Clinical Chemistry 51:3
506–515 (2005)

Intron Retention: A Common Splicing Event within the Human Kallikrein Gene Family

Iacovos P. Michael,[1,2] Lisa Kurlender,[1,2] Nader Memari,[1,2] George M. Yousef,[3]
Daisy Du,[1] Linda Grass,[1] Carsten Stephan,[4] Klaus Jung,[4] and
Eleftherios P. Diamandis[1,2*]

Background: **All human kallikrein (*KLK*) genes have at least one splice variant, some of which possess clinical utility in cancer diagnostics/prognostics. Given that introns <100 bp in length are retained in 95% of human genes and that splice variants of *KLK3* and *KLK4* retain intron III, we hypothesized that other proteins in this family, with a small intron III, may also retain it.**

Methods: **Variant-specific reverse transcription-PCRs (RT-PCRs) for *KLK1*, *KLK2*, *KLK5*, and *KLK15* were used to identify and clone the full coding sequence of intron III-containing splice variants. In addition, variant-specific RT-PCRs for the cloned *KLK3* and *KLK4* variants as well as for the "classical" forms of the six genes were used to determine their expression profiles in healthy tissues, their regulation by steroids, and their differential expression in prostate cancer.**

Results: ***KLK1*, *KLK2*, *KLK3*, *KLK4*, *KLK5*, and *KLK15* showed a common type of splice variant in which intron III is retained. Expression profiling of these splice variants revealed expression profiles similar to those of the classical mRNA forms, although the pattern of hormonal regulation was different. The *KLK15* splice variant was up-regulated in 8 of 12 cancerous prostate tissues. All encoded variant proteins were predicted to be truncated and catalytically inactive because of a lack of the serine residue of the catalytic triad.**

Conclusions: **The first six centromeric members of the *KLK* gene family have splice variants that retain intron III. Some variants show tissue-specific expression. The *KLK15* splice variant appears to be a candidate biomarker for prostate cancer.**

The mechanism of alternative splicing was first proposed by Gilbert *(1)*, who postulated that different combinations of exons could be spliced together to produce different mRNA isoforms of a gene, and was a major contribution after the discovery of exons and introns in the adenovirus *hexon* gene in 1977 *(2)*. Alternative splicing is now considered one of the main mechanisms responsible for increasing the diversity and complexity of the human genome *(3, 4)*. Genome-wide analyses indicate that 40–60% of human genes have alternative splicing forms *(5)*. Furthermore, recent reports indicate that 10–30% of alternatively spliced human genes possess tissue-specific splice forms *(6)* and that at least 300 genes possess cancer-specific mRNA variants *(7)*. Among the hallmarks of cancer are genomic mutations. It has been estimated that at least 15% of mutations that cause genetic diseases affect pre-mRNA splicing *(8)*. Indeed, numerous studies have indicated that alternative splicing occurs frequently in cancer cells, and a plethora of cancer-specific splice variants have been reported. Some of them are considered to be candidate cancer biomarkers, such as *CD44 (9)* and the Wilms tumor (*WT1*) *(10)* genes.

The human kallikrein (*KLK*)[5] genes are a family of 15 serine protease genes that map to chromosome 19q13.4 *(11–13)*. Kallikreins represent the largest cluster of protease genes of any kind within the human genome. Although the function of many of these proteases is

[1] Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, ON, Canada.
[2] Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada.
[3] Discipline of Pathology, Memorial University, St. John's, NL, Canada.
[4] Urology, University Hospital Charité, Humboldt University, Berlin, Germany.
*Address correspondence to this author at: Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Ave., Toronto, ON, Canada M5G 1X5. Fax 416-586-8628; e-mail ediamandis@mtsinai.on.ca.
Received September 7, 2004; accepted December 7, 2004.
Previously published online at DOI: 10.1373/clinchem.2004.042341

[5] Nonstandard abbreviations: *KLK*, kallikrein gene; RT-PCR, reverse transcription-PCR; EST, expressed sequence tag; pY, polypyrimidine; PTC, premature termination codon; DHT, dihydrotestosterone; and NMD, nonsense-mediated mRNA decay.

currently unknown, their clinical utilities have been successfully exploited. For example, human kallikrein 3 [prostate-specific antigen (PSA)] is the best biomarker for prostate cancer screening, diagnosis, staging, and monitoring *(14)*. Other members of this family are promising biomarkers for prostate, breast, and ovarian cancer *(15–18)*. The association of this gene family with other disorders, such as Alzheimer disease *(19)*, and physiologic processes, such as skin desquamation *(20)*, has also been suggested.

Alternative pre-mRNA splicing is a common event among members of the *KLK* gene family. To date, ~70 *KLK* splice variants have been reported, and each *KLK* gene possesses at least one variant *(13)*. Splicing events such as exon skipping, extension, truncation, cryptic exons, and intron retention have been observed in both the coding and noncoding regions of the genes. Furthermore, alternative transcriptional start sites and polyadenylation sites have also been reported. Most of these variant transcripts are predicted to encode for truncated proteins lacking one or more residues of the catalytic triad because of frameshifts. An exception is one *KLK4* isoform, which retains all of the residues of the catalytic triad but lacks the signal peptide and is thought to act intracellularly *(21)*. Despite the fact that alternative splicing substantially increases the diversity of this locus, most of these putative protein isoforms have not been isolated, with the exception of a few proteins encoded by *KLK2* and *KLK3* variants *(22)*. The association of some of these splice variants with cancer has been examined *(23–29)*. Some of them are tissue, developmental stage, stimulus, or disease specific (our unpublished data).

Common patterns of alternative splicing have been observed within several gene families. For example, a study examining the effects of alternative splicing on transcripts encoding membrane proteins revealed that a common splice form leads to the removal of the transmembrane domain of single-pass transmembrane proteins, producing a soluble protein isoform *(30)*. Another study has shown that 50 protein domain types were selectively removed by alternative splicing *(31)*.

In the present study, we examined the frequency of a common type of splicing (intron retention) within the *KLK* gene family. Intron retention has been reported for several other genes. Recently, a study examining intron retention in a set of 21 106 known human genes revealed that 14.8% retained at least one intron *(32)*. The probability of intron retention increases as intron length decreases because introns <100 bp in length are retained in 95% of cases. Among human kallikrein splice variants reported to date, two of them, Psa-rp2 for the *KLK3* gene (GenBank accession no. AJ310938) and *KLK4* variant 1 for the *KLK4* gene (GenBank accession no. AF148532), retain intron III, which is relatively short (143 and 83 bp, respectively). Because other kallikrein genes, such as *KLK1*, *-2*, *-5*, and *-15*, also possess a short intron III, we speculated that retention of intron III might be a common splicing event among members of the *KLK* family.

Materials and Methods

MATERIALS

Total RNA isolated from human adrenal gland, adult brain, colon, fetal brain, fetal liver, heart, kidney, liver, lung, mammary gland, pancreas, placenta, prostate, salivary gland, skeletal muscle, spinal cord, spleen, small intestine, stomach, testis, thymus, thyroid, trachea, and uterus was obtained from Clontech. RNA from adipose, cervical, and vaginal tissue was from Ambion, whereas RNA from skin was from Invitrogen. We used the breast cancer cell lines T-47D, ZR-75-1, BT-474, MCF-7; the prostate cancer cell lines DU 145 and PC-3; the neuroblastoma cancer cell lines SK-N-BE and SW 1088; and the ovarian cancer cell lines PA-1, ES-2, TOV-21G, and BG-1. All cell lines were obtained from ATCC except for BG-1, which was a gift from Dr. H. Rocheford (INSERM, Montpellier, France).

MATCHED NONCANCEROUS/CANCEROUS TISSUES FROM PROSTATE CANCER PATIENTS

We obtained 12 pairs of matched tissue samples (noncancerous/cancerous) from patients with a median age of 63 years who underwent radical prostatectomy for prostatic adenocarcinoma at the University Hospital Charité, Berlin, Germany. Fresh prostate tissue samples were obtained from the cancerous and noncancerous parts of prostatectomy specimens. Small pieces of tissue were gross-dissected by a pathologist immediately after prostate removal, snap-frozen, and stored in liquid nitrogen until analysis. To ensure that the tissue was either malignant or benign, histologic analysis was performed by the same pathologist, as described previously *(33)*. Only samples that were fully surrounded by malignant tissue were used. Tissue characterized as noncancerous was usually taken from the inner zone of the contralateral lobe.

The samples were collected with informed consent, and the study was approved by the Ethics Committee of the Charité Hospital.

HORMONAL REGULATION EXPERIMENTS WITH CANCER CELL LINES

Cells were cultured to near confluency in RPMI medium (Life Technologies, Inc.) supplemented with glutamine (200 mmol/L), bovine insulin (10 mg/L), fetal bovine serum (100 mL/L), antibiotics, and antimycotics. The cells were then aliquoted into 24-well tissue culture plates and cultured to ~50% confluency. Twenty-four hours before the experiments, the culture media were replaced with phenol red-free medium containing 100 mL/L charcoal-stripped fetal bovine serum. For stimulation experiments, the steroid hormones estradiol (estrogen), dihydrotestosterone (androgen), norgestrel (synthetic progestin), aldosterone (mineralocorticoid), and dexamethasone (syn-

thetic glucocorticoid) dissolved in absolute ethanol were added to the culture medium at a final concentration of 10^{-8} mol/L. Cells stimulated with ethanol were included as controls. In all cases, the final ethanol concentration was 1 mL/L. The cells were cultured for 24 h and then harvested for mRNA extraction. All experiments were performed in triplicate.

RNA EXTRACTION

Prostate tissues and other healthy human tissues (esophagus, fallopian tube, hippocampus, ovary, pituitary, and tonsil) were pulverized with a hammer under liquid nitrogen. Total RNA from these tissues, as well as from cell line pellets, was extracted with TRIzol reagent (Life Technologies) and treated with DNase I (Invitrogen) according to the manufacturers' instructions. The RNA concentration and purity were determined spectrophotometrically.

REVERSE TRANSCRIPTION

We reverse-transcribed 2 μg of total RNA into first-strand cDNA, using the SuperScript™ First-Strand Synthesis System for reverse-transcription-PCR (RT-PCR; Invitrogen). The final volume was 20 μL. We diluted 1 μL of the cDNA 100-fold and performed a PCR reaction for the housekeeping gene β-actin, as described below, to check the quality of the first-strand cDNA synthesis.

PCR

Three PCR reactions were performed for each splice variant. In the first reaction, we used the primers F1 and R1 to simultaneously amplify both the "classical" as well as the splice variant form of interest (Fig. 1). Using primers sets F2/R2 and F3/R3, we amplified and characterized the whole coding region of each splice variant (Fig. 1). Furthermore, using the primers identified by asterisks in Table 1 of the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol51/issue3/, we were able to achieve specific amplification of each splice variant alone (one of the two primers binds within intron III). To examine the tissue expression of each splice variant, we used both splice-variant specific primers (F2/R2 or F3/R3) and the F1/R1 pair of primers in different reactions. For the steroid hormone regulation experiments, we used the F1/R1 pair of primers, whereas for the prostate cancer/noncancer pairs, we used the splice variant-specific primers.

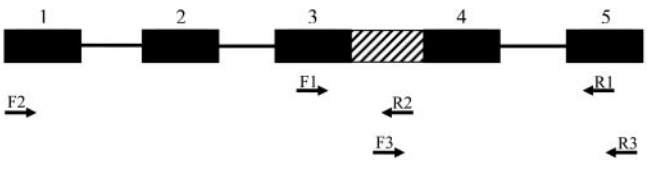

Fig. 1. Primers sets (*arrows*) used in this study.

The set F1/R1 was used to amplify both the classical form and the corresponding splice variant, whereas the sets F2/R2 and F3/R3 were used for specific amplification of the splice variants. Only coding exons (■) are shown. ▨, intron III.

Each PCR reaction was carried out in a reaction mixture containing 1 μL of cDNA, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM deoxynucleoside triphosphates, 100 ng of primers, and 2.5 U of Hot Star Taq DNA polymerase (Qiagen Inc.) on a Eppendorf thermocycler. The cycling conditions were 95 °C for 15 min to activate the *Taq* polymerase followed by 35 or 40 cycles (Table 1 in the online Data Supplement) of 94 °C for 30 s, the annealing temperature (*T*a, in °C; see Table 1 in the online Data Supplement) for 30 s, 72 °C for 30 s, and a final extension at 72 °C for 10 min. Equal amounts of PCR products were electrophoresed on 1.5% agarose gels and visualized by ethidium bromide staining.

To verify the identity of the PCR products, we purified them, using gel extraction reagents (Qiagen), and cloned them into the TOPO TA cloning vector (Invitrogen) according to the manufacturer's instructions. The inserts were sequenced from both directions by use of vector-specific primers with an automated DNA sequencer.

IN SILICO ANALYSIS

Sequence homology searching was performed with the basic local alignment research tool (BLAST), available from the National Center of Biotechnology Information, against the human expressed sequence tag (EST) database (dbEST). Sequences with >95% homology were considered as putative ESTs.

Results

CLONING OF FOUR NOVEL SPLICE VARIANTS

Given that two splice variants with intron retention were already identified, i.e., Psa-rp2 for *KLK3* (GenBank accession no. AJ310938) and *KLK4* variant 1 (AF148532), we designed primers that bound in coding exons 3 and 5 (F1 and R1 primers, respectively; Fig. 1; also see Table 1 in the online Data Supplement) to examine whether additional kallikrein genes have splice variants with a retained intron III. When we performed PCR on 36 healthy human tissue samples, we observed a second band, higher than the one expected for the classical form for the genes *KLK1*, *KLK2*, *KLK5*, and *KLK15*. After cloning and sequencing these bands, we verified that they represent splice variants with retained intron III (Fig. 2). To further characterize these splice variants, we designed two splice variant-specific RT-PCRs, with one primer of each pair binding within intron III (F3 and R2 primers; Fig. 1; also see Table 1 in the online Data Supplement). Thus, we were able to clone the complete coding sequence of each splice variant. The splice variants for the *KLK3* and *KLK4* genes, described previously, were verified by cloning and sequencing. The GenBank accession numbers, names, and mRNA lengths for the newly identified splice variants are shown in Table 1. The new variants will be referred to as *KLK1-IRIII*, *KLK2-IRIII*, *KLK5-IRIII*, and *KLK15-IRIII*, and

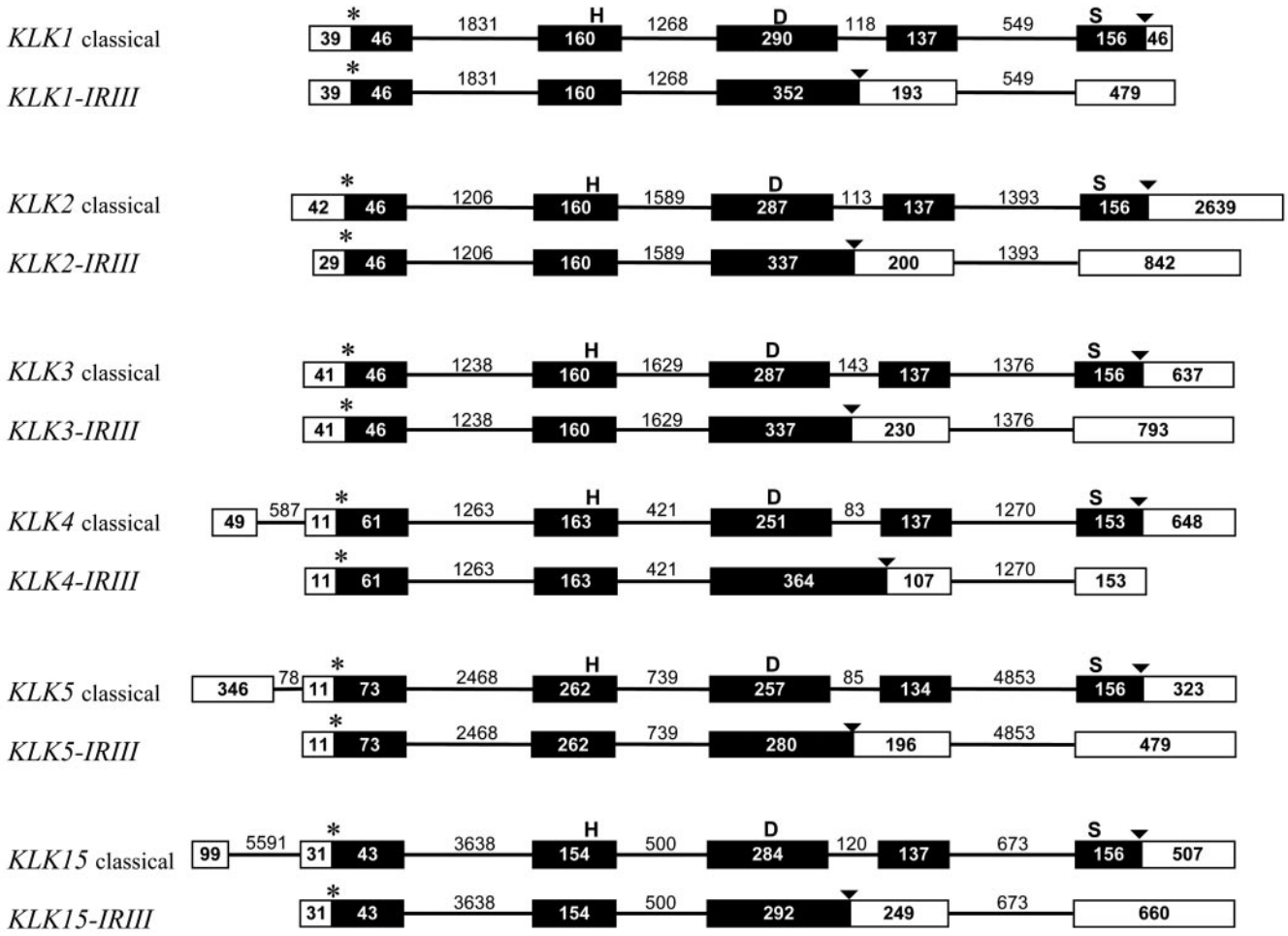

Fig. 2. Genomic organization of the classical and the corresponding splice variant of the genes *KLK1* through *-5* and *KLK15*.

■, coding exons; □, noncoding exons; ——, introns. *Numbers in* the *boxes* and *above* the *lines* represent the lengths of the exons and the introns, respectively, in base pairs. ∗, start codon; ▼, stop codon. The amino acids of the catalytic triad are indicated: *H*, histidine; *D*, aspartic acid, *S*, serine.

for those that have been published previously as *KLK3-IRIII* and *KLK4-IRIII*.

GENOMIC ORGANIZATION

Because all of these splice variants retain intron III, they have four instead of five coding exons (Fig. 2). A new exon is created by merging coding exons 3 and 4 of the classical form and the intron between them.

We defined the 5′ and 3′ splice sites [branch point, polypyrimidine (pY) tract, and acceptor site] of the classical form of the corresponding splice variant (Fig. 3) to explore the reasons and possible mechanism that gives

Table 1. Splice variants of the *KLK1* through *-5* and *KLK15* genes with retained intron III.

GenBank submission name	Adopted name	Nucleotide accession no.	mRNA length, bp	Protein accession no.	Predicted protein length, amino acids
Homo sapiens kallikrein 1 isoform 3 preproprotein (*KLK1*) mRNA, complete cds, alternatively spliced	*KLK1-IRIII*	AY429508	992	AAR10467	172
Homo sapiens kallikrein 2 isoform 5 preproprotein (*KLK2*) mRNA, complete cds, alternatively spliced	*KLK2-IRIII*	AY429509	1458	AAR10468	180
Homo sapiens mRNA for putative preproPSA-RP2 (*KLK3*) gene, transcript 2	*KLK3-IRIII*	AJ310938	1603	CAC41632	180
Homo sapiens kallikrein 4 splice variant (*KLK4*) gene, complete cds, alternatively spliced	*KLK4-IRIII*	AF148532	859	AAD38019	254
Homo sapiens kallikrein 5 isoform 3 preproprotein (*KLK5*) mRNA, complete cds; alternatively spliced	*KLK5-IRIII*	AY461805	1301	AAR23814	204
Homo sapiens kallikrein 15 isoform 5 preproprotein (*KLK15*) mRNA, complete cds; alternatively spliced	*KLK15-IRIII*	AY373373	1398	AAQ82620	162

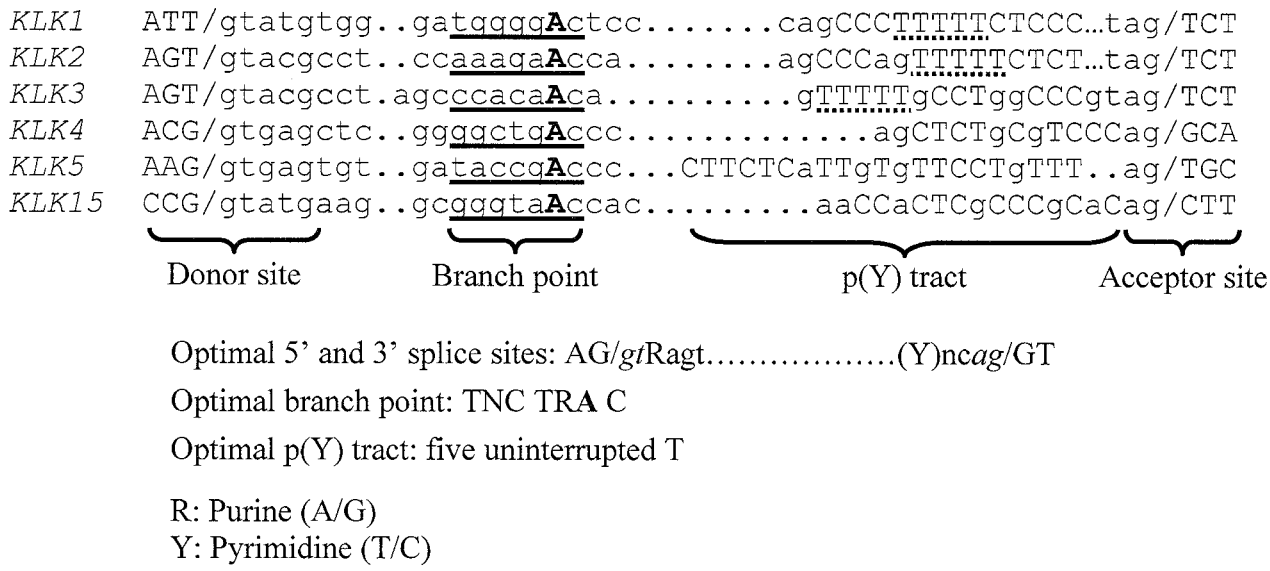

Fig. 3. The 5′ and 3′ splice sites (branch point, pY tract, and acceptor site) of the classical forms of six kallikrein genes.

The *uppercase letters* represent the exon sequences, and the *lowercase letters* represent intron III sequences. The branch point is indicated with a *solid underline* and the pY tract with a *dotted underline*. The nonconserved base pairs are *shaded* (see text for more details).

rise to this splicing pattern. Knowing that the optimal donor and acceptor sites are denoted by the sequences AG/*gt*Ragt and (Y)nc*ag*/GT, respectively, we observed that none (except *KLK5*) of the classical forms had optimal splice sites in either the donor or acceptor sites (Fig. 3). *KLK5* was the only gene that had an optimal donor site. Regarding the possible branch point (YNY YR**A** Y), again the *KLK5* gene was the only one with an optimal branch point. Finally, all of the genes have a pY tract located −18 to −40 bp before the acceptor site (Fig. 3).

Lander et al. *(3)* showed that shorter introns have a higher GC content in humans. On average, the GC content (60.8%) of intron III of kallikrein genes is higher than all other introns and is the only intron with a higher GC content than its flanking exons (data not shown). Furthermore, the GC content for sequences upstream and downstream from the premature termination codon is 66.8% and 58.2%, respectively (data not shown).

PREDICTED PROTEIN SEQUENCE

Using the open reading frame finder program (ORF Finder) from the National Center of Biotechnology Information (http://www.ncbi.nlm.nih.gov/gorf/gorf.html), we identified the predicted protein sequence for each splice variant (Table 1; also see Fig. 1 in the online Data Supplement). Comparison with *KLK3-IRIII* and *KLK4-IRIII* revealed the following features: (*a*) they retain the signal peptide, suggesting that they will be secreted proteins; (*b*) they have a premature termination codon (PTC), and thus encode truncated proteins (Fig. 2); (*c*) they lack the "Ser" residue of the catalytic triad, and thus will not function as serine proteases; and (*d*) they contain a unique C-terminal sequence encoded from the intronic sequence (Fig. 4 in the online Data Supplement).

IN SILICO ANALYSIS

Using BLAST, we identified ESTs that correspond to our experimentally defined splice variants (Table 2). We found 2 ESTs for *KLK1-IRIII*, 10 for *KLK2-IRIII*, and 1 for *KLK15-IRIII*. No EST was found for *KLK5-IRIII*. For the previously identified splice variants, we found five ESTs for *KLK3-IRIII* and one EST for *KLK4-IRIII*.

Table 2. EST clones with >95% homology with the exons of the splice variants.

Splice variant	EST	Tissue
KLK1 IR-III	BF822730	Kidney tumor
	BI832821	Pooled pancreas and spleen
KLK2 IR-III	AI547068	Prostate
	BE770927	Prostate tumor
	BE770929	Prostate tumor
	BE770943	Prostate tumor
	BE770947	Prostate tumor
	BE770950	Prostate tumor
	BE771002	Prostate tumor
	BE771007	Prostate tumor
	BQ918160	Sciatic nerve
	BQ925280	Sciatic nerve
KLK3 IR-III	AA934534	Metastatic prostate bone lesion
	BU930334	Prostate
	CB048196	Prostate
	BQ954089	Sciatic nerve
	AW973948	Unknown
KLK4 IR-III	AI557025	Prostate
KLK5 IR-III	Not found	
KLK15 IR-III	BX280958	Pooled

TISSUE EXPRESSION

The tissue expression profile of the six splice variants was elucidated by RT-PCR using total RNA from 36 healthy human tissues. By performing two PCR reactions for each splice variant (as described in the *Materials and Methods*), we were able to compare the expression of each splice variant with its corresponding classical form. Generally, the concentrations of the splice variant mRNAs were similar to the concentrations of the corresponding classical forms (Table 3). Thus, *KLK1-IRIII* was highly expressed in kidney, pancreas, salivary gland, and thyroid. The splice variants *KLK2-IRIII*, *KLK3-IRIII*, and *KLK4-IRIII* were highly expressed in the prostate. The expression of these variants appears to be more prostate specific (particularly for *KLK3-IRIII*) compared with that of the corresponding classical forms. *KLK15-IRIII* was more highly expressed in prostate, salivary gland, testis, and thyroid. Finally, *KLK5-IRIII* appears to have a broader expression pattern than that of the classical form (Fig. 2 in the online Data Supplement). High expression was seen in the fallopian tube, esophagus, and pituitary. We observed no expression of the classical form in these tissues. On the other hand, in the cervix, salivary gland, spinal cord, stomach, and thyroid, in which the classical form was expressed in high concentrations, we observed no expression of the splice variant.

HORMONAL REGULATION

We examined the hormonal regulation patterns of the splice variants and classical forms of the six kallikrein genes by RT-PCR. Splice variants *KLK2-IRIII*, *KLK3-IRIII*, *KLK4-IRIII*, and *KLK15-IRIII* followed the same pattern as

Table 3. Tissue expression of the classical kallikrein genes and the corresponding splice variants with retained intron III.

	Expression score[a]											
Tissue	***KLK1***	***KLK1 IR-III***	***KLK2***	***KLK2 IR-III***	***KLK3***	***KLK3 IR-III***	***KLK4***	***KLK4 IR-III***	***KLK5***	***KLK5 IR-III***	***KLK15***	***KLK15 IR-III***
Adipose	H	M	L		L		M	M			H	L
Adrenal gland	H	M	L	L	L		H	M	L		M	L
Bone marrow	H		L		L		L	L	M			
Brain, cerebellum	L	L	L	L	L		H	H	H	L	L	L
Brain, adult	M		L	L	L		L	L	H	L	L	L
Cervix	M		L	L	L		H	M	H			L
Colon	H	M			L		M	L	L		H	L
Esophagus	L		L	L	L					H		
Fallopian tube	L	L			L					H		
Fetal brain	M	L					H	M	M	L	L	L
Fetal liver	M	L			L				L	L		L
Heart	M	L		L	L				L			
Hippocampus								L	L	L		L
Kidney	H	H	L	L	L			M	M		M	L
Liver												
Lung	H	L	L		L				L			
Mammary gland	H		M		M		L	M	H	M		
Ovary	M									M		
Pancreas	H	H		L					L	L	L	L
Placenta	L		L		L				L			
Prostate	H	L	VH	VH	VH	VH	H	H	M		H	M
Pituitary		L								H		
Salivary gland	H	H	M	M	M	L	H	H	H		H	M
Skeletal muscle	L	L	L		L		L					
Skin	H	L	L					L	H	H		
Spinal cord	M	L	L	L	M	L	H	H	H		L	L
Spleen	H	M			L		L	L	H	H		
Small intestine	H	M			L		L	L	L		L	L
Stomach	M	L			L		L	L	H		L	L
Testis	M	M	M	L	M	L	H	H	H	L	M	M
Thymus	H	L	L		L		M	L	M		L	
Thyroid	H	H	L	M	L		H	H	H		H	M
Tonsil	M	L						L	L	L	L	
Trachea	M	L	L	L	L		M	M	L		L	L
Uterus	L	L			L			M				
Vagina	H	L			L		L	L	M	L	L	L

[a] Expression score: −, no expression; L, low; M, medium; H, high; VH, very high.

the corresponding classical forms; i.e., they were regulated mainly by the androgen dihydrotestosterone (DHT) and the androgenic progestin norgestrel. The splice variants *KLK4-IRIII*, *KLK5-IRIII*, and *KLK15-IRIII* were down-regulated by dexamethasone (Fig. 4).

Of interest is the pattern of regulation observed for *KLK5-IRIII*. Previous studies indicated that the classical form, *KLK5*, is up-regulated by estradiol and norgestrel in the cell line BT474 *(34)*. In the present study, we observed that in the cell line BT474, *KLK5-IRIII* was down-regulated by estradiol, whereas *KLK5* was up-regulated; in the cell lines PA-1 and HTB12, estradiol up-regulated *KLK5-IRIII*. In the cell line ES-2, only *KLK5-IRIII* was expressed and was down-regulated by norgestrel. Down-regulation by norgestrel was observed in the BG-1 cell line and by dexamethasone in the cell line HTB12. On the other hand, DHT up-regulated *KLK5-IRIII* in the cell line BG-1 (Fig. 4).

Differential expression in cancerous vs noncancerous matched prostate samples

We examined the expression of the six splice variants in 12 pairs of matched noncancerous and cancerous prostate tissues. We observed no differential expression for the splice variants *KLK2-IRIII*, *KLK3-IRIII*, and *KLK4-IRIII.* The *KLK1-IRIII* splice variant was down-regulated in cancer in four pairs, whereas it was up-regulated in two (data not shown). *KLK15-IRIII* was up-regulated in 8 of 12 pairs (pairs 1–5 and 9–11), whereas we observed no difference for the remaining pairs (Fig. 5).

Discussion

Alternative splicing is a major mechanism for increasing the diversity of the proteome. As many as 70 splice variants of kallikrein genes have been published to date. In this study, we examined a specific splicing event (intron retention) within the human kallikrein gene family. We found that the first six centromeric members of this family, *KLK1* through *-5* and *KLK15*, have a common splice variant derived from retention of intron III.

Intron retention is a relatively common event. For example, in *Drosophila* P elements, an intron may be either spliced or retained, depending on the cell type in which it is expressed *(35)*. In mammals, intron retention has been observed in many genes, including human tumor necrosis factor-β *(36)*, human growth hormone *(37)*, bovine growth hormone *(38)*, and rat γ-fibrinogen *(39)*. Dirksen et al. *(40)* have shown that both suboptimal 5′ and 3′ splice sites are required for intron retention of bovine growth hormone. In the present study, examination of 5′ and 3′ splice sites of the third intron of *KLK1* through *-5* and *KLK15* indicated that all of them, except for *KLK5*, have suboptimal sites. Furthermore, mutational analysis of the pY tract of introns has shown that a minimum of five uninterrupted thymidines are required for strong

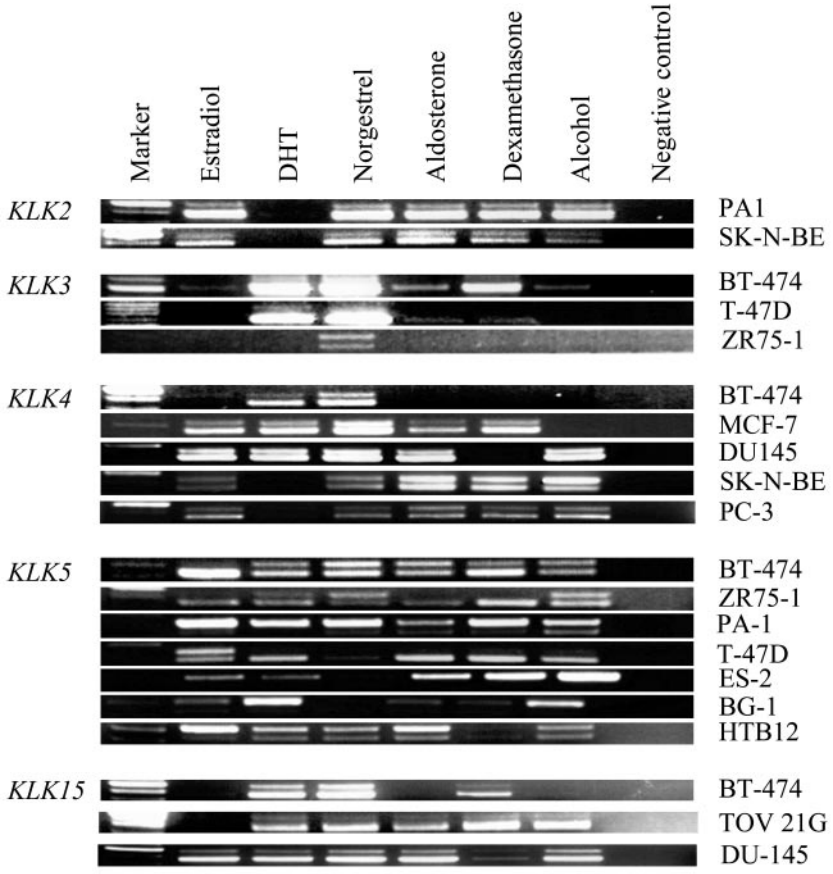

Fig. 4. Hormonal regulation of the classical forms and the corresponding splice variants of the genes *KLK1* through *-5* and *KLK15* in cell lines.

Shown are four breast cell lines (T-47D, ZR75-1, BT-474, and MCF-7), four ovarian cell lines (PA-1, ES-2, BG-1, and TOV 21G), two prostate cell lines (DU145 and PC-3), and two neuroblastoma cell lines (SK-N-BE and HTB12). Steroid hormones (estradiol, DHT, norgestrel, aldosterone, and dexamethasone) dissolved in absolute ethanol were added to the culture media at a final concentration of 10^{-8} mol/L and an ethanol concentration 1 mL/L. The *lower band* for each cell line represents the classical form, and the *upper band* the splice variant form. There was no DNA template used in the PCR negative control. The primer set used was F1/R1 (Fig. 1). For interpretations, see the text. Actin was used as a positive and nonhormonally regulated gene (not shown).

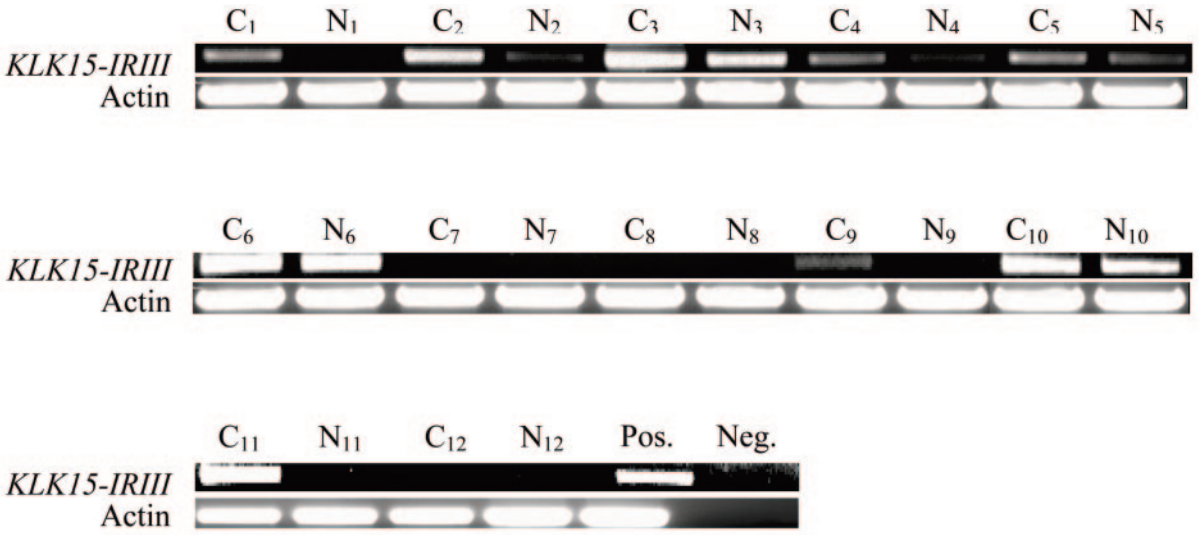

Fig. 5. *KLK15* splice variant expression in paired cancerous/healthy prostate samples, as assessed by RT-PCR.

The primer set used was F3/R3 (see Fig. 1). There was no DNA template used in the PCR negative control. Actin was used as a positive control. Note the up-regulation of *KLK15-IRIII* in 8 of 12 cancerous (*C*) tissues compared with the matched healthy tissues (*N*). For more discussion, see the text.

binding of the heterogeneous ribonucleoprotein C (hnRNP C) and for correct splicing to take place *(41)*. Only *KLK1*, *-2*, and *-3* have an uninterrupted sequence of five thymidines, whereas *KLK5* and *-15* seem to have a very weak pY tract. The branch point was also shown to be important for effective splicing *(41)*. Of the genes examined here, only *KLK5* has a fully conserved branch point. In conclusion, it seems that a suboptimal 5′ splice site (donor site) and 3′ splice site (branch point, pY tract, and acceptor site) play a major role in intron III retention of these genes.

The higher GC content of intron III compared with the rest of the introns is in agreement with the results reported by Goodall and Filipowics *(42)*, who demonstrated in plants that a higher GC content might have a lower excision rate, and with the results of a global analysis of the human transcriptome for intron retention *(32)*. High GC content is a characteristic feature of active euchromatin and transcriptional activity. This feature of intron III might indicate that this area of the gene is more susceptible to transcription factor binding and higher transcriptional activity.

Nonsense-mediated mRNA decay (NMD) is a surveillance posttranslational mechanism that controls the quality of the mRNA function by degrading all abnormal mRNA transcripts that contain a PTC *(43)*. The retention of intron III changes the reading frame and creates a PTC. It seems the splice variant transcripts reported here are immune to NMD because high concentrations of cytoplasmic mRNAs were detected. NMD-resistant mRNAs for other genes, such as β-globin *(44)*, von Willebrand factor *(45)*, cystic fibrosis transmembrane conductance regulator *(46)*, LDL receptor *(47)*, and apolipoprotein B *(48)*, have been reported. It is possible that some genes contain *cis*-acting sequences that confer resistance to NMD. Such sequences have been found in yeast *(49, 50)*.

Many proteins encoded by intron-retaining splice variants have been reported. These include isoforms of cofactor proteins CD44 and CD46 *(51, 52)*, human growth hormone *(53)*, human gonadotropin-related hormone gene *(54)*, effector cell protease receptor-1 *(55)*, and murine vitamin D receptor (VDR0) *(56)*. Interestingly, in the latter case, the vitamin D receptor isoform (VDR1) was shown to act as a dominant-negative receptor against VDR0 transactivation. The predicted proteins of the kallikrein splice variants described here will be truncated and lack the serine residue of the catalytic triad. These proteins may act in a dominant-negative manner, regulating the function of the corresponding classical forms, or may display an as yet unknown nonprotease function.

Tissue-specific splicing has been reported for 10–30% of the human genes *(6)*. We have shown here that the splice variant of the *KLK3* gene seems to be expressed exclusively in the prostate. Very high expression in the prostate was also observed for the *KLK2* gene splice variant. These data agree with our in silico analysis, according to which the *KLK2* and *KLK3* splice variants are expressed predominantly in the prostate. The finding of ESTs for these variants in sciatic nerve libraries is interesting because the classical *KLK2* and *KLK3* genes are known to be expressed almost exclusively in the prostate. The *KLK5* splice variant seems to have a pattern of expression different from that of the classical form because in some tissues, either the splice variant or the classical form is expressed. The tissue-specific expression of the *KLK2*, *KLK3*, and *KLK5* splice variants might be valuable if these genes find application as diagnostic and/or prognostic biomarkers in cancer. The differential tissue expression and hormonal regulation between the splice variants and the corresponding classical forms suggests that different *cis*- and/or *trans*-acting elements might regulate their transcription.

The connection between human kallikreins and cancer has been reported in many studies *(11)*. Numerous kallikreins are established or emerging biomarkers for the diagnosis, prognosis, and monitoring of cancer at the mRNA and/or protein level. Splice variants of these genes display cancer-specific expression, or they are differentially expressed in cancer *(22–25, 57)*. In this study, we found that the splice variant of the *KLK15* gene is up-regulated in 8 of 12 prostate cancer tissues compared with the corresponding healthy tissue samples. The classical form of the *KLK15* gene has also been shown to be up-regulated at the mRNA level in prostate cancer and was associated with more aggressive forms *(58)*.

A relationship between splice variants that retain introns and human diseases, including malignancies, has been revealed by global analysis of the human transcriptome *(32)*. Among 88 genes that generate putative truncated proteins, there are genes associated with Williams–Beuren syndrome and Batten–Spielmayer–Vogt disease. Furthermore, several genes are related to the tumorigenic process, including the p19A, tumor necrosis factor receptor, BCL2-like 11, and CDC2-like 10 genes.

Our unpublished data indicate that 3′ extension of coding exon 3 in the rest of the genes may produce similar isoforms. For example, we found that *KLK6* and *KLK13* have a splice variant with 3′-coding exon 3 extension by ~100 bp (close to the length of the retained intron III in the genes examined here). The extension also leads to a frameshift and creation of a PTC. The predicted encoded proteins will be truncated and lack the catalytic serine residue. Bioinformatic analysis of ESTs revealed that 3′-coding exon 3 extension or complete retention is universal among all kallikrein genes.

In this study, we showed a common type of alternative splicing, characterized by complete retention of intron III, in 6 (*KLK1* through *-5* and *KLK15*) of the 15 human kallikrein genes. Some of these variants may have diagnostic and/or prognostic value because they show tissue-specific expression or differential expression in comparison with the classical form. The up-regulation of the *KLK15* splice variant in prostate cancer warrants examination of the encoded protein as a biomarker of prostatic cancer.

This work was supported by a grant to E.P. Diamandis from Natural Sciences and Engineering Research Council of Canada (NSERC) and IBEX Technologies Inc., under a University/Industry program, and by a grant from Deutsche Forschungsgemeinschaft (JU 365/6-1) to K. Jung and C. Stephan.

References

1. Gilbert W. Why genes in pieces? Nature 1978;271:501.
2. Sambrook J. Adenovirus amazes at Cold Spring Harbor. Nature 1977;268:101–4.
3. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome [Erratum published in: Nature 2001;412:565]. Nature 2001;409:860–921.
4. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science 2001;291:1304–51.
5. Modrek B, Lee C. A genomic view of alternative splicing. Nat Genet 2002;30:13–9.
6. Xu Q, Modrek B, Lee C. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. Nucleic Acids Res 2002;30:3754–66.
7. Xu Q, Lee C. Discovery of novel splice forms and functional analysis of cancer-specific alternative splicing in human expressed sequences. Nucleic Acids Res 2003;31:5635–43.
8. Krawczak M, Reiss J, Cooper DN. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum Genet 1992;90:41–54.
9. Naor D, Nedvetzki S, Golan I, Melnik L, Faitelson Y. CD44 in cancer. Crit Rev Clin Lab Sci 2002;39:527–79.
10. Wagner KD, Wagner N, Schedl A. The complex life of WT1. J Cell Sci 2003;116:1653–8.
11. Yousef GM, Diamandis EP. The new human tissue kallikrein gene family: structure, function, and association to disease. Endocr Rev 2001;22:184–204.
12. Clements JA, Willemsen NM, Myers SA, Dong Y. The tissue kallikrein family of serine proteases: functional roles in human disease and potential as clinical biomarkers. Crit Rev Clin Lab Sci 2004;41:265–312.
13. Borgono CA, Michael IP, Diamandis EP. Human tissue kallikreins: physiologic roles and applications in cancer. Mol Cancer Res 2004;2:257–80.
14. Diamandis EP. Prostate-specific antigen—its usefulness in clinical medicine. Trends Endocrinol Metab 1998;9:310–6.
15. Diamandis EP, Yousef GM, Luo LY, Magklara A, Obiezu CV. The new human kallikrein gene family: implications in carcinogenesis. Trends Endocrinol Metab 2000;11:54–60.
16. Diamandis EP, Yousef GM. Human tissue kallikreins: a family of new cancer biomarkers. Clin Chem 2002;48:1198–205.
17. Sotiropoulou G, Rogakos V, Tsetsenis T, Pampalakis G, Zafiropoulos N, Simillides G, et al. Emerging interest in the kallikrein gene family for understanding and diagnosing cancer. Oncol Res 2003;13:381–91.
18. Yousef GM, Polymeris ME, Yacoub GM, Scorilas A, Soosaipillai A, Popalis C, et al. Parallel overexpression of seven kallikrein genes in ovarian cancer. Cancer Res 2003;63:2223–7.
19. Yousef GM, Kishi T, Diamandis EP. Role of kallikrein enzymes in the central nervous system. Clin Chim Acta 2003;329:1–8.
20. Caubet C, Jonca N, Brattsand M, Guerrin M, Bernard D, Schmidt R, et al. Degradation of corneodesmosome proteins by two serine proteases of the kallikrein family, SCTE/KLK5/hK5 and SCCE/KLK7/hK7. J Invest Dermatol 2004;122:1235–44.
21. Xi Z, Klokk TI, Korkmaz K, Kurys P, Elbi C, Risberg B, et al. Kallikrein 4 is a predominantly nuclear protein and is overexpressed in prostate cancer. Cancer Res 2004;64:2365–70.
22. David A, Mabjeesh N, Azar I, Biton S, Engel S, Bernstein J, et al. Unusual alternative splicing within the human kallikrein genes KLK2 and KLK3 gives rise to novel prostate-specific proteins. J Biol Chem 2002;277:18084–90.
23. Slawin KM, Shariat SF, Nguyen C, Leventis AK, Song W, Kattan MW, et al. Detection of metastatic prostate cancer using a splice variant-specific reverse transcriptase-polymerase chain reaction assay for human glandular kallikrein. Cancer Res 2000;60:7142–8.
24. Nakamura T, Stephan C, Scorilas A, Yousef GM, Jung K, Diamandis EP. Quantitative analysis of hippostasin/KLK11 gene expres-

sion in cancerous and noncancerous prostatic tissues. Urology 2003;61:1042–6.
25. Chang A, Yousef GM, Jung K, Meyts ER, Diamanids EP. Identification and molecular characterization of five novel kallikrein gene 13 (KLK13;KLK-L4) splice variants: differential expression in human testis and testicular cancer. Anticancer Res 2001;21: 3147–52.
26. Nakamura T, Mitsui S, Okui A, Kominami K, Nomoto T, Miki T, et al. Alternative splicing isoforms of hippostasin (PRSS20/KLK11) in prostate cancer cell lines. Prostate 2001;49:72–8.
27. Nakamura T, Mitsui S, Okui A, Miki T, Yamaguchi N. Molecular cloning and expression of a variant form of hippostasin/KLK11 in prostate. Prostate 2003;54:299–305.
28. Dong Y, Kaushal A, Bui L, Chu S, Fuller PJ, Nicklin J, et al. Human kallikrein 4 (KLK4) is highly expressed in serous ovarian carcinomas. Clin Cancer Res 2001;7:2363–71.
29. Dong Y, Kaushal A, Brattsand M, Nicklin J, Clements JA. Differential splicing of KLK5 and KLK7 in epithelial ovarian cancer produces novel variants with potential as cancer biomarkers. Clin Cancer Res 2003;9:1710–20.
30. Xing Y, Xu Q, Lee C. Widespread production of novel soluble protein isoforms by alternative splicing removal of transmembrane anchoring domains. FEBS Lett 2003;555:572–8.
31. Resch A, Xing Y, Modrek B, Gorlick M, Riley R, Lee C. Assessing the impact of alternative splicing on domain interactions in the human proteome. J Proteome Res 2004;3:76–83.
32. Galante PA, Sakabe NJ, Kirschbaum-Slager N, de Souza SJ. Detection and evaluation of intron retention events in the human transcriptome. RNA 2004;10:757–65.
33. Meyer A, Jung K, Lein M, Rudolph B, Schnorr D, Loening SA. Factors influencing the ratio of free to total prostate-specific antigen in serum. Int J Cancer 1997;74:630–6.
34. Yousef GM, Diamandis EP. The new kallikrein-like gene, KLK-L2. Molecular characterization, mapping, tissue expression, and hormonal regulation. J Biol Chem 1999;274:37511–6.
35. Siebel CW, Rio DC. Regulated splicing of the *Drosophila* P transposable element third intron in vitro: somatic repression. Science 1990;248:1200–8.
36. Weil D, Brosset S, Dautry F. RNA processing is a limiting step for murine tumor necrosis factor β expression in response to interleukin-2. Mol Cell Biol 1990;10:5865–75.
37. Cooke NE, Ray J, Emery JG, Liebhaber SA. Two distinct species of human growth hormone-variant mRNA in the human placenta predict the expression of novel growth hormone proteins. J Biol Chem 1988;263:9001–6.
38. Hampson RK, Rottman FM. Alternative processing of bovine growth hormone mRNA: nonsplicing of the final intron predicts a high molecular weight variant of bovine growth hormone. Proc Natl Acad Sci U S A 1987;84:2673–7.
39. Crabtree GR, Kant JA. Organization of the rat gamma-fibrinogen gene: alternative mRNA splice patterns produce the γA and γB (γ') chains of fibrinogen. Cell 1982;31:159–66.
40. Dirksen WP, Sun Q, Rottman FM. Multiple splicing signals control alternative intron retention of bovine growth hormone pre-mRNA. J Biol Chem 1995;270:5346–52.
41. Roscigno RF, Weiner M, Garcia-Blanco MA. A mutational analysis of the polypyrimidine tract of introns. Effects of sequence differences in pyrimidine tracts on splicing. J Biol Chem 1993;268: 11222–9.
42. Goodall GJ, Filipowicz W. Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. EMBO J 1991;10:2635–44.
43. Lejeune F, Li X, Maquat LE. Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities. Mol Cell 2003;12:675–87.
44. Romao L, Inacio A, Santos S, Avila M, Faustino P, Pacheco P, et al. Nonsense mutations in the human β-globin gene lead to unexpected levels of cytoplasmic mRNA accumulation. Blood 2000;96:2895–901.
45. Mohlke KL, Nichols WC, Rehemtulla A, Kaufman RJ, Fagerstrom HM, Ritvanen KL, et al. A common frameshift mutation in von Willebrand factor does not alter mRNA stability but interferes with normal propeptide processing. Br J Haematol 1996;95:184–91.
46. Rolfini R, Cabrini G. Nonsense mutation R1162X of the cystic fibrosis transmembrane conductance regulator gene does not reduce messenger RNA expression in nasal epithelial tissue. J Clin Invest 1993;92:2683–7.
47. Lehrman MA, Schneider WJ, Brown MS, Davis CG, Elhammer A, Russell DW, et al. The Lebanese allele at the low density lipoprotein receptor locus. Nonsense mutation produces truncated receptor that is retained in endoplasmic reticulum. J Biol Chem 1987;262:401–10.
48. Chan L, Chang BH, Nakamuta M, Li WH, Smith LC. Apobec-1 and apolipoprotein B mRNA editing. Biochim Biophys Acta 1997; 1345:11–26.
49. Peltz SW, Brown AH, Jacobson A. mRNA destabilization triggered by premature translational termination depends on at least three cis-acting sequence elements and one trans-acting factor. Genes Dev 1993;7:1737–54.
50. Vilela C, Linz B, Rodrigues-Pousada C, McCarthy JE. The yeast transcription factor genes YAP1 and YAP2 are subject to differential control at the levels of both translation and mRNA stability. Nucleic Acids Res 1998;26:1150–9.
51. Goodison S, Yoshida K, Churchman M, Tarin D. Multiple intron retention occurs in tumor cell CD44 mRNA processing. Am J Pathol 1998;153:1221–8.
52. Pollard AJ, Flanagan BF, Newton DJ, Johnson PM. A novel isoform of human membrane cofactor protein (CD46) mRNA generated by intron retention. Gene 1998;212:39–47.
53. MacLeod JN, Lee AK, Liebhaber SA, Cooke NE. Developmental control and alternative splicing of the placentally expressed transcripts from the human growth hormone gene cluster. J Biol Chem 1992;267:14219–26.
54. Radovick S, Wondisford FE, Nakayama Y, Yamada M, Cutler GB Jr, Weintraub BD. Isolation and characterization of the human gonadotropin-releasing hormone gene in the hypothalamus and placenta. Mol Endocrinol 1990;4:476–80.
55. Altieri DC. Splicing of effector cell protease receptor-1 mRNA is modulated by an unusual retained intron. Biochemistry 1994;33: 13848–55.
56. Ebihara K, Masuhiro Y, Kitamoto T, Suzawa M, Uematsu Y, Yoshizawa T, et al. Intron retention generates a novel isoform of the murine vitamin D receptor that acts in a dominant negative way on the vitamin D signaling pathway. Mol Cell Biol 1996;16: 3393–400.
57. Tanaka T, Isono T, Yoshiki T, Yuasa T, Okada Y. A novel form of prostate-specific antigen transcript produced by alternative splicing. Cancer Res 2000;60:56–9.
58. Yousef GM, Scorilas A, Jung K, Ashworth LK, Diamandis EP. Molecular cloning of the human kallikrein 15 gene (KLK15). Up-regulation in prostate cancer. J Biol Chem 2001;276:53–61.