transcripts (Figure W2).

We have previously shown that *TMPRSS2-ERG* is regulated by estrogen [9]. *SLC45A3* may also be similarly regulated by estrogen because the chromosome immunoprecipitation data generated by Brown et al. indicate the presence of an estrogen receptor (ER) binding site within the *SLC45A3* gene (referenced in [11]). In addition, their data also show that there is an ER binding site in the first intron of *NDRG1* (chr8:134373799-134375086) and another ER binding site (chr8:134441414-134442401) ~60 kb upstream of the start site (Figure W3). Interestingly, similar data show that FoxA1, a known ER cofactor, binding sites overlap with the ER binding sites. To test if estrogen regulates gene expression, we measured the levels of SLC45A3 or NDRG1 mRNA in LNCaP cells at different time points as a function of estrogen treatment. We observed induction of SLC45A3 mRNA 3 hours (Figure 4*C*) and that of NDRG1 mRNA 12 hours (Figure 4*D*) after 17β-estradiol treatment but not with the ERβ receptor agonist DPN similar to *IGF1R* mRNA, a known estrogen-induced gene in LNCaP cells [17] (Figure W1). These data suggest that, like *TMPRSS2-ERG*, *SLC45A3-ERG* and *NDRG1-ERG* fusion genes might also be estrogen-regulated through ERα. This would provide another mechanism for

ERG overexpression when fused to *SLC45A3* or *NDRG1*, particularly in the case of castration-resistant prostate cancer.

Discussion

The results presented here provide further evidence of the expanding variety and importance of gene fusion events in prostate cancer. Using approaches that require *a priori* selection of candidate genes (RT-PCR and FISH) and an approach that does not require *a priori* gene selection (RNA-seq), we found that there are three main gene fusions that account for ERG-overexpressing prostate cancer. Common to all three gene fusions is a hormonally regulated promoter. The novel *NDRG1-ERG* fusion is unique in that computational sequence analysis of NDRG1-ERG variant 1 cDNA suggests that this fusion, as with *BCR-ABL1* fusion gene in patients with chronic myeloid leukemia, could encode a chimeric protein containing 33 amino acids from *NDRG1* as well as the conserved protein domains of wild-type *ERG* (Sterile alpha motif/Pointed domain and ETS domain). NDRG1-ERG variant 2 mRNA is also predicted to encode a chimeric protein including the first 21 amino acids of *NDRG1* and the same conserved domains of *ERG* as in the putative protein encoded by NDRG1-ERG variant 1. A cell line model could be used to elucidate the functional role of the *NDRG1-ERG* fusion. We screened but did not find NDRG1-ERG mRNA in six widely used, commercially available prostate cancer cell lines (LNCaP, VCaP, 22Rv1, PC-3, NCI-H660, and DU145). We succeeded in transiently transfecting a NDRG1-ERG variant 1 construct into HEK-293 and BPH cells and monitored the ERG expression levels by means of quantitative RT-PCR (qRT-PCR) and fluorescent immunostaining (Figure W5). Using a commercially available antibody directed against ERG, we were able to show increased ERG expression in the transfected cells suggesting that the NDRG1-ERG fusion gives rise to a protein that shares downstream sequences with wild-type ERG (data not shown). An antibody specifically targeting the fusion protein would have to be developed. Our study differs from two recent RNA-seq studies that used either long reads alone [18] or long and short reads [19] to identify multiple non-ERG chimeric mRNA in a breast cancer line or prostate cell lines and metastatic tumor samples, respectively. Here, we relied on 30 to 36 nucleotides paired-end reads from 330 nucleotides mRNA fragments that have the advantage of directly identifying chimeric transcripts. However, short reads cannot always be mapped uniquely to the genome, making it impossible to unambiguously assign a location. As the technology advances, longer paired-end RNA-seq reads will be

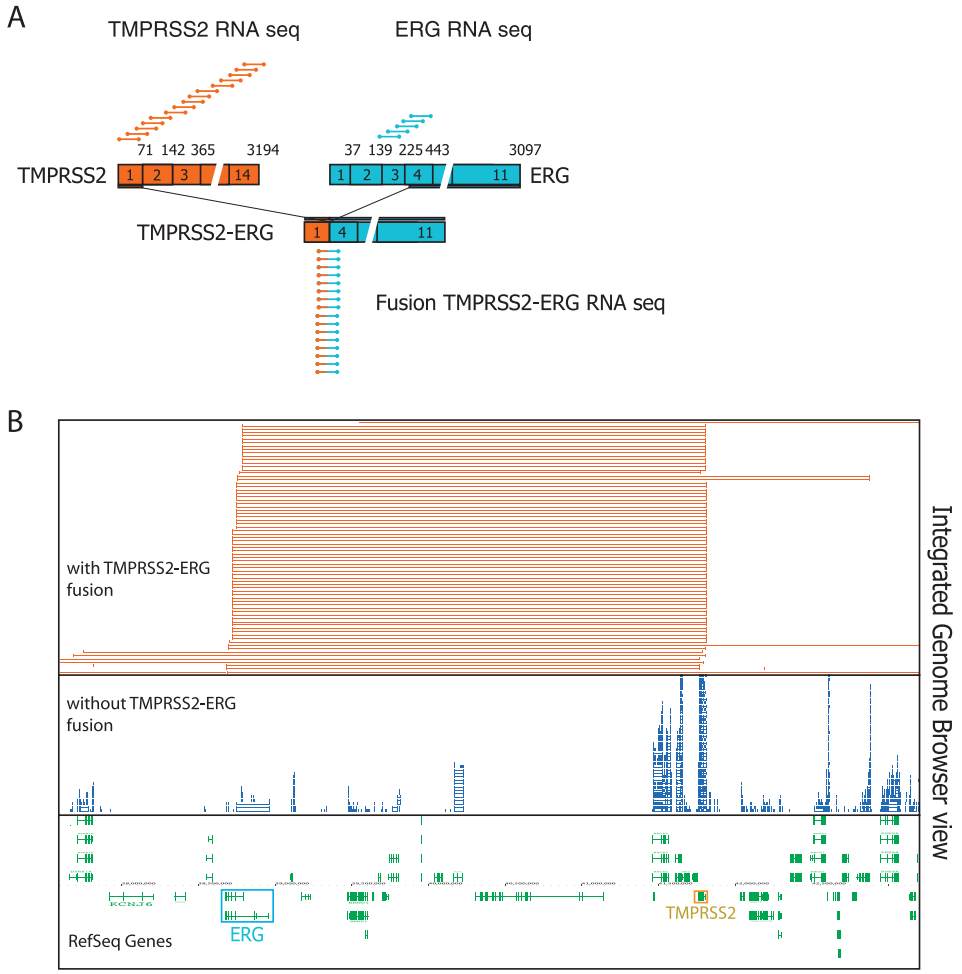

Figure 2. Identification of TMPRSS2-ERG fusion transcript using RNA-seq. (A) For wild-type alleles, the two paired-end reads will map to the exons of each gene, whereas in the presence of the fusion, one read will map to exons of *TMPRSS2* and the other read will map to exons of *ERG*, at a much bigger distance on the genome. (B) Read pairs (<2 Mb apart) are shown as connected segments (top). *TMPRSS2* and *ERG* RefSeq gene annotations/locations are given (bottom). Reads of a *TMPRSS2-ERG* fusion–negative sample (blue) are trimmed at the top for demonstration purposes.

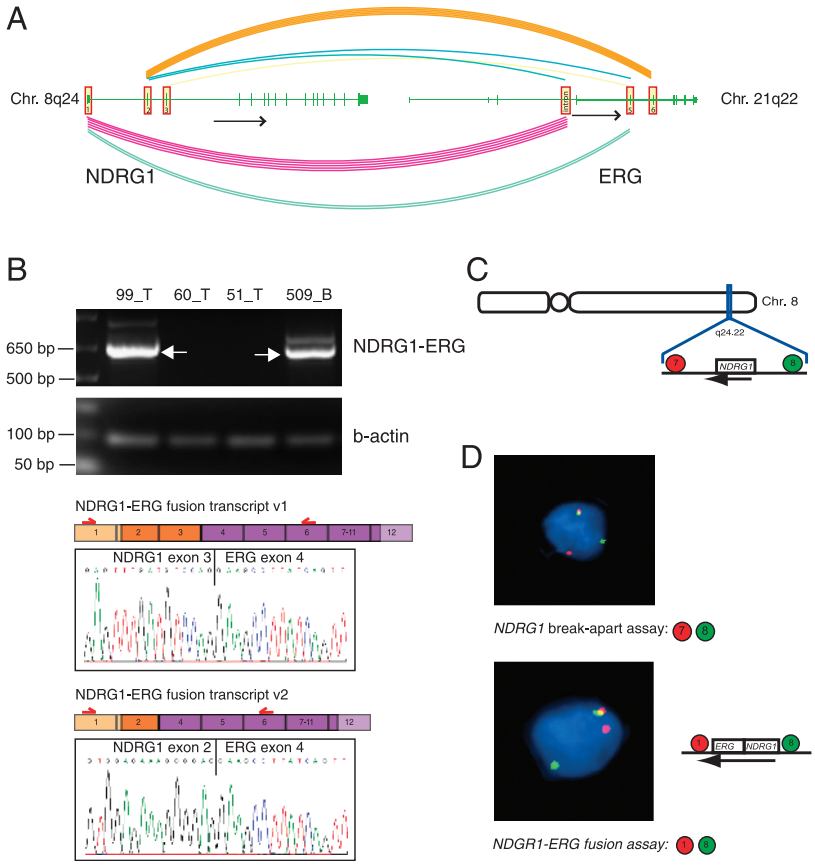

Figure 3. Identification of *NDRG1-ERG* fusion by RNA-seq. (A) The schematic shows the linear structure of *NDRG1* and *ERG*. The gene representation shows the "union" transcripts, that is, the exons of all isoforms are reported, and in the case of overlapping exons, the longest one is shown. Each arc represents one instance of paired reads where one read is mapped to *NDRG1* and the other to *ERG*. The regions of the genes involved in the fusion transcript are highlighted and numbered. (B) Reverse transcription–PCR products obtained using a forward primer targeting exon 1 of *NDRG1* and a reverse primer targeting exon 6 of *ERG* (positive control: β-actin). Arrows indicate the DNA fragments that were isolated and sequenced. Bottom panel shows sequence data from this analysis showing the NDRG1-ERG transcript exon composition and sequence covering the fusion junction for the two variant mRNA (shown schematically) identified in samples 99_T (top) and 509_B (bottom). (C) Schematic of the FISH *NDRG1* b/a and *NDRG1-ERG* fusion assays. (D) *NDRG1* rearrangement (upper) indicated by separated red and green signals and *NDRG1-ERG* fusion (lower) indicated by an overlap of the red *ERG* and the green *NDRG1* signal in a representative nucleus from case 99_T.

available, and the discovery of new and clinically relevant fusion transcripts will become more efficient and affordable. Although this is not the first RNA-seq paper to propose exploring for gene fusions, we were able to exploit paired-end sequencing to discover a novel 5′ fusion partner for *ERG*. This is significant because, unlike some of the other fusions identified to date, we believe that this is recurrent and may represent up to 5% of all *ERG*-rearranged prostate cancers. The approach used to make this discovery is being developed so that we can apply this as a pipeline for other platforms and tumor types. In conclusion, the confirmation of the *TMPRSS2-ERG* and discovery of the *NDRG1-ERG* fusion using paired-end RNA-seq provide the first proof-of-principle that this methodology can reliably discover novel gene fusions in prostate cancer tissue. Alternatively, we could have used 5′RACE as an established way of discovering gene fusions. However, this method requires the knowledge of one partner. Although the RNA-seq analysis approach that was developed can be directly applied to a number of defined regions, such as ERG, we can extend to a genome-wide search of fusions starting from paired-end reads.

The identification of *NDRG1-ERG* fusion prostate cancer has potential clinical and biologic implications. *NDRG1* is involved in cellular differentiation, repressed by the oncogenes N-*myc* and c-*myc* and is thus typically downregulated in cancer cells. Overexpression of NDRG1 has also been associated with reduced metastatic potential [20]. Therefore, hormone-induced overexpression of *NDRG1-ERG* fusion leads to high levels of the functional domains of ERG but could potentially also lead to an increased risk of metastasis due to the disruption of *NDRG1*. The occurrence of metastasis would have to be correlated with differential expression of NDRG1 protein in a larger cohort of *NDRG1-ERG* fusion–positive samples to be statistically reliable. Using only qRT-PCR techniques to indirectly monitor NDRG1 protein levels is not applicable because NDRG1 transcript levels do not necessarily reflect NDRG1 protein levels. There are more immediate clinical implications to our findings. Emerging data suggest that *ERG* rearrangement–positive prostate cancer characterizes a subclass of prostate cancers that have an aggressive natural history when left untreated. The high levels of the unique chimeric transcripts caused by the *ETS* rearrangements are only found in the disease state and not in a benign prostate gland. Thus, the detection of fusion transcripts is a logical approach to diagnose prostate cancer harboring the *ETS* fusions. Several recent studies have demonstrated the ability

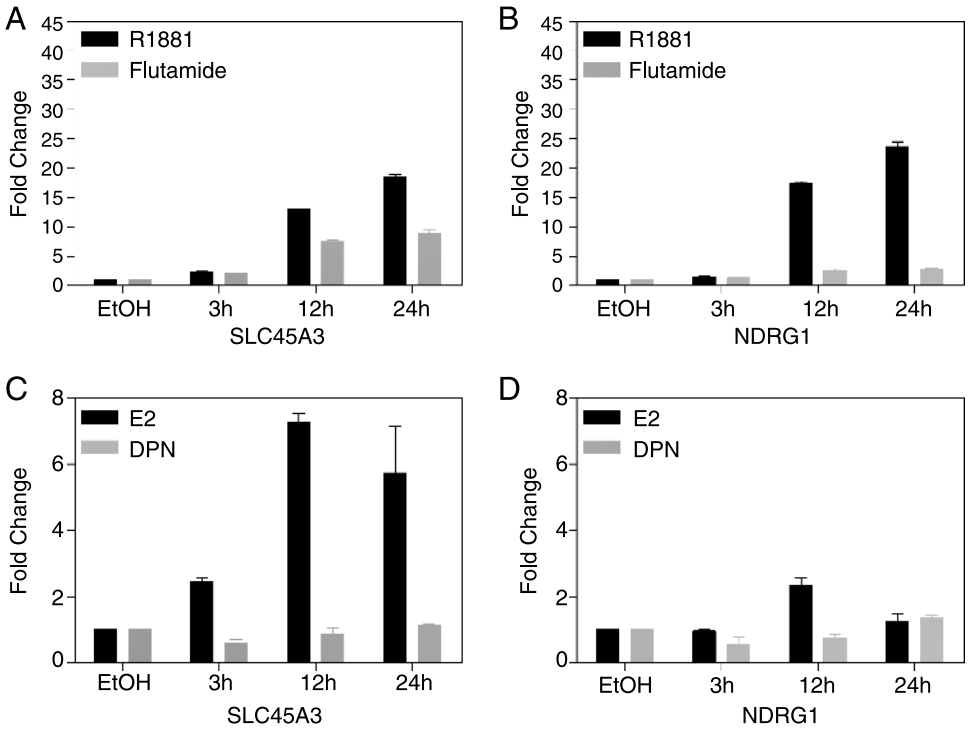

Figure 4. Hormone treatment of LNCaP cells induces SLC45A3 and NDRG1 mRNA expression. SLC45A3 (A and C) or NDRG1 (B and D) mRNA expression is induced upon stimulation with synthetic androgen (R1881) and 17β-estradiol (E2). Serum-starved LNCaP cells have been stimulated with 1 nM R1881, 1 nM R1881 in combination with 10 μM flutamide (A and B), 10 nM E2, or 10 nM DPN (C and D) for 3, 12, and 24 hours. Samples were run in triplicate and normalized against TCFL1. Columns indicate the mean fold change of induction of three biologic replicates against vehicle (ethanol)-only treated cells for the respective time points ± SEM.

to detect ETS fusion transcripts as well as PSA, PCA3, GOLPH2, and SPINK1 in urine samples [21–23]. We believe that the addition of *TMPRSS2-ERG*, *SCL45A3-ERG*, and *NDRG1-ERG* as part of a multiplex panel would have high specificity and improved sensitivity over current assays.

Acknowledgments

The authors thank Scott Tomlins and Arul M. Chinnaiyan for BAC probes to *SLC45A3*.

Conflicts of Interest

S. Perner, F. Demichelis, and M. A. Rubin are coinventors on a patent filed by The University of Michigan and The Brigham and Women's Hospital covering the diagnostic and therapeutic fields for ETS fusions in prostate cancer. The diagnostic field has been licensed to Gen-Probe, Inc. Gen-Probe has not played a role in the design and conduct of the study, in the collection, analysis or interpretation of the data, or in the preparation, review, or approval of the article. The authors mentioned disclose any financial interest.

References

[1] Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, et al. (2005). Recurrent fusion of *TMPRSS2* and *ETS* transcription factor genes in prostate cancer. *Science* **310**, 644–648.

[2] Kumar-Sinha C, Tomlins SA, and Chinnaiyan AM (2008). Recurrent gene fusions in prostate cancer. *Nat Rev Cancer* **8**, 497–511.

[3] Attard G, Clark J, Ambroisine L, Mills IG, Fisher G, Flohr P, Reid A, Edwards S, Kovacs G, Berney D, et al. (2008). Heterogeneity and clinical significance of ETV1 translocations in human prostate cancer. *Br J Cancer* **99**, 314–320.

[4] Wang J, Cai Y, Ren C, and Ittmann M (2006). Expression of variant TMPRSS2/ERG fusion messenger RNAs is associated with aggressive prostate cancer. *Cancer Res* **66**, 8347–8351.

[5] Mosquera JM, Perner S, Demichelis F, Kim R, Hofer MD, Mertz KD, Paris PL, Simko J, Collins C, Bismar TA, et al. (2007). Morphological features of *TMPRSS2-ERG* gene fusion prostate cancer. *J Pathol* **212**, 91–101.

[6] Attard G, Clark J, Ambroisine L, Fisher G, Kovacs G, Flohr P, Berney D, Foster CS, Fletcher A, Gerald WL, et al. (2008). Duplication of the fusion of TMPRSS2 to ERG sequences identifies fatal human prostate cancer. *Oncogene* **27**, 253–263.

[7] Demichelis F, Fall K, Perner S, Andren O, Schmidt F, Setlur SR, Hoshida Y, Mosquera JM, Pawitan Y, Lee C, et al. (2007). *TMPRSS2:ERG* gene fusion associated with lethal prostate cancer in a watchful waiting cohort. *Oncogene* **26**, 4596–4599.

[8] Demichelis F, Setlur SR, Beroukhim R, Perner S, Korbel JO, Lafargue CJ, Pflueger D, Pina C, Hofer MD, Sboner A, et al. (2009). Distinct genomic aberrations associated with ERG rearranged prostate cancer. *Genes Chromosomes Cancer* **48**, 366–380.

[9] Setlur SR, Mertz KD, Hoshida Y, Demichelis F, Lupien M, Perner S, Sboner A, Pawitan Y, Andren O, Johnson LA, et al. (2008). Estrogen-dependent signaling in a molecularly distinct subclass of aggressive prostate cancer. *J Natl Cancer Inst* **100**, 815–825.

[10] Han B, Mehra R, Dhanasekaran SM, Yu J, Menon A, Lonigro RJ, Wang X, Gong Y, Wang L, Shankar S, et al. (2008). A fluorescence *in situ* hybridization screen for E26 transformation-specific aberrations: identification of DDX5-ETV4 fusion protein in prostate cancer. *Cancer Res* **68**, 7629–7637.

[11] Rickman DS, Pflueger D, Moss B, VanDoren VE, Chen CX, de la Taille A, Kuefer R, Tewari AK, Setlur SR, Demichelis F, et al. (2009). SLC45A3-ELK4 is a novel and frequent erythroblast transformation–specific fusion transcript in prostate cancer. *Cancer Res* **69**, 2734–2738.

[12] Lin B, Ferguson C, White JT, Wang S, Vessella R, True LD, Hood L, and Nelson PS (1999). Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. *Cancer Res* **59**, 4180–4184.

[13] Xu J, Kalos M, Stolk JA, Zasloff EJ, Zhang X, Houghton RL, Filho AM, Nolasco M, Badaro R, and Reed SG (2001). Identification and characterization of prostein, a novel prostate-specific protein. *Cancer Res* **61**, 1563–1568.

[14] Lachat P, Shaw P, Gebhard S, van Belzen N, Chaubert P, and Bosman FT (2002). Expression of *NDRG1*, a differentiation-related gene, in human tissues. *Histochem Cell Biol* **118**, 399–408.

[15] Segawa T, Nau ME, Xu LL, Chilukuri RN, Makarem M, Zhang W, Petrovics G, Sesterhenn IA, McLeod DG, Moul JW, et al. (2002). Androgen-induced expression of endoplasmic reticulum (ER) stress response genes in prostate cancer cells. *Oncogene* **21**, 8749–8758.

[16] Tu LC, Yan X, Hood L, and Lin B (2007). Proteomics analysis of the interactome of N-*myc downstream regulated gene 1* and its interactions with the androgen response program in prostate cancer cells. *Mol Cell Proteomics* **6**, 575–588.

[17] Pandini G, Genua M, Frasca F, Squatrito S, Vigneri R, and Belfiore A (2007). 17beta-Estradiol up-regulates the insulin-like growth factor receptor through a nongenotropic pathway in prostate cancer cells. *Cancer Res* **67**, 8932–8941.

[18] Zhao Q, Caballero OL, Levy S, Stevenson BJ, Iseli C, de Souza SJ, Galante PA, Busam D, Leversha MA, Chadalavada K, et al. (2009). Transcriptome-guided characterization of genomic rearrangements in a breast cancer cell line. *Proc Natl Acad Sci USA* **106**, 1886–1891.

[19] Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, Sam L, Barrette T, Palanisamy N, and Chinnaiyan AM (2009). Transcriptome sequencing to detect gene fusions in cancer. *Nature* **458**, 97–101.

[20] Mostaghel EA, Page ST, Lin DW, Fazli L, Coleman IM, True LD, Knudsen B, Hess DL, Nelson CC, Matsumoto AM, et al. (2007). Intraprostatic androgens and androgen-regulated gene expression persist after testosterone suppression: therapeutic implications for castration-resistant prostate cancer. *Cancer Res* **67**, 5033–5041.

[21] Hessels D, Smit FP, Verhaegh GW, Witjes JA, Cornel EB, and Schalken JA (2007). Detection of TMPRSS2-ERG fusion transcripts and prostate cancer antigen 3 in urinary sediments may improve diagnosis of prostate cancer. *Clin Cancer Res* **13**, 5103–5108.

[22] Laxman B, Morris DS, Yu J, Siddiqui J, Cao J, Mehra R, Lonigro RJ, Tsodikov A, Wei JT, Tomlins SA, et al. (2008). A first-generation multiplex biomarker analysis of urine for the early detection of prostate cancer. *Cancer Res* **68**, 645–649.

[23] Laxman B, Tomlins SA, Mehra R, Morris DS, Wang L, Helgeson BE, Shah RB, Rubin MA, Wei JT, and Chinnaiyan AM (2006). Noninvasive detection of TMPRSS2:ERG fusion transcripts in the urine of men with prostate cancer. *Neoplasia* **8**, 885–888.

Supplemental Material: NDRG1-ERG Nucleic Acid and Amino Acid Sequence

Nucleic Acid Sequence Analysis of NDRG1-ERG Fusion Transcript Variant 1

This sequence has been deposited in GenBank (Accession No. FJ627786).

NDRG1 ex 1-3 – ERG ex 4-12

aacaaacctcgcctggctcccagctggtgctgaagctcgtcagttcaccatccgccctcggcttccgcggggcgctgggccgccagcctcggcaccgtccttttccttttctccctcgcgttaggcaggtgacagcagggacatgtctcgggagatgcaggatgtagacctcgctgaggtgaagccttttggtggagaaaggggagaccatcaccggcctcctgcaagagtttgatgtccaggaagcctatcagttgtgagtgaggaccagtcgttgtttgagtgtgcctacggaacgccacacctggctaagacagagatgaccgcgtcctcctccagcgactatggacagacttccaagatgagcccacgcgtccctcagcaggattggctgtctcaacccccagccagggtcaccatcaaaatggaatgtaaccctagccaggtgaatggctcaaggaactctcctgatgaatgcagtgtggccaaaggcgggaagatggtgggcagcccagacaccgttgggatgaactacggcagctacatggaggagaagcacatgccacccccaaacatgaccacgaacgagcgcagagttatcgtgccagcagatcctacgctatggagtacagaccatgtgcggcagtggctggagtgggcggtgaaagaatatggccttccagacgtcaacatcttgttattccagaacatcgatgggaaggaactgtgcaagatgaccaaggacgacttccagaggctcacccccagctacaacgccgacatccttctctcacatctccactacctcagagagactcctcttccacatttgacttcagatgatgttgataaagccttacaaaactctccacggttaatgcatgctagaaacacaggggtgcagcttttatttttccccaaatacttcagtatatcctgaagctacgcaaagaattacaactaggccagatttaccatatgagccccccaggagatcagcctggaccggtcacggccaccccacgccccagtcgaaagctgctcaaccatctccttccacagtgcccaaaactgaagaccagcgtcctcagttagatccttatcagattcttggaccaacaagtagccgccttgcaaatccaggcagtggccagatccagctttggcagttcctcctggagctcctgtcggacagctccaactccagctgcatcacctgggaaggcaccaacggggagttcaagatgacggatcccgacgaggtggcccggcgctggggagagcggaagagcaaacccaacatgaactacgataagctcagccgcgccctccgttactactatgacaagaacatcatgaccaaggtccatgggaagcgctacgcctacaagttcgacttccacgggatcgccccaggccctccagccccacccccgggagtcatctctgtacaagtaccccctcagacctcccgtacatgggctcctatcacgcccacccacagaagatgaacttttgtggcgccccacccctccagccctccccgtgacatcttccagttttttgctgccccaaacccatactggaattcaccaactgggggtatatacccccaacactaggctccccaccagccatatgccttctcatctgggcacttactactaaagacctggcggaggcttttcccatcagcgtgcattcaccagcccatcgccacaaactctatcggagaacatgaatcaaaagtgcctcaagaggaatgaaaaaagctttactggggctggggaaggaagccgggggaagagatccaaagactcttgggagggagttactgaagtcttactacagaaatgaggaggatgctaaaaatgtcacgaatatggacatatcatctgtggactgaccttgtaaaagacagtgtatgtagaagcatgaagtcttaaggacaaagtgccaaagaaagtggtcttaagaaatgtataaactttagagtagagtttggaatcccactaatgcaaactgggatgaaactaaagcaatagaaacaacacagttttgacctaacataccgtttataatgccattttaaggaaaactacctgtatttaaaaatagaaacatatcaaaaacaagagaaaagacacgagagagactgtggcccatcaacagacgttgatatgcaactgcatggcatgtgctgttttggttgaaatcaaatacattccgtttgatggacagctgtcagctttctcaaactgtgaagatgacccaaagtttccaactcctttacagtattaccgggactatgaactaaaaggtgggactgaggatgtgtatagagtgagcgtgtgattgtagacagaggggtgaagaaggaggaggaagaggcagagaaggaggagaccaggctggaaagaaacttctcaagcaatgaagactggactcaggacatttggggactgtgtacaatgagttatggagactcgagggttcatgcagtcagtgttataccaaacccagtgttaggagaaaggacacagcgtaatggagaaagggaagtagtagaattcagaaacaaaaatgcgcatctctttctttgtttgtcaaatgaaaattttaactggaattgtctgatatttaagagaaacattcaggacctcatcattatgtgggggctttgttctccacagggtcaggtaagagatggccttcttggctgccacaatcagaaatcacgcaggcattttgggtaggcggcctccagttttccttgagtcgcgaacgctgtgcgtttgtcagaatgaagtatacaagtcaatgtttttccccctttttatataataattatataacttatgcatttatacactacgagttgatctcggccagccaaagacacacgacaaaagagacaatcgataataatgtggccttgaattttaactctgtatgcttaatgtttacaatatgaagttattagttcttagaatgcagaatgtatgtaataaaataagcttggcctagcatggcaaatcagatttatacaggagtctgcatttgcacttttttagtgactaaagttgcttaatgaaaacatgtgctgaatgttgtggattttgtgttataatttactttgtccaggaacttgtgcaagggagagccaaggaaataggatgtttggcaccc

Protein Sequence Analysis of NDRG1-ERG Fusion Transcript Variant 1

The longest chimeric sequence is analyzed by InterProScan to search for protein domains.

Variant 1: NDRG1 ex 1-3 – ERG ex 4-12

MSREMQDVDLAEVKPLVEKGETITGLLQEFDVQEALSVVSEDQSLFECAYGTPHLAKTEMTASSSSDYGQTSKMSPRVPQQDWLSQPPARVTIKMECNPSQVNGSRNSPDECSVAKGGKMVGSPDTVGMNYGSYMEEKHMPPPNMTTNERRVIVPADPTLWSTDHVRQWLEWAVKEYGLPDVNILLFQNIDGKELCKMTKDDFQRLTPSYNADILLSHLHYLRETPLPHLTSDDVDKALQNSPRLMHARNTGGAAFIFPNTSVYPEATQRITTRPDLPYEPPRRSAWTGHGHPTPQSKAAQPSPSTVPKTEDQRPQLDPYQILGPTSSRLANPGSGQIQLWQFLLELLSDSSNSSCITWEGTNGEFKMTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFDFHGIAQALQPHPPESSLYKYPSDLPYMGSYHAHPQKMNFVAPHPPALPVTSSSFFAAPNPYWNSPTGGIYPNTRLPTSHMPSHLGTYY

Nucleic Acid Sequence Analysis of NDRG1-ERG Fusion Transcript Variant 2

This sequence has been deposited in GenBank (Accession No. FJ627787).

Variant 2: NDRG1 ex 1-2 – ERG ex 4-12

aacaaacctcgcctggctcccagctggtgctgaagctcgtcagttcaccatccgccctcggcttccgcggggcgctgggccgccagcctcggcaccgtccttttccttttctccctcgcgttaggcaggtgacagcagggacatgtctcgggagatgcaggatgtagacctcgctgaggtgaagccttttggtggagaaaggggagaagccttatcagttgtgagtgaggaccagtcgttgtttgagtgtgcctacggaacgccacacctggctaagacagagatgaccgcgtcctcctccagcgactatggacagacttccaagatgagcccacgcgtccctcagcaggattggctgtctcaacccccagccagggtcaccatcaaaatggaatgtaaccctagccaggtgaatggctcaaggaactctcctgatgaatgcagtgtggccaaaggcgggaagatggtgggcagcccagacaccgttgggatgaactacggcagctacatggaggagaagcacatgccacccccaaacatgaccacgaacgagcgcagagttatcgtgccagcagatcctacgctatggagtacagaccatgtgcggcagtggctggagtgggcggtgaaagaatatggccttccagacgtcaacatcttgttattccagaacatcgatgggaaggaactgtgcaagatgaccaaggacgacttccagaggctcacccccagctacaacgccgacatccttctctcacatctccactacctcagagagactcctcttccacatttgacttcagatgatgttgataaagccttacaaaactctccacggttaatgcatgctagaaacacaggggtgcagcttttatttttccccaaatacttcagtatatcctgaagctacgcaaagaattacaactaggccagatttaccatatgagccccccaggagatcagcctggaccggtcacggccaccccacgccccagtcgaaagctgctcaaccatctccttccacagtgcccaaaactgaagaccagcgtcctcagttagatccttatcagattcttggaccaacaagtagccgccttgcaaatccaggcagtggccagatccagctttggcagttcctcctggagctcctgtcggacagctccaactccagctgcatcacctgggaaggcaccaacggggagttcaagatgacggatcccgacgaggtggcccgg-cgctggggagagcggaagagcaaacccaacatgaactacgataagctcagccgcgccctccgttactactatgacaagaacatcatgaccaaggtccatgggaagcgctacgcctacaagttcgacttccacgggatcgcccaggccctccagccccacccccgggagtcatctctgtacaagtaccccctcagacctcccgtacatgggctcctatcacgcccacccacagaagatgaacttttgtggcgccccacccctccagccctccccgtgacatcttccagtttttttgctgccccaaacccatactggaattcaccaactgggggtatatacccccaacactaggctccccaccagccatatgccttctcatctgggcacttactactaaagacctggcggaggcttttcc-

catcagcgtgcattcaccagcccatcgccacaaactctatcggagaacatgaatcaaaagtgcctcaagaggaatgaaaaaagctttactggggctggggaaggaagccggggaagagatccaaagactcttgggagggagttactgaagtcttactacagaaatgaggaggatgctaaaaatgtcacgaatatggacatatcatctgtggactgaccttgtaaaagacagtgtatgtagaagcatgaagtcttaaggacaaagtgccaaagaaagtggtcttaagaaatgtataaactttagagtagagtttggaatcccactaatgcaaactgggatgaaactaaagcaatagaaacaacacagttttgacctaacataccgtttataatgccatttaaggaaaactacctgtatttaaaaatagaaacatatcaaaaacaagagaaaagacacgagagagactgtggcccatcaacagacgttgatatgcaactgcatggcatgtgctgttttggttgaaatcaaatacattccgtttgatggacagctgtcagctttctcaaactgtgaagatgacccaaagtttccaactcctttacagtattaccgggactatgaactaaaaggtgggactgaggatgtgtatagagtgagcgtgtgattgtagacagaggggtgaagaaggaggaggaagaggcagagaaggaggagaccaggctgggaaagaaacttctcaagcaatgaagactggactcaggacatttggggactgtgtacaatgagttatggagactcgagggttcatgcagtcagtgttataccaaacccagtgttaggagaaaggacacagcgtaatggagaaagggaagtagtagaattcagaaacaaaaatgcgcatctctttctttgtttgtcaaatgaaaattttaactggaattgtctgatatttaagagaaacattcaggacctcatcattatgtgggggctttgttctccacagggtcaggtaagagatggccttcttggctgccacaatcagaaatcacgcaggcattttgggtaggcggcctccagttttccttttgagtcgcgaacgctgtgcgtttgtcagaatgaagtatacaagtcaatgtttttccccctttttatataataattatataacttatgcatttatacactacgagttgatctcggccagccaaagacacacgacaaaagagacaatcgatataatgtggccttgaattttaactctgtatgcttaatgtttacaatatgaagttattagttcttagaatgcagaatgtatgtaataaaataagcttggcctagcatggcaaatcagatttatacaggagtctgcatttgcacttttttttagtgactaaagttgcttaatgaaaacatgtgctgaatgttgtggattttgtgttataatttactttgtccaggaacttgtgcaagggagagccaaggaaataggatgtttggcaccc

Protein Sequence Analysis of NDRG1-ERG Fusion Transcript Variant 2

The longest chimeric sequence is analyzed by InterProScan to search for protein domains.

Variant 2: NDRG1 ex 1-2 – ERG ex 4-12

MSREMQDVDLAEVKPLVEKGEEALSVVSEDQSLFECAYGTPHLAKTEMTASSSSDYGQTSKMSPRVPQQDWLSQPPARVTIKMECNPSQVNGSRNSPDECSVAKGGKMVGSPDTVGMNYGSYMEEKHMPPPNMTTNERRVIVPADPTLWSTDHVRQWLEWAVKEYGLPDVNILLFQNIDGKELCKMTKDDFQRLTPSYNADILLSHLHYLRETPLPHLTSDDVDKALQNSPRLMHARNTGGAAFIFPNTSVYPEATQRITTRPDLPYEPPRRSAWTGHGHPTPQSKAAQPSPSTVPKTEDQRPQLDPYQILGPTSSRLANPGSGQIQLWQFLLELLSDSSNSSCITWEGTNGEFKMTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFDFHGIAQALQPHPPESSLYKYPSDLPYMGSYHAHPQKMNFVAPHPPALPVTSSSFFAAPNPYWNSPTGGIYPNTRLPTSHMPSHLGTYY

Supplemental References

[1] Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, Eeckhoute J, Brodsky AS, Keeton EK, Fertuck KC, Hall GF, et al. (2006). Genome-wide analysis of estrogen receptor binding sites. *Nat Genet* **38**, 1289–1297.

[2] Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS, and Brown M (2008). FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**, 958–970.

Table W1. *ERG* Rearrangement and Overexpression in the Context of TMPRSS2-ERG mRNA Variants Defined Using Conventional RT-PCR and cDNA Sequencing and in the Context of the Indicated Clinical Information.

Sample ID	ERG Rearrangement	TMPRSS2-ERG Isoform	ERG Expression	Gleason Score	Stage	Surgical Margin	Seminal Vesicle Invasion
1701_A	+	III	74.04	3 + 4	IIIA	–	–
580_B	+	III, VI	66.97	3 + 4	IIC	–	–
1700_D	+	III	46.18	4 + 4	IIIA	+	–
28_T	+	III	40.40	3 + 4	IIC	–	–
45_T	+	III, VI	36.20	2 + 3	IIC	+	–
69_T	+	III, VI	32.98	4 + 3	IIIA	–	–
13_D	+	III	32.35	3 + 3	IIC	–	–
560_D	+	III, VI	30.96	4 + 3	IIIA	–	–
140_T	+	II	28.91	3 + 4	IIIA/B	–	+
435_D	+	III	28.75	3 + 4	IIIA	–	–
581_D	+	III	27.56	4 + 5	IIIB	–	+
1780_D	+	III, VI	27.10	3 + 4	IIC	–	–
431_D	+	III, VI	24.92	3 + 3	II	–	–
54_T	+	III	24.69	3 + 4	IIC	+	–
88_T	+	III	24.52	3 + 5	IIIB	–	+
522_D	+	IV	22.06	3 + 4	IIC	–	–
415_B	+	III	13.46	4 + 5	IIIA	–	–
67_T	+	II, III, VI, VIII	11.08	3 + 4	IIC	–	–
106_T	+	I, V	10.06	3 + 5	IIIB	–	+
60_T	+	–	38.40	3 + 4	IIC	+	–
145_C_M	+	–	37.54	4 + 5	IIIA	–	–
34_T	+	–	37.37	3 + 4	IIIB	–	+
150_B_M	+	–	20.06	2 + 3	IIC	–	–
51_T	+	–	15.01	3 + 4	IIIA	–	–
509_B	+	–	10.12	3 + 4	IIA	–	–
99_T	+	–	5.07	3 + 3	IIIA	+	–
1061_C	+	–	0.75	4 + 3	IIC	–	–
424_B	+	–	0.53	3 + 4	IIC	–	–
1043_B	+	–	0.50	3 + 3	IIC	–	–
1024_D	–	–	5.74	4 + 5	IIIB	–	+
97_T	–	–	2.93	2 + 3	IIC	–	–
127_T	–	–	2.72	3 + 3	IIA	–	–
1783_B	–	–	2.26	4 + 4	IIC	–	–
1023_C	–	–	1.97	3 + 3	IIC	–	–
113_T	–	–	1.77	4 + 5	IIIB	+	+
136_T	–	–	1.36	3 + 3	IIIA	+	–
134_B	–	–	1.04	3 + 2	IIB	+	–
1060_B	–	–	0.79	4 + 3	IIC	–	–
151_T	–	–	0.77	2 + 3	IIC	–	–
540_C	–	–	0.74	3 + 4	IIIA	–	–
1781_C	–	–	0.71	3 + 4	IIC	+	–
63_T	–	–	0.45	2 + 4	IIC	–	–
1702_C	–	–	0.43	3 + 4	IIIA	–	–
1765_A	–	–	0.41	3 + 4	IIA	+	–
1027_D	NA	–	1.07	NA	NA	NA	NA
1033_B	NA	–	1.04	NA	NA	NA	NA
1023_B	NA	–	1.04	NA	NA	NA	NA
1024_C	NA	–	0.90	NA	NA	NA	NA
1028_A	NA	–	0.76	NA	NA	NA	NA
1032_D	NA	–	0.73	NA	NA	NA	NA

NA indicates not applicable.

Table W2. Oligonucleotide Primers and Cycling Conditions for RT-PCR Assays.

Assay	Gene	Accession No.	Bases	Exon(s)	Primer	Sequence 5′ → 3′	Cycling
TMPRSS2-ERG exon 4	*TMPRSS2*	NM_005656.2	–4 to 17	1	TMPRSS2-ERG_f	TAGGCGCGAGCTAAGCAGGAG	94°C 2 min; 94°C 30 sec; 63°C 30 sec; 72°C 1 min 10 sec; 72°C 10 min
	ERG	NM_004449.3	276-252	5	Exon4_r1	GTAGGCACACTCAAACAACGACTGG	35 cycles
TMPRSS2-ERG exon 6	*TMPRSS2*	NM_005656.2	–4 to 17	1	TMPRSS2-ERG_f	TAGGCGCGAGCTAAGCAGGAG	94°C 2 min; 94°C 30 sec; 63°C 30 sec; 68°C 2 min; 68°C 10 min
	ERG	NM_004449.3	659-636	7	ERG_Exon 5-6_r	CCATATTCTTTCACCGCCCACTCC	40 cycles
TMPRSS2-ERG exon 9	*TMPRSS2*	NM_005656.2	–4 to 17	1	TMPRSS2-ERG_f	TAGGCGCGAGCTAAGCAGGAG	94°C 2 min; 94°C 30 sec; 55°C 30 sec; 68°C 2 min; 68°C 10 min
	ERG	NM_004449.3	945-928	10	ERG_Exon10_r	CGACTGGGGCGTGGGGTG	40 cycles
TMPRSS2-ERG exon 12	*TMPRSS2*	NM_005656.2	–4 to 17	1	TMPRSS2-ERG_f	TAGGCGCGAGCTAAGCAGGAG	94°C 2 min; 94°C 30 sec; 55°C 30 sec; 68°C 2 min; 68°C 10 min
	ERG	NM_004449.3	1575-1549	13	ERG_Exon13_r	TTAGTAGTAAGTGCCCAGATGAGAAGG	40 cycles
ERG qPCR	*ERG*	NM_004449.3	574-597	6-7	ERG_Exon 5-6_f	CGCAGAGTTATCGTGCCAGCAGAT	50°C 2 min; 95°C 15 min; 94°C 1 min; 50°C 1 min; 72°C 1 min
	ERG	NM_004449.3	659-636	7	ERG_Exon 5-6_r	CCATATTCTTTCACCGCCCACTCC	41 cycles
	SART3	NM_014706.3	635-658	2	SART3_f	GCCCGCCAGAAGATGAGTGAAATC	Melting curve: 70-90°C every 0.2°C for 1 sec
	SART3	NM_014706.3	889-866	4	SART3_r	ACCAACAGACGAGAGAGCCCTTTC	
	TCFLI/VPS72	NM_005997.1	778-801	6	TCFL1/VPS72_f	ATTGACTCCTCATGCTGGGACTGG	
	TCFLI/VPS72	NM_005997.1	948-927	6	TCFL1/VPS72_r	CGGTATAGGGCTGGACGATGGG	
TMPRSS2-ERG qPCR	*TMPRSS2*	NM_005656.2	–4 to 17	1	TMPRSS2-ERG_f	TAGGCGCGAGCTAAGCAGGAG	50°C 2 min; 95°C 15 min; 94°C 1 min; 50°C 1 min; 72°C 1 min
	ERG	NM_004449.3	276-252	5	Exon4_r1	GTAGGCACACTCAAACAACGACTGG	41 cycles
	HMBS	NM_000190.3	711-730	10	HMBS_f	CCATCATCCTGGCAACAGCT	Melting curve: 70-90°C every 0.2°C for 1 sec
	HMBS	NM_000190.3	790-772	11	HMBS_r	GCATTCCTCAGGGTGCAGG	
SLC45A3-ERG	*SLC45A3*	NM_033102.2	74-91	1	SLC45A3_f	CGCTGGCTCCGGGTGACA	94°C 2 min; 94°C 30 sec; 55°C 30 sec; 68°C 2 min; 68°C 10 min
	ERG	NM_004449.3	659-636	7	ERG_Exon 5-6_r	CCATATTCTTTCACCGCCCACTCC	40 cycles
	ERG	NM_004449.3	945-928	10	ERG_Exon10_r	CGACTGGGGCGTGGGGTG	
HERV-K-ERG	*HERV-K_22q11.23*	BC020811.1	305-327	N/A	HERV-K_f	GTCCCAAGTACGTCCACGGTCAG	94°C 2 min; 94°C 30 sec; 55°C 30 sec; 68°C 2 min; 68°C 10 min
	ERG	NM_004449.3	659-636	7	ERG_Exon 5-6_r	CCATATTCTTTCACCGCCCACTCC	40 cycles
	ERG	NM_004449.3	945-928	10	ERG_Exon10_r	CGACTGGGGCGTGGGGTG	
HNRPA2B1-ERG	*HNRPA2B1*	NM_002137.2	136-155	1	HNRPA2B1_f	TGCGGGAAATCGGGCTGAAG	94°C 2 min; 94°C 30 sec; 55°C 30 sec; 68°C 2 min; 68°C 10 min
	ERG	NM_004449.3	659-636	7	ERG_Exon 5-6_r	CCATATTCTTTCACCGCCCACTCC	40 cycles
	ERG	NM_004449.3	945-928	10	ERG_Exon10_r	CGACTGGGGCGTGGGGTG	
C15ORF21-ERG	*C15ORF21*	NM_001005266	313-336	3	C15ORF21_f	CAACTAACACTGCGGCTTCCTGAG	94°C 2 min; 94°C 30 sec; 55°C 30 sec; 68°C 2 min; 68°C 10 min
	ERG	NM_004449.3	659-636	7	ERG_Exon 5-6_r	CCATATTCTTTCACCGCCCACTCC	40 cycles
	ERG	NM_004449.3	945-928	10	ERG_Exon10_r	CGACTGGGGCGTGGGGTG	
CANT1-ERG	*CANT1 E1a*	N/A	N/A	N/A	CANT1-E1a_f	GCTGGAGAAACAAACCCTCT	94°C 2 min; 94°C 30 sec; 55°C 30 sec; 68°C 2 min; 68°C 10 min
	ERG	NM_004449.3	659-636	7	ERG_Exon 5-6_r	CCATATTCTTTCACCGCCCACTCC	40 cycles
	ERG	NM_004449.3	945-928	10	ERG_Exon10_r	CGACTGGGGCGTGGGGTG	
CANT1-ERG	*CANT1 E1*	NM_138793.2	48-65	1	CANT1-E1_f	AGCCAAGCCCCGCCGATC	94°C 2 min; 94°C 30 sec; 55°C 30 sec; 68°C 2 min; 68°C 10 min
	ERG	NM_004449.3	659-636	7	ERG_Exon 5-6_r	CCATATTCTTTCACCGCCCACTCC	40 cycles
	ERG	NM_004449.3	945-928	10	ERG_Exon10_r	CGACTGGGGCGTGGGGTG	
KLK2-ERG	*KLK2*	NM_005551.3	65-77	1	KLK2_f	TCTCTCCATCGCCTTGTCTGTG	94°C 2 min; 94°C 30 sec; 55°C 30 sec; 68°C 2 min; 68°C 10 min
	ERG	NM_004449.3	659-636	7	ERG_Exon 5-6_r	CCATATTCTTTCACCGCCCACTCC	40 cycles
	ERG	NM_004449.3	945-928	10	ERG_Exon10_r	CGACTGGGGCGTGGGGTG	
DDX5-ERG	*DDX5*	NM_004396.2	423-443	3	DDX5_f	AGAGGTCACAACTGCCCGAAG	94°C 2 min; 94°C 30 sec; 55°C 30 sec; 68°C 2 min; 68°C 10 min
	ERG	NM_004449.3	659-636	7	ERG_Exon 5-6_r	CCATATTCTTTCACCGCCCACTCC	40 cycles
	ERG	NM_004449.3	945-928	10	ERG_Exon10_r	CGACTGGGGCGTGGGGTG	
NDRG1-ERG	*NDRG1*	NM_006096.2	30-53	1	NDRG1_ex1_f	CTGAAGCTCGTCAGTTCACCATCC	94°C 2 min; 94°C 30 sec; 61°C 30 sec; 68°C 2 min; 68°C 10 min
	ERG	NM_004449.3	659-636	7	ERG_Exon 5-6_r	CCATATTCTTTCACCGCCCACTCC	40 cycles
	ERG	NM_004449.3	945-928	10	ERG_Exon10_r	CGACTGGGGCGTGGGGTG	
	ERG	NM_004449.3	1575-1549	13	ERG_Exon13_r	TTAGTAGTAAGTGCCCAGATGAGAAGG	
ERG 5′RACE	*ERG*	mRNA	601-578	6-7	ERG_GSP1	ATCCTAGACGACCGTGCTATTGAG	Reverse transcription reaction at 55°C
	ERG	NM_004449.3	571-552	6	ERG_GSP2	CGTTCGTGGTCATGTTTGGG	94°C 2 min; 94°C 30 sec; 60°C 1 min; 72°C 1 min; 72°C 7 min
	ERG	NM_004449.3	470-448	6	ERG_GSP3	GCCACACTGCATTCATCAGGAGA	35 cycles (45 cycles for GSP3-PCR)

All RT-PCR assays are listed along with the respective gene(s) investigated, GenBank, oligonucleotide primer name, its location (bases and exons), sequence (5′ to 3′), and cycling conditions. *N/A* indicates not applicable.

Table W3. TaqMan Expression Assays from Applied Biosystems and Information Regarding Available Assay ID, Reporter Dye, and Location within the Gene.

Gene	Assay ID	Reporter Dye	Location
SLC45A3	Hs00263832_m1	FAM	ex 5
TCFL1	Hs00195618_m1	FAM	ex 6
IGF1R	Hs99999020_m1	FAM	ex 2 and 3
KLK3	Hs02576345_m1	FAM	ex 1 and 2, detects all KLK3 transcript variants
NDRG1	Hs00608387_m1	FAM	ex 11 and 12

We followed the manufacturer's recommended protocol.

Table W4. Summary of the Data Obtained from Three Samples Analyzed Using RNA-Seq Showing the Number of Reads and the Number of Mappable (Including Percentage) for Each Sample.

Sample ID	No. Reads	No. Mappable Reads	% Mappable Reads	Comments
1701_A	8,542,482	3,108,222	36.39%	(T2-ERG fusion–positive)
1783_B	3,080,154	1,330,949	43.21%	(T2-ERG fusion–negative)
99_T	2,844,879	1,180,781	41.51%	(NDRG1-ERG fusion–positive)

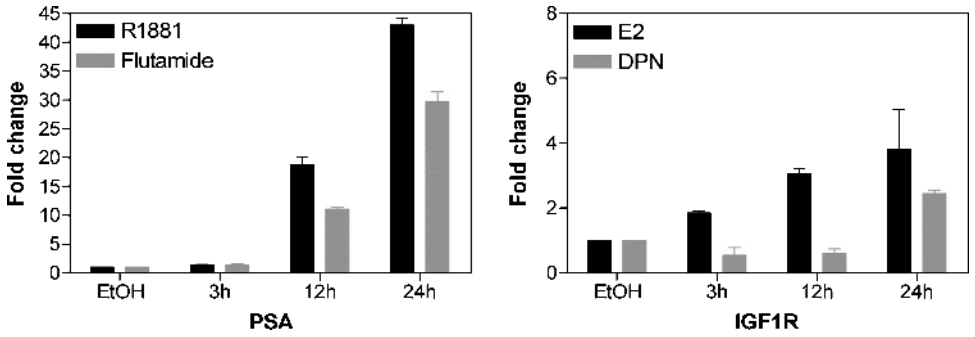

Figure W1. Androgen and 17β-estradiol (E2) signaling of known target genes in LNCaP cells. *PSA* and *IGF2R* mRNA expression is induced upon stimulation with synthetic androgen (R1881) and E2. Serum-starved LNCaP cells have been stimulated with 1 nM R1881, 1 nM R1881 in combination with 10 μM flutamide, 10 nM E2, or 10 nM DPN for 3, 12, and 24 hours. Total RNA was extracted and used for quantitative RT-PCR using TaqMan. Samples were run in triplicate and normalized against *TCFL1*. Columns indicate the mean fold change of induction of three biologic replicates against vehicle (ethanol)-only treated cells for the respective time points ± SEM.

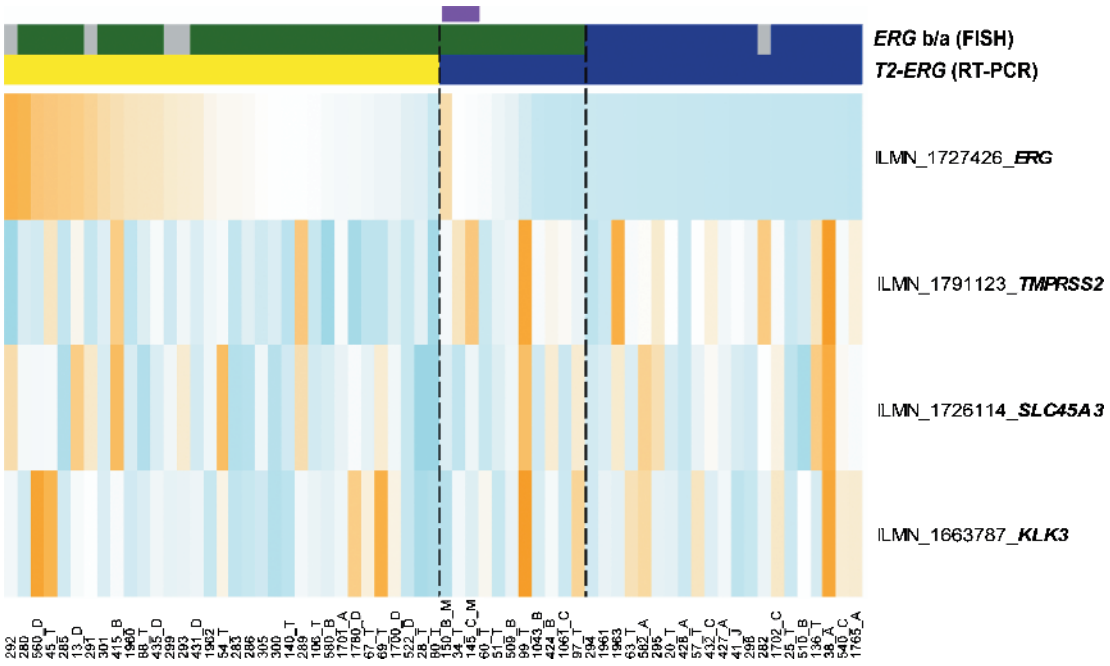

Figure W2. Expression profiling of ERG and three androgen-regulated genes. A subset of 65 of 101 samples was processed using Illumina HumanWG-6 v2.0 bead arrays. Purple indicates three samples with *SLC45A3-ERG* fusion. The heat map shows the relative expression levels of *ERG*, *TMPRSS2*, and *SLC45A3*. The gene expression levels in a given sample have been color coded where orange to blue indicates high to low levels of expression. The samples have been grouped according to *TMPRSS2-ERG* (*T2-ERG*) fusion status as determined by RT-PCR and then ordered by the level of ERG microarray feature level normalized intensity.

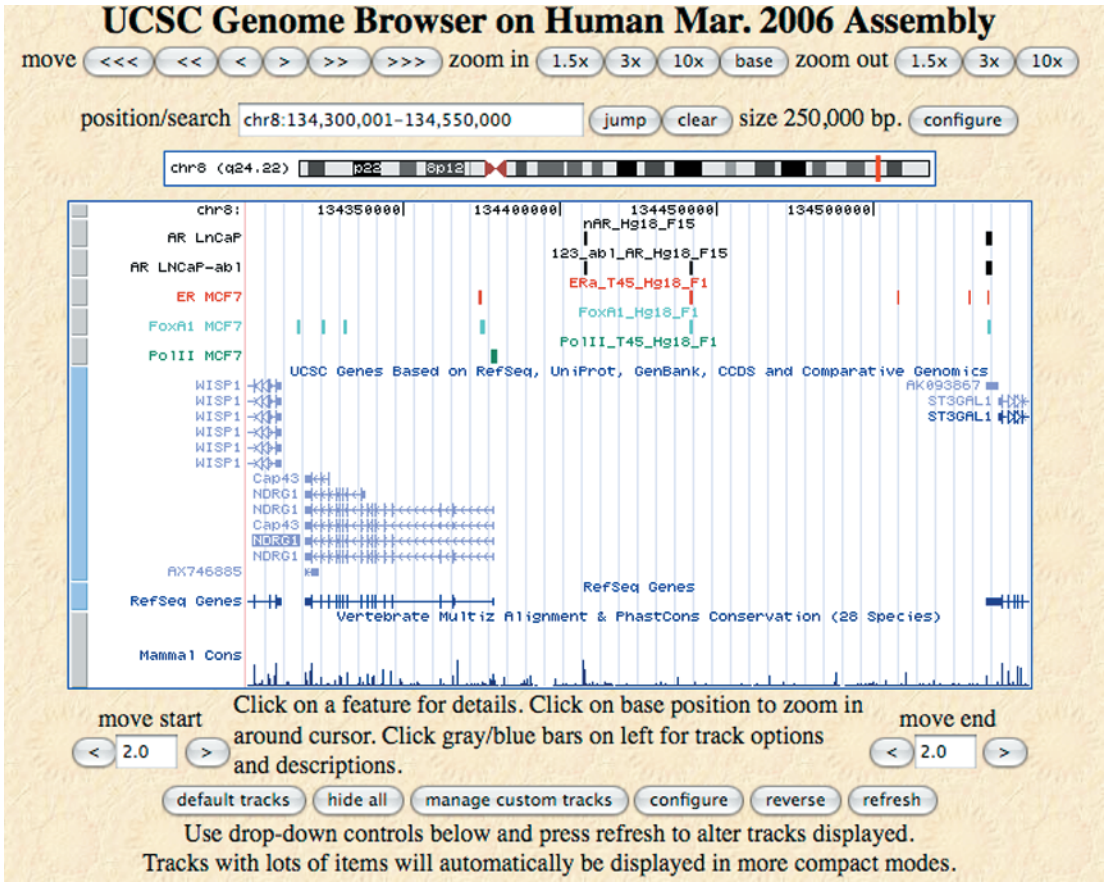

Figure W3. UCSC Genome Browser of the region of chromosome 8q24q22 that includes the location of NDRG1 exon and intron locations and upstream sequence. Shown are binding sites for androgen receptor (AR; Myles Brown, written personal communication, January 13, 2009) in LNCaP cells and ER and FoxA1 in the MCF7 cell lines [1,2].

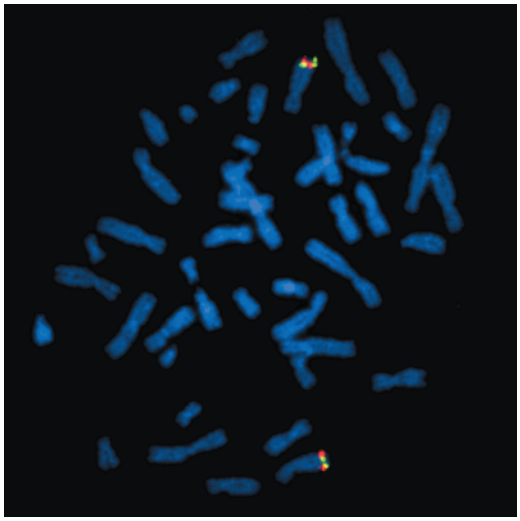

Figure W4. Representative image of a metaphase spread from normal human male lymphocytes displaying the correct chromosome 8q24.22 position of FISH BAC probes targeting the NDRG1 locus used in the b/a assay.

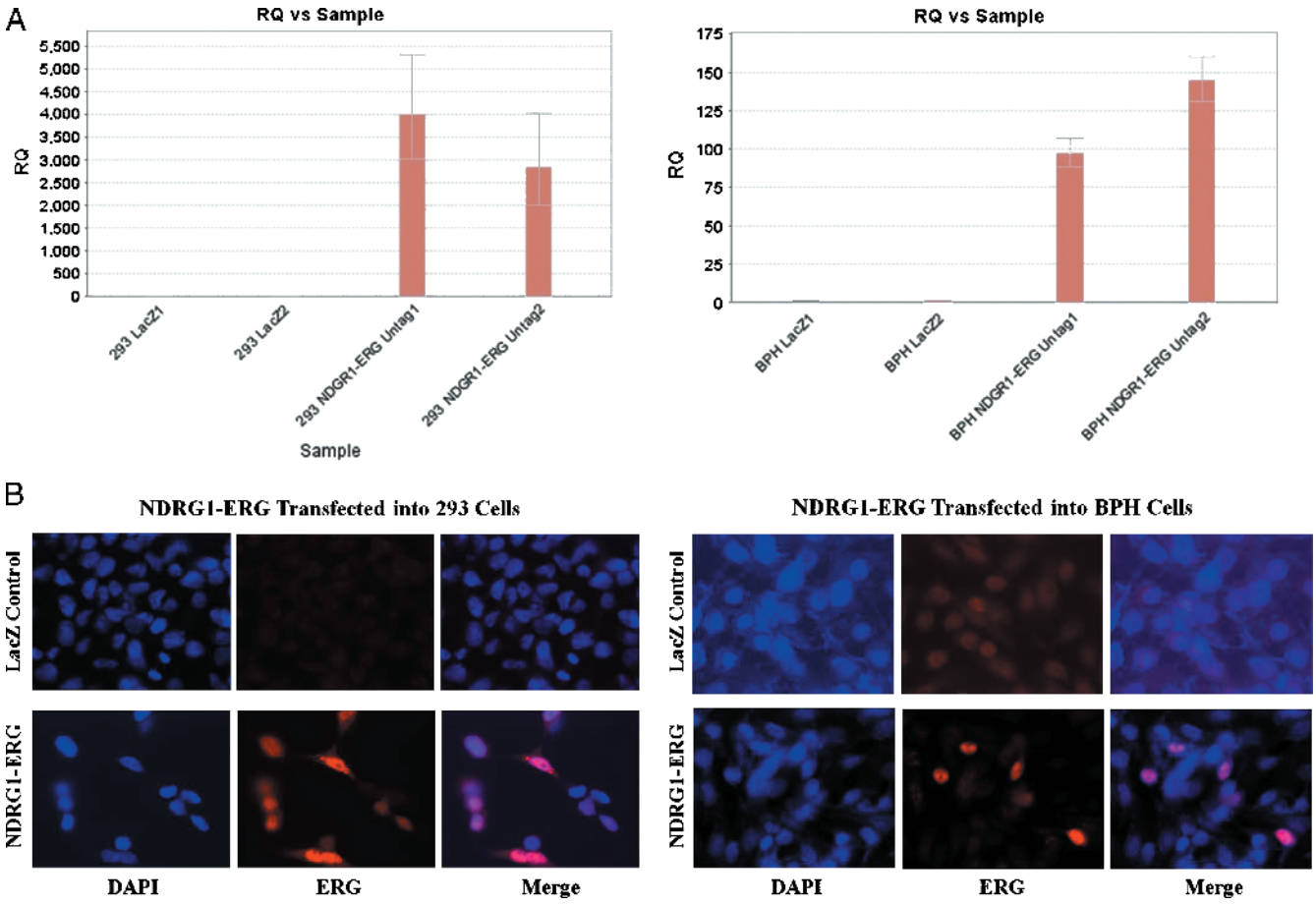

Figure W5. Evaluation of ERG expression through qRT-PCR (A) and fluorescent immunostaining (B) of HEK-293 cells (left panels) and BPH cells (right panels), respectively, transiently transfected with untagged NDRG1-ERG constructs. As a control, the cells were transfected with a lacZ construct.