TMPRSS2:ETV4 Gene Fusions Define a Third Molecular Subtype of Prostate Cancer

Scott A. Tomlins,[1] Rohit Mehra,[1] Daniel R. Rhodes,[1,3] Lisa R. Smith,[1] Diane Roulston,[1] Beth E. Helgeson,[1] Xuhong Cao,[1] John T. Wei,[2,4,5] Mark A. Rubin,[6,7] Rajal B. Shah,[1,2,4,5] and Arul M. Chinnaiyan[1,2,3,4,5]

Departments of [1]Pathology and [2]Urology, [3]Program of Bioinformatics, [4]Michigan Urology Center, and the [5]Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, Michigan and [6]Department of Pathology, Brigham and Women's Hospital, and [7]Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts

Abstract

Although common in hematologic and mesenchymal malignancies, recurrent gene fusions have not been well characterized in epithelial carcinomas. Recently, using a novel bioinformatic approach, we identified recurrent gene fusions between *TMPRSS2* and the *ETS* family members *ERG* or *ETV1* in the majority of prostate cancers. Here, we interrogated the expression of all *ETS* family members in prostate cancer profiling studies and identified marked overexpression of *ETV4* in 2 of 98 cases. In one such case, we confirmed the overexpression of *ETV4* using quantitative PCR, and by rapid amplification of cDNA ends, quantitative PCR, and fluorescence *in situ* hybridization, we show that the *TMPRSS2* (21q22) and *ETV4* (17q21) loci are fused in this case. This result defines a third molecular subtype of prostate cancer and supports the hypothesis that dysregulation of *ETS* family members through fusions with *TMRPSS2* may be an initiating event in prostate cancer development. (Cancer Res 2006; 66(7): 3396-400)

Introduction

Despite their well-defined role in hematologic and mesenchymal malignancies, recurrent gene fusions have not been well characterized in epithelial carcinomas (1–3). Recently, in an effort to nominate candidate oncogenes from DNA microarray data, we developed a novel bioinformatic approach termed cancer outlier profile analysis (COPA) to identify genes markedly overexpressed in a subset of cancers. Applying the COPA approach to a compendium of tumor gene expression data, the *ETS* family transcription factors *ERG* and *ETV1* were identified as outliers across prostate cancer profiling studies. Using a variety of molecular techniques, we characterized fusions of the 5′-untranslated region (5′UTR) of *TMPRSS2* (21q22) with *ERG* (21q22) or *ETV1* (7p21) in cases that overexpressed the respective *ETS* family member (4). *TMPRSS2* has been characterized previously as being both androgen responsive and highly expressed in the prostate, presumably through androgen response elements (ARE) in the promoter (5–7). As a possible mechanism driving the overexpression of *ETS* family members in cases with the gene fusions, we showed that androgen can induce the overexpression of *ERG* (presumably through AREs) in a *TMPRSS2:ERG*-positive cell line (4). Together, these results suggested that dysregulation of *ETS* family activity through AREs upstream of *TMPRSS2* may drive prostate cancer development. Here, we describe a rare third molecular subtype of prostate cancer, characterized by fusion of *TMPRSS2* to another *ETS* family member, *ETV4*.

Materials and Methods

ETS family expression in profiling studies. To investigate the expression of *ETS* family members in prostate cancer, we selected two prostate cancer profiling studies (8)[8] present in the Oncomine database (9). Genes with an ETS domain were identified by the Interpro filter "Ets" (Interpro ID: IPR000418). Heatmap representations were generated in Oncomine using the "median center per gene" option, and the color contrast was set to accentuate *ERG* and *ETV1* differential expression.

Samples. Prostate cancer tissues (PCA1-5) were from the radical prostatectomy series at the University of Michigan, which is part of the University of Michigan Prostate Cancer Specialized Program of Research Excellence Tissue Core. All samples were collected with informed consent of the patients and prior institutional review board approval. Total RNA was isolated with Trizol (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. A commercially available pool of benign prostate tissue total RNA (CPP, Clontech, Mountain View, CA) was also used.

Quantitative PCR. Quantitative PCR was done using SYBR Green dye on an Applied Biosystems 7300 Real-time PCR system (Applied Biosystems, Foster City, CA) as described (4). The amount of each target gene relative to the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) for each sample was reported. The relative amount of the target gene was calibrated to the relative amount from the pool of benign prostate tissue (CPP). All oligonucleotide primers were synthesized by Integrated DNA Technologies (Coralville, IA). *GAPDH* primers were as described (10). Primers for exons of *ETV4* were as follows (listed 5′ to 3′): ETV4_exon2-f, CCGGATGGAGCGGAGGATGA; ETV4_exon2-r, CGGGCGATTTGCTGCTGAAG; ETV4_exon3-f, GCCGCCCCTCGACTCTGAA; ETV4_exon4-r, GAGCCACGTCTCCTGGAAGTGACT; ETV4_exon11-f, CTGGCCGGTTCTTCTGGATGC; ETV4_exon12-r, CGGGCCGGGGAATGGAGT; ETV4_3′UTR-f, CCTGGAGGGTACCGGTTTGTCA; ETV4_3′UTR-r, CCGCCTGCCTCTGGGAACAC. Exons were numbered by alignment of the RefSeq for *ETV4* (NM_001986.1) with the May 2004 freeze of the human genome using the University of California Santa Cruz Genome Browser. For quantitative PCR confirmation of *TMPRSS2:ETV4* fusion transcripts, TMPRSS2:*ETV4*a-f (AAATAAGTTTGTAAGAGGAGCCTCAGCATC) and TMPRSS2:*ETV4*b-f (ATCGTAAAGAGCTTTTCTCCCCGC), which detects both *TMPRSS2:ETV4*a and *TMPRSS2:ETV4*b transcripts, were used with ETV4_exon4-r.

RNA ligase–mediated rapid amplification of cDNA ends. RNA ligase–mediated rapid amplification of cDNA ends (RLM-RACE) was done using the GeneRacer RLM-RACE kit (Invitrogen), according to the manufacturer's

[8] S.A. Tomlins et al., in preparation.

Note: S.A. Tomlins, R. Mehra, and D.R. Rhodes contributed equally to this report. A.M. Chinnaiyan is a Pew Biomedical Scholar. S.A. Tomlins and D.R. Rhodes are Fellows of the Medical Scientist Training Program.

Genbank Accession numbers for TMPRSS2:ETU4a and TMPRSS2:ETU4b are DQ396625-6.

Requests for reprints: Arul M. Chinnaiyan, Departments of Pathology and Urology, University of Michigan Medical School, University of Michigan, 1301 Catherine Street, MSI 4237, Ann Arbor, MI 48109-0602. Phone: 734-647-8153; Fax: 734-936-7361; E-mail: arul@umich.edu.

doi:10.1158/0008-5472.CAN-06-0168

instructions as described (4). To obtain the 5′ end of ETV4, first-strand cDNA from PCA5 was amplified using the GeneRacer 5′ Primer and ETV4_exon4-r or ETV4_exon7-r (GAAAGGGCTGTAGGGGCGACTGT). Products were cloned and sequenced as described (4). Equivalent 5′ ends of the *TMPRSS2:ETV4* transcripts were obtained from both primer pairs.

Fluorescence *in situ* hybridization. Formalin-fixed, paraffin-embedded tissue sections were used for interphase fluorescence *in situ* hybridization (FISH). Deparaffinized tissue was treated with 0.2 mol/L HCl for 10 minutes, 2× SSC for 10 minutes at 80°C and digested with Proteinase K (Invitrogen) for 10 minutes. The tissues and BAC probes were codenatured for 5 minutes at 94°C and hybridized overnight at 37°C. Post-hybridization washing was with 2× SSC with 0.1% Tween 20 for 5 minutes, and fluorescent detection was done using anti-digoxigenin conjugated to fluorescein (Roche Applied Science, Indianapolis, IN) and streptavidin conjugated to Alexa Fluor 594 (Invitrogen). Slides were counterstained and mounted in ProLong Gold Antifade Reagent with 4′,6-diamidino-2-phenylindole (Invitrogen). Slides were examined using a Leica DMRA fluorescence microscope (Leica, Deerfield, IL) and imaged with a CCD camera using the CytoVision software system (Applied Imaging, Santa Clara, CA).

All BACs were obtained from the BACPAC Resource Center (Oakland, CA), and probe locations were verified by hybridization to metaphase spreads of normal peripheral lymphocytes. For detection of *TMPRSS2:ETV4* fusion, RP11-35C4 (5′ to *TMPRSS2*) was used with multiple BACs located 3′ to *ETV4* (distal to *ETV4* to proximal: RP11-266I24, RP11-242D8, and RP11-100E5). For detection of *ETV4* rearrangements, RP11-436J4 (5′ to *ETV4*) was used with the multiple BACs 3′ to *ETV4*. For each hybridization, areas of cancerous cells were identified by a pathologist, and 100 cells were counted per sample. The reported cell count for *TMPRSS2:ETV4* fusions used RP11-242D8, and similar results were obtained with all 3′ *ETV4* BACs. To exclude additional rearrangements in PCA5, we did FISH with two probes 3′ to *ETV4* (RP11-266I24 and RP11-242D8): *ERG* split signal probes (RP11-95I21 and

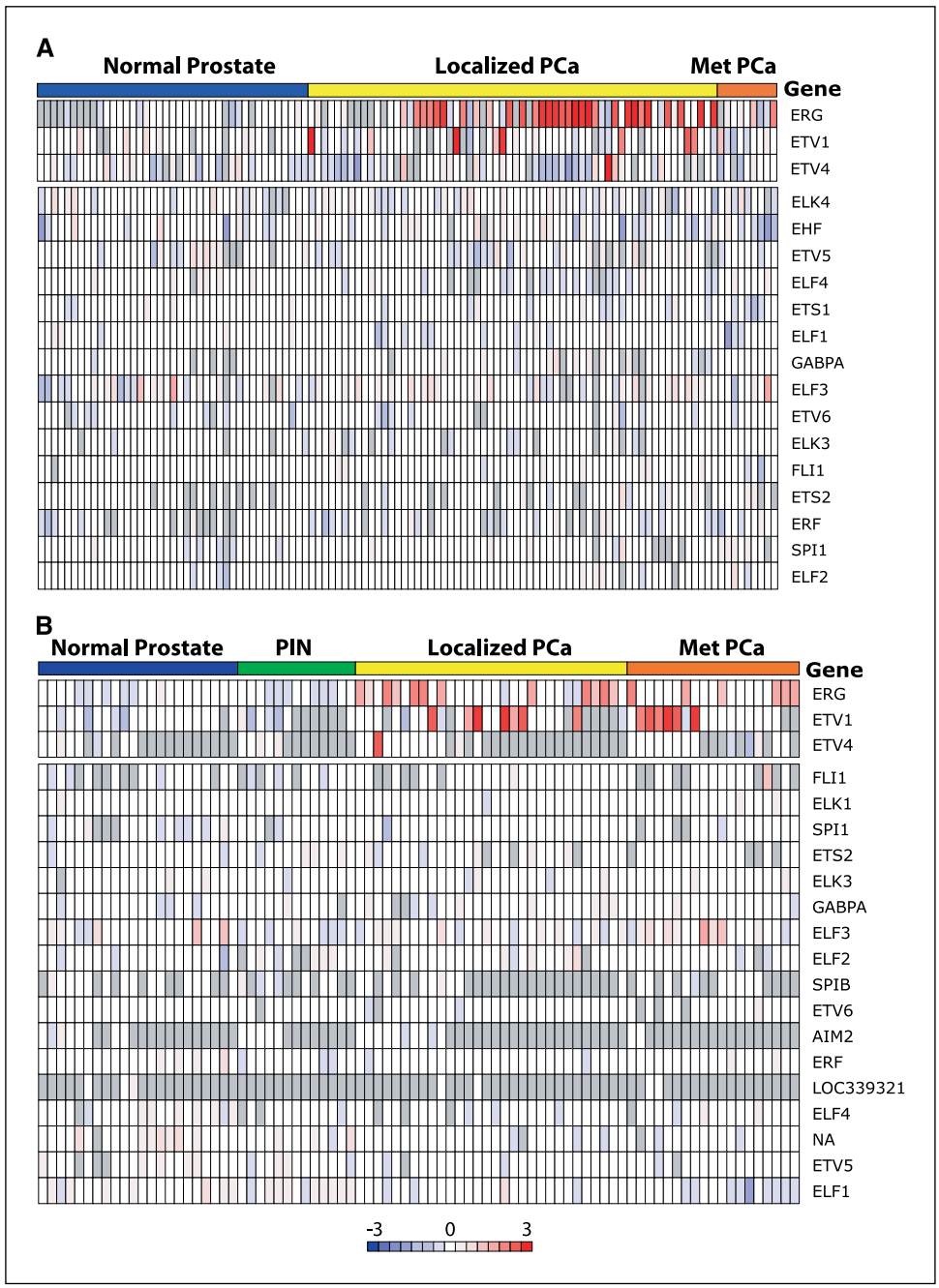

Figure 1. Overexpression of *ETS* family members in prostate cancer. Expression of all monitored *ETS* family members in profiled benign prostate (*blue*), prostatic intraepithelial neoplasia (*PIN*; *green*), clinically localized prostate cancer (*PCa*; *yellow*), and metastatic prostate cancer (*Met PCa*; *orange*) from grossly dissected tissue (*A*; ref. 8) or tissue isolated by laser capture microdissection (*B*)[8] was visualized using Oncomine (http://www.oncomine.org). *Columns,* samples of the indicated class; *rows,* indicated gene. Relative underexpression (*blue*) or overexpression (*red*; in *z*-score units, median centered per gene), as indicated by the color scale. Features that did not pass filtering in the original study (*gray*).

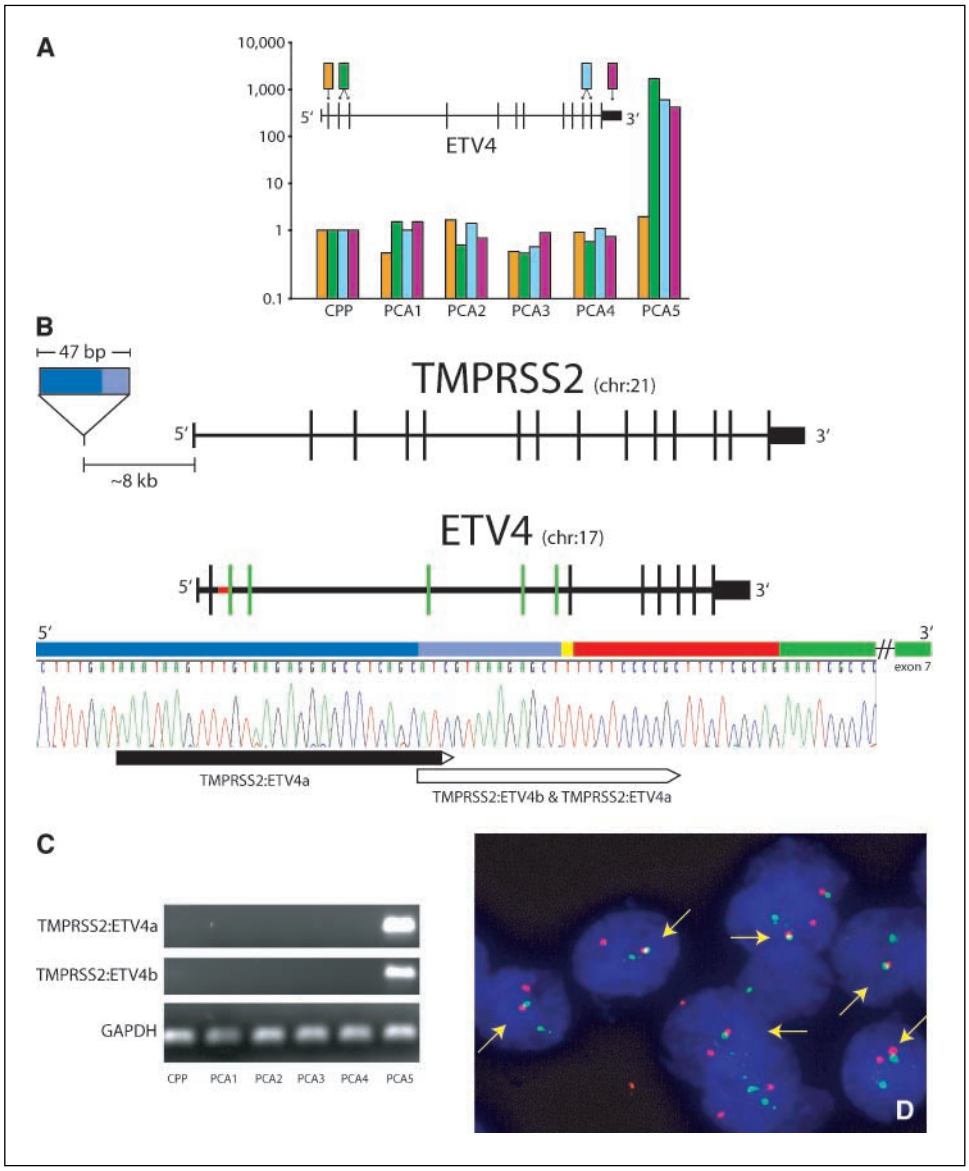

Figure 2. Fusion of *TMPRSS2* and *ETV4* loci in a prostate cancer case that overexpresses *ETV4*. *A,* quantitative PCR was used to determine the relative expression of the indicated exons or region of *ETV4* in pooled benign prostate tissue (CPP), prostate cancers that did not overexpress *ETV4* and were either *TMPRSS2:ERG* positive (PCA1-2) or negative (PCA3-4), and the prostate cancer case from our LCM cohort with *ETV4* overexpression (PCA5). The amount of *ETV4* was normalized to the amount of *GAPDH*, and the relative amount of *ETV4* for each sample was calibrated to the amount in CPP. Results are presented on a log scale. *B,* RLM-RACE reveals fusion of sequences upstream of *TMPRSS2* with *ETV4* in PCA5. RLM-RACE was done to identify the 5′ end of the *ETV4* transcript in PCA5 using a reverse primer in exon 7 of *ETV4*. Sequencing of the cloned product revealed two transcripts beginning with 47 bp (*TMPRSS2:ETV4a*, *dark* and *light blue*) or 13 bp (*TMPRSS2:ETV4b, light blue*) located ~8 kb upstream of *TMPRSS2*: a single base pair that did not map to this region or *ETV4* (*yellow*) and a contiguous region composed of the 19 terminal base pairs in the intron before exon 3 of *ETV4* (*red*) and the reference sequence of *ETV4* from exons 3 to 7 (*green*). For the structural diagrams of the *TMPRSS2* and *ETV4* loci, exons are represented by boxes, introns by horizontal lines, and regions not involved in the RLM-RACE products are in black. The same coloring for the structural diagrams is used for the nucleotide tracing shown for *TMPRSS2:ETV4a*, and the location of primers used to confirm the presence of the indicated fusion transcripts are shown below the tracing. *C,* expression of *TMPRSS2:ETV4a* and *TMPRSS2:ETV4b* in PCA5 by quantitative PCR. Quantitative PCR was done using forward primers located as shown in (*B*) for *TMPRSS2:ETV4a* or *TMPRSS2:ETV4b* and ETV4_exon4-r or GAPDH primers on the same samples as (*A*). As no amplification was detectable after 40 cycles for CPP and PCA1-4, the results could not quantified, and products were electrophoresed on a 1.2% agarose gel and visualized with ethidium bromide. *D,* interphase FISH on formalin-fixed, paraffin-embedded tissue confirms fusion of *TMPRSS2* and *ETV4* loci in PCA5. Probes for *TMPRSS2* (*red*) and *ETV4* (*green*) show fusion of the genomic loci (*yellow arrows*) in cancerous cells from PCA5.

RP11-476D17) and *TMPRSS2:ETV1* fusion probes (RP11-35C4 and RP11-124L22). BAC DNA was isolated using a QIAFilter Maxi Prep kit (Qiagen, Valencia, CA), and probes were synthesized using digoxigenin- or biotin-nick translation mixes (Roche Applied Science).

Results and Discussion

Although our initial COPA screen led to the characterization of *TMPRSS2* fusions with *ERG* or *ETV1*, we hypothesized that prostate cancers negative for these gene fusions may harbor rearrangements involving other *ETS* family members. By interrogating the expression of all *ETS* family members monitored in prostate cancer profiling studies from the Oncomine database (http://www.oncomine.org; ref. 9), we identified marked overexpression of the *ETS* family member *ETV4* in a single prostate cancer case from each of two studies: one profiling grossly dissected tissues (ref. 8; Fig. 1*A*), and the

other profiling laser capture microdissected (LCM) tissues[8] (Fig. 1*B*). As these cases did not overexpress *ERG* or *ETV1*, and no benign prostate tissues showed overexpression, we hypothesized that fusion with *TMPRSS2* may drive the overexpression of *ETV4* in these cases. Although *ELF3* was also overexpressed in a fraction of prostate cancer cases, in both studies, normal prostate tissue samples also showed marked *ELF3* overexpression, suggesting that a gene fusion driving expression in both benign and cancerous tissue is unlikely. Thus, we focused on characterizing the *ETV4* overexpressing case (designated here as PCA5) in our LCM cohort.

We isolated total RNA from PCA5 and used an exon-walking quantitative PCR strategy to confirm the overexpression of *ETV4*. Quantitative PCR showed that exons 3′ to exon 2 of *ETV4* were markedly overexpressed in this case compared with pooled benign prostate tissue (CPP; ~900-fold) and prostate cancers that did not overexpress *ETV4* and were either *TMPRSS2:ERG* positive (PCA1-2) or negative (PCA3-4; Fig. 2*A*). However, we observed a dramatic decrease (>99%) in the expression of exon 2 of *ETV4* relative to distal regions in PCA5, suggesting a possible fusion with *TMPRSS2*, as observed previously in *TMPRSS2:ERG*-positive and *TMPRSS2:ETV1*-positive cases (4).

To identify the 5′ end of the ETV4 transcript in PCA5, we did RLM-RACE using a reverse primer in exon 7. RLM-RACE revealed two transcripts, each containing 5′ ends consisting of sequence located ~8 kb upstream of *TMPRSS2* fused to sequence from *ETV4* (Fig. 2*B*). Specifically, the 5′ end of *TMPRSS2:ETV4a* consists of 47 bp from this region upstream of *TMPRSS2*, whereas the 5′ end of *TMPRSS2:ETV4b* consists of the same terminal 13 bp. These 5′ ends of both transcripts were fused to the same contiguous stretch consisting of the 9 bp of the intron immediately 5′ to exon 3 of ETV4 and the reported reference sequence of exons 3 through the reverse primer in exon 7 of ETV4.

We confirmed the existence of both transcripts in PCA5 and their absence in CPP and PCA1-4 using quantitative PCR; however, the results could not be quantified due to no detectable amplification after 40 cycles in CPP and PCA1-4 (Fig. 2*C*). To further exclude the presence of fusion transcripts involving known exons from *TMPRSS2*, we attempted quantitative PCR using a forward primer in exon 1 of *TMPRSS2* and the *ETV4* exon 4 reverse primer, and as expected, no product was detected in CPP or PCA1-5 (data not shown).

Whether other prostate cancers with *ETV4* dysregulation might contain *TMPRSS2:ETV4* fusion transcripts structurally more similar to *TMPRSS2:ERG* and *TMPRSS2:ETV1* transcripts (which involve known exons from *TMPRSS2*) is unknown. It is important to note that the *TMPRSS2:ETV4* fusions reported here would not contain the well-characterized AREs immediately upstream of TMPRSS2. However, evidence exists for androgen-responsive enhancers located upstream of the TMPRSS2 sequences present in the *TMPRSS2:ETV4* transcripts described here.[9] Nevertheless, the marked overexpression of only *ETV4* exons involved in the fusion transcript strongly suggests that the gene fusion is responsible for the dysregulation of *ETV4*. Together, the structure of the *TMPRSS2:ETV4* fusion transcripts supports the conclusion that the regulatory elements upstream of *TMPRSS2*, rather than transcribed *TMPRSS2* sequences, drive the dysregulation of ETS family members.

To confirm the fusion of the genomic loci surrounding *TMPRSS2* (21q22) and *ETV4* (17q21) as shown by RLM-RACE and quantitative PCR, we used interphase FISH. Using probes 5′ to *TMPRSS2* and 3′ to *ETV4*, we identified fusion of *TMPRSS2* and *ETV4* loci in 65% of cancerous cells from PCA5 (Fig. 2*D*). As further confirmation of the rearrangement of *ETV4*, using probes 5′ and 3′ to *ETV4*, 64% of cancerous cells from PCA5 showed split signals (data not shown). We also did FISH on PCA5 using two probes 3′ to *ETV4, ERG* split signal probes and *TMPRSS2:ETV1* fusion probes to exclude additional rearrangements, with negative results obtained for each hybridization (data not shown).

Taken together, the results from this study highlight the importance of carefully examining outlier profiles in tumor gene expression data, as most analytic methods discount profiles that do not show consistent deregulation (11–13) and would thus fail to identify *ETV4* in prostate cancer, which seems rare (2 of 98 cases). Combined with the identification of *TMPRSS2:ERG* and *TMPRSS2:ETV1* fusions, the results presented here support the hypothesis that dysregulation of *ETS* family members mediated by subversion of AREs or enhancers upstream of *TMPRSS2* is a hallmark of prostate tumorigenesis. Although the majority of *ETS* family members were represented in the profiling studies examined, other *ETS* family members that were not monitored may also be rearranged in prostate cancers for which gene fusions have not been ascribed. The reason for the observed frequencies of fusion partners with *TMPRSS2* (*ERG* > *ETV1* > *ETV4*), which are consistent across independent sample sets, is unclear, although a similar situation is present in Ewing's sarcoma, where *EWS* partners with *ETS* family members in unequal frequencies (*FLI1* > *ERG* > *ETV1*; ref. 14). Lastly, these results establish a third molecular subtype of prostate cancer, which may have prognostic and/or therapeutic relevance in the future.

[9] Q. Wang and M. Brown, personal communication.

Acknowledgments

Received 1/16/2006; revised 2/13/2006; accepted 2/17/2006.

Grant support: Early Detection Research Network Biomarker Developmental Lab grant UO1 CA111275-01 (A.M. Chinnaiyan), NIH Prostate Specialized Program of Research Excellence grant P50CA69568 (A.M. Chinnaiyan, M.A. Rubin, J.T. Wei, and R.B. Shah), grants RO1 CA97063 (A.M. Chinnaiyan) and RO1 AG21404 (M.A. Rubin), American Cancer Society grant RSG-02-179-MGO (A.M. Chinnaiyan), Department of Defense grants PC040517 (R. Mehra) and PC020322 (A.M. Chinnaiyan), Cancer Center Bioinformatics Core Support Grant 5P30 CA46592, and Cancer Biology Training Program (D.R. Rhodes).

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.

We thank Anjana Menon for technical assistance and Qianben Wang and Myles Brown for communicating unpublished results.

References

1. Mitelman F. Recurrent chromosome aberrations in cancer. Mutat Res 2000;462:247–53.
2. Rabbitts TH. Chromosomal translocations in human cancer. Nature 1994;372:143–9.
3. Rowley JD. Chromosome translocations: dangerous liaisons revisited. Nat Rev Cancer 2001;1:245–50.
4. Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005; 310:644–8.
5. 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;61:1686–92.
6. Jacquinet E, Rao NV, Rao GV, Zhengming W, Albertine KH, Hoidal JR. Cloning and characterization of the

cDNA and gene for human epitheliasin. Eur J Biochem 2001;268:2687–99.
7. Lin B, Ferguson C, White JT, et al. Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. Cancer Res 1999; 59:4180–4.
8. Lapointe J, Li C, Higgins JP, et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proc Natl Acad Sci U S A 2004;101: 811–6.
9. Rhodes DR, Yu J, Shanker K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia 2004;6:1–6.
10. Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002;3:RESEARCH0034.
11. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998;95:14863–8.
12. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286:531–7.
13. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001; 98:5116–21.
14. Arvand A, Denny CT. Biology of EWS/ETS fusions in Ewing's family tumors. Oncogene 2001;20:5747–54.

TMPRSS2:ETV4 Gene Fusions Define a Third Molecular Subtype of Prostate Cancer

Scott A. Tomlins, Rohit Mehra, Daniel R. Rhodes, et al.

Cancer Res 2006;66:3396-3400.