A variant of the *KLK4* gene is expressed as a *cis* sense–antisense chimeric transcript in prostate cancer cells

John Lai, Melanie L. Lehman, Marcel E. Dinger, et al.

Supplemental Material	http://rnajournal.cshlp.org/content/suppl/2010/03/30/rna.2019810.DC1.html
References	This article cites 40 articles, 22 of which can be accessed free at: http://rnajournal.cshlp.org/content/16/6/1156.full.html#ref-list-1
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

To subscribe to *RNA* go to:
http://rnajournal.cshlp.org/subscriptions

A variant of the *KLK4* gene is expressed as a *cis* sense–antisense chimeric transcript in prostate cancer cells

JOHN LAI,[1,2] MELANIE L. LEHMAN,[3] MARCEL E. DINGER,[4] STEPHEN C. HENDY,[3] TIM R. MERCER,[4] INGE SEIM,[2] MITCHELL G. LAWRENCE,[1,2] JOHN S. MATTICK,[4] JUDITH A. CLEMENTS,[1,2] and COLLEEN C. NELSON[1,2,3]

[1]Australian Prostate Cancer Research Centre–Queensland, Queensland University of Technology and Princess Alexandra Hospital, Woolloongabba, Queensland 4102, Australia
[2]Hormone Dependent Cancer Program, Institute of Health and Biomedical Innovation (IHBI), Queensland University of Technology, Brisbane, Queensland 4059, Australia
[3]Department of Urologic Sciences, Prostate Center, Vancouver General Hospital, University of British Columbia, Vancouver, British Columbia V6H 3Z6, Canada
[4]Institute for Molecular Bioscience, The University of Queensland, Queensland Bioscience Precinct, St. Lucia, Queensland 4072, Australia

ABSTRACT

In humans, more than 30,000 chimeric transcripts originating from 23,686 genes have been identified. The mechanisms and association of chimeric transcripts arising from chromosomal rearrangements with cancer are well established, but much remains unknown regarding the biogenesis and importance of other chimeric transcripts that arise from nongenomic alterations. Recently, a *SLC45A3–ELK4* chimera has been shown to be androgen-regulated, and is overexpressed in metastatic or high-grade prostate tumors relative to local prostate cancers. Here, we characterize the expression of a *KLK4 cis* sense–antisense chimeric transcript, and show other examples in prostate cancer. Using non-protein-coding microarray analyses, we initially identified an androgen-regulated antisense transcript within the 3′ untranslated region of the *KLK4* gene in LNCaP cells. The *KLK4 cis*-NAT was validated by strand-specific linker-mediated RT-PCR and Northern blotting. Characterization of the *KLK4 cis*-NAT by 5′ and 3′ rapid amplification of cDNA ends (RACE) revealed that this transcript forms multiple fusions with the *KLK4* sense transcript. Lack of *KLK4* antisense promoter activity using reporter assays suggests that these transcripts are unlikely to arise from a *trans*-splicing mechanism. 5′ RACE and analyses of deep sequencing data from LNCaP cells treated ±androgens revealed six high-confidence sense–antisense chimeras of which three were supported by the cDNA databases. In this study, we have shown complex gene expression at the *KLK4* locus that might be a hallmark of *cis* sense–antisense chimeric transcription.

Keywords: chimeric transcript; kallikrein 4; prostate cancer; noncoding RNA; RNA sequencing; androgens

INTRODUCTION

The Kallikrein-related peptidase 4 (*KLK4*) gene comprises five coding exons and was concomitantly cloned by three laboratories (Nelson et al. 1999; Stephenson et al. 1999; Yousef et al. 1999; Hu et al. 2000). *KLK4* is located on 19q13.4, where it is colocalized with 14 other members of the kallikrein-related peptidase gene family, which includes the *KLK3/PSA* gene (Harvey et al. 2000; Yousef et al. 2000). *KLK4* gene expression is predominantly localized to the prostate, although modest-to-low levels have been detected in other tissues including but not limited to the breast, ovary, endometrium, salivary gland, lung, adrenal gland, colon, trachea, brain, testis, spinal cord, thyroid, skin, and kidney (Nelson et al. 1999; Yousef et al. 1999; Shaw and Diamandis 2007).

The known biological role of KLK4 is derived from porcine studies that show that KLK4 is important in degradation of enamel matrix proteins during tooth maturation (Hu et al. 2007). The role of KLK4 in the prostate is less well defined, although studies have shown higher levels of KLK4 in malignant prostate cells compared to benign cells (Veveris-Lowe et al. 2005; Klokk et al. 2007). Furthermore, prostate cancer cells (PC-3 and DU145) transfected with *KLK4* have increased cellular migration, proliferation,

and colony formation (Veveris-Lowe et al. 2005; Klokk et al. 2007). KLK4 is also thought to be important in mediating migration of prostate cancer cells toward bone, a common site of prostate cancer metastasis (Gao et al. 2007). The role of KLK4 in prostate biology is thought to be mediated in part through activation of the PAR-1 and PAR-2 signaling pathways (Ramsay et al. 2008), although the precise mechanisms and importance in prostate cancer remain to be identified.

In prostate cancer cells, *KLK4* is up-regulated by the androgen receptor (*AR*) signaling axis (Nelson et al. 1999; Yousef et al. 1999; Korkmaz et al. 2001; Lai et al. 2009), which is an important pathway in prostate cancer progression. *KLK4* transcripts using alternative transcription start sites (TSSs) (Hu et al. 2000; Korkmaz et al. 2001; Lai et al. 2009) and other insertion/deletion variants have been described (Dong et al. 2001, 2005; Korkmaz et al. 2001; Kurlender et al. 2005).

Transcription studies have identified many chimeric transcripts (Kapranov et al. 2005; Akiva et al. 2006; Parra et al. 2006; Denoeud et al. 2007), suggesting that transcription and splicing events are more complex than previously recognized. Although the mechanisms and association of chimeric transcripts arising from chromosomal rearrangements with cancer are well established (Tomlins et al. 2005; Soda et al. 2007), much remains to be known regarding the biogenesis and importance of other chimeric transcripts that arise from nongenomic alterations. In humans, 31,005 nonchromosomal rearrangement chimeric transcripts originating from 23,686 genes (49% of the genome) have been identified, suggesting that chimeric gene expression occurs on a global scale (Li et al. 2009). Although few chimeric transcripts have been well characterized, an *SLC45A3–ELK4* chimera has been shown to be androgen-regulated in prostate cells and is overexpressed in metastatic or high-grade prostate tumors relative to benign prostate cancer cells (Maher et al. 2009; Rickman et al. 2009).

Currently, three models have been proposed for chimeric transcripts that arise from events other than genomic rearrangements. The *trans*-splicing model proposes that independently transcribed RNAs are spliced together using the same machinery as *cis*-splicing due to the presence of GU-AG motifs at the junctions of the two chimeras (Horiuchi and Aigaki 2006). The transcriptional slippage model suggests that transcribed RNAs dissociate from the RNA polymerase II (Pol II) complex, then hybridizes with another DNA template at a different locus using short homologous sequences, where transcription resumes (Li et al. 2009). The third model is described by the RNA-dependent RNA polymerase (RdRP) activity of human telomerase reverse transcriptase (TERT) transcribing the antisense RNA using the sense RNA as template (Maida et al. 2009).

In this study, we initially identified a novel androgen-regulated *KLK4* antisense transcript by high-throughput microarray analysis designed to detect novel androgen-regulated transcription. Sequencing of this transcript revealed

that this *KLK4* antisense transcript forms a chimera with the sense transcript. Data from this study suggest that the transcriptional slippage mechanism might mediate formation of the *KLK4* sense–antisense chimera due to the lack of *KLK4* antisense promoter activity, and the imperfect nature of the sense–antisense chimeras. We also show other examples of sense–antisense chimeras in a human LNCaP RNA-seq data set, and by 5′ rapid amplification of cDNA ends (RACE) using unidirectional PCR to target *cis* sense–antisense chimeras.

RESULTS

Identification and validation of the *KLK4* antisense transcript

Using the Agilent 244K microarray, a transcript located antisense to the *KLK4* 3′ untranslated region (UTR) (CGH probe A_14_P104372) was found to be 3.4-fold up-regulated in synthetic androgen (1 nM R1881) treated cells at 48 h relative to vehicle control–treated cells (Fig. 1A). To overcome the nonspecificity of RT-PCR in validating the *KLK4* antisense transcript due to RNA self-priming in reverse
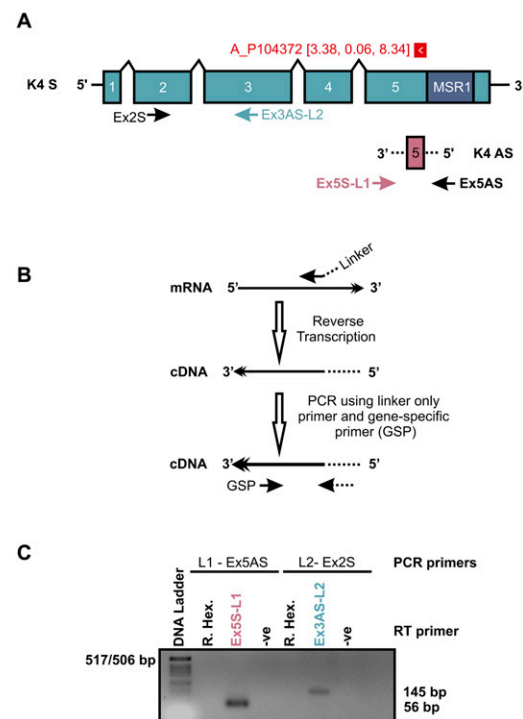

FIGURE 1. Validation of the *KLK4* antisense transcript. (*A*) A microarray probe (red box with white arrowhead) detected a 3.38-fold androgen up-regulated *KLK4* antisense transcript (pink box) in LNCaP prostate cancer cells. Primers used in subsequent figures are shown relative to the classical *KLK4* sense transcript, which is encoded by five exons (blue boxes). Dotted lines represent unknown antisense sequences. (*B*) Diagram of the RT-PCR approach taken to differentiate overlapping sense antisense transcripts. (*C*) Validation of the *KLK4* sense and antisense transcripts using strand-specific primers conjugated to synthetic linker sequences. PCR was carried out at 30 cycles using linker 1 (L1) and linker 2 (L2) primers, and the specified *KLK4* primer.

transcription (Supplemental Fig. 1A) and the co-localization of both sense and antisense transcripts to the same genomic locus, a synthetic linker sequence conjugated to the strand-specific primer was used in the RT reaction. The strand of the RNA can then be determined as products only amplify from cDNA synthesized from strand-specific linker primers due to the use of a linker-only primer, and a gene-specific primer in subsequent PCRs (Fig. 1B). Consequently, cDNA generated from random oligonucleotides or RNA self-priming cannot be amplified in PCR as these cDNAs lack linker sequences for PCR. Using this approach, both the *KLK4* sense and antisense transcripts were detected in 1 nM R1881-treated LNCaP cells (Fig. 1C). The strand specificity of this approach is shown by the amplification of both *KLK4* sense and antisense transcripts from cDNA synthesized from strand-specific linker primers, and not from cDNA synthesized using random hexameric oligonucleotides (Fig. 1C; Supplemental Fig. 1B).

The 5′ end of the *KLK4* antisense transcript forms a chimera with the *KLK4* sense transcript

5′ RNA ligase–mediated (RLM)-RACE was carried out to identify the 5′ end of the *KLK4* antisense transcript.

Initially, two bands of 608 bp and 541 bp were detected by PCR from RLM-RACE cDNA using an internal primer within the CGH microarray probe sequence that was initially used to identify the *KLK4* antisense transcript (Fig. 2A). RNA was extracted from LNCaP cells treated with R1881 for 24 h. However, the TSS of the *KLK4* chimeras was mapped to exon 2 sense (Fig. 2B). As shown in Figure 2B and Supplemental Figure 2A, the point of fusion between the sense and antisense transcripts occurs prior to the end of exon 3 (sense) and within the MSR1 repeat of exon 5 (antisense). Other chimeras were also detected by 5′ RACE, including an exon 2 deleted but partial intron 2 and 3 retained variant, and an exon 3 deleted variant (Fig. 2C). The strands of the chimera were determined using the 5′-RACE adapter oligonucleotide as reference. These products are unlikely to be ligation artifacts from 5′ RACE as they were also detected in cDNA generated from nonligated RNA, and PCR using *KLK4* primers (Supplemental Fig. 2B, upper panel). Furthermore, the observation of these products in other prostate cancer cell lines (Supplemental Fig. 2B, lower panel) and the absence in genomic DNA (data not shown) suggest that these chimeras are not the result of chromosomal DNA rearrangements.
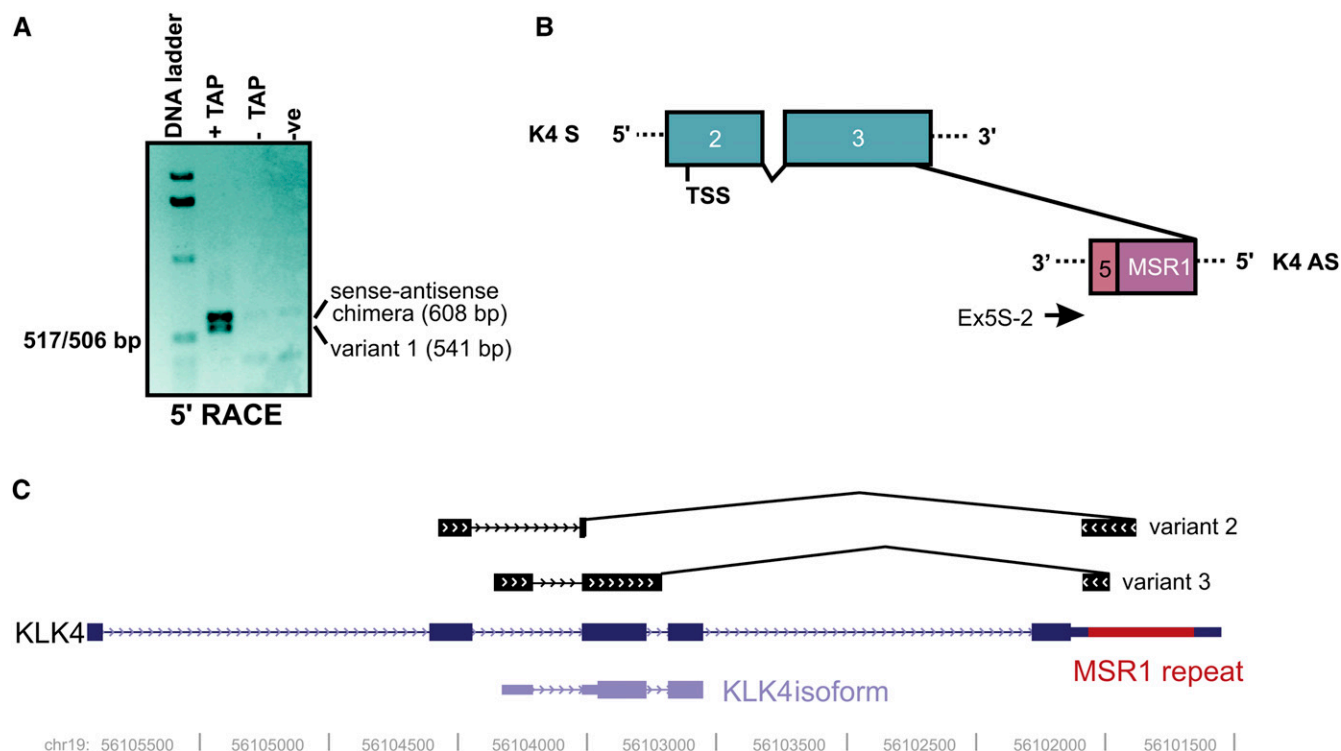

FIGURE 2. Mapping of the *KLK4* antisense transcription start site (TSS) in LNCaP cells. (*A*) Two transcripts were initially identified for the *KLK4* antisense transcript by 5′ RLM-RACE. (*B*) Exon usage of the sense–antisense chimeric transcripts identified by 5′ RLM-RACE. Also shown is the location and direction of the internal primer (Ex5S-2) used in 5′ RACE. (*C*) BLAT results of other sense–antisense variants identified by 5′ RACE.

The 3′ end of the *KLK4* antisense transcript extends to exon 2 and forms a secondary fusion point with a sense transcript

3′ RACE was carried out to identify the 3′ end of the *KLK4* antisense transcript, and PCR was performed using an internal primer sequence that is located within the CGH probe. One band of 778 bp in size was detected (Fig. 3A). The intron and exon usage of this chimeric transcript is shown in Figure 3B and Supplemental Figure 2C. The MSR1 repeat was not found in this transcript, but was detected in other 3′-RACE products using different combinations of internal and external PCR primers (Fig. 3C). Interestingly,

exon 4 was present in all 3′-RACE transcripts and not in any of the 5′-RACE transcripts, suggesting that these amplicons are not RT second-strand artifacts (Mader et al. 2001; Zaphiropoulos 2002; Cocquet et al. 2006; Roy and Irimia 2008). Furthermore, the 5′ and 3′ chimeric transcripts are unlikely to be the result of sense–antisense cDNA hybridization artifacts in the PCR as this would result in the identification of the same transcripts in 5′ and 3′ RACE. Northern blot analysis targeting the *KLK4* antisense detected an androgen-regulated transcript of ~1.5 kb, which would correspond to the size of a transcript that comprises both sequences from 5′- and 3′-RACE products (Fig. 3D). This potential transcript is shown in Figure 3E. The strands of the
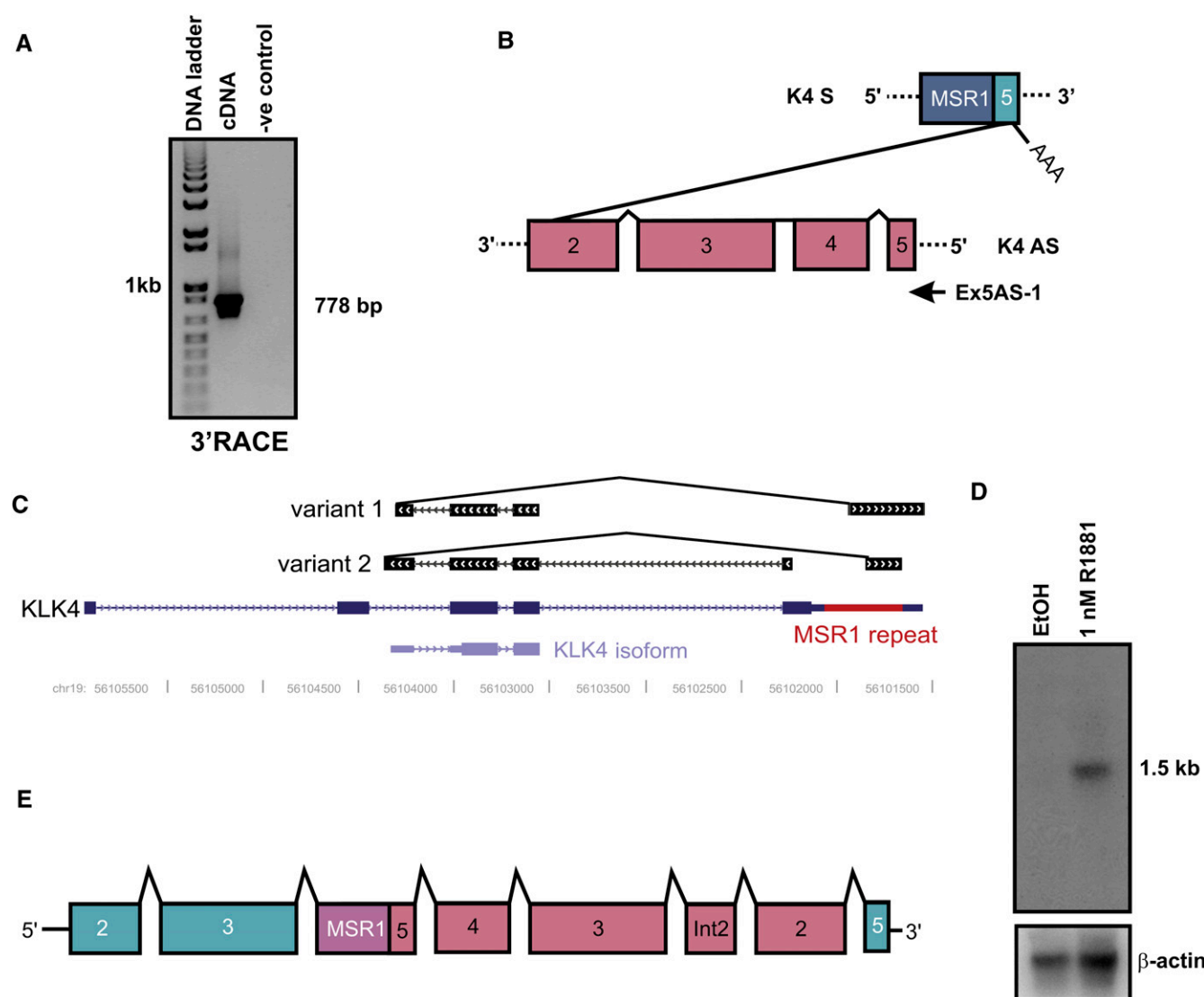

FIGURE 3. Mapping of the 3′ end of the *KLK4* antisense transcript in LNCaP cells. (*A*) One transcript of 778 bp was initially detected by 3′ RACE. (*B*) Diagram of the exon and intron usage of the 3′-RACE product. The arrow represents the location and direction of the internal *KLK4* primer (Ex5AS-1) used in 3′ RACE. (*D*) BLAT results of other sense–antisense chimeras identified in 3′ RACE. (*E*) Northern blot targeting the antisense transcript detects an ~1.5-kb transcript in cells treated with androgens (R1881) but not in mock-treated cells (ethanol). (*F*) Diagram showing the potential exon (numbered boxes) and intron (Int2) usage of a transcript comprising both 5′- and 3′-RACE sequences.

3′-RACE chimeras were determined using the genomic position of the oligo(dT) primers as reference.

The *KLK4* locus is transcriptionally complex and encodes other *trans*-chimeras

3′ RACE using primers targeting exon 2 sense that would detect the classical *KLK4* transcript were used to further understand gene transcription at the *KLK4* locus. Of the three more readily detectable transcripts, one represented the classical *KLK4* transcript (Fig. 4, upper panel), one was a premature truncated variant using a polyadenylation signal in intron 4 (Fig. 4, middle panel), and the third was a *KLK4–KLP1* chimera (Fig. 4, lower panel).

The *KLK4* antisense transcript extends past the MSR1 repeat, but the antisense promoter is transcriptionally inactive

PCR using linker-primed cDNA revealed that the *KLK4* sense transcript extends up to 760 bp downstream from the MSR1 repeat in LNCaP cells, and that the antisense transcript starts within −234 bp to −759 bp after the MSR1 repeat (Supplemental Fig. 3). The extension of the *KLK4* antisense transcript past the MSR1 repeat suggested that the antisense transcript may be independently transcribed, and that the antisense promoter might be transcriptionally active. In silico analyses of the *KLK4* antisense promoter showed a high concentration of retrotransposon elements (LINEs and SINEs), as well as a cluster of

predicted Sp1 and TSSs between −1883 bp and −2285 bp from the potential end of the antisense transcript (Fig. 5A). As such, three promoter deletion constructs for the *KLK4* antisense promoter were synthesized and tested for transcriptional activity in 22Rv1 cells (Fig. 5A). A construct (K4.2875-Luc) encompassing 2875 bp of promoter for the sense–antisense chimeras identified in 5′ RACE was also used to compare promoter activities. The *PSA* promoter and enhancer (PSA-5.8-A-Luc) was used as a positive control. Figure 5B shows that the *KLK4* antisense promoter spanning −3357 bp upstream of the potential start site of the antisense transcript is transcriptionally inactive. The *KLK4* sense promoter for the 5′-RACE transcript was transcriptionally active (Fig. 4B).

Analysis of other MSR1 genes and deep sequencing tags identifies various loci that produce sense–antisense chimeras

We hypothesized that the MSR1 repeat is important in mediating expression of the *KLK4* chimera, as this is the point of fusion between the sense and antisense transcripts. Consequently, 5′ RACE was carried out using primers that would detect sense–antisense chimeric transcripts for prostate expressed genes that harbor the MSR1 repeat. Of the five genes tested (*KLK3/PSA*, *KLK14*, *OSCAR*, *ATPBD3*, *ISOC2*), a chimera was found for the *ISOC2* gene (Fig. 6A). We further investigated the prevalence of sense–antisense chimeric transcripts by analyzing recently published strand-specific RNA-seq data from prostate cancer cells (LNCaP) after DHT
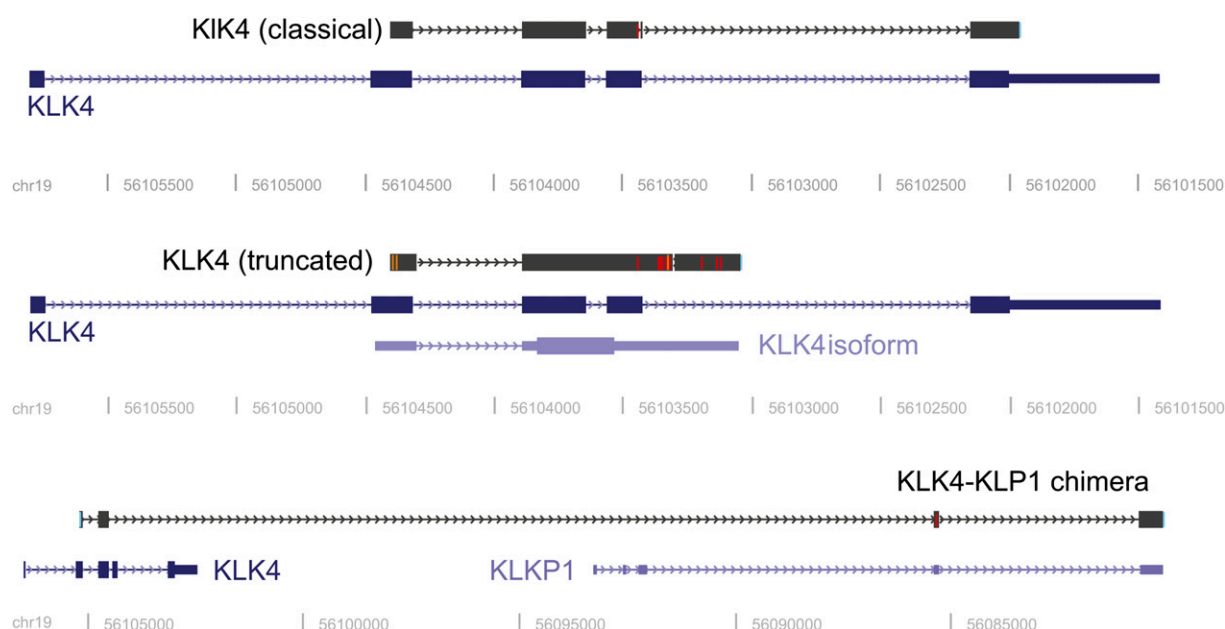

FIGURE 4. Further analysis of gene transcription at the *KLK4* locus. Shown are the BLAT results from 3′ RACE using primers targeting the sense transcript. The three more readily detected transcripts include the classical *KLK4* transcript, a premature intron 4 truncated transcript that is already in the cDNA databases, and a *KLK4–KLP1* chimera.
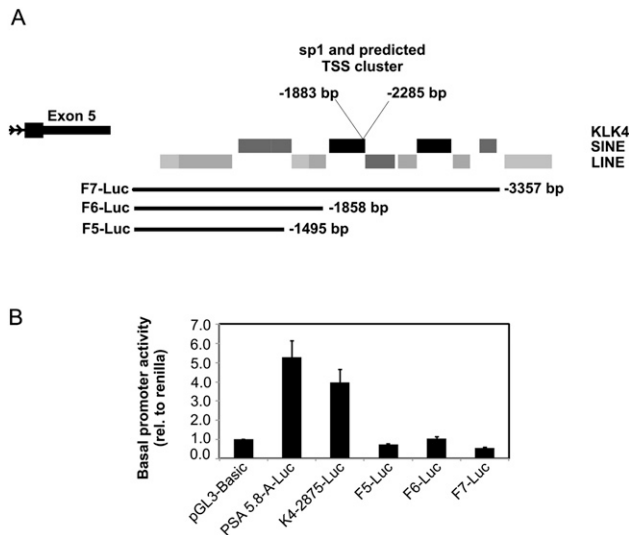
FIGURE 5. Assessing for basal promoter activity of the *KLK4* antisense transcript for evidence of independent *KLK4* antisense transcription. (*A*) Diagram showing retrotransposon elements (SINE and LINE) and a cluster of predicted sp1 and transcription start sites (TSS) located within the putative antisense promoter. Also shown are the locations of the three inserts used in luciferase reporter assays. The nucleic acid positions are relative to the reverse primer used to generate the promoter constructs. (*B*) Assessing basal promoter activity of the antisense promoter using reporter assays. Data are represented as SEM from three independent experiments that were each carried out in quadruplicate. Data were normalized to pGL3-Basic activity.

or mock treatment (Li et al. 2008). These strand-specific tags have been stringently filtered to omit artifacts of strand switching by reverse transcriptase during cDNA library preparation (Perocchi et al. 2007) and therefore provide an amenable data set for identifying chimeric transcripts.

The frequency distribution of tags that mapped to chimeric junctions increased as a function of distance to the junction (Fig. 6B). The increased mapping frequency relative to the exon junction is likely to reflect the increasing likelihood of tags mapping by chance; therefore, we omitted any tags that did not overlap either exon by at least 8 nt. This minimum threshold was empirically determined from a background of random interchromosomal splice junctions (see Materials and Methods). Using this approach, we identified six novel chimeric sense–antisense splice events of which three were supported by chimeric cDNA transcripts (Fig. 6C). Given the stringent requirements imposed by tag size (33-mers), sequence depth, exact mapping requirements, and restriction to well-annotated canonical splice junctions, it is likely that these events represent a small fraction of sense–antisense chimeric events.

DISCUSSION

In this study, we report the expression of a *KLK4* antisense transcript that forms chimeras with the *KLK4* sense

transcript in prostate cancer cells. The *KLK4* antisense transcript was initially identified in androgen-treated prostate cancer cells (LNCaP) using CGH arrays to detect novel transcription randomly across the human genome. The CGH probe detecting the *KLK4* antisense transcript was adjacent to an MSR1 repetitive sequence located within exon 5. We have validated the *KLK4* antisense transcript using Northern blotting and a linker-mediated RT-PCR approach (Moseley et al. 2006) due to the confounding effects of RNA self-priming in RT-PCR (Haddad et al. 2007).

We then carried out 5′ RACE to characterize the TSS, as known sequence data for the *KLK4* antisense transcript was limited to 53 bp from the CGH probe. Results from the 5′-RACE experiments and subsequent unidirectional PCRs show that the *KLK4* antisense transcript forms a non-genomic rearrangement chimera with *KLK4* sense. The *KLK4* chimera starts in exon 2 sense, similar to previous observations reporting that the exon 2 TSSs are more active for the classical *KLK4* sense transcript (Nelson et al. 1999; Korkmaz et al. 2001; Lai et al. 2009). Sequencing results from 3′-RACE products revealed that the 3′ end of the *KLK4* antisense transcript formed a secondary fusion point with the classical *KLK4* sense transcript.

We suggest that the 5′- and 3′-RACE chimeras form one transcript, as there were overlaps between the primers used in 5′ and 3′ RACE. Indeed, our Northern blot analyses targeting the *KLK4* antisense transcript identified an androgen-regulated transcript of ~1.5 kb that would correlate with a transcript comprising both the 5′- and 3′-RACE transcripts. Furthermore, a recent genome-wide study of chimeric transcripts revealed that 2% of chimeric transcripts arise from multiple fusions (Li et al. 2009). The transcript identified in the Northern blot was lowly expressed and only detected after long exposure. Of note, part of the predicted 3′ UTR (Nelson et al. 1999), including the proposed polyadenylation signal, was, in fact, detected in our 3′ RACE chimeric transcripts. Interestingly, our 3′ RACE targeting the *KLK4* sense transcript showed that the *KLK4* locus is transcriptionally complex and produces a truncated transcript that utilizes a polyadenylation signal within intron 4, as well as a *trans*-chimeric transcript with the next *KLK* gene (*KLIP1*). It is tempting to speculate whether such complexities at particular loci are reflective of sense–antisense transcription.

Our identification of a sense–antisense transcript for the *ISOC2* gene that also harbors the MSR1 repeat suggests that MSR1 might be important in mediating formation of *cis* sense–antisense transcripts. Furthermore, our detection of six high-confidence sense–antisense deep sequencing tags, three of which are confirmed by cDNA databases, from LNCaP cells suggest that the sense–antisense event at the *KLK4* locus is not a unique phenomenon.

Currently, there are three models that describe the formation of chimeric transcripts. However, both the *trans*-splicing
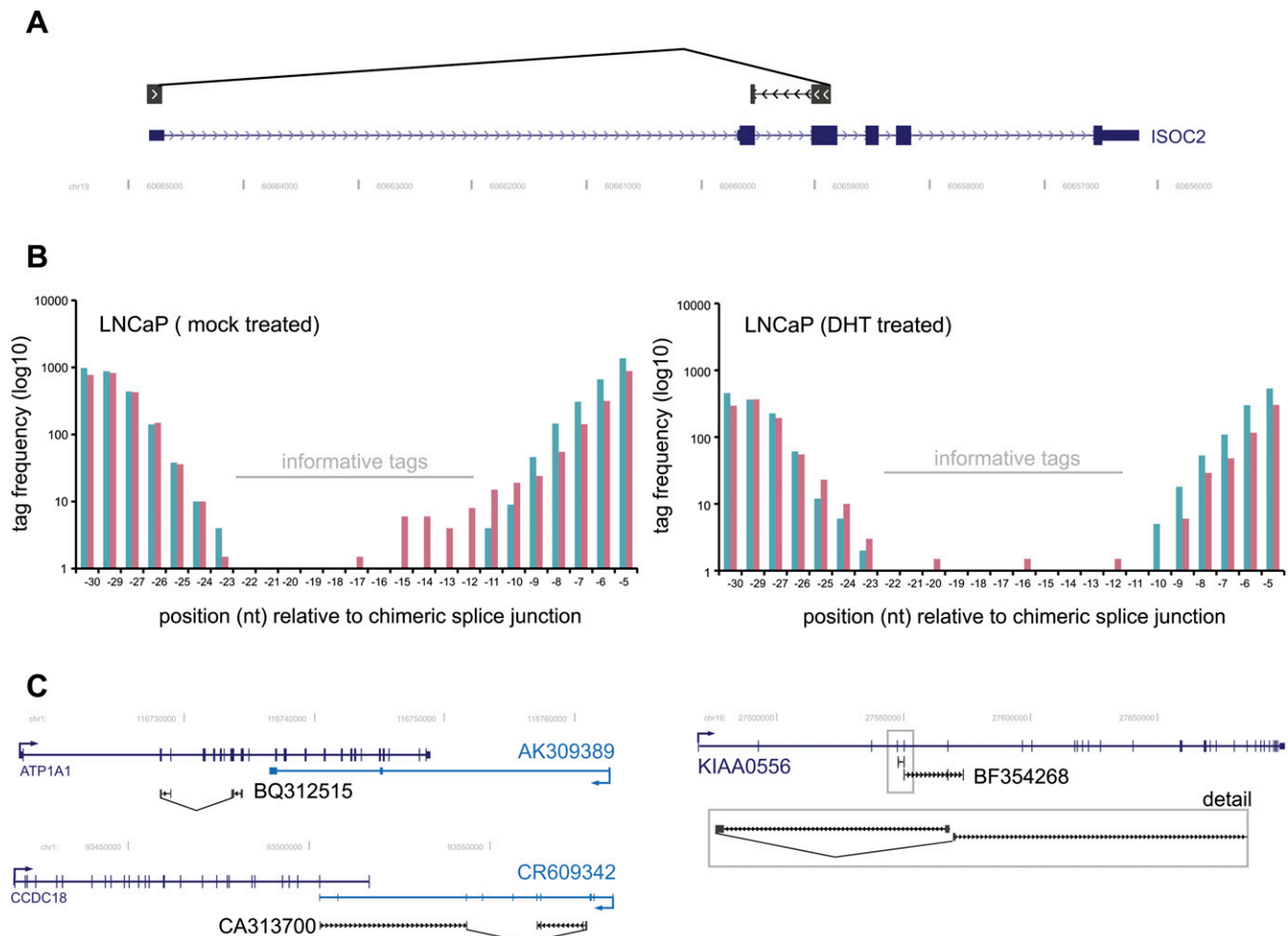
FIGURE 6. Genome-wide identification of sense–antisense chimeric splice events. (*A*) Identification of a *cis* sense–antisense transcript in another MSR1 gene (*ISOC2*) gene by 5′ RACE. (*B*) Frequency distribution of LNCaP mock- (*left*) and DHT- (*right*) treated RNA-seq tags mapping across sense–antisense chimeric exon–exon junctions. (Red bars) Frequencies of sense–antisense chimeric exon–exon junctions (RefSeq exon paired with antisense mRNA/EST exon within 10 kb); (blue bars) frequencies of background chimeric exon–exon junction mappings (RefSeq exon paired with random mRNA/EST exon). Informative tags (indicated) were defined within regions where no background mapping (i.e., mapping by chance) occurred. (*C*) Three genes (dark blue) identified from RNA-seq analysis as sharing exon–exon junctions with antisense transcripts (light blue) with corroborating cDNA transcripts (black). Transcriptional direction and splice junctions are indicated.

and RNA-dependent RNA polymerase (RdRP) activity models have limitations in describing transcripts that are similar to the *KLK4 cis* sense–antisense chimeras described in this study. The *trans*-splicing model (Horiuchi and Aigaki 2006) would require an active promoter for an independently transcribed antisense transcript, which we have failed to detect for *KLK4*. The RdRP model (Maida et al. 2009) is limited in describing the *KLK4* chimeras in that many of the transcripts are not perfect sense–antisense chimeras, which is also supportive of our hypothesis that these transcripts are not experimental artifacts. Furthermore, the RdRP model would require that the antisense transcript is first transcribed in order to provide a template for the sense sequence for the 5′-RACE products. The transcriptional slippage model (Li et al. 2009) at this stage best reflects the *cis* sense–antisense chimeras detected for *KLK4* in this study, due to the presence

of direct repeats of short homologous sequences at the sense–antisense junctions (data not shown). However, this requires that transcription of *KLK4* sense by Pol II would need to either (1) stop then continue transcription in the opposite direction using the antisense strand as template; or (2) use the antisense strand from the other chromosome 19 to continue transcription in the same direction.

There are many interesting aspects to chimeric gene transcription. Particularly, the question of whether chimeric transcripts represent normal or aberrant gene transcription is one of great importance. For example, if chimeric transcripts were the result of dysregulation of normal transcription and/or splicing processes, then it would be valuable to understand the mechanisms that mediate these events. Furthermore, global higher expression of chimeric transcripts may reflect an "unhappy cell" and consequently

be used as a diagnostic and/or prognostic marker for complex diseases such as cancer that are the result of the accumulation of many genetic and epigenetic events. Conversely, chimeric gene transcription and/or splicing may simply reflect nature's maximization of the coding potential of exons and promoters as highlighted in a recent review (Gingeras 2009).

Regardless of whether chimeric transcripts are indicative of normal physiology or pathophysiology, the implications for chimeric transcription/splicing are far reaching. For example, genome-wide single nucleotide polymorphism (SNP) association studies will be confounded by the fact that an SNP at one locus may now have a phenotypic effect on a transcript originating from a distant locus, or possibly at many loci. Second, our understanding of the factors that mediate gene transcription will be redefined, as downstream exons of chimeric transcripts may now co-opt promoters from upstream exons from a different locus. This is highlighted by the adoption of the androgen-responsive *SLC45A3* promoter by the ETS transcription factor *ELK4* (Maher et al. 2009; Rickman et al. 2009). Importantly, the *SLC45A3–ELK4* chimera is associated with high-grade and metastatic prostate cancers relative to benign prostate cancers (Maher et al. 2009; Rickman et al. 2009). Indeed, these aspects may now confer biological significance to chimeric transcripts even when they are lowly expressed. On a practical level, *cis* sense–antisense chimeras will confound results from experiments such as in situ hybridization that often use the antisense strand as a negative control.

In conclusion, we have identified a novel *KLK4* sense–antisense chimera that to our knowledge is one of few (Bartsch et al. 2004; Maida et al. 2009) well-characterized reports of endogenous *cis* sense–antisense chimeric transcripts. Open reading frame analysis (data not shown) suggests that the *KLK4* chimera is unlikely to encode a secreted and proteolytic active serine peptidase due to the loss of the pre-pro domain in this variant, and the catalytic serine residue located within exon 5. Future inhibition studies that selectively target the *KLK4* chimera will further our understanding of the biological significance of this transcript, especially given the high concentration of MSR1 on chromosome 19 and within the *KLK* locus (Das et al. 1987), and the potential of these chimeras to be processed into short double-stranded RNA. Preliminary studies of the *SLC45A3–ELK4* chimera (Maher et al. 2009; Rickman et al. 2009) show potential for other chimeric transcripts to function as a prognostic marker for prostate cancer.

MATERIALS AND METHODS

Cell culture

All cell lines used were obtained from the American Type Culture Collection (ATCC). LNCaP, 22Rv1, DU145, and PC-3 were maintained in RPMI 1640 media (Invitrogen) supplemented with

10% fetal calf serum (FCS) (Invitrogen). RWPE-1 cells were grown in keratinocyte serum-free medium with 50 μg/mL bovine pituitary extract and 5 ng/mL recombinant human epidermal growth factor (EGF) (Invitrogen). MDA-PCa-2B cells were maintained in BRFF-HPC1 media from AthenaES (Sapphire Biosciences) supplemented with 20% FCS. All media was supplemented with 50 U/mL penicillin G and 50 μg/mL streptomycin (CSL Biosciences). For steroid treatment experiments, cells were cultured until ~70% confluent, and then maintained in media containing 10% charcoal-stripped serum (HyClone) for 48 h followed by the addition of 1 nM R1881.

Microarray analyses

cDNA was prepared and end-labeled using the 3DNA Array 350 kit (Genisphere) from RNA isolated from both LNCaP cells treated with 1 nM R1881 for 48 h and an untreated ethanol vehicle control. The cDNA was hybridized to 244K Agilent CGH arrays—arrays that are normally used to detect anomalies at the DNA level but used here to detect random transcription across the genome. Agilent's two-color microarray-based gene expression protocol was followed using Agilent hybridization and wash buffers, scanner, and feature extraction software. The raw data from three independent experiments were processed with the LIMMA Bioconductor package (Smyth 2004) using quantile between-array normalization. A Bayesian-adjusted *t*-statistic from a linear model built using LIMMA was used to determine differential expression between R1881 treated and untreated samples. Benjamini-Hochberg multiple test correction was used to control for a false discovery rate of 10%.

RNA extraction, cDNA synthesis, and RT-PCR

Total RNA was extracted from cells using TRI-Reagent (Sigma Chemical Co.) according to the manufacturer's protocol and treated with DNase I (Invitrogen). Complementary DNA (cDNA) was synthesized from 2 μg of total RNA using SuperScript III Reverse Transcriptase (Invitrogen), and random hexamers (Proligo) or gene/strand-specific primers targeting the *KLK4* and *β2-microglobulin* genes (Supplemental Table 1). In some experiments, a synthetic linker sequence was conjugated to gene/strand-specific primers for use in the RT as indicated. The cDNA was treated with 1 U of RNase H (Invitrogen) for 20 min at 37°C prior to use in PCR. For nested PCR experiments, 1 μL of the first-round PCR was used as template DNA for the second-round PCR.

5′ and 3′ RNA ligase–mediated rapid amplification of cDNA ends (RLM-RACE)

RNA was extracted as per RT-PCR from cells after 24 h of treatment with 1 nM R1881 with an additional purification of the RNA through an RNeasy column (QIAGEN). An amount of 10 μg (5′) and 1 μg (3′) of total RNA was used in RLM-RACE (FirstChoice RLM-RACE kit; Ambion) as per the manufacturer's instructions, except Superscript III (instead of AMV) reverse transcriptase was used for cDNA synthesis. The 5′ RACE outer and Ex5S-1 (first-round PCR), and 5′ RACE inner and Ex5S-2 (second round PCR) primers were used for 5′-RLM-RACE. The 3′ RACE outer and Ex5AS or Ex2S (first-round PCR), and 3′ RACE inner and Ex5AS-1 or Ex2S (second-round PCR) primers

were used in 3′ RACE. All primer sequences are detailed in Supplemental Table 1. PCR amplicons were cloned into the pGEMT easy vector (Promega), and clones were sequenced at the Australia Genome Research Facility, University of Queensland, Brisbane, Australia.

Northern blot analysis

LNCaP cells were cultured in 5% charcoal/dextran stripped FBS (CSS) for 48 h followed by 48 h of treatment with either 1 nM R1881 or vehicle control (ethanol). Total RNA was extracted using the MirVana RNA Isolation kit (Ambion), and subsequent poly(A) RNA was isolated using the poly(A) purist kit (Ambion). Northern blots were carried out using standard techniques. Briefly, 10 μg of poly(A) RNA was resolved on 1% agarose-formaldehyde denaturing gels, transferred to Hybond-N membranes (GE Healthcare) by capillary action using 10× SSC (1.5 M NaCl, 0.15 M sodium citrate at pH 7.0) and cross-linked by UV irradiation. Membranes were prehybridized for 1 h at 68°C, then hybridized to an [α-^{32}P]-labeled riboprobe targeting the *KLK4* antisense 3′ UTR using the Ultrahyb hybridization buffer (Ambion) overnight at 68°C. Blots were washed according to the Ultrahyb protocol (Ambion) and imaged by either autoradiography or PhosphorImaging. Riboprobes were generated by in vitro transcription as described (Maniatis et al. 1982) using a DNA template prepared by introducing the T7 RNA polymerase promoter upstream of a 45-mer derived from the microarray probe sequence used to detect the *KLK4* antisense transcript: (5′-AACAGATATTCCTGAATTCCTTCCGCAGGATGTATTTGGGG GTCA-3′). The PCR primers (T7 KLK4 3′ UTR and T7 KLK4 L) are detailed in Supplemental Table 1. *KLK4* antisense bands were normalized to *β-Actin* by reprobing the blots using a probe generated by primer extension. The *β-Actin* labeling template was generated from LNCaP cDNA using primers AC1 and AC2 (Supplemental Table 1). *KLK4* antisense blots were then reprobed for *β-Actin* using probes generated by primer extension using primer AC2.

Genome-wide analysis of sense–antisense chimeras in prostate cancer cells

Identification of further instances of sense–antisense chimeras was performed using a recently published RNA-seq data set from prostate cancer cells (LNCaP) after DHT or mock treatment (Li et al. 2008). As well as the relevant biological context of these data to the current study, this data set was selected for analysis because, unlike many other RNA-seq data sets, contained only strand-specific tags that had been stringently filtered to omit artifacts of strand switching by reverse transcriptase during cDNA library preparation (Perocchi et al. 2007). To identify sense-chimeric transcripts within this data set, an index of all possible sense–antisense exon–exon junctions within a 10-kb region (upstream and downstream) for all mouse RefSeq genes (as at October 2009) was built. That is, for each RefSeq gene, sequences were built comprising every combination of the last 30 nt of each exon (3′ end) paired with first 30 nt (5′ end) of every exon of any antisense transcript (as defined by UCSC all mRNA and all EST tracks) within 10 kb of the gene (Supplemental Fig. 4). This library of combinations was then interrogated for exact matching tags from

the RNA-seq data set using ZOOM (Lin et al. 2008). RNA-seq tags that mapped exactly to either the genome or conventional splice junctions had been omitted from the set. To determine the minimum number of nucleotides that a tag must span across the sense–antisense exon–exon junction, a control analysis exactly as described above was performed, except that the pairing of RefSeq exons with antisense mRNA/EST exons was randomized, such that RefSeq exons could be paired with any mRNA/EST exon throughout the genome. This approach has the advantage that it provides an empirical basis upon which to determine the threshold exon–exon junction overlap, where a mapped tag represented the presence of a real transcript rather than a chance mapping event and did not require any assumptions to be made regarding variable nucleotide frequencies across the genome and sequencing error rates in the RNA-seq tags. RNA-seq tags indicative of sense–antisense chimeras could then be detected using this empirically determined threshold. Finally, to confirm the presence of these chimeric transcripts, mouse cDNA sequence databases were queried with the identified sequences of the sense–antisense exon–exon junctions.

In silico analyses of the *KLK4* antisense promoter

Retrotransposon elements were identified using the RepeatMasker track on the UCSC Genome Browser, March 2006 (hg18) assembly (http://genome.ucsc.edu). TSSs were predicted using the Neural Network Promoter Prediction software from the Berkeley *Drosophila* Genome Project (http://www.fruitfly.org/seq_tools/promoter.html). Transcriptional regulatory element clusters were predicted using the Cis-element Cluster Finder, CISTER algorithm (http://zlab.bu.edu/~mfrith/cister.shtml).

Transfection and luciferase reporter assays

LNCaP genomic DNA was used as template to generate three deletion constructs encompassing 3357 bp of the *KLK4* antisense promoter beginning at −234 bp from the MSR1 repeat. SacI and HindIII restriction sites were added to the primers to facilitate orientation-specific cloning into the pGL3-Basic vector (Promega). Forward (F5, F6, and F7) and reverse (2ASR-Hind III) primers are detailed in Supplemental Table 1. The *KLK4* promoter inserts were amplified using high-fidelity platinum *Taq* (Invitrogen) and subcloned into the pGEM-T Easy vector prior to cloning into pGL3-Basic. All constructs were sequenced as described above. *KLK4* luciferase promoter constructs are designated in the Results section according to their corresponding forward primer name.

22Rv1 cells were seeded in 24-well plates, and culture medium was changed to antibiotic-free RPMI 1640, supplemented with 10% FCS (basal media) for 48 h prior to transfection. Transient transfection of pGL3-Basic, F5-Luc, F6-Luc, and F7-Luc was carried out at equimolar amounts equivalent to 0.5 μg of K4.2875-Luc (Lai et al. 2009) using 2 μL of Lipofectamine 2000/well (Invitrogen). *Renilla* (0.3 μg) was used as an internal control for transfection efficiency. After 6 h of transfection, media was replaced with basal media for a further 24 h. Luciferase activity was measured using the Dual-Luciferase Reporter Assay System (Promega) on a PolarStar plate reader (BMG; Labtech). Data are expressed as luciferase activity normalized to *Renilla* activity and represented as the standard error of the mean (SEM) from three independent treatments that were each carried out in triplicate.

SUPPLEMENTAL MATERIAL

Supplemental material can be found at http://www.rnajournal.org.

REFERENCES

Akiva P, Toporik A, Edelheit S, Peretz Y, Diber A, Shemesh R, Novik A, Sorek R. 2006. Transcription-mediated gene fusion in the human genome. *Genome Res* **16:** 30–36.

Bartsch H, Voigtsberger S, Baumann G, Morano I, Luther HP. 2004. Detection of a novel sense–antisense RNA–hybrid structure by RACE experiments on endogenous troponin I antisense RNA. *RNA* **10:** 1215–1224.

Cocquet J, Chong A, Zhang G, Veitia RA. 2006. Reverse transcriptase template switching and false alternative transcripts. *Genomics* **88:** 127–131.

Das HK, Jackson CL, Miller DA, Leff T, Breslow JL. 1987. The human apolipoprotein C-II gene sequence contains a novel chromosome 19-specific minisatellite in its third intron. *J Biol Chem* **262:** 4787–4793.

Denoeud F, Kapranov P, Ucla C, Frankish A, Castelo R, Drenkow J, Lagarde J, Alioto T, Manzano C, Chrast J, et al. 2007. Prominent use of distal 5′ transcription start sites and discovery of a large number of additional exons in ENCODE regions. *Genome Res* **17:** 746–759.

Dong Y, Kaushal A, Bui L, Chu S, Fuller PJ, Nicklin J, Samaratunga H, Clements JA. 2001. Human kallikrein 4 (KLK4) is highly expressed in serous ovarian carcinomas. *Clin Cancer Res* **7:** 2363–2371.

Dong Y, Bui LT, Odorico DM, Tan OL, Myers SA, Samaratunga H, Gardiner RA, Clements JA. 2005. Compartmentalized expression of kallikrein 4 (KLK4/hK4) isoforms in prostate cancer: Nuclear, cytoplasmic and secreted forms. *Endocr Relat Cancer* **12:** 875–889.

Gao J, Collard RL, Bui L, Herington AC, Nicol DL, Clements JA. 2007. Kallikrein 4 is a potential mediator of cellular interactions between cancer cells and osteoblasts in metastatic prostate cancer. *Prostate* **67:** 348–360.

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

Haddad F, Qin AX, Giger JM, Guo H, Baldwin KM. 2007. Potential pitfalls in the accuracy of analysis of natural sense–antisense RNA pairs by reverse transcription-PCR. *BMC Biotechnol* **7:** 21. doi: 10.1186/1472-6750-7-21.

Harvey TJ, Hooper JD, Myers SA, Stephenson SA, Ashworth LK, Clements JA. 2000. Tissue-specific expression patterns and fine mapping of the human kallikrein (KLK) locus on proximal 19q13.4. *J Biol Chem* **275:** 37397–37406.

Horiuchi T, Aigaki T. 2006. Alternative *trans*-splicing: A novel mode of pre-mRNA processing. *Biol Cell* **98:** 135–140.

Hu JC, Zhang C, Sun X, Yang Y, Cao X, Ryu O, Simmer JP. 2000. Characterization of the mouse and human PRSS17 genes, their relationship to other serine proteases, and the expression of PRSS17 in developing mouse incisors. *Gene* **251:** 1–8.

Hu JC, Chun YH, Al Hazzazzi T, Simmer JP. 2007. Enamel formation and amelogenesis imperfecta. *Cells Tissues Organs* **186:** 78–85.

Kapranov P, Drenkow J, Cheng J, Long J, Helt G, Dike S, Gingeras TR. 2005. Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res* **15:** 987–997.

Klokk TI, Kilander A, Xi Z, Waehre H, Risberg B, Danielsen HE, Saatcioglu F. 2007. Kallikrein 4 is a proliferative factor that is overexpressed in prostate cancer. *Cancer Res* **67:** 5221–5230.

Korkmaz KS, Korkmaz CG, Pretlow TG, Saatcioglu F. 2001. Distinctly different gene structure of KLK4/KLK-L1/prostase/ARM1 compared with other members of the kallikrein family: Intracellular localization, alternative cDNA forms, and regulation by multiple hormones. *DNA Cell Biol* **20:** 435–445.

Kurlender L, Borgono C, Michael IP, Obiezu C, Elliott MB, Yousef GM, Diamandis EP. 2005. A survey of alternative transcripts of human tissue kallikrein genes. *Biochim Biophys Acta* **1755:** 1–14.

Lai J, Myers SA, Lawrence MG, Odorico DM, Clements JA. 2009. Direct progesterone receptor and indirect androgen receptor interactions with the kallikrein-related peptidase 4 gene promoter in breast and prostate cancer. *Mol Cancer Res* **7:** 129–141.

Li H, Lovci MT, Kwon YS, Rosenfeld MG, Fu XD, Yeo GW. 2008. Determination of tag density required for digital transcriptome analysis: Application to an androgen-sensitive prostate cancer model. *Proc Natl Acad Sci* **105:** 20179–20184.

Li X, Zhao L, Jiang H, Wang W. 2009. Short homologous sequences are strongly associated with the generation of chimeric RNAs in eukaryotes. *J Mol Evol* **68:** 56–65.

Lin H, Zhang Z, Zhang MQ, Ma B, Li M. 2008. ZOOM! Zillions of oligos mapped. *Bioinformatics* **24:** 2431–2437.

Mader RM, Schmidt WM, Sedivy R, Rizovski B, Braun J, Kalipciyan M, Exner M, Steger GG, Mueller MW. 2001. Reverse transcriptase template switching during reverse transcriptase-polymerase chain reaction: Artificial generation of deletions in ribonucleotide reductase mRNA. *J Lab Clin Med* **137:** 422–428.

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

Maida Y, Yasukawa M, Furuuchi M, Lassmann T, Possemato R, Okamoto N, Kasim V, Hayashizaki Y, Hahn WC, Masutomi K. 2009. An RNA-dependent RNA polymerase formed by TERT and the RMRP RNA. *Nature* **461:** 230–235.

Maniatis T, Sambrook J, Fritsch EF. 1982. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Moseley ML, Zu T, Ikeda Y, Gao W, Mosemiller AK, Daughters RS, Chen G, Weatherspoon MR, Clark HB, Ebner TJ, et al. 2006. Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nat Genet* **38:** 758–769.

Nelson PS, Gan L, Ferguson C, Moss P, Gelinas R, Hood L, Wang K. 1999. Molecular cloning and characterization of prostase, an androgen-regulated serine protease with prostate-restricted expression. *Proc Natl Acad Sci* **96:** 3114–3119.

Parra G, Reymond A, Dabbouseh N, Dermitzakis ET, Castelo R, Thomson TM, Antonarakis SE, Guigo R. 2006. Tandem chimerism as a means to increase protein complexity in the human genome. *Genome Res* **16:** 37–44.

Perocchi F, Xu Z, Clauder-Munster S, Steinmetz LM. 2007. Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Res* **35:** e128. doi: 10.1093/nar/gkm683.

Ramsay AJ, Dong Y, Hunt ML, Linn M, Samaratunga H, Clements JA, Hooper JD. 2008. Kallikrein-related peptidase 4 (KLK4) initiates intracellular signaling via protease-activated receptors (PARs). KLK4 and PAR-2 are co-expressed during prostate cancer progression. *J Biol Chem* **283:** 12293–12304.

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.

Roy SW, Irimia M. 2008. When good transcripts go bad: Artifactual RT-PCR 'splicing' and genome analysis. *Bioessays* **30:** 601–605.

Shaw JL, Diamandis EP. 2007. Distribution of 15 human kallikreins in tissues and biological fluids. *Clin Chem* **53:** 1423–1432.

Smyth GK. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3:** Article 3. doi: 10.2202/1544-6115.1027.

Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, et al. 2007. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* **448:** 561–566.

Stephenson SA, Verity K, Ashworth LK, Clements JA. 1999. Localization of a new prostate-specific antigen-related serine protease gene, KLK4, is evidence for an expanded human kallikrein gene family cluster on chromosome 19q13.3-13.4. *J Biol Chem* **274:** 23210–23214.

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.

Veveris-Lowe TL, Lawrence MG, Collard RL, Bui L, Herington AC, Nicol DL, Clements JA. 2005. Kallikrein 4 (hK4) and prostate-specific antigen (PSA) are associated with the loss of E-cadherin and an epithelial-mesenchymal transition (EMT)-like effect in prostate cancer cells. *Endocr Relat Cancer* **12:** 631–643.

Yousef GM, Obiezu CV, Luo LY, Black MH, Diamandis EP. 1999. Prostase/KLK-L1 is a new member of the human kallikrein gene family, is expressed in prostate and breast tissues, and is hormonally regulated. *Cancer Res* **59:** 4252–4256.

Yousef GM, Chang A, Scorilas A, Diamandis EP. 2000. Genomic organization of the human kallikrein gene family on chromosome 19q13.3-q13.4. *Biochem Biophys Res Commun* **276:** 125–133.

Zaphiropoulos PG. 2002. Template switching generated during reverse transcription? *FEBS Lett* **527:** 326. doi: 10.1016/S0014-5793(02)03239-8.