Discovery of non-ETS gene fusions in human prostate cancer using next-generation RNA sequencing

Dorothee Pflueger,[1,12] Stéphane Terry,[1,12] Andrea Sboner,[2,3,12] Lukas Habegger,[3] Raquel Esgueva,[1] Pei-Chun Lin,[1] Maria A. Svensson,[1] Naoki Kitabayashi,[1] Benjamin J. Moss,[1] Theresa Y. MacDonald,[1] Xuhong Cao,[4] Terrence Barrette,[4,5] Ashutosh K. Tewari,[6,7] Mark S. Chee,[8] Arul M. Chinnaiyan,[4,5,9] David S. Rickman,[1] Francesca Demichelis,[1,10] Mark B. Gerstein,[2,3,11,13] and Mark A. Rubin[1,13,14]

[1]Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, New York 10021, USA; [2]Molecular Biophysics and Biochemistry Department, Yale University, New Haven, Connecticut 06510, USA; [3]Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut 06510, USA; [4]Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, Michigan 48109, USA; [5]Department of Pathology, University of Michigan, Ann Arbor, Michigan 48109, USA; [6]Department of Urology, Weill Cornell Medical College, New York, New York 10021, USA; [7]Institute of Prostate Cancer, Weill Cornell Medical College, New York Presbyterian Hospital, New York, New York 10021, USA; [8]Prognosys Biosciences, Inc., La Jolla, California 92037, USA; [9]Howard Hughes Medical Institute, Department of Urology, and Comprehensive Cancer Center, University of Michigan, Ann Arbor, Michigan 48109, USA; [10]Institute for Computational Biomedicine, Weill Cornell Medical College, New York, New York 10021, USA; [11]Department of Computer Science, Yale University, New Haven, Connecticut 06510, USA

Half of prostate cancers harbor gene fusions between *TMPRSS2* and members of the ETS transcription factor family. To date, little is known about the presence of non-ETS fusion events in prostate cancer. We used next-generation transcriptome sequencing (RNA-seq) in order to explore the whole transcriptome of 25 human prostate cancer samples for the presence of chimeric fusion transcripts. We generated more than 1 billion sequence reads and used a novel computational approach (FusionSeq) in order to identify novel gene fusion candidates with high confidence. In total, we discovered and characterized seven new cancer-specific gene fusions, two involving the ETS genes *ETV1* and *ERG*, and four involving non-ETS genes such as *CDKN1A* (p21), *CD9*, and *IKBKB* (IKK-beta), genes known to exhibit key biological roles in cellular homeostasis or assumed to be critical in tumorigenesis of other tumor entities, as well as the oncogene *PIGU* and the tumor suppressor gene *RSRC2*. The novel gene fusions are found to be of low frequency, but, interestingly, the non-ETS fusions were all present in prostate cancer harboring the *TMPRSS2–ERG* gene fusion. Future work will focus on determining if the ETS rearrangements in prostate cancer are associated or directly predispose to a rearrangement-prone phenotype.

[Supplemental material is available online at http://www.genome.org. The data sets used in this study have been submitted to dbGaP (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap) under accession no. phs000310.v1.p1. Sequences of the novel gene fusion transcripts from this study have been submitted to GenBank (http://www.ncbi.nlm.nih.gov/Genbank/) under accession nos. HM245385–HM245396.]

Systematic review of common recurrent translocations in human epithelial cancers, leukemias, and sarcomas implicates a discrete repertoire of genes and gene families (Mitelman et al. 2007; Prensner and Chinnaiyan 2009). Emerging data confirm that recurrent chromosomal rearrangements may be driven by nuclear transcription factors such as the ligand-bound androgen receptor in prostate cancer (Lin et al. 2009). Indeed, prostate cancer seems to be prone to recurrent gene fusions involving androgen-regulated genes (e.g., *TMPRSS2*, *SLC45A3*, and *NDRG1*) and ETS transcription factors (e.g., *ERG*, *ETV1*) (Tomlins et al. 2005; Kumar-Sinha et al. 2008) and can be detected in ~45% of all prostate specific antigen (PSA)-screened cancers (Mosquera et al. 2009).

The advent of next-generation RNA sequencing (RNA-seq) opens opportunities to interrogate the transcriptome as described by Levin et al. (2009) and Berger et al. (2010). Through either targeted RNA-seq of a preselected set of cancer-related genes in a chronic myeloid leukemia (CLL) cell line (K562) (Levin et al. 2009) or whole transcriptome sequencing of 10 melanoma samples (patient-derived primary cell cultures and cell lines) (Berger et al. 2010), the authors demonstrate the use of transcriptome data to identify chimeric transcripts and alternative splicing of genes and point mutations, and simultaneously infer both overall gene expression levels and allele-specific expression. Berger and colleagues demonstrate rare genomic rearrangements in human melanoma samples. These likely represent bystander genomic alterations in a malignancy widely assumed to be primarily driven by somatic mutations (e.g., *BRAF*). It is expected that many gene fusions resulting from chromosomal rearrangements may represent "passenger" events (Stratton et al. 2009), and therefore functional data will be requisite to help distinguish "driving" from "passenger" events. Mounting data support ETS fusions as a "gatekeeper"

---

lesion in prostate cancer, necessary but not sufficient to the implementation of transforming processes (Kumar-Sinha et al. 2008). In vitro studies have shown that ETS fusions influence the cellular program of invasiveness and initiate malignant transformation that leads to preneoplasic lesions also referred to as prostate intraepithelial neoplasia (PIN) in mouse model systems (Klezovitch et al. 2008; Tomlins et al. 2008). Because of the inability of ETS fusions to form fully developed cancer in vivo, Tomlins et al. (2008) and Carver et al. (2009) hypothesize the presence of a second hit necessary to provide the cells with full neoplastic potential. With the goal of identifying novel genomic rearrangements, Maher et al. (2009a,b) reported on using RNA-seq to interrogate the whole cellular transcriptome of breast cancer cell lines and a small number of prostate cancer cell lines and tumor samples. Their study identified 11 novel prostate cancer-specific gene fusions (six in prostate cancer cell lines and five in primary patient samples). A subsequent study by Pflueger et al. (2009) identified an additional ETS fusion using RNA-seq, and Palanisamy et al. (2010) recently published rearrangements of *RAF* genes (*BRAF* and *CRAF*) in ~1%–2% of prostate cancer and other cancers. The results of these studies are summarized in Supplemental Table S1. Given the known heterogeneous nature of prostate cancer (Sboner et al. 2010a), the limited number of human samples analyzed so far are unlikely to be sufficient to capture the total repertoire of genomic translocation events in prostate cancer. Therefore, to investigate the extent to which novel gene fusions or chimeric read-through transcripts are involved in prostate cancer, we used whole transcriptome sequencing of 25 prostate cancer samples enriched for ETS fusion negative samples and three benign prostate tissues using the Illu-

mina Genome Analyzer II (GAII). In total, this represents the preparation and processing of 69 lanes of Illumina Genome Analyzer II and the generation of more than 1 billion reads (Table 1). About 600 million mapped reads derived from the paired-end (PE) RNA-seq data were processed through a novel computational tool called FusionSeq (an openly available resource at http://rnaseq.gersteinlab.org/fusionseq/; Sboner et al. 2010b) developed to specifically nominate high-confident chimeric RNA transcripts by accounting for sources of noise that can introduce artifacts (Fig. 1A; Supplemental Fig. S1). Mandatory attribution of confidence scores by FusionSeq enabled us to prioritize chimeric RNA transcripts for validation by reverse-transcription polymerase chain reaction (RT-PCR) and fluorescence in situ hybridization (FISH). Subsequently, we performed a set of selected functional follow-up studies on confirmed novel gene fusions.

Results

Computational approach

We used a novel computational approach, FusionSeq (Sboner et al. 2010b; also described in the Supplemental material), to identify chimeric transcripts based on PE reads where the two ends are mapped to different genes (Fig. 1A; Supplemental Fig. S1). In order to discard artifactual candidates, a set of filters is then applied (the various filters are summarized in Supplemental material). The application of these filters reduces the initial list of candidate fusion transcripts of ~96% on average, thus enriching for potentially real fusion transcripts (Supplemental Table S2). Also, similar to

Table 1. Summary statistics of data generation with next-generation RNA sequencing

Sample	Tissue type	ETS fusion	No. of lanes	Read length	No. of reads	No. of mapped reads	Percent of mapped reads	No. of reads/lane	No. of mapped reads/lane	Nucleotides mapped
STID410_C	PCa	No	4	50	85,735,958	57,736,970	67%	21,433,990	14,434,243	2,886,848,500
STID420_D	PCa	No	2	51/54	40,427,051	23,244,395	58%	20,213,526	11,622,198	1,185,464,145
STID580_B	PCa	Yes	1	54	43,363,383	23,558,632	54%	43,363,383	23,558,632	1,272,166,128
STID581_D	PCa	Yes	4	50/54	54,796,238	34,954,299	64%	13,699,060	8,765,641	1,853,251,180
STID1024_D	PCa	No	1	54	33,869,075	20,265,731	60%	33,869,075	20,265,731	1,094,349,474
STID1043_B	PCa	No	2	36/54	13,673,454	6,066,051	44%	6,836,727	3,033,026	302,056,182
STID1700_D	PCa	Yes	2	51/54	45,122,208	26,856,664	59%	22,561,104	13,428,332	1,427,372,037
STID1783_B	PCa	No	4	50	68,297,519	32,873,206	48%	17,074,380	8,218,302	1,643,660,300
STID2525_A	PCa	No	4	50/51	41,159,746	22,456,506	55%	10,289,937	5,614,127	1,122,825,300
STID2620_D	PCa	Yes	4	50/51	57,263,072	32,484,916	57%	14,315,768	8,121,229	1,624,245,800
STID2660_B	PCa	No	2	54	25,141,860	11,486,658	46%	12,570,930	5,743,329	620,279,532
STID2661_D	PCa	No	2	54	17,053,903	8,501,341	50%	8,526,952	4,250,671	459,072,414
STID2740_A	PCa	No	2	54	39,198,207	15,974,309	41%	19,599,104	7,987,155	862,612,686
STID2743_D	PCa	Yes	3	54	42,715,472	21,561,194	50%	14,238,491	7,187,065	1,164,304,476
STID2762_D	PCa	No	2	54	55,913,548	26,388,959	47%	27,956,774	13,194,480	1,425,003,786
STID2849_D	PCa	No	2	54	50,794,664	21,493,260	42%	25,397,332	10,746,630	1,160,636,040
STID2858_C	PCa	Yes	2	54	38,444,691	19,052,852	50%	19,222,346	9,526,426	1,028,854,008
STID2872_D	PCa	No	2	54	42,552,504	18,618,017	44%	21,276,252	9,309,009	1,005,372,918
STID3023_B62	PCa	Yes	4	50	49,936,734	30,710,795	62%	12,484,184	7,677,699	1,535,539,750
STID3027_B57	PCa	No	4	50	52,824,859	29,991,462	57%	13,206,215	7,497,866	1,499,573,100
STID3042_H51	PCa	No	2	50	26,442,544	15,069,117	57%	13,221,272	7,534,559	753,455,850
STID3043_B56	PCa	No	2	50	24,704,009	13,698,071	55%	12,352,005	6,849,036	684,903,550
STID3071_B51	PCa	No	2	54	53,957,766	25,073,054	46%	26,978,883	12,536,527	1,353,944,916
STID3127_B56	PCa	No	2	54	42,869,486	25,252,807	59%	21,434,743	12,626,404	1,363,651,578
STID3134_B58	PCa	No	2	54	26,092,816	13,883,568	53%	13,046,408	6,941,784	749,712,672
STID2762_A	Benign prostate	No	2	54	43,791,420	20,153,525	46%	21,895,710	10,076,763	1,088,290,350
STID2743_C	Benign prostate	No	2	54	47,408,441	20,567,529	43%	23,704,221	10,283,765	1,110,646,566
STID2740_C	Benign prostate	No	2	54	45,247,018	19,908,070	44%	22,623,509	9,954,035	1,075,035,780
	Total		69		1,208,797,646	637,881,958		533,392,281	276,984,664	33,353,129,018

Summary of sequence data generation for 25 prostate cancer samples (PCa) and three benign prostate tissue samples analyzed by next-generation whole transcriptome sequencing on Genome Analyzer II (Illumina).
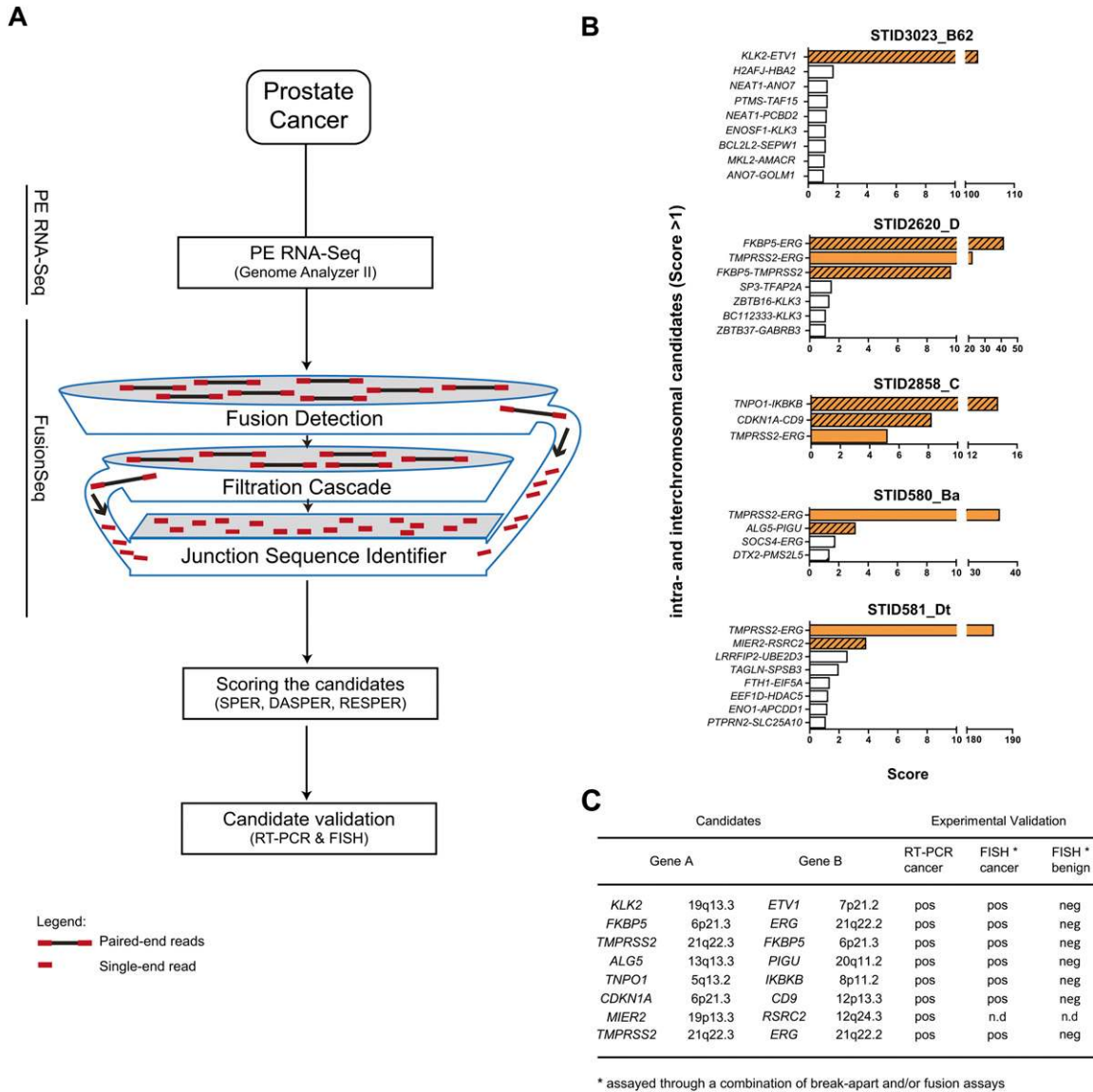
Figure 1. FusionSeq identifies seven high-scoring intra- and interchromosomal candidates that could be validated positively in five prostate cancer samples. (*A*) Schematic of the computational processes composing FusionSeq. (*B*) List of candidates within five samples. Candidates with negative DASPER scores were removed, and the remaining candidates were sorted anticlimactic by RESPER (referred to as Score) with a cutoff of 1. True gene fusions (solid orange, known *TMPRSS2–ERG* fusions; striated orange, novel gene fusions) score higher than candidates that appear down the list (white bars). (*C*) The novel gene fusions were experimentally validated using RT-PCR and FISH.

previous studies, the remaining candidates are categorized as interchromosomal, intrachromosomal (connecting distant genes on the same chromosome), read-through (joining neighboring genes located on the same strand), or *cis* events (bridging genes located on the same chromosome but in different strands). In addition, FusionSeq attributes scores in order to sort and prioritize the candidates for experimental follow-up.

Validation of candidates in the read-through and *cis* categories

To begin with, we sought to validate selected read-through and *cis* candidate chimeras. Nine of the 11 highest scoring read-through chimeras were validated by RT-PCR (Supplemental Table S3). Read-through transcripts form from similarly oriented neighboring genes. While the mechanism is unknown, evidence from several

studies using high-resolution oligonucleotide SNP array analysis suggest that there is neither gross genomic rearrangement nor loss of intervening genomic DNA required to the formation of this specific class of fusion transcripts (for review, see Gingeras 2009). Our results suggest that read-through events are not cancer-restricted, as several positively validated chimeric read-through transcripts were nominated in both prostate cancer and benign prostate tissues (Supplemental Table S3) including the previously described *ZNF649–ZNF577* and *SLC45A3–ELK4* read-through events (Maher et al. 2009a; Rickman et al. 2009). Although this analysis does not completely rule out possible contamination by epithelial cancer cells in the benign tissue, our data provide evidence for an extended complexity of the cellular transcriptome and support a broader definition of a gene that might include utilization of exons from neighboring genes (Gerstein et al. 2007; Gingeras

2009). It is still unknown whether the formation of read-through transcripts is part of a normal biological activity or if it rather defines a cancer-related process that could appear in adjacent benign tissue of a cancerous prostate. None of the two *cis* chimera candidates were validated by RT-PCR, suggesting that they might be nominated due to an incomplete gene annotation, as exemplified by *PICK1–SLC16A8* (Supplemental material; Supplemental Fig. S2).

Identification of novel gene fusions in prostate cancer

Seven high-scoring inter/intrachromosomal cancer-specific chimeric candidates were nominated for experimental validation (Fig. 1B,C). From two *TMPRSS2–ERG* fusion negative tumor samples (STID3023_B62 and STID2620_D), we identified and validated two novel gene fusions involving ETS family members, *KLK2–ETV1* and *FKBP5–ERG* (Fig. 2; Supplemental Fig. S3; Supplemental material). Both of these 5′ partners are known androgen-regulated genes (Makkonen et al. 2009), and *KLK2* has been reported as a 5′ gene fusion partner to *ETV4* (Hermans et al. 2008). To our knowledge, this is the first description of a *KLK2–ETV1* fusion event. At the transcript level, the fusion was successfully validated using RT-PCR followed by Sanger sequencing (Fig. 2B). We confirmed the *KLK2–ETV1* rearrangement at the DNA level by FISH break-apart assays (Fig. 2C). In the second case, empirical validation of fusions involving *FKBP5* led to the discovery of a complex triple fusion event with *FKBP5* joined to *TMPRSS2* and *ERG* (Supplemental Fig. S3). Interestingly, two novel fusion candidates (*CDKN1A–CD9* and *TNPO1–IKBKB*) were nominated in a tumor sample that was also *TMPRSS2–ERG* gene fusion positive (STID2858_C) (Figs. 3, 4). Both fusions were validated by RT-PCR (Figs. 3B, 4B), and the presence of genomic rearrangements was confirmed by FISH assays (Figs. 3C, 4C). The gene fusions only occur in the cancer focus, not in the adjacent benign tissue. The fusion event between *CDKN1A* and *CD9* appears to influence the expression of *CDKN1A*, which is low in this sample (Fig. 3D). Evaluation of CD9 protein expression in the tissue derived from this patient showed partial loss of plasma membrane staining in prostate cancer cells as compared to adjacent benign prostate ep-

ithelial cells or to a prostate cancer without the *CDKN1A–CD9* rearrangement (Fig. 3E). Examination of the *CDKN1A–CD9* fusion open reading frame (ORF) sequence reveals that the fusion brings into frame a downstream initiating methionine of the CD9 and therefore is predicted to produce an N-terminally truncated CD9 that lacks one transmembrane domain and the small extracellular loop (Fig. 3F). To determine the basis underlying altered expression of CD9, we transiently transfected HEK293 cells with either wild-type (WT)–CD9 or CDKN1A–CD9 fusion expression vector and examined the expression of the protein products. Western blot analysis indicated lower protein expression from the CDKN1A–CD9 fusion construct in HEK293 cells as compared with the WT–CD9 construct (Supplemental Fig. S4A). This is most likely due to the utilization of an AUG start site that has a weaker Kozak consensus sequence than the one used to produce WT–CD9. Immunofluorescence staining revealed that in contrast to WT–CD9, truncated CD9 exhibits weak to absent membranous expression (Fig. 3G). In a prostate model (benign prostate epithelial RWPE), we confirmed the lack of membranous expression in cells transfected with truncated CD9 (Supplemental Fig. S4B). The truncated CD9 protein was unable to transform cells in foci formation assays using NIH3T3 cells (data not shown). We screened various prostate cancer cell lines for endogenous expression of *CD9* (Fig. 3H). The lowest expression was found in DU145 cells. The androgen receptor (AR) negative DU145 line is derived from a brain metastasis and represents an aggressive stage of prostate cancer (van Bokhoven et al. 2003). DU145 is normally invasive in vitro, and, therefore, we set out to further characterize the contribution of *CD9* loss to the induction of invasion. Stably reintroducing high WT–CD9 levels in DU145 resulted in a significant reduction in the invasive behavior of DU145 cells (Fig. 3I–K).

We also detected a fusion placing *IKBKB* (IKK-beta) next to *TNPO1* (transportin 1) regulatory sequences (Fig. 4). The presence of the *TNPO1–IKBKB* fusion was associated with an IKK-beta gene expression level more than ninefold higher than the median expression level of other prostate cancer samples (Fig. 4D). The fusion gene maintains the native translation start site in exon 2 of *IKBKB* and therefore is predicted to encode for wild-type IKK-beta (the
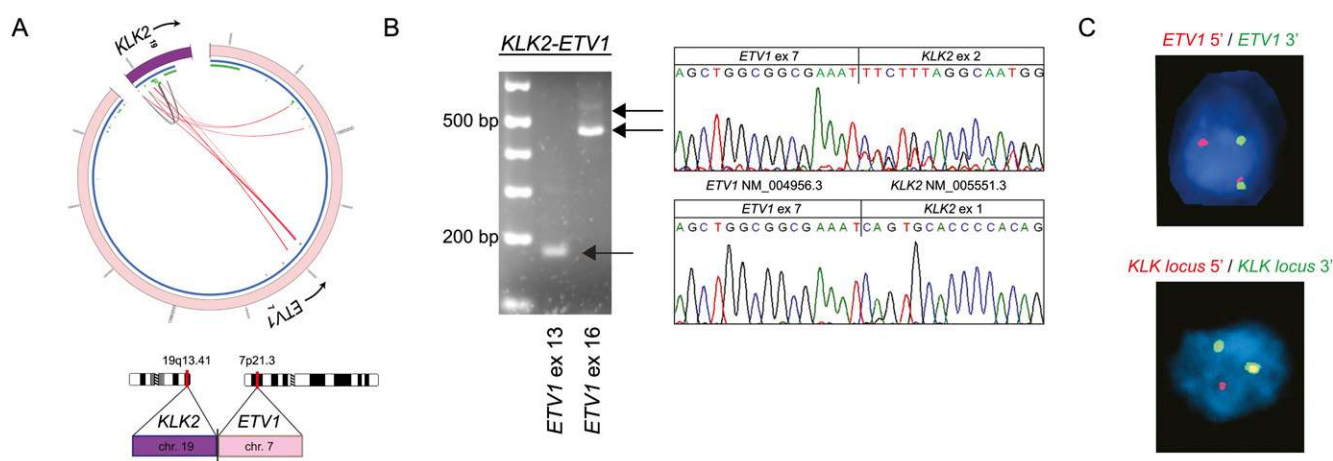
Figure 2. Identification of *KLK2* as novel 5′ fusion partner of *ETV1*. (*A*) Circos plot of *KLK2–ETV1* rearrangement. *Outer* ring: (purple) chromosome 19; (pink) chromosome 7. *Inner* ring: (blue) genes; (green) exons. *Within* the *inner* ring, lines denote PE reads with both reads belonging to *KLK2* (gray) and reads connecting *ETV1* and *KLK2* (red). (*B*) RT-PCR and Sanger sequencing of the resulting fusion transcript reveals the expression of two different fusion transcripts. PCR products were sequenced using the reverse primer, so sequence traces are given in reverse orientation. (*C*) *ETV1* and the *KLK* locus are rearranged as determined by FISH break-apart assays. Of note, *KLK2*-specific FISH is not feasible due to the small size of the *KLK2* gene and its close location to neighboring genes of the *KLK* locus. To address this, genomic rearrangement of the *KLK2* gene is inferred from a rearrangement in the genomic locus of the *KLK* gene family.
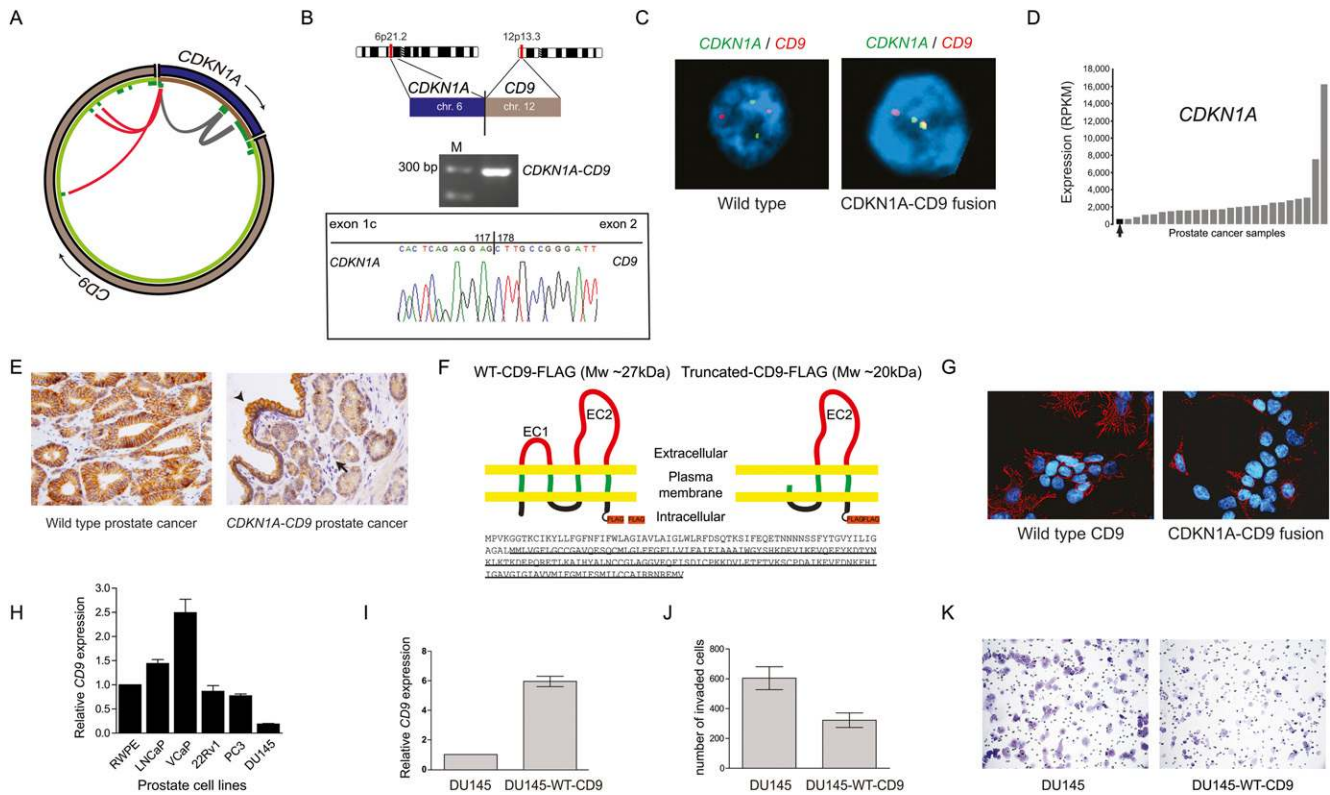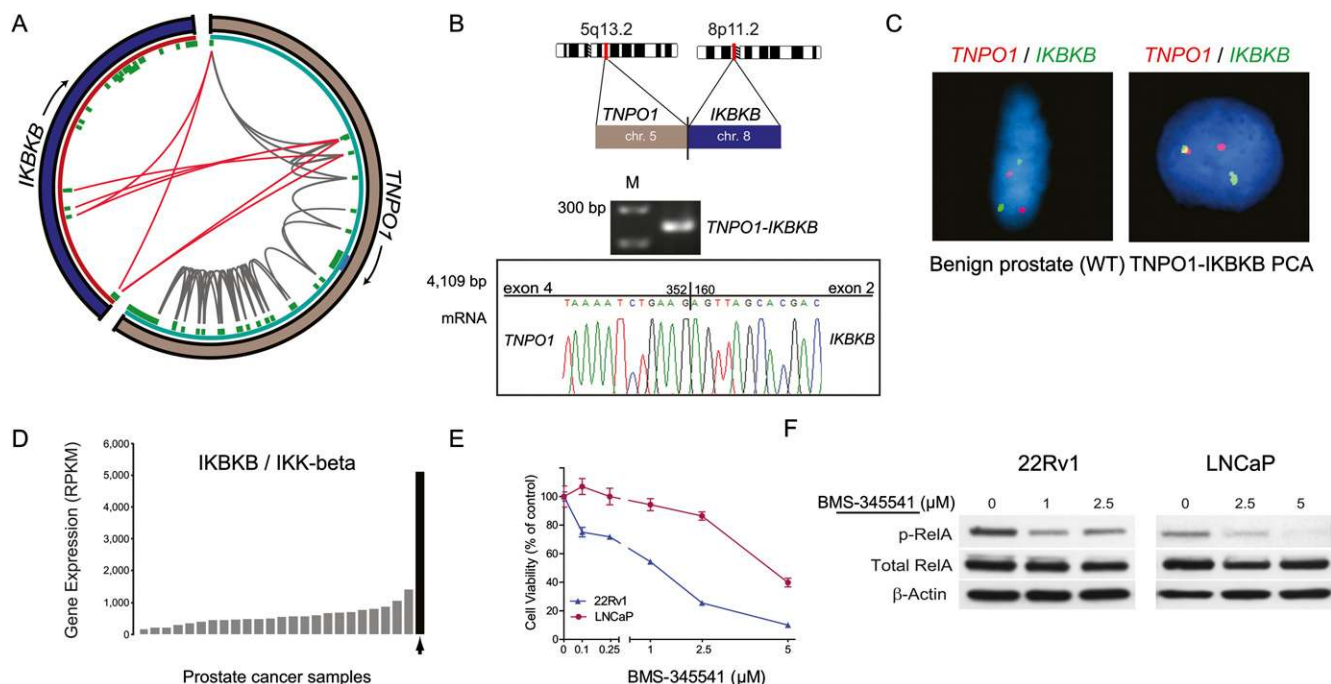
Figure 3. Characterization of the *CDKN1A–CD9* gene fusion candidate. (*A*) Mapping of PE RNA-seq reads to *CDKN1A* with either both ends mapping to *CDKN1A* (gray) or one end mapping to *CD9* (red) and thereby connecting *CDKN1A* exon 1 to 3′ exons of *CD9*. (*B*) Experimental validation of the gene fusion transcript by RT-PCR and subsequent Sanger sequencing. (*C*) FISH validation of the *CDKN1A–CD9* fusion in the index case (*right*) but not in another control cancer (*left*). (*D*) The fusion positive prostate cancer sample (black) has the lowest *CDKN1A* expression levels across 25 prostate cancers. Expression levels are indicated in reads per kilobase of exon model per million mapped reads (RPKM) as defined previously (Mortazavi et al. 2008). (*E*) Immuno-detection of CD9 by immunohistochemistry. (*Left*) Strong CD9 membranous expression detected in malignant glands from a prostate cancer sample without a detectable *CD9* rearrangement. (*Right*) Malignant glands in *CDKN1A–CD9* fusion positive case (arrow) showing weak membranous expression in comparison to adjacent benign areas (arrowhead). (*F*) Schematic illustration of Flag-tagged WT–CD9 and Flag-tagged CDKN1A–CD9 fusion protein. The fusion event leads to a truncated CD9 protein with loss of two transmembrane domains and the small extracellular domain (EC1). (*Bottom*) The amino acid sequence of WT–CD9 protein with the truncated CD9 protein version underlined. (*G*) Immunofluorescence staining revealed the expression of WT–CD9 and truncated CD9 in HEK293 cells stained with anti-Flag (red) antibody. Nuclei were stained with DAPI (blue). (*H*) Expression levels of *CD9* mRNA transcripts in prostate cell lines. (*I*) *CD9* expression from control or stably CD9-expressing DU145 cells. (*J*) Bar graph comparing the invasiveness of the two lines as assessed by Boyden chamber assays. (*K*) Representative images of invaded cells for each cell line.

predicted ORF resulting from the fusion is provided in the Supplemental material). We examined whether IKK-beta expression is essential to maintain the viability of human prostate cancer cells LNCaP and 22Rv1 by incubating them with BMS-345541, a potent selective allosteric site-binding inhibitor of IKK-beta (Burke et al. 2003). Within 72 h, cell viability in the BMS-345541-treated cells was significantly compromised compared to vehicle-treated cultures (Fig. 4E). Apoptotic cells were manifest as early as 24 h following treatment at concentrations of 1 µM and 5 µM, respectively (data not shown). Within 14 h, reduced levels of phospho-RelA were observed indicating that BMS-345541 can inhibit NFKB signaling in these two prostate cancer cell lines through its ability to antagonize IKK-beta activity (Fig. 4F).

Interestingly, another potential oncogene, *PIGU*, was nominated in one case harboring *TMPRSS2–ERG* rearrangement (STID580_Ba). Paired-end reads connecting *PIGU* (chr 20) and *ALG5* (chr 13) (Fig. 5A) indicated the presence of a reciprocal balanced translocation event between *ALG5* and *PIGU* giving rise to two novel gene fusions confirmed by RT-PCR and FISH (Fig. 5B,C). As none of the common prostate cancer models (i.e., LNCaP, DU145, VCaP, 22Rv1) harbor this translocation (as determined by

RT-PCR), we transiently expressed *ALG5–PIGU* and *PIGU–ALG5* in HEK293 cells. While both mRNA transcripts were detectable (Fig. 5D), only the ALG5–PIGU fusion protein is produced with a molecular weight of <50 kDa (Fig. 5E; Supplemental Fig. S5A). A sequence-based prediction of protein domains further indicates that the functional domain of PIGU is maintained in the ALG5–PIGU fusion protein (Supplemental material). The expression of PIGU in the tumor with this fusion event did not, however, appear significantly different from other samples that did not harbor the fusion (Supplemental Fig. S5B). In vitro experiments using siRNA (Fig. 5F,G; Supplemental Fig. S5C) demonstrate that PIGU expression is required for maintaining anchorage-independent growth and proliferative capacity of prostate cancer cells. Since a previous *PIGU* translocation was described as a germline event in a patient with bladder cancer (Guo et al. 2004), we also examined this case for a similar germline alteration by using FISH, but no alterations were observed in normal adjacent prostate tissue (Supplemental Fig. S5D).

Interestingly, none of the newly identified 5′ gene fusion partners (i.e., *ALG5*, *PIGU*, *CDKN1A*, and *TNPO1*) are androgen-regulated (Supplemental Fig. S6) in contrast to the most common

Figure 4. Characterization of the *TNPO1–IKBKB* gene fusion candidate. (*A*) Mapping of PE RNA-seq reads to *TNPO1* with either or both ends mapping to *TNPO1* (gray) or one end mapping to *IKBKB* (red). (*B*) Experimental validation by RT-PCR and Sanger sequencing verifies the expression of a *TNPO1–IKBKB* fusion transcript. (*C*) Fusion specific FISH assays confirm the existence of *TNPO1–IKBKB* fusion in cancer tissue (*right*) but not in adjacent benign tissue (*left*). (*D*) *IKBKB* expression levels in a set of 25 prostate cancers. *IKBKB* levels are highest in the fusion positive sample (black). (*E*) Dose response curve assessing the effect of the IKK-beta inhibitor BMS-345541 on viability of LNCaP and 22Rv1 prostate cancer cells. (*F*) Immunoblots depicting the effect of BMS-345541 exposure on RelA phosphorylation in LNCaP and 22Rv1 cells.

5′ prostate cancer fusion partners (i.e., *TMPRSS2*, *SLC45A3*, and *NDRG1*). Finally, we detected a fusion between *MIER2* and *RSRC2* in a third *TMPRSS2–ERG* fusion positive sample (STID581_Dt). *RSRC2* has been proposed previously as a tumor suppressor gene in the setting of esophageal cancer (Kurehara et al. 2007). RT-PCR validation and subsequent Sanger sequencing revealed a fusion transcript connecting exon 3 of *MIER2* with exon 11 of *RSRC2* (Supplemental Fig. S7A). The predicted open reading frame resulting from the *MIER2–RSRC2* fusion encodes a 114-amino-acid fusion protein that might retain 81 amino acids of the N-terminal portion of MIER2 and 59 amino acids from the small C-terminal portion of RSRC2 (Supplemental Fig. S7B). Thus, this fusion likely impairs the proper function of these two genes. In addition, examination of mRNA expression levels showed that this sample had relatively low levels of *RSRC2* and *MIER2* (Supplemental Fig. S7C,D). We screened additional prostate cancer samples by FISH on tissue microarrays for the presence of rearrangements involving *CDKN1A* (n = 87), *CD9* (n = 77), *IKBKB* (n = 83), or *PIGU* (n = 74). In all, we identified one additional case displaying *CD9* rearrangement and one with *IKBKB* rearrangement. Interestingly, these two fusions occur in ETS rearrangement positive cases. Neither *CDKN1A* nor *TNPO1* were the respective gene fusion partners in these cases. The lack of availability of frozen material for this sample prevented us from nominating any potential 5′ fusion partners. However, these data suggest that, as exemplified by the plethora of ETV1 gene fusion partners (Tomlins et al. 2007), there may be additional unknown 5′ fusion partners to *CD9* and *IKBKB*. In another set of 110 prostate cancer cases from the University of Michigan, we failed to identify additional instances of the novel fusions presented here.

Validation of scoring system in FusionSeq

In order to better estimate the accuracy of our scoring system that may have failed in detecting gene fusions in the 17 non-ETS rearranged cases, we tested for the presence of chimeric transcripts in two instances of interchromosomal fusion candidates with low or negative FusionSeq scores. The prostate tissue-specific gene and known 5′ gene fusion partner *KLK2* with *P4HB*, and the ubiquitously expressed housekeeping gene and known 5′ gene fusion partner *HNRNPA2B1* with *PLA2G2A* failed experimental validation despite the fact that they were nominated in more than one sample. A likely explanation for this is that chimeric RNA transcripts involving two genes that are highly and ubiquitously expressed might be the result of experimental artifacts created during sample preparation by random ligation of abundantly available gene transcripts. Further support of this finding comes from the low or negative score of those chimeric candidates. Collectively, these results support the hypothesis that our scoring system enriches for real fusion transcripts.

Discussion

This study significantly expands the scope of known gene fusions in human prostate cancer to now include fusions involving both ETS and non-ETS genes. We identified seven novel gene fusions in human tissue prostate cancer samples, five of which are rare and involve genes not previously implicated in gene fusions.

The complex *TMPRSS2–FKBP5–ERG* fusion could not have been anticipated using paired-end RNA-seq without RT-PCR and FISH validation. This fusion represents a novel class of complex
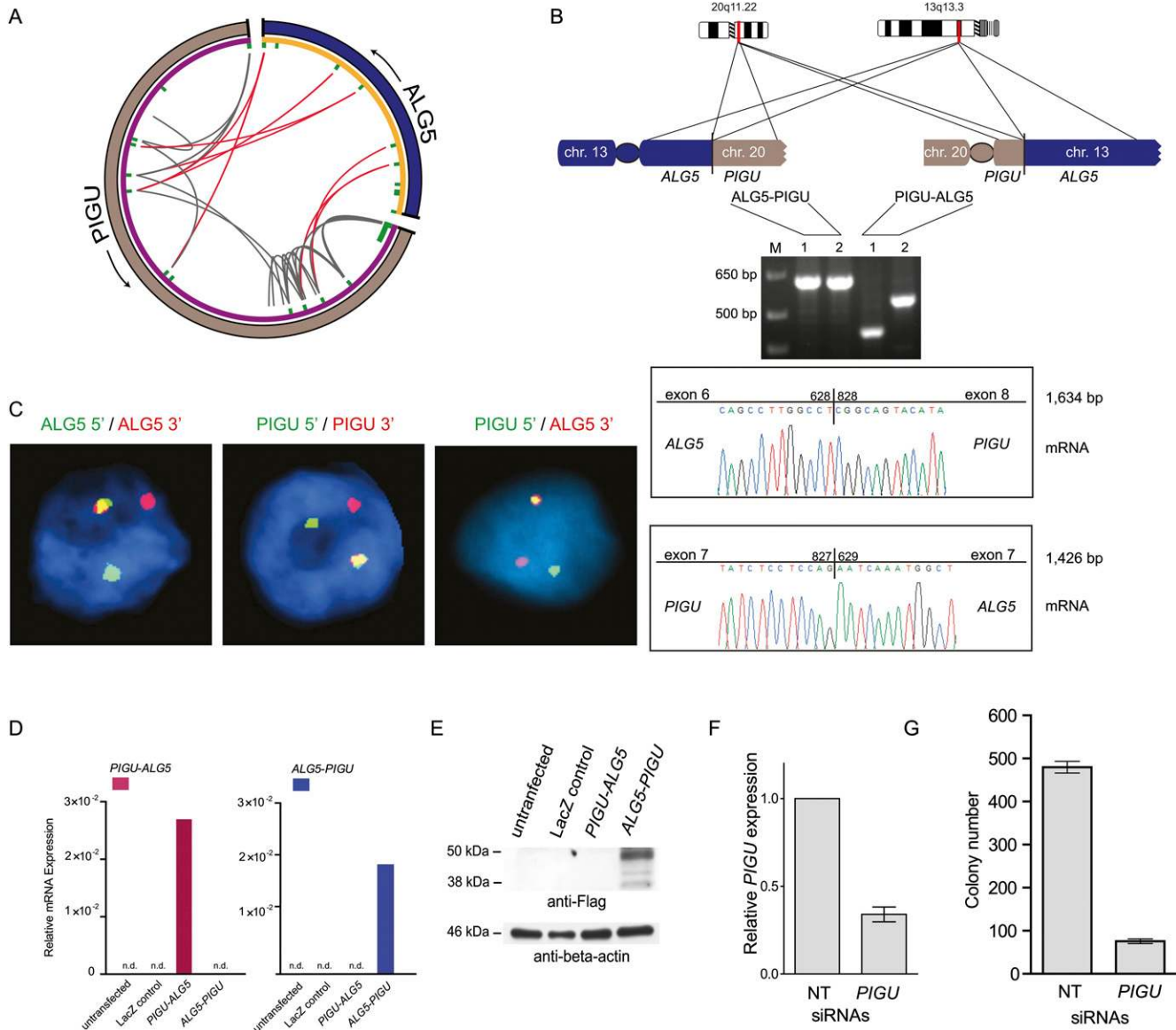
Figure 5. Characterization of the reciprocal balanced translocation event involving *PIGU* and *ALG5*. (*A*) Circos plot of *ALG5–PIGU* rearrangement. In the *outer* ring, (gray) chromosome 20 and (blue) chromosome 13. Genes are represented in purple (*PIGU*) and orange (*ALG5*) on the *inner* ring. *Within* the *inner* ring, each line denotes paired-end reads with either both ends mapping to *PIGU* (gray) or one end mapping to *PIGU* and the other end to *ALG5* (red). The red lines connect 5′ *ALG5* to 3′ *PIGU* and vice versa, thereby indicating a balanced translocation event. (*B*) Experimental validation of the two resulting gene fusions by RT-PCR and subsequent Sanger sequencing of the resulting PCR products. Two different primer pairs were used for verification of each gene fusion transcript indicated as 1 and 2. (*C*) FISH validation of the *ALG5–PIGU* fusion in the index case by break-apart assays (*left* and *middle*) and a fusion assay (*right*). (*D*) Expression of *ALG5–PIGU* and *PIGU–ALG5* messages, as determined by quantitative PCR following transfection of the indicated constructs in HEK293 cells. (*E*) Immunoblot analysis on the transfected HEK293 cells showing protein expression only in *ALG5–PIGU* transfected cells. (*F,G*) LNCaP cells were treated with nontargeting (NT) siRNAs or siRNAs against *PIGU* to assay for effect of anchorage-independent growth in soft agar. (*F*) Bar graph showing expression of *PIGU* after siRNA treatment. (*G*) LNCaP cells treated with *PIGU* siRNAs show reduced colony formation ability.

fusions that may have functional relevance in conferring a growth advantage to neoplastic cells.

Altered CD9 expression has been reported in several human carcinomas originating from the colon, breast, cervix, lung, and prostate, and has been consistently associated with adverse prognosis (Higashiyama et al. 1995; Huang et al. 1998; Mori et al. 1998; Sauer et al. 2003; Wang et al. 2007). Our findings suggest that CD9 is an important mediator in regulating invasiveness and that its loss or reduction might contribute to increasing the metastatic competence of prostate cancer cells. The *CDKN1A–CD9* fusion

likely belongs to a class of inactivating fusion events whose main consequence is to compromise the expression of the wild-type form of the fusion partners. However, it is assumed that in most cases, the remaining wild-type allele(s) of a "gene-inactivating" fusion event can compensate for such loss. In some circumstances, however, heterozygous inactivation of a gene may be sufficient to have functional consequence, alone or in conjunction with additional events as supported by loss of *PTEN* alone or in conjunction with ETS rearrangements (Yoshimoto et al. 2008; Carver et al. 2009; Han et al. 2009). As another example for inactivating fusion

events, we have identified an intrachromosomal *RYBP–FOXP1* fusion in NCI-H660 (*TMPRSS2–ERG* fusion positive cell line). Both genes reside at chromosome 3p14, a locus involved in frequent deletions in *TMPRSS2–ERG* positive prostate cancer, which has recently been suggested to contain a tumor suppressor (Supplemental Fig. S8; Taylor et al. 2010).

IKK-beta kinase is one subunit of the IkB kinase (IKK) complex that is a major component of the NFKB pathway, but recent evidence also demonstrated that IKK-beta might have multiple substrates extending its functions to NFKB-independent pathways (Chariot 2009). Such substrates include genes that are especially pertinent to homeostasis and cancer. Despite the growing body of literature showing that IKK-beta exerts crucial function in inflammation, immunity, and cancer, the role of IKK-beta in human prostate cancer cells remains largely elusive. Recent evidence using small molecule inhibitors of IKK such as BMS-345541, MLX 105, or PS1145 suggests IKK activity as a potential target for therapy in various malignancies (Lam et al. 2005; Yang et al. 2006; Yemelyanov et al. 2006; Idris et al. 2009; Jinawath et al. 2010). Our analysis now provides evidence that IKK-beta activity is important for maintaining the integrity of prostate cancer cells. If these findings are confirmed, this might offer the opportunity for therapeutic intervention by anti-IKK treatment, especially in prostate cancer patients harboring high expression and/or activity of IKK-beta as we observed in this patient presenting with a *TNPO1–IKBKB* fusion.

PIGU is overexpressed in bladder cancers and has been proposed as an oncogene in bladder (Guo et al. 2004) and breast cancer (Wu et al. 2006). PIGU protein has been found to be up-regulated in prostate cancer (Nagpal et al. 2008). Indeed, the fact that we see no obvious change in PIGU expression in the fusion positive case when compared to the entire data set stands in contrast to the previous fusions that we identified. One explanation for this could be that *PIGU* expression might play a role in cell growth, and thus a significant activity of the wild-type PIGU remains elevated for certain tumor cells.

Interestingly, all of the new interchromosomal fusions (*CDKN1A–CD9*, *TNPO1–IKBKB*, *ALG5–PIGU*, *PIGU–ALG5*, and *MIER–RSRC2*) identified in the present study are found in *TMPRSS2–ERG* rearranged prostate cancer rather than in the ETS fusion negative class. Although the number of samples is limited, this observation does raise the intriguing possibility that the presence of *TMPRSS2–ERG* rearrangements could be a prerequisite for subsequent oncogenic events. In fact, corroborated by the discovery by Lin et al. (2009) that there are mechanisms in place mediating nonrandom chromosomal rearrangements by targeted transcription factor binding, one might anticipate that a class of prostate cancer is predisposed to harbor gene rearrangement, where *TMPRSS2–ERG* arises first and additional gene rearrangements follow. Recently, we discovered the presence of multiple ETS gene rearrangements in prostate cancer samples (Svensson et al. 2010). Assessing the mutual exclusivity of these events in situ using FISH, we observed instances wherein 90% of cancer cells in a given prostate cancer focus harbor a *TMPRSS2–ERG* rearrangement and 10% of these rearranged cancer cells demonstrate another ETS rearrangement, suggesting that the second ETS rearrangement emerged later as part of tumor progression (Svensson et al. 2010).

Recent genome-wide sequencing projects have shed light on the landscape of somatic alterations in several cancer genomes including kidney, breast, and colorectal cancer (Sjoblom et al. 2006; Stephens et al. 2009; Dalgliesh et al. 2010). Many more so-

matic alterations are detectable than expected to be requisite for the transformation of normal cells, which led to the assumption that many are of no functional consequence and considered "passengers" (Stratton et al. 2009). However, a minimal number of five to six alterations is suggested to be essential in driving cancer behavior. These "driving" mutations most likely involve a myriad of different genes giving each cancer case its individual "face" but are ultimately converging onto the modification of a finite number of critical pathways. Secondary alterations are well known to occur in hematological malignancies (Bergsagel and Kuehl 2001; Sulong et al. 2009). The novel gene fusions presented here might represent secondary lesions that occur at a later time point in prostate oncogenesis. After screening more than 200 additional cases, we only identified two recurrent rearrangements from the seven identified by RNA-seq in this study. This suggests that many of these fusion events may be "private events," occurring only in one patient. We have yet to fully characterize if they represent "passenger" or "driving" mutations. However, we hypothesize that a subset of these fusions do add to the biologic nature of an individual's tumor. Some may be highly targetable such as the recently discovered *RAF* fusions (Palanisamy et al. 2010) or the *IKBKB* (IKK-beta) fusion identified in this study. These types of targetable fusions may play a role in a personalized approach to cancer care.

In summary, we describe novel ETS and non-ETS prostate cancer fusions clustering in ETS rearranged prostate cancer. We further demonstrate that some of these fusions may act as additional drivers of tumor progression. These findings also support an emerging view that ETS overexpression may predispose to highly specific DNA breakpoints that are distinct from those observed in fusion negative prostate cancers.

Methods

Samples

The prostate samples (see Supplemental Tables S4 and S5 for clinical information) have been collected as part of an IRB approved protocol at Weill Cornell Medical College. The samples derived from patients with localized or locally advanced disease that were treated with radical prostatectomy as monotherapy. H&E slides of frozen tissue blocks were examined by the study's pathologists (R.E. and M.A.R.) to select for high-density cancer foci with <10% stroma or other contaminating noncancerous material. The *ERG* rearrangement status was examined on the frozen cancer focus using a FISH break-apart assay as previously described (Tomlins et al. 2005; Perner et al. 2006). Simultaneously, RNA was extracted from the frozen cancer tissue using TRIzol (Invitrogen) according to the manufacturer's protocol. For the purpose of molecular characterization, RNA was reverse-transcribed using the cDNA High Capacity kit (Applied Biosystems) and subjected to *TMPRSS2–ERG* PCRs to type for *TMPRSS2–ERG* fusion transcript expression (as described in Tomlins et al. 2005). In the case of RNA extraction from benign tissue, we identified a frozen tissue block with no evidence of tumor tissue to minimize the possibility of tumor contamination in the benign tissue sample.

RNA-seq sample prep and sequencing

Total RNA was prepared in accordance with Illumina's sample preparation protocol for PE sequencing of mRNA unless described otherwise. In brief, 5–10 μg of total RNA was fragmented by heat between 2 and 3 min based on the desired insert size, reverse-transcribed using Superscript II (Invitrogen), and transformed to double-stranded cDNA. To improve PE RNA-seq data quality, we

introduced an additional gel-based size selection step after cDNA double-strand synthesis and before the ligation of the PE adapters. This was postulated by Quail et al. (2008) as a means to reduce the inclusion of artifactual chimeric transcripts that are composed of two cDNA fragments into the sequencing library. We also integrated the use of T4 ligase (Enzymatics Inc.) to improve the efficiency of adapter ligation. Over the course of the study, we increased the library size range from 250 bp to 450 bp. The gel dissolutions of all gel-based purification steps were conducted at room temperature under slight agitation as described by Quail et al. (2008). After the enrichment of cDNA template by PCR, the concentrations and the sizes of the libraries were measured using a Qubit fluorometer (Invitrogen) and DNA 1000 kit (Agilent Technologies) on an Agilent 2100 Bioanalyzer, respectively. PE RNA-seq was performed with the Genome Analyzer II (Illumina) increasing the read size of the PE reads from 36 to 54 bp over the course of the study. Additionally, Illumina introduced improved sequencing reagents and upgraded imaging software over time to increase data quality and sequencing coverage.

Processing of RNA-seq data through FusionSeq

PE reads were aligned to the human genome (hg18) using ELAND, part of the standard software suite from Illumina. PE reads that mapped with only up to two mismatches were then processed through FusionSeq. In brief, FusionSeq consists of a compendium of modules that:

1. Classify the reads in groups depending on if the respective PE reads belong to the same gene, to different genes, or if they are not mappable.
2. Filter all the chimeric PE reads, that is, those where the two reads map to different genes, in order to discard artifactual candidates such as those involving homologous genes as well as those joining highly expressed genes, such as ribosomal, that might have been generated from sample prep. It also filters candidates that may result from PCR artifacts as well as gene annotation issues. This module also categorizes candidates and attributes various scores in order to sort and prioritize them for experimental follow-up.
3. Identify the sequence of the junction between the mRNAs coming from the two different genes with the aim to find overall support for the chimeric transcript.

Validation experiments by RT-PCR

PCR validation has been performed using primers outside of the region that is covered by PE reads to introduce a second step of quality control for the further reduction of the possibility of detecting false-positive chimeric transcripts due to a regional high homology between partner genes. A second round of PCR, using another primer pair, was employed if the results from the first primer pair were inconclusive (Supplemental Table S6).

Validation experiments by FISH

All FISH experiments were performed as described (Tomlins et al. 2005; Perner et al. 2006) on the same cancer focus that was subjected to RNA extraction for RNA-seq and PCRs. In a FISH breakapart assay, a nucleus without gene rearrangement shows two pairs of red–green signals, often forming two yellow signals, indicating two intact alleles. A nucleus with gene rearrangement shows the split of one yellow signal into two distinct red and green signals for the rearranged allele and one remaining juxtaposed red–green (yellow) signal indicating the intact allele. In a FISH fusion assay, a nucleus with fusion shows one red and one green distinct signal coming together to form a yellow signal and indicating the fusion allele. The FISH probes are listed in Supplemental Table S7. For screening analysis, FISH experiments were carried out on two tissue microarrays (TMA) containing 41 and 47 samples, respectively.

Immunohistochemistry

Immunohistochemical staining of CD9 (clone 72F6, 1:250 dilution; Abcam) was performed using the Bond Max Autostainer (Leica Microsystems). Formalin-fixed, paraffin-embedded tissue sections were deparaffinized, and endogenous peroxidase was inactivated. Antigen retrieval was accomplished using the Bond Epitope Retrieval Solution 2 (ER2) for 20 min at 99°C–100°C (Leica Microsystems). Following retrieval, the sections were incubated sequentially with the primary antibody for 25 min, post-primary for 15 min, and polymer for 25 min ending with colorimetric development with diaminobenzidine (DAB) for 10 min (Bond Polymer Refine Detection; Leica Microsystems).

Cell cultures and reagents

Hormonal treatment

The human prostate cancer cell lines LNCaP (clone FGC) 22Rv1, VCaP, and DU145 were obtained from the American Type Culture Collection (Manassas, VA). Cells were maintained in either DMEM or RPMI 1640 supplemented with 10% fetal bovine serum (FBS) and penicillin/streptomycin according to manufacturer's instructions. For androgen treatment, cells were cultured in phenol red-free DMEM or RPMI supplemented with 10% charcoal-stripped FBS for 24 h in the presence or absence of 1 nM R1881. BMS-345541 was purchased from EMD4 Biosciences and dissolved in DMSO, and 10 mM stock solution was stored at −20°C before use.

RT–PCR

RNA was extracted using the TRIzol reagent (Invitrogen), subjected to DNase treatment (DNA-*free* kit; Applied Biosystems) according to the manufacturer's instructions, and used in quantitative RT-PCR. Quantitative RT-PCR was performed using the ABI 7900 Real-Time PCR System (Applied Biosystems) following the manufacturer's RNA-to-CT 1-step protocol. Each target was run in triplicate, and expression levels relative to the housekeeping gene *TBP* were determined on the basis of the comparative threshold cycle CT method ($2^{-\Delta\Delta CT}$). The primer sequences used in these experiments are given in Supplemental Table S6.

Invasion assay

For the invasion assay, 2×10^4 DU145 cells and DU145 stably transfected with WT-CD9 were resuspended in 0.5 mL of RPMI-1640 medium containing 1% FBS and placed into Matrigel-coated Transwell inserts containing 8-μm filters (BD Falcon). The bottom wells contained RPMI 10% with epidermal growth factor (10 ng/mL). After 36 h, the cells on the upper surface of the filters were removed with a cotton swab. The filters were fixed and stained with Crystal Violet 0.5% for 30 min. The migrated cells were quantified by counting the numbers of cells that penetrated the membrane in five microscopic fields (at 20× objective magnification) per filter. There were no noticeable morphological differences between the control DU145 cells and CD9-transfected DU145 cells. Nor were there any evident differences in growth rates noted during standard passage of the subline compared to the parental cells with an estimated cell population doubling time of 29.5 ± 2.3 h and 30.3 ± 1.5 h, respectively.

Anchorage–independent growth

Cell cultures of LNCaP cells were transfected with either 50 nM of nontargeting control (ON-TARGETplus Non-targeting Pool, Thermo Scientific) or 50 nM siRNAs directed against *PIGU* (ON-TARGETplus SMARTpool) using Lipofectamine 2000 (Invitrogen). Twenty-four hours later, cells were trypsinized and processed for soft-agar assays performed on six-well plates in triplicate. For each well, 1.5×10^4 cells were resuspended in 1 mL of 0.35% agar in RPMI containing 10% FBS. This upper layer was seeded into six-well plates coated with 0.6% agar in RPMI containing 10% FBS. Each well was allowed to solidify and was subsequently covered with 2 mL of the corresponding medium, which was refreshed every 3 d. After 2 wk, colonies were stained with 0.005% (w/v) *p*-iodonitrotetrazodium violet (Sigma) and counted using an inverse microscope at 40× magnification.

Cell viability and cell proliferation assays

For drug sensitivity assays following BMS-345541 treatment, LNCaP and 22Rv1 cells (1×10^4 per well) were seeded on 96-well tissue culture plates. The next day, cells were treated with either BMS-345541 at growing concentrations or vehicle (DMSO) for 72 h. Cell viability was determined by performing WST-1 assay (Roche) reading absorbance at 450 nm according to the manufacturer's instructions. Values from four wells were obtained for each treatment. Data are expressed as percentage of control group without BMS-345541. For siRNA transfection RWPE (7.5×10^3 per well), 22Rv1 (1×10^4 per well), DU145 (2.5×10^3 per well), DU145-ERG (2.5×10^3 per well), and LNCaP (1×10^4 per well) cells were seeded on 96-well tissue culture plates. On the next day, cells were transfected with 100 nM *PIGU*, *MYC*, or nontargeting (NT) siRNAs (ON-TARGETplus; Thermo Scientific) using Lipofectamine 2000 (Invitrogen). At J0 and 72 h, growth was assessed by performing WST-1 assay reading absorbance at 450 nm according to the manufacturer's instructions. Values from five wells were obtained for each treatment. The efficacy of the siRNA knockdown was assessed in several independent experiments by qRT-PCR (data not shown), and the optimal amount of siRNA used for transfection was determined as being 100 nM.

In vitro expression of WT–CD9, CDKN1A–CD9, PIGU–ALG5, and ALG5–PIGU

Using cDNA derived from the human prostate cancer tissue sample, Flag-tagged wild-type ORF of *CD9*, Flag-tagged *CDKN1A–CD9*, *PIGU–ALG5*, and *ALG5–PIGU* fusion sequences were cloned by means of gateway technology (Invitrogen) into pENTR/D-TOPO vector. All the clones generated in the entry vector were sequenced to confirm proper orientation and sequence of cloning and were later recombined into Gateway pLenti6.3/V5–DEST expression vector (Invitrogen) by LR Clonase II enzyme reaction following manufacturer's instructions. C terminus double Flag-tag was generated for initial verification of protein expression in HEK293 cells.

Immunofluorescent staining

HEK293 or benign prostate epithelial RWPE cells were plated onto eight-well Lab-Tek chambered cover slides and allowed to grow overnight. Aliquots of *WT–CD9* or *CDKN1A–CD9* fusion expression vectors (320 μg) were mixed with calcium phosphate and applied to the cultures according to the manufacturer's instructions (Agilent Technologies). Forty-eight hours later, cells were fixed with 4% formaldehyde in PBS for 10 min at room temperature, permeabilized with 0.05% Triton X-100 in PBS for 5 min, and then blocked with 1% BSA in PBS for 30 h. Cells were incubated with the monoclonal anti-Flag antibody (Sigma) in 0.5% BSA overnight at 4°C. After washing with PBS, cells were incubated with the anti-mouse secondary antibodies coupled with Alexa Fluor (Invitrogen) in 0.5% BSA for 30 min at room temperature, washed, and mounted using mounting medium containing DAPI (Invitrogen). Microscopic images were obtained under a 60× oil immersion objective using an Olympus BX-51 fluorescence microscope.

Protein extraction and Western blot analysis

HEK293 cells were transfected with either *LacZ* control, WT-*CD9*, *CDKN1A–CD9* fusion expression vectors or control vector in six-well plates and cultured for 48 h. *PIGU–ALG5* and *ALG5–PIGU* fusion expression vectors were also transfected under these conditions. For BMS-345541 treatment, subconfluent cultures of LNCaP and 22Rv1 cells were treated during 14 h. Protein lysates were prepared in the RIPA buffer (radioimmunoprecipitation assay lysis buffer) supplemented with protease inhibitor cocktail and phosphatase inhibitors (Thermo Scientific). The total protein concentration of the soluble extract was determined using the BCA protein assay Kit (Thermo Scientific). Each protein sample (30 μg) was resolved to SDS-PAGE, transferred onto a polyvinylidene difluoride membrane (Millipore), and incubated overnight at 4°C with primary antibodies. Mouse monoclonal anti-Flag antibody was obtained from Sigma, mouse monoclonal anti-NFKB p65 was from Santa Cruz (sc-8008) to monitor total RelA levels, and rabbit anti-phospho-NFKB p65 (Ser536) was purchased from Cell Signaling to monitor RelA phosphorylation status. Following three washes with TBS-T, the blot was incubated with horseradish peroxidase–conjugated secondary antibody, and immune complexes were visualized by enhanced chemiluminescence detection (ECL plus kit, GE Healthcare). The blot was reprobed with monoclonal antibody against beta-actin (Sigma).

References

Berger MF, Levin JZ, Vijayendran K, Sivachenko A, Adiconis X, Maguire J, Johnson LA, Robinson J, Verhaak RG, Sougnez C, et al. 2010. Integrative analysis of the melanoma transcriptome. *Genome Res* **20:** 413–427.

Bergsagel PL, Kuehl WM. 2001. Chromosome translocations in multiple myeloma. *Oncogene* **20:** 5611–5622.

Burke JR, Pattoli MA, Gregor KR, Brassil PJ, MacMaster JF, McIntyre KW, Yang X, Iotzova VS, Clarke W, Strnad J, et al. 2003. BMS-345541 is a highly selective inhibitor of IκB kinase that binds at an allosteric site of the enzyme and blocks NF-κB-dependent transcription in mice. *J Biol Chem* **278:** 1450–1456.

Carver BS, Tran J, Gopalan A, Chen Z, Shaikh S, Carracedo A, Alimonti A, Nardella C, Varmeh S, Scardino PT, et al. 2009. Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate. *Nat Genet* **41:** 619–624.

Chariot A. 2009. The NF-κB-independent functions of IKK subunits in immunity and cancer. *Trends Cell Biol* **19:** 404–413.

Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, Davies H, Edkins S, Hardy C, Latimer C, et al. 2010. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463:** 360–363.

Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, Emanuelsson O, Zhang ZD, Weissman S, Snyder M. 2007. What is a gene, post-ENCODE? History and updated definition. *Genome Res* **17:** 669–681.

Gingeras TR. 2009. Implications of chimaeric non-co-linear transcripts. *Nature* **461:** 206–211.

Guo Z, Linn JF, Wu G, Anzick SL, Eisenberger CF, Halachmi S, Cohen Y, Fomenkov A, Hoque MO, Okami K, et al. 2004. *CDC91L1* (PIG-U) is a newly discovered oncogene in human bladder cancer. *Nat Med* **10:** 374–381.

Han B, Mehra R, Lonigro RJ, Wang L, Suleman K, Menon A, Palanisamy N, Tomlins SA, Chinnaiyan AM, Shah RB. 2009. Fluorescence in situ hybridization study shows association of PTEN deletion with ERG rearrangement during prostate cancer progression. *Mod Pathol* **22:** 1083–1093.

Hermans KG, Bressers AA, van der Korput HA, Dits NF, Jenster G, Trapman J. 2008. Two unique novel prostate-specific and androgen-regulated fusion partners of ETV4 in prostate cancer. *Cancer Res* **68:** 3094–3098.

Higashiyama M, Taki T, Ieki Y, Adachi M, Huang CL, Koh T, Kodama K, Doi O, Miyake M. 1995. Reduced motility related protein-1 (*MRP-1/CD9*) gene expression as a factor of poor prognosis in non-small cell lung cancer. *Cancer Res* **55:** 6040–6044.

Huang CI, Kohno N, Ogawa E, Adachi M, Taki T, Miyake M. 1998. Correlation of reduction in *MRP-1/CD9* and *KAI1/CD82* expression with recurrences in breast cancer patients. *Am J Pathol* **153:** 973–983.

Idris AI, Libouban H, Nyangoga H, Landao-Bassonga E, Chappard D, Ralston SH. 2009. Pharmacologic inhibitors of IκB kinase suppress growth and migration of mammary carcinosarcoma cells in vitro and prevent osteolytic bone metastasis in vivo. *Mol Cancer Ther* **8:** 2339–2347.

Jinawath N, Vasoontara C, Jinawath A, Fang X, Zhao K, Yap KL, Guo T, Lee CS, Wang W, Balgley BM, et al. 2010. Oncoproteomic analysis reveals co-upregulation of RELA and STAT5 in carboplatin resistant ovarian carcinoma. *PLoS ONE* **5:** e11198. doi: 10.1371/journal.pone.0011198.

Klezovitch O, Risk M, Coleman I, Lucas JM, Null M, True LD, Nelson PS, Vasioukhin V. 2008. A causal role for ERG in neoplastic transformation of prostate epithelium. *Proc Natl Acad Sci* **105:** 2105–2110.

Kumar-Sinha C, Tomlins SA, Chinnaiyan AM. 2008. Recurrent gene fusions in prostate cancer. *Natl Rev* **8:** 497–511.

Kurehara H, Ishiguro H, Kimura M, Mitsui A, Ando T, Sugito N, Mori R, Takashima N, Ogawa R, Fujii Y, et al. 2007. A novel gene, RSRC2, inhibits cell proliferation and affects survival in esophageal cancer patients. *Int J Oncol* **30:** 421–428.

Lam LT, Davis RE, Pierce J, Hepperle M, Xu Y, Hottelet M, Nong Y, Wen D, Adams J, Dang L, et al. 2005. Small molecule inhibitors of IκB kinase are selectively toxic for subgroups of diffuse large B-cell lymphoma defined by gene expression profiling. *Clin Cancer Res* **11:** 28–40.

Levin JZ, Berger MF, Adiconis X, Rogov P, Melnikov A, Fennell T, Nusbaum C, Garraway LA, Gnirke A. 2009. Targeted next-generation sequencing of a cancer transcriptome enhances detection of sequence variants and novel fusion transcripts. *Genome Biol* **10:** R115. doi: 10.1186/gb-2009-10-10-r115.

Lin C, Yang L, Tanasa B, Hutt K, Ju BG, Ohgi K, Zhang J, Rose DW, Fu XD, Glass CK, et al. 2009. Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell* **139:** 1069–1083.

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

Maher CA, Palanisamy N, Brenner JC, Cao X, Kalyana-Sundaram S, Luo S, Khrebtukova I, Barrette TR, Grasso C, Yu J, et al. 2009b. Chimeric transcript discovery by paired-end transcriptome sequencing. *Proc Natl Acad Sci* **106:** 12353–12358.

Makkonen H, Kauhanen M, Paakinaho V, Jaaskelainen T, Palvimo JJ. 2009. Long-range activation of FKBP51 transcription by the androgen receptor via distal intronic enhancers. *Nucleic Acids Res* **37:** 4135–4148.

Mitelman F, Johansson B, Mertens F. 2007. The impact of translocations and gene fusions on cancer causation. *Natl Rev* **7:** 233–245.

Mori M, Mimori K, Shiraishi T, Haraguchi M, Ueo H, Barnard GF, Akiyoshi T. 1998. Motility related protein 1 (MRP1/CD9) expression in colon cancer. *Clin Cancer Res* **4:** 1507–1510.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5:** 621–628.

Mosquera JM, Mehra R, Regan MM, Perner S, Genega EM, Bueti G, Shah RB, Gaston S, Tomlins SA, Wei JT, et al. 2009. Prevalence of TMPRSS2-ERG fusion prostate cancer among men undergoing prostate biopsy in the United States. *Clin Cancer Res* **15:** 4706–4711.

Nagpal JK, Dasgupta S, Jadallah S, Chae YK, Ratovitski EA, Toubaji A, Netto GJ, Eagle T, Nissan A, Sidransky D, et al. 2008. Profiling the expression pattern of GPI transamidase complex subunits in human cancer. *Mod Pathol* **21:** 979–991.

Palanisamy N, Ateeq B, Kalyana-Sundaram S, Pflueger D, Ramnarayanan K, Shankar S, Han B, Cao Q, Cao X, Suleman K, et al. 2010. Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. *Nat Med* **16:** 793–798.

Perner S, Demichelis F, Beroukhim R, Schmidt FH, Mosquera JM, Setlur S, Tchinda J, Tomlins SA, Hofer MD, Pienta KG, et al. 2006. TMPRSS2:ERG fusion-associated deletions provide insight into the heterogeneity of prostate cancer. *Cancer Res* **66:** 8337–8341.

Pflueger D, Rickman DS, Sboner A, Perner S, LaFargue CJ, Svensson MA, Moss BJ, Kitabayashi N, Pan Y, de la Taille A, et al. 2009. *N-myc downstream regulated gene 1* (*NDRG1*) is fused to *ERG* in prostate cancer. *Neoplasia* **11:** 804–811.

Prensner JR, Chinnaiyan AM. 2009. Oncogenic gene fusions in epithelial carcinomas. *Curr Opin Genet Dev* **19:** 82–91.

Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, Swerdlow H, Turner DJ. 2008. A large genome center's improvements to the Illumina sequencing system. *Nat Methods* **5:** 1005–1010.

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

Sauer G, Windisch J, Kurzeder C, Heilmann V, Kreienberg R, Deissler H. 2003. Progression of cervical carcinomas is associated with down-regulation of CD9 but strong local re-expression at sites of transendothelial invasion. *Clin Cancer Res* **9:** 6426–6431.

Sboner A, Demichelis F, Calza S, Pawitan Y, Setlur SR, Hoshida Y, Perner S, Adami HO, Fall K, Mucci LA, et al. 2010a. Molecular sampling of prostate cancer: A dilemma for predicting disease progression. *BMC Med Genomics* **3:** 8. doi: 10.1186/1755-8794-3-8.

Sboner A, Habegger L, Pflueger D, Terry S, Chen D, Rozowsky J, Tewari AK, Kitabayashi N, Moss BJ, Chee MS, et al. 2010b. FusionSeq: A modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. *Genome Biol* (in press).

Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, et al. 2006. The consensus coding sequences of human breast and colorectal cancers. *Science* **314:** 268–274.

Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, Stebbings LA, Leroy C, Edkins S, Mudie LJ, et al. 2009. Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462:** 1005–1010.

Stratton MR, Campbell PJ, Futreal PA. 2009. The cancer genome. *Nature* **458:** 719–724.

Sulong S, Moorman AV, Irving JA, Strefford JC, Konn ZJ, Case MC, Minto L, Barber KE, Parker H, Wright SL, et al. 2009. A comprehensive analysis of the *CDKN2A* gene in childhood acute lymphoblastic leukemia reveals genomic deletion, copy number neutral loss of heterozygosity, and association with specific cytogenetic subgroups. *Blood* **113:** 100–107.

Svensson MA, LaFargue CJ, MacDonald T, Pflueger D, Kitabayashi N, Santa-Cruz A, Garsha K, Sathyanarayana U, Riley J, Yun S, et al. 2010. Testing mutual exclusivity of ETS rearranged prostate cancer. *Lab Invest.* doi: 10.1038/labinvest.2010.179.

Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B, et al. 2010. Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18:** 11–22.

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

Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, Cao X, Morris DS, Menon A, Jing X, Cao Q, Han B, et al. 2007. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* **448:** 595–599.

Tomlins SA, Laxman B, Varambally S, Cao X, Yu J, Helgeson BE, Cao Q, Prensner JR, Rubin MA, Shah RB, et al. 2008. Role of the *TMPRSS2-ERG* gene fusion in prostate cancer. *Neoplasia* **10:** 177–188.

van Bokhoven A, Varella-Garcia M, Korch C, Johannes WU, Smith EE, Miller HL, Nordeen SK, Miller GJ, Lucia MS. 2003. Molecular characterization of human prostate carcinoma cell lines. *Prostate* **57:** 205–225.

Wang JC, Begin LR, Berube NG, Chevalier S, Aprikian AG, Gourdeau H, Chevrette M. 2007. Down-regulation of CD9 expression during prostate carcinoma progression is associated with *CD9* mRNA modifications. *Clin Cancer Res* **13:** 2354–2361.

Wu G, Guo Z, Chatterjee A, Huang X, Rubin E, Wu F, Mambo E, Chang X, Osada M, Sook Kim M, et al. 2006. Overexpression of glycosylphosphatidylinositol (GPI) transamidase subunits phosphatidylinositol glycan class T and/or GPI anchor attachment 1 induces tumorigenesis and contributes to invasion in human breast cancer. *Cancer Res* **66:** 9829–9836.

Yang J, Amiri KI, Burke JR, Schmid JA, Richmond A. 2006. BMS-345541 targets inhibitor of κB kinase and induces apoptosis in melanoma: Involvement of nuclear factor κB and mitochondria pathways. *Clin Cancer Res* **12:** 950–960.

Yemelyanov A, Gasparian A, Lindholm P, Dang L, Pierce JW, Kisseljov F, Karseladze A, Budunova I. 2006. Effects of IKK inhibitor PS1145 on NF-κB function, proliferation, apoptosis and invasion activity in prostate carcinoma cells. *Oncogene* **25:** 387–398.

Yoshimoto M, Joshua AM, Cunha IW, Coudry RA, Fonseca FP, Ludkovski O, Zielenska M, Soares FA, Squire JA. 2008. Absence of *TMPRSS2:ERG* fusions and *PTEN* losses in prostate cancer is associated with a favorable outcome. *Mod Pathol* **21:** 1451–1460.