Modern Pathology (2007) 20, 467–473

www.modernpathology.org

A variant *TMPRSS2* isoform and *ERG* fusion product in prostate cancer with implications for molecular diagnosis

Jacques Lapointe[1,2,*,†], Young H Kim[1,*], Melinda A Miller[3], Chunde Li[4,5], Gulsah Kaygusuz[1,‡], Matt van de Rijn[1], David G Huntsman[3], James D Brooks[2] and Jonathan R Pollack[1]

[1]*Department of Pathology, Stanford University, Stanford, CA, USA;* [2]*Department of Urology, Stanford University, Stanford, CA, USA;* [3]*Department of Pathology, British Columbia Cancer Agency, Vancouver, BC, Canada;* [4]*Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden and* [5]*Department of Oncology, Karolinska University Hospital, Stockholm, Sweden*

Prostate cancer is the most commonly diagnosed cancer among men in the United States. Recently, fusion of *TMPRSS2* with ETS family oncogenic transcription factors has been identified as a common molecular alteration in prostate cancer, where most often the rearrangement places *ERG* under the androgen-regulated transcriptional control of *TMPRSS2*. Here, we carried out rapid amplification of cDNA ends (RACE) on a prostate cancer specimen carrying an atypical aberration discovered by array-based comparative genomic hybridization (array CGH), suggesting an alternative fusion partner of *ERG*. We identified novel transcribed sequences fused to *ERG*, mapping 4 kb upstream of the *TMPRSS2* start site. The sequences derive from an apparent second *TMPRSS2* isoform, which we found also expressed in some prostate tumors, suggesting similar androgen-regulated control. In a reverse transcription-polymerase chain reaction (RT-PCR)-based survey of 63 prostate tumor specimens (54 primary and nine lymph node metastases), 44 (70%) cases expressed either the known or novel variant *TMPRSS2-ERG* fusion, 28 (44%) expressed both, 10 (16%) expressed only the known, and notably six (10%) expressed only the variant isoform fusion. In this specimen set, the presence of a *TMPRSS2-ERG* fusion showed no statistical association with tumor stage, Gleason grade or recurrence-free survival. Nonetheless, the discovery of a novel variant *TMPRSS2* isoform-*ERG* fusion adds to the characterization of ETS-family rearrangements in prostate cancer, and has important implications for the accurate molecular diagnosis of *TMPRSS2-ETS* fusions.
Modern Pathology (2007) **20,** 467–473. doi:10.1038/modpathol.3800759; published online 2 March 2007

Keywords: prostate cancer; *TMPRSS2-ERG* fusion; molecular diagnosis

Prostate cancer is the most frequently diagnosed cancer among men in the United States,[1] with one in six men being diagnosed in their lifetime. Localized prostate cancer can be treated by surgical resection or radiation, but recurs in approximately a quarter of patients.[2] Advanced prostate cancer is treated by androgen-ablation therapy, though hormone-refractory prostate cancer invariably recurs within 1–2 years.[3] Although taxane-based therapies have recently shown promise in prolonging survival,[4] new therapeutic targets and molecularly directed therapies for end-stage prostate cancer are urgently needed.

Recently, by examining outlier values of gene expression in human cancers, Tomlins *et al*[5] discovered elevated expression of the ETS family members *ERG* (v-ets erythroblastosis virus E26 oncogene like) and *ETV1* (Ets variant gene 1) to be a common feature of prostate cancer. Elevated expression resulted from chromosomal rearrangement fusing *TMPRSS2* (transmembrane protease, serine 2) (chr 21q22.3) to *ERG* (chr 21q22.2), an apparent intra-chromosomal deletion of ~3 Mb on chromosome 21, or less frequently to *ETV1*, an inter-chromosomal rearrangement between chr 21q22.3 and chr 7p21.2.

Correspondence: Dr JR Pollack, MD, PhD, Department of Pathology, Stanford University School of Medicine, 269 Campus Drive, CCSR-3245A, Stanford, CA 94305-5176, USA.
E-mail: pollack1@stanford.edu
*These authors contributed equally to this work.
†Current address: Urology Division, Department of Surgery, McGill University, Montreal, QC, Canada.
‡Current address: Department of Pathology, Ankara University School of Medicine, Ankara, Turkey.
Received 29 November 2006; revised 16 January 2007; accepted 17 January 2007; published online 2 March 2007

ETS transcription factors are both positive and negative regulators of gene expression, and are involved in various biological processes including cell proliferation, differentiation, development, transformation and apoptosis.[6] In human cancers, ETS genes are frequently found amplified, overexpressed or rearranged, the latter including chimeric fusion products characteristic of Ewing's sarcoma and certain leukemias.[6] *TMPRSS2*, which encodes a serine protease with a transmembrane domain,[7] is highly expressed in both normal and neoplastic prostate epithelium,[8,9] and its expression in prostate cancer cells is androgen-regulated.[8,10] In prostate tumors harboring rearrangements, most frequently the non-coding first exon of *TMPRSS2* has been found fused to the fourth exon of *ERG* in the expressed transcript. The rearrangement thereby places *ERG* under the androgen-regulated transcriptional control of *TMPRSS2*.[5] This finding has immediately suggested an additional possible mechanism for the androgen-dependent growth of therapy-naïve prostate tumors.

Here, we describe the discovery in prostate tumors of novel sequences fused to *ERG*, which we define as a variant *TMPRSS2* isoform-*ERG* fusion. Our findings further our knowledge of ETS family rearrangements in prostate cancer, and provide for the more accurate molecular diagnosis of this lesion for future molecularly directed therapies.

Materials and methods

Prostate Tumor Specimens

The prostate tumor specimens included in this study, along with detailed clinicopathological annotations and procedures for specimen processing, were described previously.[11] Briefly, 54 freshly frozen primary prostate tumor specimens (obtained from radical prostatectomy) and nine therapy-naive metastatic pelvic lymph node specimens (from aborted surgery) were scalpel-dissected such that $\geq$50% of cells were epithelial, and $\geq$90% of epithelial cells were neoplastic. Total RNA was extracted using the Trizol (Invitrogen, Carlsbad, CA, USA) method.

RACE

5′ RNA ligase-mediated rapid amplification of cDNA ends (RACE) was performed using the SMART RACE cDNA Amplification kit (Clontech, Mountain View, CA, USA) according to the manufacturer's instructions, using a gene-specific primer from exon 7 of *ERG* (GCTGCACCCCCTGTGTTTCTAGCATG CATTAACCG). The PCR products were resolved by electrophoresis on a 1% agarose gel, excised and TOPO-TA cloned into pCR4-TOPO (Invitrogen). Purified plasmid DNA from eight independent colonies was bi-directionally sequenced using M13 forward and reverse primers on an ABI 3100 automated capillary DNA sequencer by the Stanford PAN facility.

RT-PCR

RT-PCR was performed as described previously[12] to characterize expression of *TMPRSS2* isoforms, and the known and variant *TMPRSS2-ERG* fusions. Briefly, cDNA was synthesized by oligo(dT)-priming using SuperScript First-Strand Synthesis System (Invitrogen). PCR was then carried out using 1/25th the RT reaction product for template, 1 $\times$ AmpliTaq PCR buffer, 200 μM dNTPs, 1.5 mM $MgCl_2$, 10 pmol of each gene-specific primer pair (Table S1, Supplementary Information), and 2 U AmpliTaq DNA polymerase in a 50 μl reaction for 35–40 cycles (see Table S1, Supplementary Information for primer-specific annealing/extension conditions), followed by electrophoresis of 1/10th the PCR product on a 1% TAE agarose gel, and imaging using an AlphaImager 2200 (Alpha Innotech, San Leandro, CA, USA).

Fluorescence *In Situ* Hybridization

Fluorescence *in situ* hybridization (FISH) was carried out on paraffin sections of selected prostate tumor cases arrayed in tissue microarray (TMA) format, as described previously.[12] A 6 μM TMA section was baked overnight at 60°C, then deparaffinized in xylene and denatured in 100% ethanol. The TMA was then subjected to sequential pretreatment washes: 0.2 N HCl for 20 min, ddH_2O for 10 min, 2 $\times$ SSC for 3 min and 1 M NaSCN at 80°C for 30 min. Slides were then protease-treated at 37°C for 10 min, dehydrated in 100% ethanol and air-dried. FISH analysis was performed as described (Rajput *et al*[13]) using the following BACs (BACPAC Resources Centre, Children's Hospital Oakland Research Institute, Oakland, CA, USA): RP11-95I21 (5′ *ERG*), RP11-476D17 (3′ *ERG*) and RP11-35C4 (telomeric to *TMPRSS2*). BACs RP11-95I21 and RP11-35C4 were directly labeled by nick translation using Spectrum Green and Orange, respectively (Vysis, Downer's Grove, IL, USA). BAC RP11-476D17 was indirectly labeled using a modified protocol with Cy5 (MetaSystems, Belmont, MA, USA) using the BioPrime DNA labeling system (Invitrogen). FISH signals were visualized on a Zeiss Axioplan epifluorescent microscope, and captured using the ISIS FISH imaging software (MetaSystems, Belmont, MA, USA); signals were counted for 50 tumor cells per specimen.

Results

In an array-based comparative genomic hybridization (array CGH) study of prostate tumors (paper in

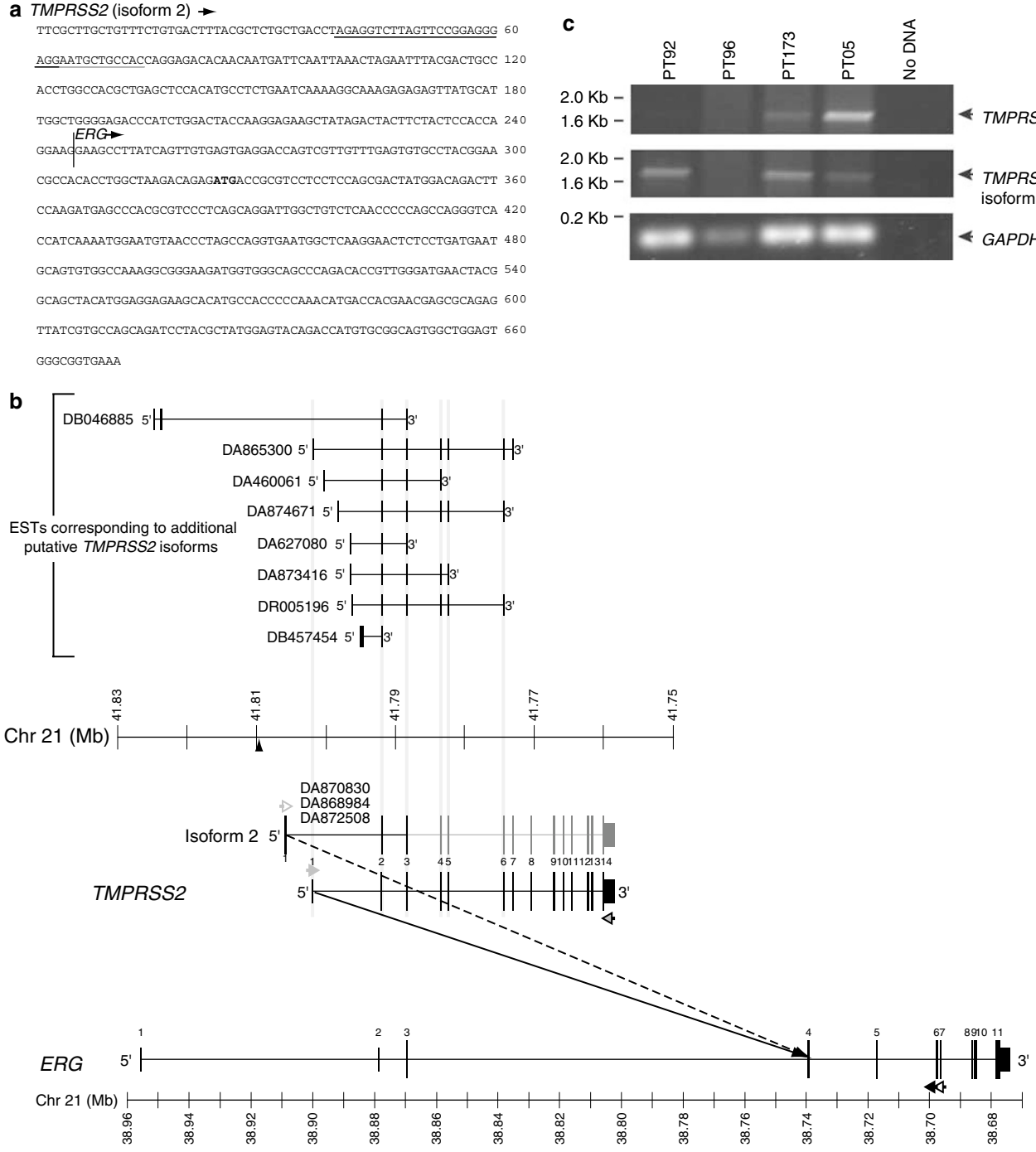

Figure 1 Variant *TMPRSS2* isoform and *ERG* fusion product. (**a**) RACE identifies novel DNA sequences fused to *ERG*. Novel sequences derive from a *TMPRSS2* isoform with alternate first exon. Junction of *TMPRSS2* (isoform 2) and *ERG* sequences is indicated. RT-PCR primers used to validate expression of *TMPRSS2* isoform 2 (black line) and *ERG* fusion (black/gray line) are indicated. Numbering is from the first nucleotide of RACE product. Predicted start site of translation, an in-frame ATG in exon 4 of *ERG*, is in bold (**b**) Genomic structure of *ERG* (isoform 2),[19] *TMPRSS2*, and *TMPRSS2* (isoform 2), and exonic organization of the novel *TMPRSS2* (isoform 2,exon1)-*ERG* (exon4) fusion (dashed connecting line), and the most commonly reported known *TMPRSS2* (exon1)-*ERG* (exon4) fusion (solid connecting line), drawn to scale from genome assembly (http://genome.ucsc./edu). Position of primers is indicated by arrows: RACE/3′-*TMPRSS2* (isoform 2)-*ERG* fusion (black-open), 3′-*TMPRSS2-ERG* (black-filled), 5′-*TMPRSS2* (gray-filled), 5′-*TMPRSS2* isoform 2 (gray-open), 3′-*TMPRSS2* (black/gray-filled). Mapped ESTs representing additional putative *TMPRSS2* variants (see text) are indicated, top. Location of sequences reported by Tomlins *et al*[18] fused to *ETV4* is indicated by upwards-pointing arrowhead. (**c**) RT-PCR analysis of *TMPRSS2* (upper panel), *TMPRSS2* isoform 2 (middle panel), and control *GAPDH* (lower panel) expression in four primary prostate tumors (PT) along with no-template control. Molecular sizes (kb) are indicated.

preparation), we identified a single specimen (PT187) with an apparent single-copy terminal deletion of chr 21q with the break point at *ERG* (21q22. 2), suggesting a possible inter-chromosomal rearrangement and alternative fusion partner of *ERG*. To identify the 5′ fusion partner of *ERG* in this specimen, we carried out RACE, using a 3′ primer within the *ERG* gene. The resultant PCR products, a predominant band (~0.9 kb) and a lesser band (~0.8 kb), were separately gel-purified, cloned and sequenced. The lower molecular-weight (MW) band corresponded to the previously reported *TMPRSS2* (exon1)-*ERG* (exon4) fusion,[5] whereas four independent clones from the higher-MW band represented an identical novel sequence (Figure 1a) fused to the fourth exon of *ERG*. A BLAST search of this sequence to the human expressed sequence tag (EST) database yielded a match with the first exon of three overlapping ESTs (DA872508, DA868984 and DA870830; cloned from prostate tissue), mapping 4 kb upstream (ie distal on chr 21) of exon 1 of *TMPRSS2* (Figure 1b).

The genome annotation (http://genome.ucsc.edu) suggested these sequences to encompass an alternative start site of a variant of *TMPRSS2* that included an alternate noncoding exon 1, along with known exons 2 and 3 of *TMPRSS2*. To define further the novel transcript, we carried out RT-PCR analysis using a 5′ primer within the novel sequences (Figure 1a), and a 3′ primer from the terminal exon (exon 14) of *TMPRSS2*. We identified an expressed transcript in three of four prostate tumors analyzed (Figure 1c), with a size consistent with its comprising an alternative exon 1 but sharing the known exons 2–14 of *TMPRSS2* (Figure 1b). We therefore propose to name the novel non-rearranged transcript variant, *TMPRSS2* isoform 2.

In the index prostate tumor specimen on which we carried out RACE, we identified noncoding exon 1 of the variant *TMPRSS2* isoform 2 fused to the fourth exon of *ERG*. To confirm the expression and assess the frequency of the variant *TMPRSS2* isoform-*ERG* fusion in prostate cancer, we used RT-PCR primer sets (Figure 1b; Table S1, Supplementary Information) designed to amplify specifically either the known or variant fusion transcripts. Amplification of *GAPDH* served as a positive control. Altogether, we surveyed 63 prostate tumors, including 54 primary prostate tumors and nine unmatched (ie from different patients) therapy-naïve lymph node metastases, for which we had previously profiled gene expression by cDNA microarray.[11] Exemplary data are shown in Figure 2. Of the 63 specimens, 44 (70%) cases expressed either the known or novel variant *TMPRSS2-ERG* fusion, and 28 (44%) expressed both. Ten cases (16%) expressed only the known fusion, and notably six cases (10%) expressed only the novel variant isoform fusion (eg Figure 2, specimens PT112 and PL114).

The co-expression of both the known and variant fusions in a subset of specimens might reflect alternative transcriptional start sites from the same rearranged chromosome, or alternatively disparate start sites from two different rearranged chromosomes. To distinguish between these possibilities, we carried out FISH analysis on paraffin sections corresponding to four prostate tumor specimens co-expressing the known and variant fusions. In two of the four specimens analyzed, only a single (rather than multiple) *TMPRSS2-ERG* chromosomal rearrangement was detected in the majority of cells (Figure 3), indicating that both transcripts could be expressed from the same rearranged locus.

Scoring the known and variant *TMPRSS2-ERG* fusions together, *TMPRSS2-ERG* expression showed no statistical association with specimen type (primary *vs* metastasis), tumor stage or Gleason grade (Table 1). We also found no association with recurrence-free survival time (Figure S1, Supplementay Information), though the number of specimens with clinical follow-up ($n = 28$), and the median clinical follow up time (2 years) were both modest.

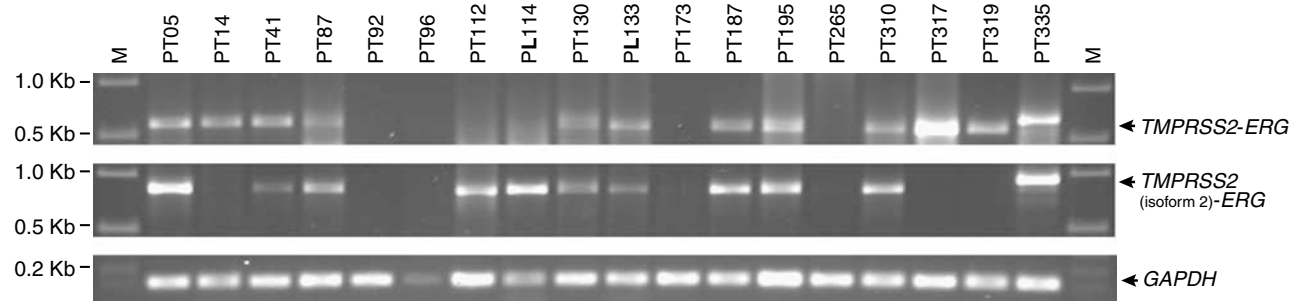

Figure 2 Expression of the variant *TMPRSS2* (isoform 2)-*ERG* fusion in prostate cancer. Shown are representative RT-PCR assays scoring the expression of known (top panel) and variant isoform (middle panel) *TMPRSS2-ERG* fusion transcripts in prostate cancer, along with *GAPDH* control (lower panel). Molecular sizes (kb) are indicated. PT and PL indicate primary tumor and lymph node metastasis samples, respectively. PCR product sizes are consistent with, and limited DNA sequence analysis confirms that most fusions represent *TMPRSS2* (exon1)-*ERG* (exon4) transcripts (top panel) and *TMPRSS2* (isoform 2, exon1)-*ERG* (exon4) transcripts (middle panel). The higher-MW PCR product from sample PT335 corresponds to fusion of *TMPRSS2* (exon1) to *ERG* (exon3), and a lower-MW PCR product from sample PT316 (not shown) corresponds to fusion of *TMPRSS2* (exon2) to *ERG* (exon5); both variants have been previously reported.[15,17]

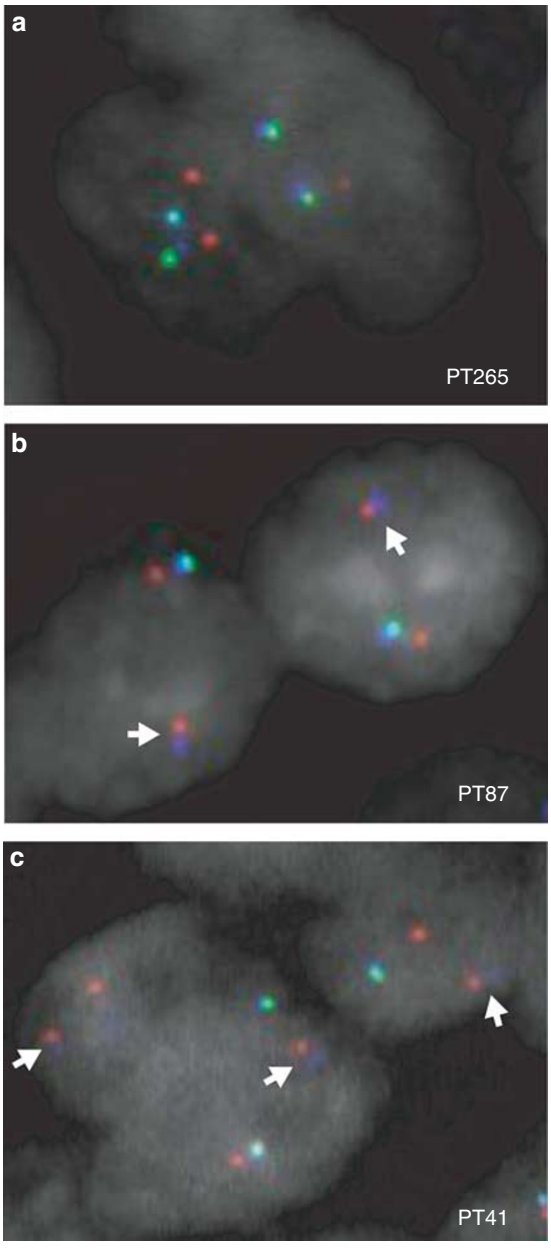

Figure 3 FISH analysis of *TMPRSS2-ERG* rearrangement. Representative FISH images from prostate tumors determined by RT-PCR to express (**a**) no *TMRPSS2-ERG* fusion (PT265), and (**b**, **c**) both known and novel variant isoform *TMPRSS2-ERG* fusion transcripts (PT87, PT41). FISH signals in these pseudocolored images are red (just telomeric to *TMPRSS2*; RP11-35C4), green (5′ *ERG*; RP11-95I21) and blue (3′ *ERG*; RP11-476D17). Rearrangement (arrows) is evident by juxtaposition of red and blue signals, typically with loss of green signal. Note that in non-rearranged chromosomes the proximity of green and blue signals can result in an aqua-colored signal.

Discussion

Our report represents an independent confirmation of the high frequency of *TMPRSS2-ERG* fusion in prostate cancer. Scored together, the known and/or variant fusion was expressed in 70% of cases assayed. These findings are comparable to the original report by Tomlins *et al*[5] (79% of cases, scored by FISH), and recent confirmations by Soller *et al*[14] (78%, RT-PCR), Yoshimoto *et al*[15] (40%, RT-PCR), Perner *et al*[16] (49%, FISH), and Wang *et al*[17] (59%, RT-PCR). However, it should be noted that our set of cases was selected in part for sufficient scalpel-dissected tumor material for expression profiling (before the availability of robust methods for RNA amplification). Therefore, it is biased towards larger specimens, and our result should not be regarded as a true population estimate.

In our set of specimens, we found no relationship between *TMPRSS2-ERG* expression and tumor stage, Gleason grade or tumor recurrence. Although the *TMPRSS2* (exon 1)-*ERG* (exon 4) fusion is the most common, other expressed variants have been reported,[15,17] differing with respect to the exon break point sites in *TMPRSS2* (exons 1–5) and *ERG* (exons 2–5). Depending on the exon break point, protein translation is predicted to start either at the native ATG of *ERG* (exon 3), the native ATG of *TMPRSS2* (exon 2), or an in-frame ATG within exon 4 or 5 of *ERG*; in the latter cases the N terminus of ERG would be truncated. Interestingly, the expression of selected *TMPRSS2-ERG* break point variants has recently been linked to specific clinicopathological features of prostate tumors.[17] Although we had sequenced only a subset of fusion products in our study (see Figure 2 legend), the PCR product sizes of most are consistent with their representing the common *TMPRSS2* (exon1)-*ERG* (exon4) and the novel *TMPRSS2* (isoform2, exon1)-*ERG* (exon4) transcripts, for both of which protein translation is predicted to initiate from an in-frame ATG in *ERG* exon 4.

Given that *TMPRSS2* isoform 2 is expressed in prostate tumors, and that a fraction of prostate tumors harbors rearrangement of the variant *TMPRSS2* (isoform 2)-*ERG* fusion alone (where it is a presumptive pathogenic alteration), it is likely that *TMPRSS2* isoform 2 and the resultant *ERG* fusion are under similar androgen-regulated transcriptional control as is *TMPRSS2* (isoform 1) and its fusion. However, further studies are needed to elucidate the roles of the known and variant *ERG* fusions in the pathogenesis and androgen dependency (and subsequent independency) of prostate cancer.

Interestingly, also mapped on the genome assembly (Figure 1b, top) are additional ESTs that, like the ESTs (DA872508, DA868984 and DA870830) corresponding to *TMPRSS2* isoform 2, appear likely to comprise alternative first exons and transcriptional start sites of *TMPRSS2* variants. The first exons of these distinct transcripts map both upstream (DB046885; ~23 kb upstream of *TMPRSS2*) and downstream (DA865300, DA460061, DA874671, DA627080, DA873416, DR005196 and DB457454) of *TMPRSS2* exon 1. It is possible that these putative *TMPRSS2* isoforms also serve as androgen-regulated

Table 1 Histopathological associations of *TMPRSS2-ERG* fusion

Histopathological parameter	*No* TMPRSS2-ERG *fusion*	TMPRSS2-ERG *fusion*	
Sample type			
Primary ($n=54$)	15 (28%)	39 (72%)	$P=0.43$[a]
LN metastasis ($n=9$)	4 (44%)	5 (56%)	
Tumor stage[b,c]			
Local ($\leq$T2) ($n=24$)	9 (38%)	15 (62%)	$P=0.22$[a]
Advanced ($\geq$T3) ($n=25$)	5 (20%)	20 (80%)	
Gleason grade[b]			
<7 ($n=24$)	9 (38%)	15 (62%)	$P=0.32$[d]
$=7$ ($n=17$)	4 (24%)	13 (76%)	
>7 ($n=13$)	2 (15%)	11 (85%)	

Abbreviation: LN, lymph node.
[a]Fisher's exact test (two-tail).
[b]Primary tumors only.
[c]Information not available on all specimens.
[d]χ^2 test.

fusion partners for *ERG* (or *ETV1*) rearrangements. Indeed, our preliminary data (not shown) identify the additional expression of a DA460061 (exon1)-ERG (exon 4) fusion in the index specimen (PT187). To note, Tomlins *et al*[18] recently reported infrequent cases of prostate cancer harboring novel genomic sequences 8 kb upstream of *TMPRSS2* (Figure 1b, arrowhead) fused to *ETV4*, also an ETS family member. These sequences, which do not match any known EST, are distinct from any of those we report here, but may also result from androgen-regulated expression.

In the index prostate tumor specimen for which we carried out RACE, an apparent terminal deletion identified by array CGH suggested an inter-chromosomal rearrangement. However, RACE-PCR identified only the known and variant isoform *TMPRSS2-ERG* fusion reported here. A likely explanation is that *TMPRSS2-ERG* fusion in this specimen results from an unbalanced inter-chromosomal rearrangement between two copies of chr 21. Indeed, a recent study,[16] and our own data (unpublished), indicate that only half or less of *TMPRSS2-ERG* fusion events are associated with intra-chromosomal deletion between the *TMPRSS2* and *ERG* loci on chr 21. Whether other prostate tumor specimens might contain rearrangements fusing *ERG* to genes other than *TMPRSS2*, and its isoforms described here, remains to be investigated.

Our findings underscore the pathogenetic relevance of altered *ERG* expression in prostate cancer, as well as the potential for novel therapeutic opportunities. Importantly, 10% of specimens in our series expressed only the variant isoform *TMPRSS2-ERG* fusion, which would be missed using published RT-PCR primers from the known *TMPRSS2* transcript.[5,14,15,17] Our discovery, therefore, has important implications for the accurate detection of *TMPRSS2-ERG* fusion transcript, which is desirable for defining the molecular and pathological correlates of the fusion in tumor specimens, and may be useful in the future for molecular diagnosis. Specifically, accurate detection of *TMPRS22-ERG* fusion transcripts in prostate tissue, or in fluids such as urine, prostatic secretions, seminal fluid, ejaculate or blood, may prove useful in the diagnosis of prostate cancer, in prognostication and selection of appropriately aggressive therapies, and possibly in the future in selection of patients for new directed therapies targeting the oncogenic fusion.

Acknowledgement

We thank members of the Pollack lab for helpful discussion. This work was supported in part by grants from the NIH (CA111782; JDB). Nucleotide sequences for the *TMPRSS2* (isoform 2)-*ERG* fusion have been deposited at GenBank with accession number EF194202.

References

1 Jemal A, Siegel R, Ward E, *et al.* Cancer statistics, 2006. CA Cancer J Clin 2006;56:106–130.
2 Khan MA, Partin AW. Management of patients with an increasing prostate-specific antigen after radical prostatectomy. Curr Urol Rep 2004;5:179–187.
3 Kasamon KM, Dawson NA. Update on hormone-refractory prostate cancer. Curr Opin Urol 2004;14: 185–193.
4 Petrylak DP, Tangen CM, Hussain MH, *et al.* Docetaxel and estramustine compared with mitoxantrone and prednisone for advanced refractory prostate cancer. N Engl J Med 2004;351:1513–1520.

5 Tomlins SA, Rhodes DR, Perner S, *et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005;310:644–648.
6 Seth A, Watson DK. ETS transcription factors and their emerging roles in human cancer. Eur J Cancer 2005;41:2462–2478.
7 Paoloni-Giacobino A, Chen H, Peitsch MC, *et al.* Cloning of the TMPRSS2 gene, which encodes a novel serine protease with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3. Genomics 1997;44:309–320.
8 Lin B, Ferguson C, White JT, *et al.* Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. Cancer Res 1999; 59:4180–4184.
9 Afar DE, Vivanco I, Hubert RS, *et al.* Catalytic cleavage of the androgen-regulated TMPRSS2 protease results in its secretion by prostate and prostate cancer epithelia. Cancer Res 2001;61:1686–1692.
10 Jacquinet E, Rao NV, Rao GV, *et al.* Cloning and characterization of the cDNA and gene for human epitheliasin. Eur J Biochem 2001;268:2687–2699.
11 Lapointe J, Li C, Higgins JP, *et al.* Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proc Natl Acad Sci USA 2004;101: 811–816.
12 Kim H, Lapointe J, Kaygusuz G, *et al.* The retinoic acid synthesis gene ALDH1a2 is a candidate tumor suppressor in prostate cancer. Cancer Res 2005;65: 8118–8124.
13 Rajput AB, Miller MA, De Luca A, *et al.* Frequency of the TMPRSS2:ERG gene fusion is increased in moderate to poorly differentiated prostate cancers. J Clin Pathol 2007, 26 January [E-pub ahead of print].
14 Soller MJ, Isaksson M, Elfving P, *et al.* Confirmation of the high frequency of the TMPRSS2/ERG fusion gene in prostate cancer. Genes Chromosomes Cancer 2006;45:717–719.
15 Yoshimoto M, Joshua AM, Chilton-Macneill S, *et al.* Three-color FISH analysis of TMPRSS2/ERG fusions in prostate cancer indicates that genomic microdeletion of chromosome 21 is associated with rearrangement. Neoplasia 2006;8:465–469.
16 Perner S, Demichelis F, Beroukhim R, *et al.* TMPRSS2:ERG Fusion-Associated Deletions Provide Insight into the Heterogeneity of Prostate Cancer. Cancer Res 2006;66:8337–8341.
17 Wang J, Cai Y, Ren C, *et al.* Expression of variant TMPRSS2/ERG fusion messenger RNAs is associated with aggressive prostate cancer. Cancer Res 2006;66: 8347–8351.
18 Tomlins SA, Mehra R, Rhodes DR, *et al.* TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. Cancer Res 2006;66: 3396–3400.
19 Owczarek CM, Portbury KJ, Hardy MP, *et al.* Detailed mapping of the ERG-ETS2 interval of human chromosome 21 and comparison with the region of conserved synteny on mouse chromosome 16. Gene 2004;324: 65–77.

Supplementary Information accompanies the paper on the Modern Pathology website (http://www.nature.com/modpathol)