



ELSEVIER

See related Commentary on page 2608The American Journal of
PATHOLOGY

ajp.amjpathol.org

TUMORIGENESIS AND NEOPLASTIC PROGRESSION**Novel Fusion Transcripts Associate with Progressive Prostate Cancer**

Yan P. Yu,* Ying Ding,[†] Zhanghui Chen,* Silvia Liu,[†] Amantha Michalopoulos,* Rui Chen,[†] Zulfiqar G. Gulzar,[‡] Bing Yang,[§] Kathleen M. Cieply,* Alyssa Luvison,* Bao-Guo Ren,* James D. Brooks,[‡] David Jarrard,[§] Joel B. Nelson,[†] George K. Michalopoulos,* George C. Tseng,[†] and Jian-Hua Luo*

From the Departments of Pathology,* Biostatistics,[†] and Urology,[‡] University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; the Department of Urology,[‡] Stanford University School of Medicine, Stanford, California; and the Department of Urology,[§] University of Wisconsin School of Medicine at Madison, Madison, Wisconsin

CME Accreditation Statement: This activity ("ASIP 2014 AJP CME Program in Pathogenesis") has been planned and implemented in accordance with the Essential Areas and policies of the Accreditation Council for Continuing Medical Education (ACCME) through the joint sponsorship of the American Society for Clinical Pathology (ASCP) and the American Society for Investigative Pathology (ASIP). ASCP is accredited by the ACCME to provide continuing medical education for physicians.

The ASCP designates this journal-based CME activity ("ASIP 2014 AJP CME Program in Pathogenesis") for a maximum of 48 AMA PRA Category 1 Credit(s)[™]. Physicians should only claim credit commensurate with the extent of their participation in the activity.

CME Disclosures: The authors of this article and the planning committee members and staff have no relevant financial relationships with commercial interests to disclose.

Accepted for publication
June 30, 2014.

Address correspondence to
Jian-Hua Luo, M.D., Ph.D.,
Department of Pathology,
University of Pittsburgh School
of Medicine, 3550 Terrace St,
Scaife S-760, Pittsburgh,
PA 15261. E-mail: luoj@upmc.edu.

The mechanisms underlying the potential for aggressive behavior of prostate cancer (PCa) remain elusive. In this study, whole genome and/or transcriptome sequencing was performed on 19 specimens of PCa, matched adjacent benign prostate tissues, matched blood specimens, and organ donor prostates. A set of novel fusion transcripts was discovered in PCa. Eight of these fusion transcripts were validated through multiple approaches. The occurrence of these fusion transcripts was then analyzed in 289 prostate samples from three institutes, with clinical follow-up ranging from 1 to 15 years. The analyses indicated that most patients [69 (91%) of 76] positive for any of these fusion transcripts (*TRMT11-GRIK2*, *SLC45A2-AMACR*, *MTOR-TP53BP1*, *LRRCS9-FLJ60017*, *TMEM135-CCDC67*, *KDM4-ACO11523.2*, *MAN2A1-FER*, and *CCNH-C5orf30*) experienced PCa recurrence, metastases, and/or PCa-specific death after radical prostatectomy. These outcomes occurred in only 37% (58/157) of patients without carrying those fusion transcripts. Three fusion transcripts occurred exclusively in PCa samples from patients who experienced recurrence or PCa-related death. The formation of these fusion transcripts may be the result of genome recombination. A combination of these fusion transcripts in PCa with Gleason's grading or with nomogram significantly improves the prediction rate of PCa recurrence. Our analyses suggest that formation of these fusion transcripts may underlie the aggressive behavior of PCa. (*Am J Pathol* 2014, 184: 2840–2849; <http://dx.doi.org/10.1016/j.ajpath.2014.06.025>)

Despite a high incidence,^{1,2} only a fraction of men diagnosed with prostate cancer (PCa) develop metastases and even fewer die from the disease. Most prostate cancers remain asymptomatic and clinically indolent. The precise mechanisms for the development of progressive, clinically relevant PCa remain elusive. Furthermore, the inability to predict potential aggressiveness of PCa has resulted in

Supported by NIH grants RO1 CA098249 (J.-H.L.) and 1U01CA152737-01 (J.D.B.), American Cancer Society grant RSG-08-137-01-CNE (Y.P.Y.), and a University of Pittsburgh Cancer Institute grant.

Y.P.Y., Y.D., and Z.C. contributed equally to this work.

Y.P.Y., J.B.N., G.C.T., and J.-H.L. contributed equally as supervisors of this work.

Disclosures: None declared.

significant overtreatment of the disease. The dichotomous nature of PCa—a subset of life-threatening malignancies in the larger background of histological alterations lacking the clinical features implicit with that label—is a fundamental challenge in disease management.

To identify genome markers for PCa, tumor (T), adjacent normal prostate tissue (AT), and peripheral blood (B) samples were obtained from five prostate cancer patients who experienced recurrence with fast increase of prostate-specific antigen doubling time (PSADT; <4 months). In one patient, normal adjacent prostate tissue was not available (Supplemental Table S1). For whole genome sequencing, an average of 200 GB was sequenced per sample to achieve 33-fold coverage of the entire genome. Total RNA from all T and AT samples was sequenced to achieve >1333-fold (average, 400 million reads per sample) coverage per gene. Total RNA from four age-matched, entirely histologically benign prostate tissues, harvested from organ donors, was similarly sequenced as a tissue control. The sequencing data were aligned to human reference genome HG19.³ Fusion transcripts were then identified, filtered, and validated. It was our hypothesis that the presence of these fusion transcripts in the primary tumor would be associated with disease recurrence, development of metastatic disease, or prostate cancer-specific death. Therefore, the fusion transcripts were analyzed on 90 PCa samples from men with known clinical outcomes and 10 benign prostates harvested at organ donation. A prediction model for PCa recurrence and short postoperative PSADT was built. This model was then applied to 89 additional PCa samples from the University of Pittsburgh Medical Center (UPMC; Pittsburgh, PA), 30 samples from Stanford University Medical Center (Stanford, CA), and 36 samples from the University of Wisconsin Madison Medical Center (Madison, WI), with follow-up ranging from 1 to 15 years. A total of 127 of these samples are from patients who experienced PCa recurrence after radical prostatectomy, and 106 are from patients with no evidence of recurrence for at least 5 years after the surgery. The remaining 46 samples are from patients who had <5 years of follow-up and had not yet experienced biochemical recurrence. The association of fusion transcript expression with PCa recurrence was analyzed.

Materials and Methods

Tissue Samples

A total of 218 specimens of PCa, 4 matched ATs, 5 matched Bs, and 14 organ donor prostates were obtained from the University of Pittsburgh Tissue Bank in compliance with institutional regulatory guidelines (Supplemental Tables S1 and S2). Procedures of microdissection of PCa samples and DNA extraction were previously described.^{4–6} The protocols of tissue procurement and procedure were approved by the Institution Board of Review of the University of Pittsburgh. PCa samples of the Stanford University Medical

Center cohort (Supplemental Table S3) and the University of Wisconsin Madison Medical Center cohort (Supplemental Table S4) were obtained from corresponding institutional tissue banks and approved by Institutional Review Boards. Information about PSADT and time to recurrence was not available for the Wisconsin cohort.

Genome and Transcriptome Sequencing Library Preparation and Sequencing

To prepare the genomic DNA libraries, 50 ng of DNA was subjected to the tagmentation reactions using the NEXTERA DNA sample prep kit (Madison, WI) for 5 minutes at 55°C. The DNA was then amplified with adaptor and sequencing primers for nine cycles of the following procedure: 95°C for 10 seconds, 62°C for 30 seconds, and 72°C for 3 minutes. The PCR products were purified with Ampure beads (Beckman-Coulter, Brea, CA). The quality of genomic DNA libraries was then analyzed with quantitative PCR using Illumina sequencing primers (Illumina, Inc., San Diego, CA) and quantified with an Agilent 2000 bioanalyzer. For transcriptome sequencing, total RNA was extracted from prostate samples using TRIzol, and treated with DNaseI (Life Technologies, Grand Island, NY). Ribosomal RNA was then removed from the samples using a RIBO-Zero Magnetic kit (Epicentre, Madison, WI). The RNA was reverse transcribed to cDNA and amplified using TruSeq RNA Sample Prep Kit version 2 from Illumina, Inc. The library preparation process, such as adenylation, ligation, and amplification, was performed following the manual provided by the manufacturer. The quantity and quality of the libraries were assessed as those described above. The procedure of 200-cycle paired-end sequencing in Illumina HiSeq2000 (Illumina, Inc.) followed the manufacturer's manual.

Read Alignment

Whole genome DNA sequencing reads from five PCas, four ATs, and five Bs were aligned by Burrows-Wheeler Aligner³ version 1.4.1 against the University of California, Santa Cruz, hg19 human reference genome, allowing a maximal two-base mismatches per (100-nucleotide) read. After alignment, the average coverage of whole genome was >30× for all 14 samples. Picard tool (<http://picard.sourceforge.net>, last accessed October 11, 2013) was applied to remove duplicate reads after the alignment. RNA-seq reads (from five Ts, four matched ATs, and four organ donor prostate samples) were at an average of 1333× coverage. A maximum of two mismatches per read was allowed.

Fusion Transcript Detection

To identify fusion transcript events, we applied the Fusioncatcher version 0.97 algorithm⁷ to the RNA-seq samples. Embedded in Fusioncatcher, BOWTIE and BLAT were used to align sequences to the reference genome. The preliminary

Table 1 Primer Sequences for RT-PCR

Gene(s)	Forward	Reverse
<i>TMEM135-CCDC67</i>	5'-TTGGCATGATAGACCAGTCCC-3'	5'-CAGCACCAAGGGAATGTGTAG-3'
<i>Mtor-TP53BP1</i>	5'-TTGGCATGATAGACCAGTCCC-3'	5'-CAGCACCAAGGGAATGTGTAG-3'
<i>TRMT11-GRIK2</i>	5'-GCGCTGTCGTGTACCCTAAC-3'	5'-GGTAAGGGTAGTATTGGGTAGC-3'
<i>CCNH-C5orf30</i>	5'-CCAGGGCTGGAATTACTATGG-3'	5'-AAGCACCAGTCTGCACAATCC-3'
<i>SLC45A2-AMACR</i>	5'-TTGATGTCTGCTCCCATCAGG-3'	5'-TGATATCGTGGCCAGCTAACC-3'
<i>KDM4B-AC011523.2</i>	5'-AACACGCCCTACCTGTACTTC-3'	5'-CTGAGCAAAGACAGCAACACC-3'
<i>MAN2A1-FER</i>	5'-TGGAAGTTCAAGTCAGCGCAG-3'	5'-GCTGTCTTTGTGTGCAAACCTCC-3'
<i>LRRC59-FLJ60017</i>	5'-GTGACTGCTTGGATGAGAAGC-3'	5'-CCAGCATGCAGCTTTTCTGAG-3'
<i>TMPRSS2-ERG</i>	5'-AGTAGGCGCGAGCTAAGCAGG-3'	5'-GGGACAGTCTGAATCATGTCC-3'
<i>ACTB</i> (β -Actin)	5'-TCAAGATCATTTGCTCCTCCTGAGC-3'	5'-TGCTGTACCTTCACCGTTCCAGT-3'

list of candidate fusion transcripts is filtered in Fusioncatcher on the basis of the existing biological knowledge of the literature, including the following: i) if the genes are known to be the other's paralog in Ensembl, ii) if one of the fusion transcripts is the partner's pseudogene, iii) if one of the fusion transcripts is micro/transfer/small-nuclear RNA, iv) if the fusion transcript is known to be a false-positive event (eg, Conjoin gene database⁸), v) if it has been found in healthy samples (Illumina Body Map 2.0; <http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-513>, last accessed September 22, 2013), and vi) if the head and tail genes are overlapping with each other on the same strand. Fusion genes were visualized with CIRCOS software version 0.66 (BC Cancer Research Centre, Vancouver, British Columbia, Canada) (Supplemental Figure S1).⁹

Machine Learning Techniques to Predict Recurrence Status

Eight fusion genes from five tumor samples validated by RT-PCR, Sanger sequencing, and fluorescence *in situ* hybridization (FISH) analyses were used as features to predict nonrecurrence versus recurrence and the nature of the recurrence (PSADT, <4 versus ≥ 15 months or nonrecurrent). We applied linear discriminant analysis (LDA) to construct the prediction model. In light of relatively rare occurrence of the fusion transcripts (4.4% to 9.0%) in our 90-sample Pittsburgh training cohort, we also applied a simple prediction rule on the basis of the presence in any subset of the eight fusion genes (ie, a patient is predicted as recurrence if any fusion transcript in a designated subset exists). Leave-one-out cross validation was first applied in the 90-sample Pittsburgh training cohort to construct the model and estimate the accuracy. The model with the highest Youden index (sensitivity + specificity - 1)¹⁰ is selected from the training cohort, and is then evaluated in an 89-sample Pittsburgh test cohort, a 21-sample Stanford test cohort, and a 30-sample Wisconsin test cohort. To compare the statistical significance of the area under the curve (AUC) difference between two models, a bootstrap test is used to generate *P* values.¹¹ To compare the accuracy of two models, a test for equal proportions using prop.test in R (<http://www.r-project>) was applied.

To demonstrate the potential translational predictive value of these fusion transcripts, information on nomogram-estimated,

5-year, prostate-specific antigen (PSA)-free survival probability and Gleason scores of the patients was incorporated into our prediction models. The following models were generated: i) eight fusion transcripts alone, ii) Gleason scores alone, iii) nomogram values alone, iv) Gleason scores plus eight fusion transcripts, and v) nomogram values plus eight fusion transcripts. Complete information on prediction accuracy, sensitivity, specificity, and Youden index for these eight models is available in Supplemental Tables S2–S13.

RT-PCR and FISH

Double-stranded cDNA was synthesized as described previously.^{12,13} PCR assays were performed with the primers indicated in Table 1 using the following conditions: 94°C for 5 minutes, followed by 30 cycles of 94°C for 30 seconds, 61°C for 1 minute, and 72°C for 2 minutes. The procedures of probe preparation (Supplemental Table S5) and FISH were described previously.^{14,15}

Results

Fusion Transcripts Discovered by RNA and Whole Genome Sequencing

To identify fusion transcripts, analysis of RNA sequencing was performed on five PCa samples. A total of 76 RNA fusion events were identified using the Fusioncatcher⁷ program. Thirteen of these fusion events were confirmed by genome sequencing. To control for tissue-based normal fusion transcript events, fusion transcripts present in any of the four age-matched organ donor prostate tissues were eliminated. Furthermore, fusion transcripts with <20 kb between each element and read in the *cis*-direction were also eliminated. As a result of this filtering, 28 of 76 fusion transcript events were identified as PCa specific (Supplemental Table S6 and Supplemental Figure S1). Among these fusion events, *TMPRSS2-ERG*, the most common PCa fusion transcript,^{16–18} was found in two PCa samples. Most of the fusion events identified were novel. No fusion transcripts were identified in any of the AT samples, suggesting the somatic nature of these fusion transcripts. To validate these fusion transcripts, RT-PCR was performed using primers specific for fusion transcript regions encompassing the fusion breakpoints, and the PCR products were

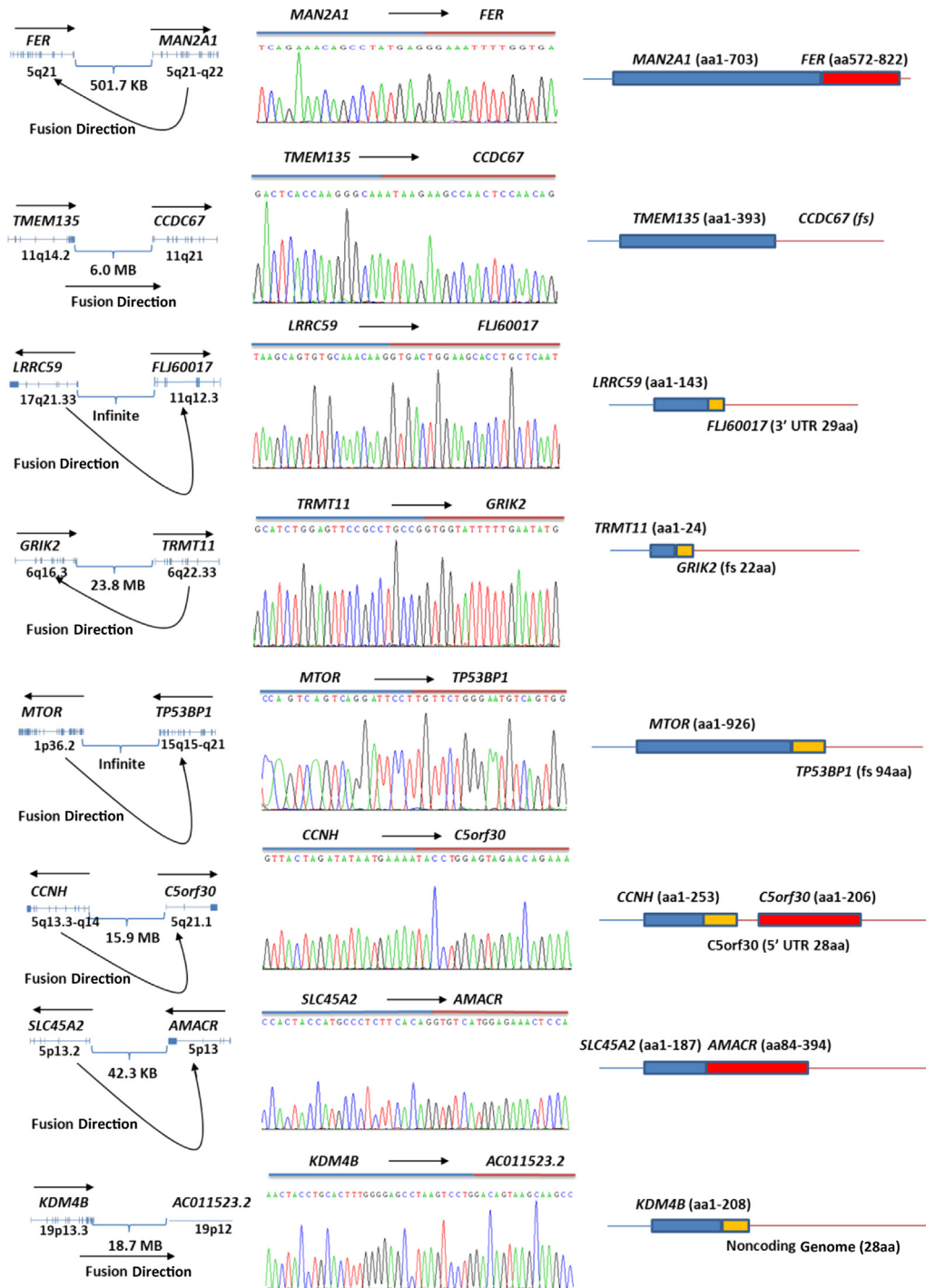


Figure 1 Unique fusion gene events. Miniature diagrams of genome of the fusion genes, the transcription directions, the distances between the joining genes, and the directions of the fusions. Representative sequencing chromatograms of fusion transcripts. The joining gene sequences are indicated. Diagrams of translation products of fusion transcripts are given. Blue head, gene translation product; red tail, gene translation product; orange, novel translation products due to frameshift or translation products from a nongene region. UTR, untranslated region.

sequenced. Eight of these fusion transcript events were validated through sequencing (Figure 1).

Five of the eight fusion events resulted in truncation of a head gene and frameshift in translation of a tail gene. One of the fusion transcripts produced a truncated cyclin H and an independent open reading frame of a novel protein whose function is not known. Two fusion events, however, are

predicted to produce chimeras that possibly retain at least partial function of both genes. For example, a fusion transcript between the N-terminus 703 amino acids of α -mannosidase 2A (*MAN2A1*) and the C-terminus 250 amino acids of FER, a feline tyrosine kinase, retains the glycoside hydrolase domain of *MAN2A1* but replaces the mannosidase domain with the tyrosine kinase domain from FER. Another

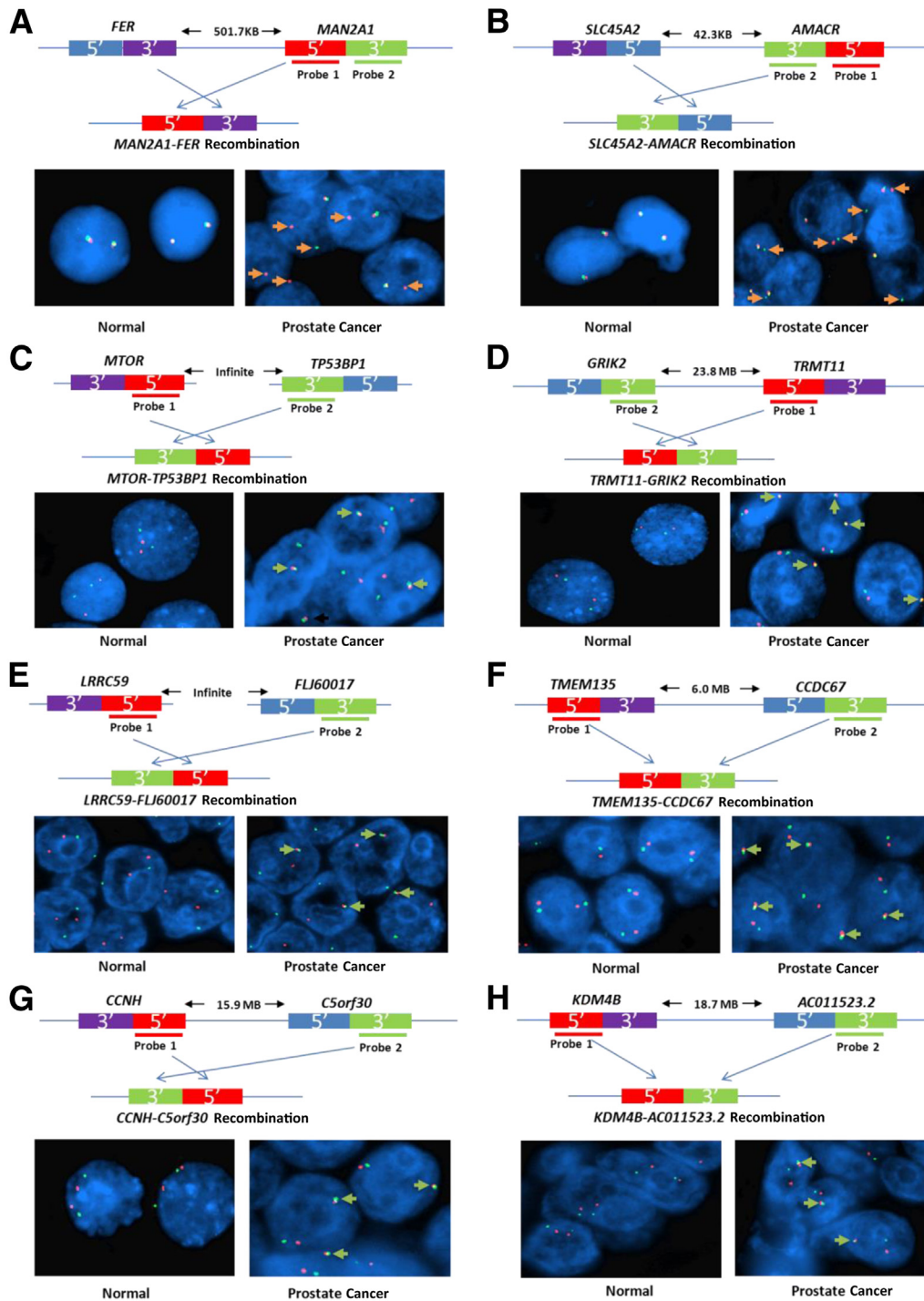


Figure 2 FISH suggests genome recombination in prostate cancer cells. Schematic diagram of *MAN2A1* and *FER* (A), *SLC45A2* and *AMACR* (B), *MTOR* and *TP53BP1* (C), *GRIK2* and *TRMT11* (D), *LRRC59* and *FLJ60017* (E), *TMEM135* and *CCDC67* (F), *CCNH* and *C5orf30* (G), and *KDM4B* and *AC011523.2* (H) genome recombination and FISH probe positions. Representative FISH images are shown for normal prostate epithelial cells and cancer cells positive for *MAN2A1-FER* (A), *SLC45A2-AMACR* (B), *MTOR-TP53BP1* (C), *TRMT11-GRIK2* (D), *LRRC59-FLJ60017* (E), *TMEM135-CCDC67* (F), *CCNH-C5orf30* (G), and *KDM4B-AC011523.2* (H) fusions. **Orange dots** denote probe 1; **green dots**, probe 2. Break-apart signals are indicated by **orange arrows**; fusion joining signals, **green arrows**.

fusion transcript couples 5 of 10 transmembrane domains of the membrane transporter protein *SLC45A2* with the methyl-acyl CoA transferase domain from *AMACR*. Interestingly, both *MAN2A1-FER* and *SLC45A2-AMACR* fusions are in the *trans*-direction, eliminating the possibility of a fusion

event from simple chromosome deletion or collapse of an extremely large RNA transcript.

The most frequent fusion events observed in PCa were *TRMT11-GRIK2* [22 (7.9%) of 279] and *SLC45A2-AMACR* [20 (7.2%) of 279] (Supplemental Figures S2-S4). *TRMT11-*

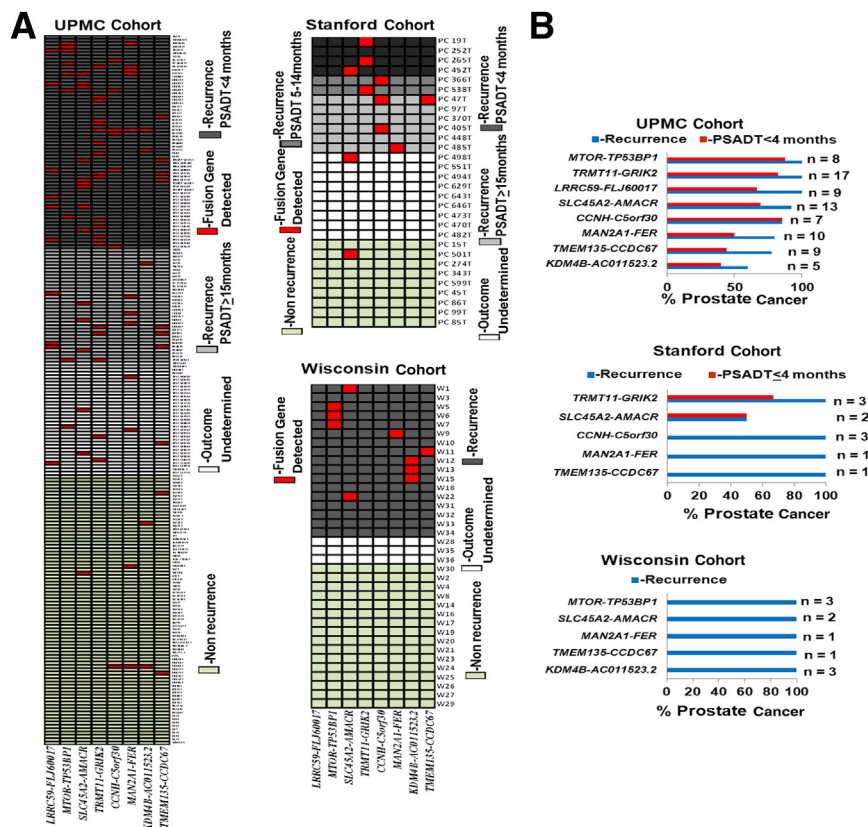


Figure 3 Fusion transcripts expressed in prostate cancer. **A:** Distribution of eight indicated fusion transcripts in 213 prostate cancer samples from UPMC, 30 samples from Stanford University Medical Center, and 36 samples from University of Wisconsin Madison Medical Center. Samples from patients who experienced recurrence were indicated with light gray (PSADT ≥ 15 months), or intermediate gray (PSADT = 5 to 14 months); samples from patients who have no recurrence for at least 5 years, green; and samples from patients whose clinical follow-up is ongoing but <5 years, white (undetermined). **B:** Correlation of fusion transcript events with prostate cancer recurrence. The percentage of prostate cancer experiencing recurrence from samples positive for fusion transcripts is plotted for each fusion transcript from specified cohorts.

GRIK2 fusion represents a giant truncation of *TRMT11*, a tRNA methyltransferase, and elimination of *GRIK2*, a glutamate receptor but reported to possess tumor-suppressor activity.¹⁹ Indeed, when *GRIK2* expression was examined in 14 *TRMT11-GRIK2*-positive PCa samples, it was undetectable, whereas it was detected in organ donor prostate samples (Supplemental Figure S5). Only 4 of 14 samples with *TRMT11-GRIK2* expressed full-length *TRMT11* transcripts. Thus, the fusion event of *TRMT11-GRIK2* likely produces a loss of function.

FISH Suggests Genome Recombination Underlying Fusion Transcript Formation

To investigate the mechanism of these fusion events, FISH was performed on PCa tissues where the fusion transcripts were present. By using the probes surrounding the *MAN2A1* breakpoint, a physical separation of signals between 5' and 3' *MAN2A1* in cancer cells containing the fusion was observed, whereas the wild-type alleles in normal prostate epithelial cells showed overlapping fluorescent signals (Figure 2). Similar break-apart hybridization occurred in *SLC45A2-AMACR*-positive PCa samples (Figure 2B). These findings indicate that *MAN2A1-FER* and *SLC45A2-AMACR* fusions are the result of chromosomal recombination events and validate the fusion transcripts found by RNA-seq. Interestingly, in PCa cells containing break-apart signals of *MAN2A1*, only 31% of the cells retained the 3' end signal, suggesting that the recombination event results in truncation of the

C-terminus of *MAN2A1* in most PCa cells. A similar collateral loss of the N-terminus of *AMACR* was found in PCa cells expressing the *SLC45A2-AMACR* fusion transcript (29% retaining the N-terminus signal of *AMACR*). Other FISH analyses confirmed that genome translocations occur in cancer cells expressing *TRMT11-GRIK2*, *MTOR-TP53BP1*, *LRRCS9-FLJ60017*, *TMEM137-CCDC67*, *CCNH-C5orf30*, and *KDM4B-AC011523.2* fusion transcripts (Figure 2, C–G). These fusion transcripts are either separated widely on a single chromosome (*TRMT11-GRIK2*, *TMEM135-CCDC67*, *CCNH-C5orf30*, and *KDM4B-AC011523.2*) or located on separate chromosomes (*MTOR-TP53BP1* and *LRRCS9-FLJ60017*). The overlapping signals of hybridizations in PCa cells offered additional validation of these fusion events. Finally, genomic breakpoints were identified in three fusion pairs through Sanger sequencing of the cancer genomic DNA (*CCNH-C5orf30*, *TMEM135-CCDC67*, and *LRRCS9-FLJ60017*) (Supplemental Figure S6).

Fusion Transcripts Associate with PCa Recurrence

To investigate the clinical and biological significance of the fusion transcripts, their presence was assessed in PCa specimens obtained from 213 men and in histologically confirmed benign prostate tissues obtained from 10 organ donors free of urological disease (aged 20 to 70 years). For 179 of the 223 PCa samples, clinical outcome data after radical prostatectomy were available, and 81 had no detectable PSA recurrence after a minimum of 5 years of follow-up, whereas 98 developed

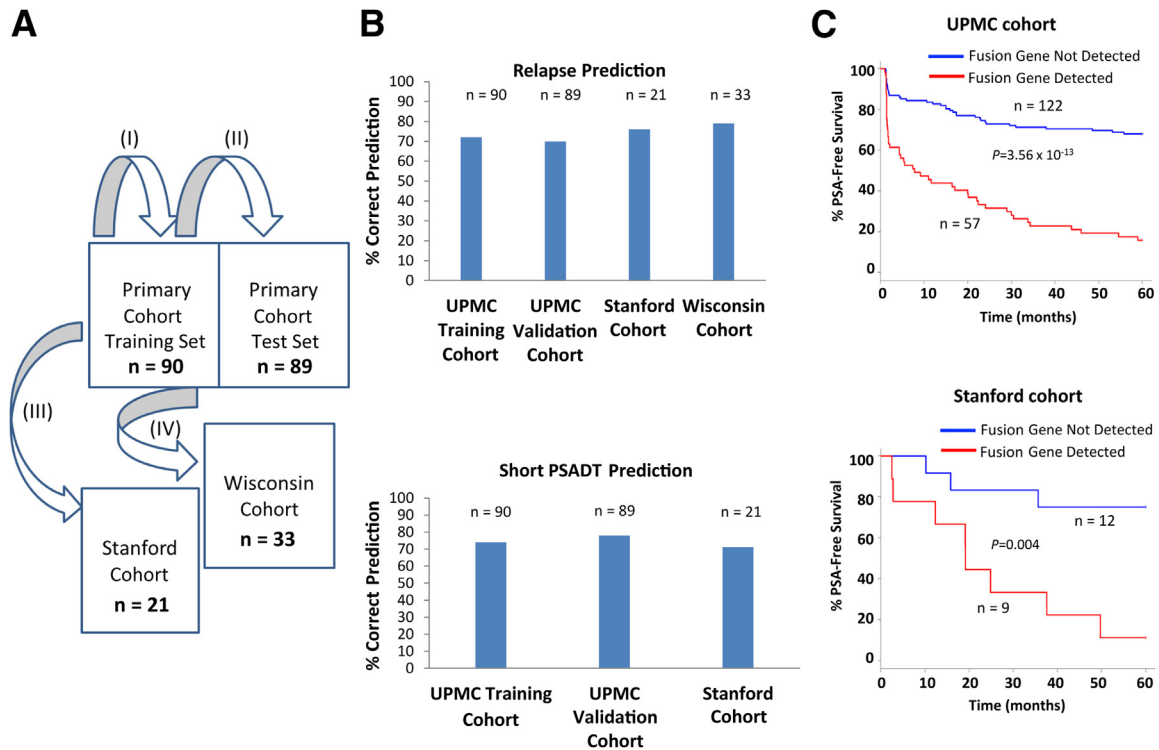


Figure 4 Fusion genes predict recurrence of prostate cancer. **A:** Scheme of training and validation steps in building fusion gene prediction models for prostate cancer recurrence and short PSA-DT. (I) The algorithm of fusion gene prediction of prostate cancer recurrence and PSA-DT <4 months was obtained from 90 randomly assigned prostate cancer samples from UPMC. (II) The algorithm was then applied to 89 samples from UPMC. (III) The algorithm was applied to 21 samples from Stanford University Medical center. (IV) The algorithm was applied to 33 samples from University of Wisconsin at Madison Medical Center. **B:** Prediction rate of prostate cancer recurrence and PSA-DT <4 months using prostate cancer sample cohorts from UPMC, Stanford Medical Center, and University of Wisconsin at Madison Medical Center, on the basis of the algorithm obtained from the 90-training sample cohort. **C:** Kaplan-Meier analysis of patients who were positive for any of *TRMT11-GRIK2*, *SLC45A2-AMACR*, *MTOR-TP53BP1*, *LRRC59-FLJ60017*, *TMEM135-CCDC67*, and *CCNH-C5orf30* versus those who were negative for these fusion events. Kaplan-Meier analysis of prostate cancer sample cohort from UPMC and Stanford University Medical Center; P value is indicated for the significant difference in survival between the group that is positive for at least one fusion transcript and the group that is negative for at least one fusion transcript.

biochemical recurrence (defined as a measurable PSA ≥ 0.2 ng/mL). In the patients without recurrence, only 7.4% (6/81) primary prostate cancers expressed one of the fusion transcripts. In contrast, 52% (51/98) primary prostate cancers expressed at least one fusion in patients who developed biochemical recurrence (Figure 3 and Supplemental Figure S2A). No fusion transcripts were detected in benign prostate tissues obtained from healthy organ donors (Supplemental Figure S2B). Three fusion events were observed exclusively in recurrent PCa after radical prostatectomy (*TRMT11-GRIK2*, *MTOR-TP53BP1*, and *LRRC59-FLJ60017*) (Figure 3, A and B).

Fisher’s exact test showed a significant difference in recurrent status between patients with at least one of the eight fusion transcripts and those without this transcript ($P = 6.8 \times 10^{-16}$). In the combined University of Pittsburgh Medical Center (UPMC), Stanford, and Wisconsin data sets, 91% (69/76) of patients positive for one of the fusion transcripts experienced prostate cancer recurrence in 5 years after prostate resection. On the basis of the hypothesis that the presence of at least one of the eight fusion transcripts would indicate a recurrence for a prostate cancer patient, a PCa prediction model was built and tested, using 90 randomly selected PCa samples from UPMC (training set). This training cohort yielded an accuracy of PCa recurrence prediction of 71%, with 89% specificity and 58%

sensitivity ($P < 0.005$) (Supplemental Figure S7A and Supplemental Table S7). When this model was applied to a separate cohort of 89 samples (test set), the model correctly predicted recurrence in 70% of patients. To further validate this model, we tested its performance in a 30-patient (21 with qualified clinical follow-up) cohort from Stanford University Medical Center and a 36-patient (30 with qualified clinical follow-up) cohort from University of Wisconsin at Madison Medical Center (Figure 3 and Supplemental Figures S4 and S5). Once again, the model correctly predicted recurrence, with 76.2% accuracy, 89% specificity, and 67% sensitivity on the PCa cohort from Stanford, and 80% accuracy, 100% specificity, and 63% sensitivity on the Wisconsin cohort (Supplemental Table S8).

In itself, recurrence does not signal an aggressive prostate cancer, because many patients with PSA recurrence do not develop metastases or die from their disease. A PSA-DT <4 months after radical prostatectomy is strongly associated with the early development of metastatic disease and prostate cancer-specific death, whereas these events are rare and remote in men with a PSA-DT >15 months.^{20,21} The presence of one or more fusion transcripts in the PCa tissue showed a strong association with PSA-DT <4 months ($P = 6 \times 10^{-9}$). To examine whether these fusion transcripts have prognostic

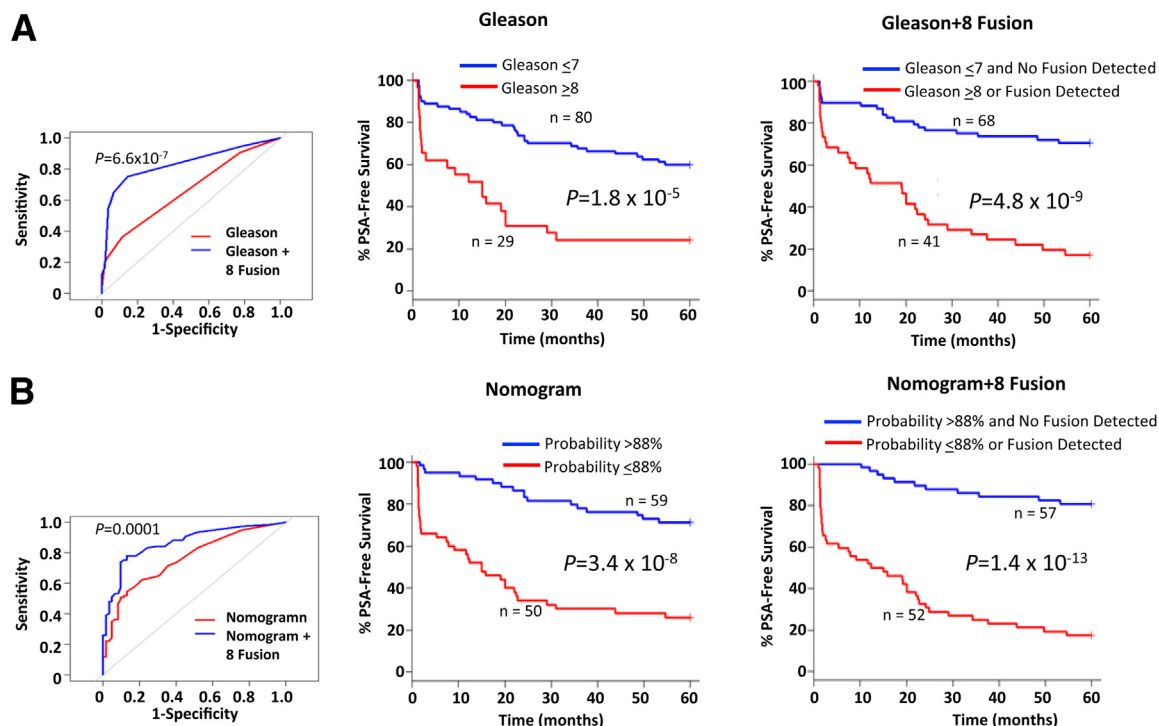


Figure 5 Combining status of fusion transcript and clinical/pathological parameter to improve prediction of prostate cancer recurrence. **A:** Combining Gleason's grading and the status of eight fusion transcripts in prostate cancer samples using LDA technique to predict the recurrence of prostate cancer. ROC curve analysis of Gleason alone or Gleason plus the presence of fusion transcripts using LDA technique in the prediction of prostate cancer recurrence; P value (permutation test) is indicated for the significant difference between the ROC curve generated by Gleason alone and the curve generated by Gleason plus the presence of fusion transcripts using LDA technique. Kaplan-Meier analysis of PSA-free survival of prostate cancer patients with Gleason score ≥ 8 versus < 8 from combined UPMC testing, Wisconsin, and Stanford data sets; P value (log-rank test) is indicated for the significant difference in survival between the group that has Gleason score ≥ 8 and the group that has score < 8 . Kaplan-Meier analysis of PSA-free survival of prostate cancer patients with Gleason score ≥ 8 or positive for any of the eight fusion transcripts in the prostate cancer samples versus those with Gleason score < 8 and negative for fusion transcripts using LDA from combined UPMC testing, Wisconsin, and Stanford data sets. P value (log-rank test) is indicated for the significant difference in survival between the group that is positive for at least one fusion transcript or has Gleason score ≥ 8 and the group that is negative for fusion transcript and has Gleason score < 8 . **B:** Combining nomogram and the status of eight fusion transcripts in prostate cancer samples using LDA technique to predict the recurrence of prostate cancer. ROC analysis of nomogram alone or nomogram plus the presence of fusion transcripts using LDA technique in the prediction of prostate cancer recurrence. P value (permutation test) is indicated for the significant difference between the ROC curve generated by nomogram alone and the curve generated by nomogram plus the presence of fusion transcripts using LDA technique. Kaplan-Meier analysis of PSA-free survival of prostate cancer patients with probability $>88\%$ versus $\leq 88\%$ from combined UPMC testing, Wisconsin, and Stanford data sets; P value (log-rank test) is indicated for the significant difference in survival between the group that has probability $>88\%$ PSA-free survival and the group that has $\leq 88\%$ probability. Kaplan-Meier analysis of PSA-free survival of prostate cancer patients with nomogram $\leq 88\%$ or positive for any of the eight fusion transcripts in the prostate cancer samples versus those $>88\%$ and negative for fusion transcripts using LDA from combined UPMC testing, Wisconsin, and Stanford data sets. P value (log-rank test) is indicated for the significant difference in survival between the group that is negative for fusion transcript and has probability $>88\%$ PSA-free survival and the group that is positive for fusion transcript or has $\leq 88\%$ probability.

value for PCa clinical outcome, a prediction model was built using the optimized weight of each fusion transcript calculated by LDA classifier on the basis of the P value of each fusion transcript with short PSADT. The panel of eight fusion transcripts correctly predicted 74.4% for PSA doubling time in the 90-sample training cohort (Supplemental Figure S6). When the same algorithm was applied to a separate 89-sample test set from UPMC and a 21-sample cohort from Stanford University Medical Center, the prediction rate for PSADT ≤ 4 months was found to be 78% and 71%, respectively (Figure 4B). To examine the impact of fusion transcripts on patients' PSA-free survival, a Kaplan-Meier analysis was performed on the PCa cohort from University of Pittsburgh. Patients (84.2%) had an observed disease recurrence within 5 years of radical prostatectomy if they carried any of the eight fusion transcripts (Figure 4C). No patient survived 5 years

without recurrence if his or her primary PCa contained a *TRMT11-GRIK2* or *MTOR-TP53BP1* transcript fusion. In contrast, 68% of the patients were free of disease recurrence if none of the fusion transcripts was detected in their primary PCa. Similar findings were also identified in the Stanford cohort: 88.9% of the patients experienced recurrence of PCa if they carried any fusion transcript, whereas 66.7% were free of the disease recurrence if they were negative for the transcript.

Combining Detection of Fusion Transcripts and Clinical/Pathological Parameters Improves Prediction Rate of Prostate Cancer Recurrence

Prostate cancer samples with at least one fusion transcript correlate with more advanced stage of prostate cancer ($P = 0.004$), lymph node involvement status ($P = 0.005$), and

lower nomogram scores ($P = 0.0003$) (Supplemental Table S9). Gleason grading alone produces a prostate cancer recurrence prediction rate of 61.1%, with 85.7% specificity and 39.6% sensitivity in the 90-sample UPMC training cohort, when a Gleason score ≥ 8 was used as cutoff to predict prostate cancer recurrence. The Gleason model yielded prediction accuracy ranging from 57% to 60% in three separate testing cohorts (Supplemental Tables S10 and S11). However, when fusion transcript status was combined with Gleason grade ≥ 8 , improvement of prediction was found for all four cohorts: 72%, 74%, 76%, and 90% for the UPMC training, UPMC test, Stanford, and Wisconsin cohorts, respectively. The receiver operating characteristic (ROC) curve showed a significantly larger AUC (0.84 versus 0.67; $P = 6.6 \times 10^{-7}$) and a higher testing accuracy (77.7% versus 59.7%; $P = 0.0019$) (Figure 5A) when Gleason score was combined with detection of any of the eight fusion transcripts. Similarly, nomogram prediction of prostate cancer recurrence has the best accuracy of 76%, with 68.8% sensitivity and 83.3% specificity in the analysis of a 90-sample UPMC training cohort (Supplemental Table S12). When this model was applied to UPMC testing, Stanford, and Wisconsin cohorts independently, the results showed that the prediction accuracy ranged from 60% to 75% among the three cohorts (Supplemental Table S13). When nomogram was combined with the status of eight fusion transcripts using the LDA technique to build a classifier, the accuracy of prediction improves to 81% to 83% among the testing cohorts (Supplemental Table S13). The ROC curve showed an increase of AUC from 0.76 to 0.87 ($P = 0.0001$) and an improvement of accuracy from 69% to 81% ($P = 0.026$) (Figure 5B). As a result, we concluded that classifier combining nomogram and the eight fusion gene panel generated the best prediction accuracy, which outperforms each diagnostic tool alone.

Discussion

Transcriptome sequencing revealed numerous fusion RNA transcripts occurring not just in PCa but also in benign adjacent tissues and histologically normal organ donor prostate samples (Supplemental Table S14). Whether these transcripts encode functional fusion proteins is not known. Deep sequencing of the whole genome and RNA revealed a significant number of prostate cancer–specific fusion transcripts. These fusions are not detectable in either benign organ donor prostate or benign prostate tissues from PCa patients. Most of these fusion transcripts appear to express in low abundance, with only an average of 6.6 reads of these fusion transcripts detected in $>1333\times$ sequencing. Indeed, when the coverage was reduced to $600\times$ in simulation studies, only *MTOR-TP53BP1* was detected consistently. One general characteristic of these fusion transcripts is that they either have a large distance between the gene targets or are oriented in *trans*-, features that could only occur as a result of chromosome recombination events. In either scenario, the fusions must be the product of significant structural DNA rearrangements.

Although the association between the eight novel fusion transcripts and prostate cancer recurrence is striking, the biological roles of these fusion transcripts are not yet elucidated. Given the known function of the genes contributing to the fusion transcripts, their formation may have impact on several cell pathways, such as RNA stability²² (*TRMT11-GRIK2*), protein glycosylation²³ (*MAN2A1-FER*), cell cycle progression^{24–26} (*CCNH-C5orf50* and *MTOR-TP53BP1*), fibroblast growth factor nuclear import²⁷ (*LRRC59-FLJ60017*), histone demethylation²⁸ (*KDM4B-AC011523.2*), and fatty acid metabolism²⁹ (*SLC45A2-AMACR*). Many of these pathways appear to be fundamental to cell growth and survival.

Two of the fusion genes are of particular interest: *MAN2A1-FER* and *SLC45A2-AMACR*. First, *MAN2A1* is a mannosidase critical in glycosylation of proteins.²³ It is usually located in the Golgi apparatus. The truncation in *MAN2A1-FER* replaces the mannosidase domain with a tyrosine kinase domain from *FER*,³⁰ whereas it leaves the glycosyl transferase domain intact. The chimera protein likely loses the mannosidase function. The new kinase domain in *MAN2A1-FER* may confer the chimera protein a tyrosine kinase activity. Thus, the impact of this fusion gene could be profound: abnormal glycosylation and phosphorylation in hundreds of secreted or plasma membrane proteins. It may affect cell-cell interactions and signal transduction and generate a new immune response to the cancer cells. Second, *AMACR* is a racemase that catalyzes 2R stereoisomers of phytanic and pristanic acids to their S counterparts. *AMACR* is essential for β oxidation of branch fatty acid in mitochondria. *SLC45A2* is a transmembrane solute carrier known for its protective role in melanoma. *SLC45A2-AMACR* chimeric protein has five transmembrane domains of *SLC45A2* truncated and is replaced with a largely intact racemase. *SLC45A2-AMACR* also loses the mitochondria target site in *AMACR*. Presumably, the fusion protein would be located in the plasma membrane. It is of interest that most of the prostate cancer samples with *SLC45A2-AMACR* fusion proved highly aggressive. Identification of the signaling pathways of this chimeric protein may gain critical insight into the behavior of prostate cancer.

Although the prevalence of each fusion transcript in prostate cancer samples is low (ranging from 2.9% to 7.9%), up to 60% of prostate cancers that later recurred and had a short PSADT were positive for at least one of these fusion transcripts. The specificity of these fusion transcripts in predicting prostate cancer recurrence appears remarkably high, ranging from 89% to 100% among four separate prediction cohorts. There were no long-term recurrence-free survivors if the primary tumor contained *TRMT11-GRIK2*, *MTOR-TP53BP1*, or *LRRC59-FLJ60017* fusion transcripts.

To our knowledge, this is the first report showing that a set of fusion transcripts is strongly associated with PCa prognosis. This set of fusion transcripts is of significant clinical value because of its high prediction rate of prostate cancer clinical outcomes when they are positive, and their

roles in improving the existing clinical prediction models when combining these fusion transcripts with Gleason's grading or nomogram. Addition of *TMPRSS2-ERG* does not improve the prediction model (Supplemental Tables S7, S8, S11, and S13). This discovery may prove useful in clinical practice given the current limitations of serum PSA, Gleason grading, and clinical stage to accurately predict disease course. If these fusion transcripts are found to have a biological role in prostate cancer progression, this may provide new targets for therapeutic intervention.

Acknowledgment

J.-H.L. had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Supplemental Data

Supplemental material for this article can be found at <http://dx.doi.org/10.1016/j.ajpath.2014.06.025>.

References

- Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D: Global cancer statistics. *CA Cancer J Clin* 2011, 61:69–90
- Siegel R, Naishadham D, Jemal A: Cancer statistics, 2012. *CA Cancer J Clin* 2012, 62:10–29
- Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009, 25:1754–1760
- Yu YP, Ding Y, Chen R, Liao SG, Ren BG, Michalopoulos A, Michalopoulos G, Nelson J, Tseng GC, Luo JH: Whole-genome methylation sequencing reveals distinct impact of differential methylations on gene transcription in prostate cancer. *Am J Pathol* 2013, 183:1960–1970
- Luo JH, Ding Y, Chen R, Michalopoulos G, Nelson J, Tseng G, Yu YP: Genome-wide methylation analysis of prostate tissues reveals global methylation patterns of prostate cancer. *Am J Pathol* 2013, 182:2028–2036
- Yu YP, Song C, Tseng G, Ren BG, Laframboise W, Michalopoulos G, Nelson J, Luo JH: Genome abnormalities precede prostate cancer and predict clinical relapse. *Am J Pathol* 2012, 180:2240–2248
- Edgren H, Murumagi A, Kangaspeska S, Nicorici D, Hongisto V, Kleivi K, Rye IH, Nyberg S, Wolf M, Borresen-Dale AL, Kallioniemi O: Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome Biol* 2011, 12:R6
- Prakash T, Sharma VK, Adati N, Ozawa R, Kumar N, Nishida Y, Fujikake T, Takeda T, Taylor TD: Expression of conjoined genes: another mechanism for gene regulation in eukaryotes. *PLoS One* 2010, 5:e13284
- Zeng W, Fu C-W, Arisona SM, Qu H: Visualizing interchange patterns in massive movement data. *Computer Graphics Forum, Proceedings of EuroVis* 2013, 32:3, part 3, 271–280
- Youden WJ: Index for rating diagnostic tests. *Cancer* 1950, 3:32–35
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M: pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011, 12:77
- Luo JH, Yu YP, Cieply K, Lin F, DeFlavia P, Dhir R, Finkelstein S, Michalopoulos G, Becich M: Gene expression analysis of prostate cancers. *Mol Carcinog* 2002, 33:25–35
- Yu YP, Landsittel D, Jing L, Nelson J, Ren B, Liu L, McDonald C, Thomas R, Dhir R, Finkelstein S, Michalopoulos G, Becich M, Luo JH: Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. *J Clin Oncol* 2004, 22:2790–2799
- Ren B, Yu G, Tseng GC, Cieply K, Gavel T, Nelson J, Michalopoulos G, Yu YP, Luo JH: MCM7 amplification and overexpression are associated with prostate cancer progression. *Oncogene* 2006, 25:1090–1098
- Yu YP, Yu G, Tseng G, Cieply K, Nelson J, Defrances M, Zarnegar R, Michalopoulos G, Luo JH: Glutathione peroxidase 3, deleted or methylated in prostate cancer, suppresses prostate cancer growth and metastasis. *Cancer Res* 2007, 67:8043–8050
- Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM: Recurrent fusion of *TMPRSS2* and *ETS* transcription factor genes in prostate cancer. *Science* 2005, 310:644–648
- Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, et al: The genomic complexity of primary human prostate cancer. *Nature* 2011, 470:214–220
- Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, et al: Punctuated evolution of prostate cancer genomes. *Cell* 2013, 153:666–677
- Sinclair PB, Sorour A, Martineau M, Harrison CJ, Mitchell WA, O'Neill E, Foroni L: A fluorescence in situ hybridization map of 6q deletions in acute lymphocytic leukemia: identification and analysis of a candidate tumor suppressor gene. *Cancer Res* 2004, 64:4089–4098
- Freedland SJ, Humphreys EB, Mangold LA, Eisenberger M, Dorey FJ, Walsh PC, Partin AW: Death in patients with recurrent prostate cancer after radical prostatectomy: prostate-specific antigen doubling time subgroups and their associated contributions to all-cause mortality. *J Clin Oncol* 2007, 25:1765–1771
- Antonarakis ES, Zahurak ML, Lin J, Keizman D, Carducci MA, Eisenberger MA: Changes in PSA kinetics predict metastasis-free survival in men with PSA-recurrent prostate cancer treated with nonhormonal agents: combined analysis of 4 phase II trials. *Cancer* 2012, 118:1533–1542
- Towns WL, Begley TJ: Transfer RNA methyltransferases and their corresponding modifications in budding yeast and humans: activities, predilections, and potential roles in human health. *DNA Cell Biol* 2012, 31:434–454
- Misago M, Liao YF, Kudo S, Eto S, Mattei MG, Moremen KW, Fukuda MN: Molecular cloning and expression of cDNAs encoding human alpha-mannosidase II and a previously unrecognized alpha-mannosidase IIx isozyme. *Proc Natl Acad Sci U S A* 1995, 92:11766–11770
- Fisher RP, Morgan DO: A novel cyclin associates with MO15/CDK7 to form the CDK-activating kinase. *Cell* 1994, 78:713–724
- Yang H, Rudge DG, Koos JD, Vaidialingam B, Yang HJ, Pavletich NP: mTOR kinase structure, mechanism and regulation. *Nature* 2013, 497:217–223
- Wang H, Luo K, Tan LZ, Ren BG, Gu LQ, Michalopoulos G, Luo JH, Yu YP: p53-Induced gene 3 mediates cell death induced by glutathione peroxidase 3. *J Biol Chem* 2012, 287:16890–16902
- Zhen Y, Sorensen V, Skjerpens CS, Haugsten EM, Jin Y, Walchli S, Olsnes S, Wiedlocha A: Nuclear import of exogenous FGF1 requires the ER-protein LRRC59 and the importins Kpnalpha1 and Kpnbeta1. *Traffic* 2012, 13:650–664
- Yang J, Jubb AM, Pike L, Buffa FM, Turley H, Baban D, Leek R, Gatter KC, Ragoussis J, Harris AL: The histone demethylase JMJD2B is regulated by estrogen receptor alpha and hypoxia, and is a key mediator of estrogen induced growth. *Cancer Res* 2010, 70:6456–6466
- Savolainen K, Kotti TJ, Schmitz W, Savolainen TI, Sormunen RT, Ilves M, Vainio SJ, Conzelmann E, Hiltunen JK: A mouse model for alpha-methylacyl-CoA racemase deficiency: adjustment of bile acid synthesis and intolerance to dietary methyl-branched lipids. *Hum Mol Genet* 2004, 13:955–965
- Krolewski JJ, Lee R, Eddy R, Shows TB, Dalla-Favera R: Identification and chromosomal mapping of new human tyrosine kinase genes. *Oncogene* 1990, 5:277–282