LETTERS

Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer

Scott A. Tomlins[1]*, Bharathi Laxman[1]*, Saravana M. Dhanasekaran[1]*, Beth E. Helgeson[1], Xuhong Cao[1], David S. Morris[2], Anjana Menon[1], Xiaojun Jing[1], Qi Cao[1], Bo Han[1], Jindan Yu[1], Lei Wang[1], James E. Montie[2,4], Mark A. Rubin[5,6], Kenneth J. Pienta[2,3,4], Diane Roulston[1], Rajal B. Shah[1,2,4], Sooryanarayana Varambally[1,4], Rohit Mehra[1,4] & Arul M. Chinnaiyan[1,2,4]

Recently, we identified recurrent gene fusions involving the 5′ untranslated region of the androgen-regulated gene *TMPRSS2* and the ETS (E26 transformation-specific) family genes *ERG*, *ETV1* or *ETV4* in most prostate cancers[1,2]. Whereas *TMPRSS2–ERG* fusions are predominant, fewer *TMPRSS2–ETV1* cases have been identified than expected on the basis of the frequency of high (outlier) expression of *ETV1* (refs 3–13). Here we explore the mechanism of *ETV1* outlier expression in human prostate tumours and prostate cancer cell lines. We identified previously unknown 5′ fusion partners in prostate tumours with *ETV1* outlier expression, including untranslated regions from a prostate-specific androgen-induced gene (*SLC45A3*) and an endogenous retroviral element (*HERV-K_22q11.23*), a prostate-specific androgen-repressed gene (*C15orf21*), and a strongly expressed housekeeping gene (*HNRPA2B1*). To study aberrant activation of *ETV1*, we identified two prostate cancer cell lines, LNCaP and MDA-PCa 2B, that had *ETV1* outlier expression. Through distinct mechanisms, the entire *ETV1* locus (7p21) is rearranged to a 1.5-megabase prostate-specific region at 14q13.3–14q21.1 in both LNCaP cells (cryptic insertion) and MDA-PCa 2B cells (balanced translocation). Because the common factor of these rearrangements is aberrant *ETV1* overexpression, we recapitulated this event *in vitro* and *in vivo*, demonstrating that *ETV1* overexpression in benign prostate cells and in the mouse prostate confers neoplastic phenotypes. Identification of distinct classes of ETS gene rearrangements demonstrates that dormant oncogenes can be activated in prostate cancer by juxtaposition to tissue-specific or ubiquitously active genomic loci. Subversion of active genomic regulatory elements may serve as a more generalized mechanism for carcinoma development. Furthermore, the identification of androgen-repressed and insensitive 5′ fusion partners may have implications for the anti-androgen treatment of advanced prostate cancer.

Recurrent chromosomal rearrangements have been causally implicated in haematological and mesenchymal malignancies; although predicted to occur in common epithelial carcinomas, they have not been well characterized[14,15]. Using a bioinformatics strategy to nominate genes showing high (outlier) expression in a subset of cancers, we identified fusions of the 5′-untranslated region of *TMPRSS2* (21q22) to *ERG* (21q22), *ETV1* (7p21) or *ETV4* (17q21) in cases that overexpressed the respective ETS family member[1,2]. *TMPRSS2* had previously been characterized as androgen-regulated[16], and its androgen-responsive regulatory elements drive ETS family member outlier expression[1,17]. Thus, fusions between *TMPRSS2* and ETS family members are functionally similar to haematological malignancy rearrangements in which tissue-specific promoter or enhancer elements of one gene are juxtaposed to proto-oncogenes[15,18].

Multiple studies have confirmed the presence of *TMPRSS2–ERG* fusions in 36–78% of prostate cancers from prostate-specific-antigen-screened surgical cohorts (Supplementary Table 1). Approximately 90% of samples with *ERG* outlier expression harbour *TMPRSS2–ERG* fusions[1,19], confirming this as the predominant mechanism driving *ERG* overexpression. In contrast, although microarray studies show *ETV1* outlier expression in 6–16% of prostate cancer samples, only 2 of 205 (1.0%) samples analysed harboured *TMPRSS2–ETV1* fusions (Supplementary Table 1).

Here we addressed this discrepancy between *ETV1* outlier and *TMPRSS2–ETV1* frequencies. By quantitative PCR (Q-PCR) across 2 cohorts, 26 and 3 of 54 localized prostate cancer samples showed *ERG* (48%) and *ETV1* (5.5%) outlier expression, respectively (Supplementary Fig. 1). Additionally, two hormone-refractory metastatic prostate cancer samples, MET26 (our *TMPRSS2–ETV1* index case[1]) and MET23, showed *ETV1* outlier expression. However, other than MET26, no samples expressed *TMPRSS2–ETV1* fusion transcripts.

To characterize the *ETV1* transcript in outlier cases, we performed 5′-RNA-ligase-mediated rapid amplification of complementary DNA ends (RLM-RACE). Instead of 5′ exons from *TMPRSS2*, the other four samples contained unique 5′ sequences (Fig. 1a). In PCa_ETV1_1, exons 1–4 of *ETV1* were replaced with two exons from 22q11.23 that had homology to human endogenous retrovirus family K (referred to as *HERV-K_22q11.23*). In PCa_ETV1_2, exon 1 of *ETV1* was replaced with exon 1 of *HNRPA2B1* (7p15), whereas in PCa_ETV1_3, exons 1–4 of *ETV1* were replaced with a 5′-extended exon 1 of *SLC45A3* (1q32). In MET23, exons 1–5 of *ETV1* were replaced with exons 1–2 from *C15orf21* (15q21) (Fig. 1a). We confirmed these fusion transcripts by Q-PCR and genomic fusions by fluorescent *in situ* hybridization (FISH; Supplementary Figs 2 and 3). Additional information about the 5′ partners and the FISH results is described in Supplementary Discussion.

HERV-K_22q11.23–ETV1, *SLC45A3–ETV1* and *C15orf21–ETV1* fusions contain no predicted translated sequences from the 5′ partner, and *HNRPA2B1* would only contribute two residues to a *HNRPA2B1–ETV1* fusion protein. Because their regulatory elements probably drive aberrant *ETV1* expression, we characterized the tissue specificity and androgen regulation of these 5′ partners by microarray or massively parallel signature sequencing (MPSS, as described

[1]Michigan Center for Translational Pathology, Department of Pathology, [2]Department of Urology, [3]Department of Internal Medicine, and [4]Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. [5]Department of Pathology, Brigham and Women's Hospital, [6]Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA.
*These authors contributed equally to this work.

in the Methods). Similar to *TMPRSS2*, *SLC45A3* showed marked overexpression in prostate cancer (median is 2.45 standard deviations above the median value per array) compared to other tumour types (median = 0.33, $P = 2.4 \times 10^{-7}$) in a large DNA microarray study. *C15orf21* showed similar overexpression in prostate cancer ($P = 3.4 \times 10^{-6}$). In contrast, *HNRPA2B1* showed high expression in prostate and other tumour types (median = 2.36 versus 2.41, $P > 0.05$) (Fig. 1b). By MPSS, *HERV-K_22q11.23* was shown to be highly expressed in normal prostate (94 transcripts per million) compared to the 31 other normal tissues (median = 9 transcripts per million; Fig. 1b). By Q-PCR, endogenous expression of *SLC45A3* (21.6-fold, $P = 6.5 \times 10^{-4}$) and *HERV-K_22q11.23* (7.8-fold, $P = 2.4 \times 10^{-4}$) in the LNCaP prostate cancer cell line was shown to be greatly increased by the synthetic androgen R1881, similar to *TMPRSS2* (14.8-fold, $P = 9.95 \times 10^{-7}$). Conversely, the expression of *C15orf21* was significantly decreased (1.9-fold, $P = 0.0012$) and the expression of *HNRPA2B1* was not affected by R1881 stimulation (1.17-fold, $P = 0.29$) (Fig. 1c).

We next sought to identify cell-line models of *ETV1* outlier expression. Previously, we reported that the LNCaP cell line markedly overexpressed *ETV1*, however RLM-RACE revealed expression of only the wild-type transcript[1]. We proposed that LNCaP may harbour a previously unknown rearrangement affecting the expression of *ETV1*, and used a split-probe FISH strategy to look for gross rearrangements (Fig. 2). On LNCaP metaphase spreads, this assay revealed two pairs of co-localizing signals at the *ETV1* locus (7p) and two split signals in which the 5′ signals remained on 7p whereas the 3′ probes (overlapping the *ETV1* locus) were inserted into another chromosome (Fig. 2b). As described in the Supplementary Discussion and Supplementary Figs 4–8, we identified this rearrangement as a cryptic insertion of a minimal region around *ETV1* into an intronic sequence from the *MIPOL1* locus at 14q13.3–14q21.1 in the LNCaP cell line (Fig. 2d).

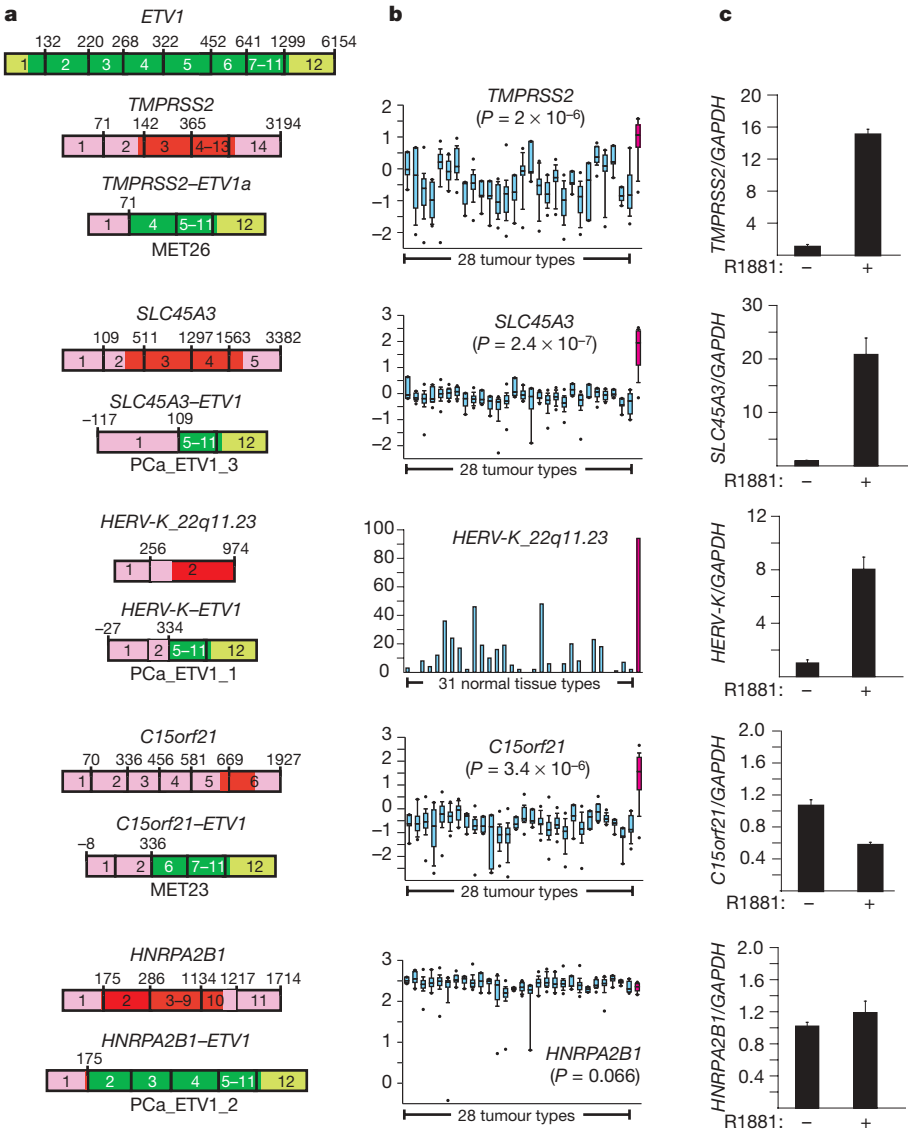

Figure 1 | Identification of prostate-specific or ubiquitously active regulatory elements fused to *ETV1*. **a**, Structure of new 5′ partners fused to *ETV1* in outlier cases. Structures of *ETV1* and all 5′ partners are based on sequences listed in Supplementary Table 2. The numbers in the boxes represent exons. The numbers above the boxes indicate the last base of each exon. Untranslated regions are in lighter shades (pink and light green). **b**, Tissue specificity of 5′ fusion partners was determined in normal tissues or cancers (blue) and in normal prostate or prostate cancer (magenta) (see Methods). See Supplementary Table 4 for tissue and tumour classes. The significance of prostate cancer versus all other tumours is indicated. Box and whisker plots show median +/− 90th%/10th%. **c**, Assessment of androgen regulation of 5′ fusion partners. Endogenous expression of 5′ fusion partners (normalized to the housekeeping gene *GAPDH*) was assessed by Q-PCR in LNCaP cells with (+) or without (–) stimulation by the synthetic androgen R1881 (mean ($n = 4$) plus s.e.).

By screening additional prostate cancer cell lines for *ETV1* overexpression, we identified *ETV1* outlier expression in MDA-PCa 2B (Supplementary Fig. 9). A previous analysis of MDA-PCa 2B demonstrated the presence of a balanced t(7;14)(p21;q21)[20] translocation corresponding to the locations of the *ETV1* and *MIPOL1* loci. We demonstrate that MDA-PCa 2B also harbour a rearrangement involving *ETV1*, because the *ETV1* locus translocates to the d14 (Fig. 2c). The 1.5-Mb 14q13.3–14q21.1 region is the partner of this balanced translocation, because the telomeric 14q13.3–14q21.1 probe localizes to the d7 (Supplementary Fig. 7).

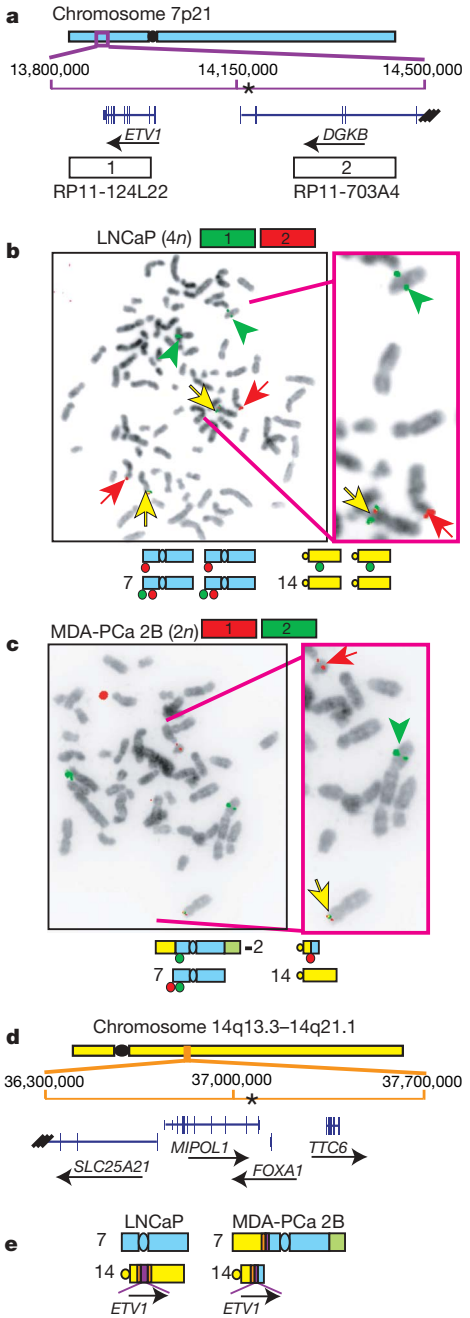

Figure 2 | *ETV1* is rearranged to 14q13.3–14q21.1 in LNCaP and MDA-PCa 2B. a, Schematic showing the *ETV1* locus (purple) on chromosome 7 (blue) and the BACs (rectangles) used for FISH (adapted from the UCSC Genome Browser). Genes are shown and the direction of transcription is indicated by arrows. The LNCaP breakpoint (see Supplementary Figs 4–8) is indicated by an asterisk. **b**, **c**, FISH using the indicated BACs on LNCaP (**b**, tetraploid) and MDA-PCa 2B (**c**, diploid) metaphase spreads. Co-localized signals are indicated by yellow arrows, and separate red and green signals (indicating rearrangements) are indicated by red arrows and green arrowheads, respectively. Schematics of probe localization and chromosome structures as determined by spectral karyotyping[20,28] are indicated. **d**, Schematic of 14q13.3–14q21.1 (orange) on chromosome 14 (yellow). **e**, Structure of *ETV1* and 14q13.3–14q21.1 in LNCaP and MDA-PCa 2B.

The existence of mechanistically distinct rearrangements resulting in the localization of *ETV1* to 14q13.3–14q21.1 (Fig. 2e) in prostate cancer cell lines with *ETV1* outlier expression suggests that elements in this region mediate the aberrant expression of *ETV1*. By characterizing the tissue specificity and androgen regulation of the four contiguous genes at the 14q13.3–14q21.1 breakpoint (Fig. 2d; *SLC25A21*, *MIPOL1*, *FOXA1* and *TTC6*) as well as that of *ETV1* in LNCaP and its androgen-insensitive derivative C4-2B, we demonstrate that this region is both prostate-specific and coordinately regulated by androgen (Supplementary Figs 10–12 and Supplementary Discussion).

The 5′ partners do not contribute a coding sequence to the *ETV1* transcript, therefore the common result of the different *ETV1* rearrangements in clinical samples and prostate cancer cell lines is aberrant overexpression of truncated *ETV1*. We recapitulated this event *in vitro* and *in vivo* to determine the role of aberrant ETS family member expression in prostate cancer. We designed adenoviral and lentiviral constructs to overexpress *ETV1*, as expressed in our index *TMPRSS2–ETV1* fusion-positive case MET26 (Supplementary Fig. 13a). In RWPE and PrEC cells, *ETV1* overexpression had no detectable effect on proliferation (Supplementary Fig. 13b, c). *ETV1* overexpression had no effect on the percentage of RWPE cells in the S phase of the cell cycle (Supplementary Fig. 13d) and was not sufficient for transformation (Supplementary Fig. 13e). However, *ETV1* overexpression markedly increased invasion in a modified basement membrane invasion assay in RWPE (3.4-fold, $P = 0.0005$) (Fig. 3a) and PrEC (6.3-fold, $P = 0.0006$) (Supplementary Fig. 14a). Additionally, *ETV1* knockdown in LNCaP using either short interfering (si)RNA or short hairpin (sh)RNA inhibited invasion (Fig. 3b and Supplementary Fig. 14b–d), consistent with previous work[21]. To investigate the transcriptional programme regulated by *ETV1*, we profiled stable RWPE-*ETV1* cells and analysed the expression signatures using the Oncomine Concepts Map (OCM, http://www.oncomine.org). The OCM is a resource to look for associations between more than 20,000 biologically related gene sets by disproportionate overlap[10,22]. As shown in Fig. 3c, OCM analysis identified a network of molecular concepts related to cell invasion that were enriched in our *ETV1*-overexpressed signature, consistent with the phenotypic effects described above. Specific examples are described in Supplementary Discussion.

We next determined the effects of *ETV1* overexpression *in vivo*. We generated transgenic mice expressing a Flag-tagged, truncated version of *ETV1* under the control of the modified probasin promoter (ARR2Pb-*ETV1*) (Supplementary Fig. 13a), which drives strong transgene expression exclusively in the prostate under androgen regulation[23]. This transgene is functionally analogous to the androgen-induced gene fusions of *ETV1* we identified in human prostate cancer. By 12–14 weeks of age, 6 of 8 (75%) ARR2Pb-*ETV1* mice developed mouse prostatic intraepithelial neoplasia (mPIN) (Fig. 4 and Supplementary Table 5). Consistent with the definition of mPIN[24], we observed focal proliferative lesions contained within normal glands in the prostates of ARR2Pb-*ETV1* mice (Fig. 4a–d); these lesions exhibited nuclear atypia, including stratification, hyperchromasia and macronucleoli. mPIN was observed in all three prostatic lobes (anterior, ventral and dorsolateral) of ARR2Pb-*ETV1* mice, most commonly in the ventral lobe (7 of 11, 63.6%) (Supplementary Table 5). By immunohistochemistry in ARR2Pb-*ETV1* mice, we observed strong ETV1–Flag expression exclusively in mPIN foci and not in benign glands (data not shown). Although we have not observed progression to invasive carcinoma in ARR2Pb-*ETV1* mice, we have only characterized 4 mice older than 19 weeks of age, 3 of which (75%) also had mPIN (Supplementary Table 5); this suggests that additional genetic lesions are required for the development of carcinoma. Combined with our *in vitro* observations, these

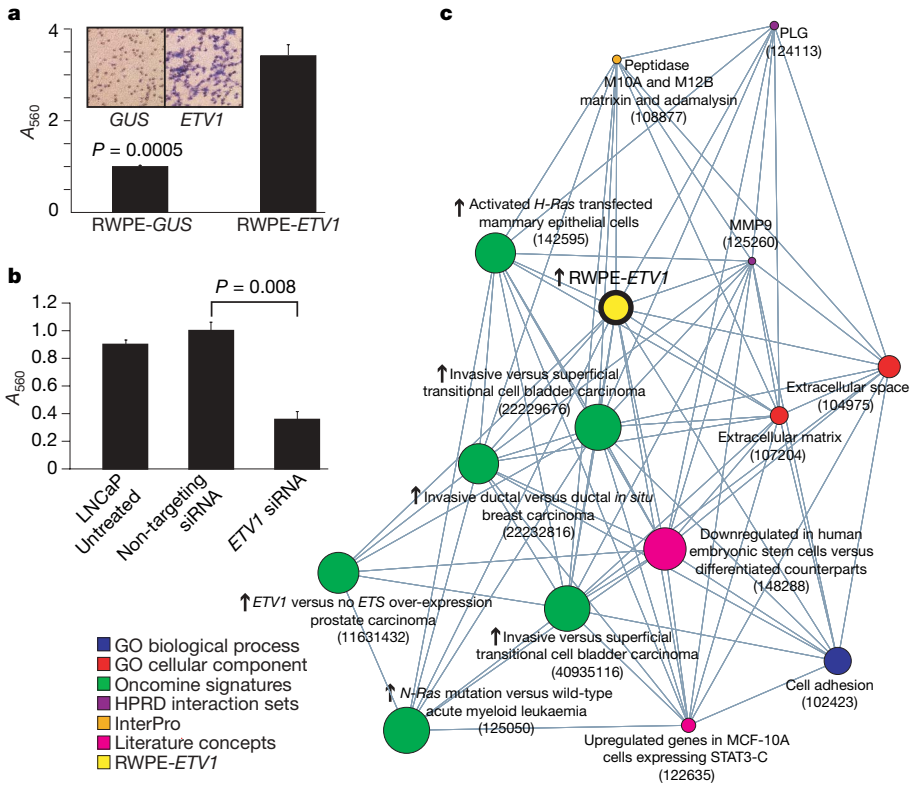

Figure 3 | *ETV1* overexpression in prostate cells confers invasiveness. **a**, We infected the benign prostate cell line RWPE with lentiviruses expressing *ETV1* (exon 4 to the reported stop codon) or control (*GUS*). Stable clones were assayed for invasion through a modified basement membrane (mean ($n = 3$) plus s.e.). Photomicrographs of invaded cells are shown. **b**, LNCaP cells were treated with transfection reagent alone (untreated) or were transfected with non-targeting or *ETV1* siRNA, and assessed for invasion as in **a** (mean ($n = 3$) plus s.e.). **c**, Oncomine concepts map of genes overexpressed in RWPE-*ETV1* compared to RWPE-*GUS* cells (yellow node). Nodes represent molecular concepts (biologically related gene sets). Node size is proportional to the number of genes in the concept, and the colour indicates the concept type. Each edge represents a significant enrichment ($P < 0.005$). HPRD interaction sets for MMP9 and PLG represent known interactors for each gene product. Oncomine concept map IDs are provided for all concepts.

results demonstrate that *ETV1* induces a neoplastic phenotype in the mouse prostate and supports an oncogenic role for ETS gene fusions in human prostate cancer.

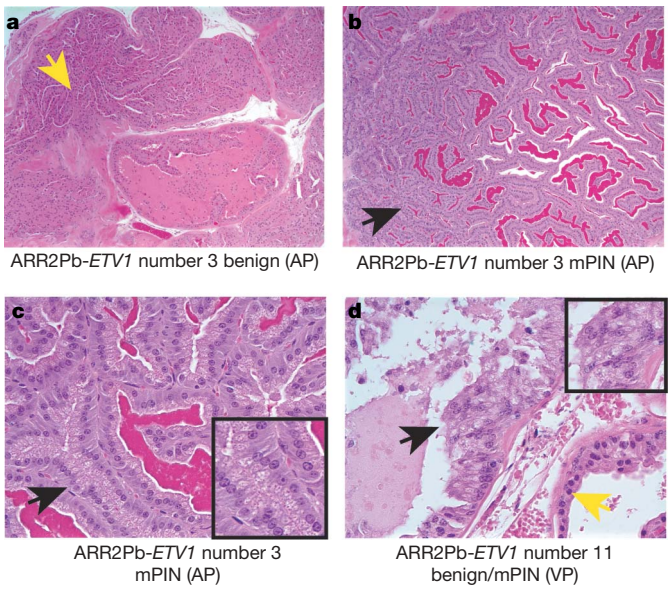

Figure 4 | Transgenic mice expressing *ETV1* develop mPIN. We generated transgenic mice expressing *ETV1* under the control of the modified probasin promoter (ARR2Pb-*ETV1*) (Supplementary Fig. 13a). Mice were killed at 12–43 weeks and mPIN was observed in 75% of ARR2Pb-*ETV1* mice (Supplementary Table 5). **a–d**, Haematoxylin-and eosin-stained ARR2Pb-*ETV1* prostates. Benign epithelia and areas of mPIN are indicated by yellow and black arrows, respectively. Consistent with the definition of mPIN, normal areas (**a**) and mPIN (**b**) were observed in the anterior prostate (AP) of an ARR2Pb-*ETV1* mouse. **c**, High-power view of **b**. **d**, Normal glands and mPIN foci in the ventral prostate (VP) of an ARR2Pb-*ETV1* mouse. Original magnification for **a** and **b** is ×100, and for **c** and **d** is ×400. Insets in **c** and **d** show prominent macronucleoli.

Including fusions between *TMPRSS2* and ETS family members, we have now identified five classes of ETS rearrangements in prostate cancer (Supplementary Fig. 15). Identification of untranslated regions from the prostate-specific androgen-induced gene *TMPRSS2* has provided a mechanism for aberrant ETS family member expression. Thus, fusions between *TMPRSS2* and ETS family members (class I) represent the predominant class of ETS rearrangements in prostate cancer. Rearrangements involving fusions with untranslated regions from other prostate-specific androgen-induced 5′ partner genes (class IIa) or endogenous retroviral elements (class IIb) are probably functionally similar to *TMPRSS2*–ETS rearrangements. Similar to 5′ partners in class I and II ETS rearrangements, *C15orf21* is markedly overexpressed in prostate cancer; however, because *C15orf21* is repressed by androgen, this represents a previously unknown class of rearrangements (class III) involving prostate-specific androgen-repressed 5′ partners.

In contrast, *HNRPA2B1*, which encodes a member of the ubiquitously expressed heteronuclear ribonuclear proteins, did not show prostate-specific expression or androgen responsiveness. Thus, *HNRPA2B1–ETV1* represents a previously unknown class of ETS rearrangements (class IV), in which non-tissue-specific promoter elements drive ETS expression. Whereas class I–III ETS rearrangements are functionally analogous to *IGH–MYC* rearrangements in B-cell malignancies, *HNRPA2B1–ETV1* is more analogous to inv(3)(q21q26) and t(3;3)(q21;q26) in acute myeloid leukaemia, which are thought to place *EVI1* (ecotropic viral integration site 1) under the control of enhancer elements of the constitutively expressed *RPN1* gene[25,26].

By screening prostate cancer cell lines that have *ETV1* outlier expression, we identified rearrangements in LNCaP and MDA-PCa 2B that result in the localization of *ETV1* to 14q13.3–14q21.1. Because this aberration is recurrent in prostate cancer cell lines, we propose that characterizing additional prostate cancer cases with *ETV1* outlier expression will identify clinical specimens with similar rearrangements (class V), in which the entire ETS family gene is rearranged to prostate-specific regions. The identification of distinct

classes of 5′ fusion partners has implications for the detection of gene fusions in prostate cancer and may be important for management, particularly with regard to the effects of androgen ablation on the expression of the different ETS rearrangement classes (Supplementary Fig. 16), as described in Supplementary Discussion.

Multiple classes of gene rearrangements in prostate cancer suggest a generalized role for chromosomal rearrangements in common epithelial cancers. For example, tissue-specific promoter elements may be fused to oncogenes in other hormone-driven cancers, such as breast cancer. Additionally, whereas prostate-specific fusions would not provide a growth advantage and be selected for in other epithelial cancers, fusions involving strong promoters of ubiquitously expressed genes, such as *HNRPA2B1*, could result in the aberrant expression of oncogenes across tumour types. This study supports a role for chromosomal rearrangements in common epithelial tumour development through a variety of mechanisms, similar to haematological malignancies.

METHODS SUMMARY

Q-PCR, RLM-RACE for *ETV1* fusions, androgen stimulation of LNCaP cells and interphase FISH were performed essentially as described[1,2] using indicated oligonucleotide primers (Supplementary Table 2) and bacterial artificial chromosome (BAC) probes (Supplementary Table 3). Tissue-specific expression of 5′ fusion partners was determined using the International Genomics Consortium's expO data set accessed in the Oncomine database[27] and the Lynx Therapeutics MPSS data set (GSE1747). Expression profiling was performed using Agilent Whole Human Genome Oligo Microarrays. Adenoviruses and lentiviruses expressing *ETV1* were generated by the University of Michigan Vector Core. Transgenic ARR2Pb-*ETV1* mice were generated by the University of Michigan Transgenic Animal Model Core.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 April; accepted 16 June 2007.

1. Tomlins, S. A. *et al.* Recurrent fusion of *TMPRSS2* and ETS transcription factor genes in prostate cancer. *Science* **310,** 644–648 (2005).
2. Tomlins, S. A. *et al. TMPRSS2:ETV4* gene fusions define a third molecular subtype of prostate cancer. *Cancer Res.* **66,** 3396–3400 (2006).
3. Cerveira, N. *et al. TMPRSS2–ERG* gene fusion causing ERG overexpression precedes chromosome copy number changes in prostate carcinomas and paired HGPIN lesions. *Neoplasia* **8,** 826–832 (2006).
4. Glinsky, G. V., Glinskii, A. B., Stephenson, A. J., Hoffman, R. M. & Gerald, W. L. Gene expression profiling predicts clinical outcome of prostate cancer. *J. Clin. Invest.* **113,** 913–923 (2004).
5. Hermans, K. G. *et al. TMPRSS2:ERG* fusion by translocation or interstitial deletion is highly relevant in androgen-dependent prostate cancer, but is bypassed in late-stage androgen receptor-negative prostate cancer. *Cancer Res.* **66,** 10658–10663 (2006).
6. Lapointe, J. *et al.* Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc. Natl Acad. Sci. USA* **101,** 811–816 (2004).
7. Mehra, R. *et al.* Comprehensive assessment of *TMPRSS2* and *ETS* family gene aberrations in clinically localized prostate cancer. *Mod. Pathol.* **6,** 1177–1187 (2007).
8. Perner, S. *et al. TMPRSS2:ERG* fusion-associated deletions provide insight into the heterogeneity of prostate cancer. *Cancer Res.* **66,** 8337–8341 (2006).
9. Soller, M. J. *et al.* Confirmation of the high frequency of the *TMPRSS2/ERG* fusion gene in prostate cancer. *Genes Chromosom. Cancer* **45,** 717–719 (2006).
10. Tomlins, S. A. *et al.* Integrative molecular concept modeling of prostate cancer progression. *Nature Genet.* **39,** 41–51 (2007).
11. Winnes, M., Lissbrant, E., Damber, J. E. & Stenman, G. Molecular genetic analyses of the *TMPRSS2–ERG* and *TMPRSS2–ETV1* gene fusions in 50 cases of prostate cancer. *Oncol. Rep.* **17,** 1033–1036 (2007).
12. Yoshimoto, M. *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* **8,** 465–469 (2006).
13. Yu, Y. P. *et al.* Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. *J. Clin. Oncol.* **22,** 2790–2799 (2004).
14. Mitelman, F., Johansson, B. & Mertens, F. Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. *Nature Genet.* **36,** 331–334 (2004).
15. Rowley, J. D. Chromosome translocations: dangerous liaisons revisited. *Nature Rev. Cancer* **1,** 245–250 (2001).
16. Lin, B. *et al.* Prostate-localized and androgen-regulated expression of the membrane-bound serine protease *TMPRSS2*. *Cancer Res.* **59,** 4180–4184 (1999).
17. Mertz, K. D. *et al.* Molecular characterization of the *TMPRSS2–ERG* gene fusion in the NCI-H660 prostate cancer cell line—a new perspective for an old model. *Neoplasia* **9,** 200–206 (2007).
18. Rabbitts, T. H. Chromosomal translocations in human cancer. *Nature* **372,** 143–149 (1994).
19. Demichelis, F. *et al. TMPRSS2:ERG* gene fusion associated with lethal prostate cancer in a watchful waiting cohort. *Oncogene* advance online publication, doi: 10.1038/sj.onc.1210237 (22 January 2007).
20. van Bokhoven, A. *et al.* Spectral karyotype (SKY) analysis of human prostate carcinoma cell lines. *Prostate* **57,** 226–244 (2003).
21. Cai, C. *et al. ETV1* is a novel androgen receptor-regulated gene that mediates prostate cancer cell invasion. *Mol. Endocrinol.* advance online publication, doi:10.1210/me.2006-0480 (15 May 2007).
22. Rhodes, D. R. *et al.* Molecular concepts analysis links tumors, pathways, mechanisms, and drugs. *Neoplasia* **9,** 443–454 (2007).
23. Ellwood-Yen, K. *et al.* Myc-driven murine prostate cancer shares molecular features with human prostate tumors. *Cancer Cell* **4,** 223–238 (2003).
24. Shappell, S. B. *et al.* Prostate pathology of genetically engineered mice: definitions and classification. The consensus report from the Bar Harbor meeting of the Mouse Models of Human Cancer Consortium Prostate Pathology Committee. *Cancer Res.* **64,** 2270–2305 (2004).
25. Suzukawa, K. *et al.* Identification of a breakpoint cluster region 3′ of the ribophorin I gene at 3q21 associated with the transcriptional activation of the EVI1 gene in acute myelogenous leukemias with inv(3)(q21q26). *Blood* **84,** 2681–2688 (1994).
26. Wieser, R. Rearrangements of chromosome band 3q21 in myeloid leukemia. *Leuk. Lymphoma* **43,** 59–65 (2002).
27. Rhodes, D. R. *et al.* Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9,** 166–180 (2007).
28. Beheshti, B., Karaskova, J., Park, P. C., Squire, J. A. & Beatty, B. G. Identification of a high frequency of chromosomal rearrangements in the centromeric regions of prostate cancer cell lines by sequential giemsa banding and spectral karyotyping. *Mol. Diagn.* **5,** 23–32 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Rhodes, S. Kalyana-Sundaram and T. Barrette for support of the OCM, L. Smith for cytogenetics assistance, E. Keller and J. Macoska for prostate cancer cell lines, J. Moran for discussions regarding endogenous retroviral elements, J. Presner for technical assistance, and R. Craig and L. Stoolman for the FACS analysis. We thank the UM Transgenic Animal Model Core for generating transgenic mice and the UM Vector core for virus generation. This work was supported in part by Department of Defense (to R.M., A.M.C. and S.V.), the National Institutes of Health (to K.J.P., A.M.C., R.B.S., K.J.P. and A.M.C.), the Early Detection Research Network (to A.M.C.), the Prostate Cancer Foundation (to A.M.C.), and Gen-Probe Incorporated (to A.M.C.). A.M.C. is supported by a Clinical Translational Research Award from the Burroughs Wellcome Foundation. S.A.T. is supported by a Rackham Predoctoral Fellowship. K.J.P. is supported as an American Cancer Society Clinical Research Professor. S.A.T. is a Fellow of the Medical Scientist Training Program.

Author Information The primary microarray data have been deposited in NCBI's Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) under the GEO series accession numbers GSE7701 and GSE7702. Sequences of the *ETV1* fusion transcript junctions identified by RACE have been deposited in GenBank under accession numbers EF632109–EF632112. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.M.C. (arul@umich.edu).

METHODS

Samples and cell lines. Prostate tissues were from the radical prostatectomy series at the University of Michigan and from the Rapid Autopsy Program[29], both of which are part of University of Michigan Prostate Cancer Specialized Program of Research Excellence Tissue Core. All samples were collected with informed consent of the patients and previous institutional review board approval.

The benign immortalized prostate cell line RWPE and the prostate cancer cell lines LNCaP, Du145 NCI-H660 and PC3 were obtained from the American Type Culture Collection. Primary benign prostatic epithelial cells (PrEC) were obtained from Cambrex Bio Science. The prostate cancer cell lines C4-2B, LAPC4 and MDA-PCa 2B were provided by E. Keller. The prostate cancer cell line 22-RV1 was provided by J. Macoska. VCaP was derived from a vertebral metastasis from a patient with hormone-refractory metastatic prostate cancer[30].

For androgen stimulation experiments, LNCaP cells were grown in charcoal-stripped serum containing media for 24 h, before treatment for 24 h with 1% ethanol or 1 nM of methyltrienolone (R1881, NEN Life Science Products) dissolved in ethanol. For all samples, total RNA was isolated with Trizol (Invitrogen) according to the manufacturer's instructions.

Quantitative PCR. Q-PCR was performed using Power SYBR Green Mastermix (Applied Biosystems) on an Applied Biosystems 7300 Real Time PCR system as described[1,2]. All oligonucleotide primers were synthesized by Integrated DNA Technologies and are listed in Supplementary Table 2. HMBS and GAPDH[31], and PSA[32] primer sequences were as described. Androgen stimulation reactions were performed in quadruplicate, siRNA knockdown reactions were performed in triplicate and all other reactions were performed in duplicate.

RNA-ligase-mediated rapid amplification of cDNA ends. RLM-RACE was performed using the GeneRacer RLM-RACE kit (Invitrogen), according to the manufacturer's instructions as described[1,2]. To obtain the 5′ end of *ETV1*, first-strand cDNA was amplified with Platinum Taq High Fidelity (Invitrogen) using the GeneRacer 5′ primer and ETV1_exon4-5-r. For amplification from MET23, ETV1_exon7-r was used with the GeneRacer 5′ primer. Products were cloned and sequenced bidirectionally as described[1,2]. RLM-RACE cDNA was not used for other assays.

Fluorescence *in situ* hybridization. Interphase FISH on formalin-fixed paraffin-embedded tissue sections was performed as described[2]. A minimum of 50 nuclei per assay were evaluated. For metaphase FISH, spreads of LNCaP and MDA-PCa 2B were prepared using standard cytogenetic techniques. Slides were pre-treated in 2 × SSC for 2 min, 70% ethanol for 2 min and 100% ethanol for 2 min, and then air-dried. Slides and probes were co-denatured at 75 °C for 2 min, and hybridized overnight at 37 °C. Post-hybridization was in 0.5 × SSC at 42 °C for 5 min, followed by 3 washes in 1 × phosphate buffered saline with 0.1% Tween-20 (PBST). Fluorescent detection was performed using anti-digoxigenin conjugated to fluorescein (Roche Applied Science) and streptavidin conjugated to Alexa Fluor 594 (Invitrogen). Slides were counterstained and mounted in ProLong Gold Antifade Reagent with DAPI (Invitrogen). Slides were examined using a Zeiss Axio Imager Z1 fluorescence microscope (Zeiss) and imaged with a CCD camera using ISIS software (Metasystems). At least five metaphases were assessed, and reported aberrations were observed in all interpretable spreads. BACs (listed in Supplementary Table 3) were obtained from the BACPAC Resource Center, and probes were prepared as described[2]. Pre-labelled chromosome 7 centromere and 7p telomeric probes were obtained from Vysis. The integrity and correct localization of all probes were verified by hybridization to metaphase spreads of normal peripheral lymphocytes.

Tissue-specific expression. To determine the tissue-specific expression of 5′ fusion partners and genes at 14q13–q21, we interrogated the International Genomics Consortium's expO data set (http://expo.intgen.org/expo/public/downloaddata.jsp), consisting of expression profiles from 630 tumours of 29 distinct types, using the Oncomine database (http://www.oncomine.org)[27]. To interrogate the expression of *HERV-K_22q11.23*, which is not monitored by commercial array platforms, we queried the Lynx Therapeutics normal tissue MPSS data set (GSE1747) with the MPSS tag 'GATCTTTGTGACCTACT', which unambiguously identifies *HERV-K_22q11.23*, as described[33]. Descriptions of tumour types from the expO data set and the normal tissue types from the MPSS data set are provided in Supplementary Table 4.

Expression profiling. Expression profiling of LNCaP, C4-2B, RWPE-*ETV1* and RWPE-*GUS* cells was performed using the Agilent Whole Human Genome Oligo Microarray. Total RNA isolated using Trizol was purified using the Qiagen RNAeasy Micro kit. One microgram of total RNA was converted to cRNA and labelled according to the manufacturer's protocol (Agilent). Hybridizations were performed for 16 h at 65 °C, and arrays were scanned using an Agilent DNA microarray scanner. Images were analysed and data were extracted using Agilent Feature Extraction Software 9.1.3.1, with linear and lowess normalization performed for each array. For the LNCaP and C4-2B hybridizations, the reference for each cell line was pooled benign prostate total RNA (Clontech). A dye flip for each cell line was also performed. Features were ranked by average expression (log ratio) in the two LNCaP arrays divided by the average expression in the two C4-2B arrays after correction for the dye flip. For RWPE cells, four hybridizations were performed (duplicate RWPE-*ETV1*/RWPE-*GUS* and RWPE-*GUS*/RWPE-*ETV1* hybridizations). Over- and under-expressed signatures were generated by filtering to include only features with significant differential expression (P-value logratio < 0.01) in all four hybridizations and twofold average over- or under-expression (log ratio) after correction for the dye flip. Over- and under-expressed RWPE-*ETV1*/RWPE-*GUS* signatures were loaded into the Molecular Concepts Map[22], resulting in concepts containing 527 and 558 unique genes, respectively. Each signature was tested against all contained concepts in the Molecular Concepts Map for association using Fisher's exact test as described[10,22].

Southern hybridization. Genomic DNA (10 μg) from LNCaP, VCaP, pooled normal human male DNA (Promega) and normal placental DNA (Promega) was digested with *Eco*RI or *Pst*I (New England Biologicals) overnight. Fragments were resolved on a 0.8% agarose gel at 40 V overnight, transferred to Hybond NX nylon membrane, pre-hybridized, hybridized with probe and washed according to standard protocols. A series of 22 probes spanning the region of chromosome 7 implicated by FISH (between RP11-313C20 and RP11-703A4) were generated by PCR amplification with Platinum Taq High Fidelity on pooled normal human male genomic DNA (Supplementary Table 2 and Supplementary Fig. 5). Twenty-five nanograms of each probe was labelled with dCTP-^{32}P and used for hybridization.

Inverse PCR. To identify the *ETV1* breakpoint in LNCaP cells, we used an inverse PCR strategy based on the rearrangement identified by Southern blotting (probe A, Supplementary Table 2) as described previously[34] and shown in Supplementary Fig. 6. Primers A1, A2 and A3, which are reverse complemented from the wild-type sequence and are divergent to primers B1, B2 and B3, were used for inverse PCR on *Pst*I-digested and religated (to promote intramolecular ligation) LNCaP genomic DNA template. Nested PCRs were performed in the following order of primer combinations: A1–B1, A2–B2 and A3–B3. The Expand 20 kbplus PCR system (Roche) was used for amplifying the fusion product according to the manufacturer's suggestions. The enriched 3-kb band observed in nested PCRs was cloned into pCR8/GW/TOPO (Invitrogen), miniprep DNA was screened for inserts, and positive clones were sequenced (University of Michigan DNA Sequencing Core). The *ETV1* insertion was confirmed by PCR with Platinum Taq High Fidelity using primers from chromosomes 7 and 14 (Supplementary Table 2).

***In vitro* overexpression of *ETV1*.** cDNA of *ETV1*, as present in the *TMPRSS2*–*ETV1* fusion to the reported stop codon of *ETV1* (269–1521, NM_004956.3), was amplified by RT–PCR from MET26 (ref. 1) and TOPO cloned into the Gateway entry vector pCR8/GW/TOPO (Invitrogen), yielding pCR8-*ETV1*. To generate adenoviral and lentiviral constructs, pCR8-*ETV1* and a control entry clone (pENTR-*GUS*) were recombined with pAD/CMV/V5 (Invitrogen) and pLenti6/CMV/V5 (Invitrogen), respectively, using LR Clonase II (Invitrogen). Control pAD/CMV/*LACZ* clones were obtained from Invitrogen. Adenoviruses and lentiviruses were generated by the University of Michigan Vector Core. The benign immortalized prostate cell line RWPE was infected with lentiviruses expressing *ETV1* or *GUS*, and stable clones were generated by selection with blasticidin (Inivtrogen). Benign PrEC were infected with adenoviruses expressing *ETV1* or *LACZ*, because stable lines could not be generated in primary PrEC cells. Cell counts were estimated by treating cells with trypsin, and analysis was performed by a Coulter counter at the indicated time points in triplicate. For invasion assays, PrEC-*ETV1* and PrEC-*LACZ* (48 h after infection) or stable RWPE-*ETV1* and RWPE-*GUS* cells were used. Representative results from three separate experiments are shown in Fig. 3a and Supplementary Fig. 14.

***ETV1* knockdown.** For siRNA knockdown of *ETV1* in LNCaP cells, the individual siRNAs composing the Dharmacon SMARTpool against *ETV1* (MU-003801-01) were tested for *ETV1* knockdown by Q-PCR, and the most effective single siRNA (D-003801-05) was used for further experiments. siCONTROL non-targeting siRNA number 1 (D-001210-01) or siRNA against ETV1 was transfected into LNCaP cells using Oligofectamine (Invitrogen). After 24 h we performed a second identical transfection, and cells were harvested 24 h later for RNA isolation and invasion assays as described below. For shRNA knockdown of *ETV1* in LNCaP cells, the microRNA-adapted shRNA construct against *ETV1* from the pMS2 retroviral vector (V2HS_61929, Open Biosystems) was cloned into an empty pGIPZ lentiviral vector (RHS4349, Open Biosystems) according to the manufacturer's protocol. pGIPZ lentiviruses with microRNA-adapted shRNAs against *ETV1* or a non-silencing control (RHS4346) were generated by the University of Michigan Vector Core. LNCaP cells were infected with lentiviruses, and 48 h later cells were used for invasion assays as described below. Representative results from six independent experiments are reported.

Invasion assays. Equal numbers of the indicated cells were seeded onto the basement membrane matrix (EC matrix, Chemicon) present in the insert of a 24-well culture plate, with fetal bovine serum added to the lower chamber as a chemoattractant. After 48 h, non-invading cells and the EC matrix were removed by a cotton swab. Invaded cells were stained with crystal violet and photographed. The inserts were treated with 10% acetic acid and absorbance was measured at 560 nm.

FACS cell cycle analysis. RWPE-*ETV1* and RWPE-*GUS* cells were assessed by FACS for cell-cycle characterization. Cells were washed with $2\times$ PBS, and approximately 2×10^6 cells were resuspended in PBS before fixation in 70% ethanol. Pelleted cells were washed and treated with RNase (100 μg ml^{-1} final concentration) and propidium iodide (10 μg ml^{-1} final concentration) at 37 °C for 30 min. Stained cells were analysed on a LSR II flow cytometer (BD Biosciences) running FACSDivia, and cell-cycle phases were calculated using ModFit LT (Verity Software House).

Soft agar assay. A 0.6% (w/v) bottom layer of low-melting-point agarose in normal medium was prepared in six-well culture plates. On top, a layer of 0.3% agarose containing 1×10^4 RWPE-*GUS*, RWPE-*ETV1* or *DU145* (positive control) cells was placed. After 12 d, foci were stained with crystal violet and were counted.

Transgenic *ETV1* mice. For *in vivo* overexpression of *ETV1*, a carboxy-terminal 3×Flag-epitope-tagged construct was generated by PCR using pCR8-*ETV1* as the template, with the reverse primer encoding a triple Flag tag before the stop codon. The product was TOPO cloned into pCR8. To generate a prostate-specific *ETV1* transgenic construct, 3×Flag-*ETV1* was inserted into pBSII (Stratagene) downstream of a modified small composite probasin promoter (ARR2Pb) and upstream of a bovine growth hormone polyA site (PA-BGH). The ARR2Pb sequence contains the original probasin sequence Pb (426/+28) plus two additional androgen response elements[23]. The construct was sequenced and androgen was used to test for promoter inducibility in LNCaP cells on transient transfection before microinjection into FVB mouse eggs. The ARR2Pb-*ETV1* plasmid was linearized with *Pvu*I/*Kpn*I//*Sac*II and microinjected into fertilized FVB mouse eggs, and was then surgically transplanted into a pseudo-pregnant female by the University of Michigan Transgenic Animal Model Core. Transgenic founders were screened by PCR using genomic DNA isolated from tail snips. Multiple ARR2Pb-*ETV1* transgenic founders were obtained and crossed with FVB mice, and transgene-positive male mice offspring were killed at various time points.

Prostates from transgenic mice were dissected using a Nikon dissection scope, fixed in 10% buffered formalin and embedded in paraffin. Five-micrometre sections were stained with haematoxylin and eosin, and evaluated by three pathologists (R.M., M.A.R. and R.B.S.) according to the criteria provided in the Consensus Report from the Bar Harbor Meeting of the Mouse Models of Human Cancer Consortium Prostate Pathology Committee[24].

29. Rubin, M. A. *et al.* Rapid (''warm'') autopsy study for procurement of metastatic prostate cancer. *Clin. Cancer Res.* **6**, 1038–1045 (2000).
30. Korenchuk, S. *et al.* VCaP, a cell-based model system of human prostate cancer. *In Vivo* **15**, 163–168 (2001).
31. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT–PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, RESEARCH0034 (2002).
32. Specht, K. *et al.* Quantitative gene expression analysis in microdissected archival formalin-fixed and paraffin-embedded tumor tissue. *Am. J. Pathol.* **158**, 419–429 (2001).
33. Stauffer, Y., Theiler, G., Sperisen, P., Lebedev, Y. & Jongeneel, C. V. Digital expression profiles of human endogenous retroviral families in normal and cancerous tissues. *Cancer Immun.* **4**, 2 (2004).
34. Wiemels, J. L. & Greaves, M. Structure and possible mechanisms of *TEL–AML1* gene fusions in childhood acute lymphoblastic leukemia. *Cancer Res.* **59**, 4075–4082 (1999).