Cloning of a New Member of the Human Kallikrein Gene Family, *KLK14*, Which Is Down-Regulated in Different Malignancies

George M. Yousef, Angeliki Magklara, Albert Chang, Klaus Jung, Dionyssios Katsaros, and Eleftherios P. Diamandis[1]

Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada [G. M. Y., A. M., A. C., E. P. D.]; Department of Laboratory Medicine and Pathobiology, University of Toronto, Ontario M5G 1LS, Canada [G. M. Y., A. M., A. C., E. P. D.]; Department of Urology, University Hospital Charite, Humboldt University, Berlin D-10098, Germany [K. J.]; and Department of Gynecologic Oncology, Institute of Obstetrics and Gynecology, University of Turin, Turin 10128, Italy [D. K.]

ABSTRACT

Kallikreins (KLKs) belong to the serine protease family of proteolytic enzymes. Human pancreatic/renal KLK (*KLK1*) encodes for an enzyme that is involved in posttranslational processing of polypeptide precursors. The function of the other members of this gene family is currently unknown, but growing evidence suggests that many KLKs are implicated in carcinogenesis. By using the positional candidate approach, we were able to identify a new human KLK-like gene, *KLK14* (also known as *KLK-L6*). This new gene maps to chromosome 19q13.3–q13.4 and is formed of seven exons (two untranslated and five coding exons) and six intervening introns. *KLK14* was defined as a KLK gene based on structural and mapping criteria, in relation to other known KLK genes. *KLK14* is expressed in a variety of tissues, but the highest levels of *KLK14* are found in the central nervous system, including brain, cerebellum, and spinal cord. Our preliminary results show that *KLK14* is down-regulated, at the mRNA level, in breast, testicular, prostatic, and ovarian cancer.

INTRODUCTION

New gene discovery has been significantly accelerated with the availability of the EST[2] database (dbEST) and the development of powerful bioinformatics software for gene prediction (1). KLKs are subgroup of serine proteases with diverse physiological functions (2). The human tissue KLK gene family was, until recently, thought to consist of only three members: (*a*) the pancreatic/renal KLK (*KLK1*; Ref. 3); (*b*) the glandular KLK (*KLK2*; Ref. 4); and (*c*) PSA (PSA or *KLK3*; Ref. 5). However, in the last 2–3 years, many new members of the human KLK gene family have been identified (6). Recent advances in this field facilitated a better understanding of the role of KLKs in diverse physiological and pathological conditions related to the cardiovascular, hematological, endocrine, sensory, and muscular systems (7).

In addition, growing evidence suggests a role of human KLK genes in cancer. PSA testing has revolutionized the diagnosis and management of patients with prostate cancer (8). hK2 (encoded by the *KLK2* gene) is also gaining popularity as an adjuvant diagnostic marker for prostate cancer (9). The normal epithelial cell-specific 1 gene (*NES1* or *KLK10*; Ref. 10), *zyme/protease M/neurosin/KLK6* (11), *neuropsin/TADG14* (tumor-associated differentially expressed gene 14)/*KLK8* (12), and the stratum corneum chymotryptic enzyme (*HSCCE/KLK7*; Ref. 13) are differentially expressed in various malignancies (6).

In our efforts to identify new KLK-L genes that might be involved in malignancy, we have examined a 300-kb genomic region around human chromosome 19q13.3–q13.4, which is known to contain the

KLK gene locus and cloned a new gene, *KLK14*. Here, we describe the cloning of the new gene, its genomic and mRNA structure, its precise location in relation to other known KLKs, and its tissue expression pattern. *KLK14* is highly expressed in tissues of the CNS including the brain, cerebellum, and spinal cord. Our preliminary results suggest that this gene is down-regulated, at the mRNA level, in breast, testicular, ovarian, and prostate cancer.

MATERIALS AND METHODS

Analysis of Genomic Sequences around the Human KLK Gene Locus. We have obtained sequencing data of approximately 300 kb of nucleotides on chromosome 19q13.3–q13.4 from the web site of the Lawrence Livermore National Laboratory.[3] This sequence was in the form of nine contigs of variable lengths. Restriction analysis study, long PCR reactions, bacterial artificial chromosome and P1-derived artificial chromosome library screening, and end sequencing were used to construct a contiguous (~300-kb) region representing the putative human KLK locus. Computer programs were used to predict the presence of putative new genes in the genomic area of interest, as described previously (14).

EST Searching. The predicted exons of the putative new gene, *KLK14,* were subjected to homology search using the BLASTN algorithm (15) on the National Center for Biotechnology Information web server[4] against the human EST database, dbEST. A clone with 99% identity was obtained from the I.M.A.G.E. consortium (16) through Research Genetics Inc. (Huntsville, AL). The clone was propagated, purified, and sequenced from both directions with an automated sequencer, using insert-flanking vector primers.

Tissue Expression. Total RNA isolated from 27 different human tissues was purchased from Clontech (Palo Alto, CA). We prepared cDNA as described below and used it for PCR reactions. Tissue cDNAs were amplified at various dilutions. Due to the high degree of homology with other human KLKs, and to exclude nonspecific amplification, the PCR products of the *KLK14* gene were cloned into the pCR 2.1-TOPO vector (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. The inserts were then sequenced from both directions using vector-specific primers, with an automated DNA sequencer.

RT-PCR. Total RNA was extracted from normal and malignant breast tissues using Trizol reagent (Life Technologies, Inc., Gaithersburg, MD) following the manufacturer's instructions. RNA concentration was determined spectrophotometrically. Two μg of total RNA were reverse-transcribed into first-strand cDNA using the Superscript preamplification system (Life Technologies, Inc.). The final volume was 20 μl. Based on the combined information obtained from the predicted genomic structure of the new gene and the EST sequences, two gene-specific primers were designed (Table 1), and PCR 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, 150 ng of primers, and 2.5 units of AmpliTaq Gold DNA polymerase (Roche Molecular Systems, Branchburg, NJ) on a Perkin-Elmer 9600 thermal cycler. The cycling conditions were 94°C for 9 min to activate the Taq Gold DNA polymerase, followed by 43 cycles of 94°C for 30 s, 63°C for 1 min, and a final extension step at 63°C for 10 min. Equal amounts of PCR products were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining. All primers for RT-PCR spanned at least two exons to avoid contamination by genomic DNA.

Table 1 *Primers used for RT-PCR analysis*

Gene	Primer name	Sequence[a]	Product size (bp)
KLK14	L6-F1	TCA-GGT-TGC-CCT-GGG-CAA-GC	647 (with L6-R1)
	L6-F2	GCA-CCT-AGC-AGA-TCC-CAA-GC	1133 (with L6-R1)
	L6-R1	CCT-GGA-CTC-CTG-GGT-CTG-AG	
PSA	PSAS	TGC-GCA-AGT-TCA-CCC-TCA	754
	PSAAS	CCC-TCT-CCT-TAC-TTC-ATC-C	
Actin	ACTINS	ACA-ATG-AGC-TGC-GTG-TGG-CT	372
	ACTINAS	TCT-CCT-TAA-TGT-CAC-GCA-CGA	

[a] All nucleotide sequences are given in the 5′→3′ orientation.

Normal and Malignant Breast Tissues. Normal breast tissues were obtained from women undergoing reduction mammoplasties. Breast tumor tissues were obtained from female patients at participating hospitals of the Ontario Provincial Steroid Hormone Receptor Program. The normal and tumor tissues were immediately frozen in liquid nitrogen after surgical resection and stored in this manner until extracted. The tissues were pulverized with a hammer under liquid nitrogen, and RNA was extracted as described above, using Trizol reagent.

Normal and Breast Cancer Cell Lines. The MCF-12A normal breast cell line and the BT-474 and T-47D breast cancer cell lines were purchased from the American Type Culture Collection (Manassas, VA). Cells were cultured in RPMI 1640 (Life Technologies, Inc.) supplemented with glutamine (200 mmol/liter), bovine insulin (10 mg/liter), fetal bovine serum (10%), antibiotics, and antimycotics in plastic flasks to near confluence. The cells were then aliquoted into 24-well tissue culture plates and cultured to 50% confluence. Twenty-four h before the experiments, the culture media were changed into phenol red-free media containing 10% charcoal-stripped fetal bovine serum. The cells were cultured for 24 h and then harvested for mRNA extraction.

Normal and Malignant Prostate Tissues. Prostate tissue samples were obtained from 10 patients who had undergone radical retropubic prostatectomy for prostatic adenocarcinoma at the Charite University Hospital (Berlin, Germany). The patients did not receive any hormonal therapy before surgery. The use of these tissues for research purposes was approved by the Ethics Committee of the Charite Hospital. Fresh prostate tissue samples were obtained from the cancerous and noncancerous parts of the same prostates that had been removed. Small pieces of tissue were dissected immediately after removal of the prostate and stored in liquid nitrogen until analysis. Histological analysis from all of the tissue pieces was performed as described previously (17) to ensure that the tissue was either malignant or benign. The tissues were pulverized with a hammer under liquid nitrogen, and RNA was extracted using RNeasy Mini Spin columns (Qiagen Inc., Valencia, CA) following the manufacturer's instructions. Total RNA concentration and quality were determined spectrophotometrically.

Normal and Malignant Testicular Tissues. Ten pairs of normal and cancerous testicular tissues were obtained from 10 patients who had undergone radical orchiectomy for testicular tumors at the Charite Hospital. All patients had a histologically confirmed diagnosis of primary testicular cancer and received no treatment before surgery. Matched cancerous and normal tissues were obtained from the cancerous and noncancerous parts of the same testis. Histological analysis from all of the tissue pieces was performed as described previously. Tissue specimens were homogenized, and total RNA was extracted using RNeasy Mini Spin columns (Qiagen Inc.) following the manufacturer's instructions. Total RNA concentration and quality were determined spectrophotometrically.

Normal and Malignant Ovarian Tissues. Normal ovarian cDNA was obtained from Gene Pool cDNA collection (Invitrogen). Ovarian tumor tissues were obtained from patients who had undergone surgery for primary ovarian carcinoma at the Department of Obstetrics and Gynecology, Gynecological Oncology Unit, University of Turin (Turin, Italy). Tissue specimens with different types of ovarian tumors were confirmed histologically, and 50–100 mg of tumor tissue were pulverized on dry ice, followed by total RNA extraction using the Trizol method (Life Technologies, Inc.). After measurement of total RNA concentration by spectrophotometry, 4 μg of total RNA from each tissue were used to carry out first-strand cDNA synthesis using the SuperScript Preamplification System, as prescribed by the manufacturer (Life Technologies, Inc.). To test the success of cDNA synthesis, 1 μl of the reverse transcription product was amplified using PCR with primers specific for actin. Product was visualized on a 2% agarose gel stained with ethidium bromide.


```
tctttgcatc tccagctctc actttgtctc tGCACCTAGC AGATCCCAAG CTGGGGAATG
CCAGTTCTGG CACCAACCTT CCTGCTCCCT GCTGGGGCCT CTGCTCCCCC ATCTCTCAGG
AGTCGAAAGT GAGAAAGCAA Ggtgg ......Intron I......caagACATCAAGGA GGGACCTGTG
CCCTGCTCCA CATCCTCCCA CCTGCCGCCC GCAGAGCCTG CAGGCCCCGC CCCCCTCGTC
TCTGGTCCCT ACCTCTCTGC TGTGTCTTCA TGTCCCTGAG GGTCTTGGGC TCTGG gtaa
Intron II ...........acag GACCTGGCCC TCAGCCCCTA AA ⟨ATG⟩ TTC CTC CTG CTG
                                                    M   F   L   L   L
ACA GCA CTT CAA GTC CTG GCT ATA G gtaagaga......Intron III.......tag
 T   A   L   Q   V   L   A   I
CC  ATG ACA CAG AGC CAA GAG GAT GAG AAC AAG ATA ATT GGT GGC CAT
 A   M   T   Q   S   Q   E   D   E   N   K   I   I   G   G   H
ACG TGC ACC CGG AGC TCC CAG CCG TGG CAG GCG GCC CTG CTG GCG GGT
 T   C   T   R   S   S   Q   P   W   Q   A   A   L   L   A   G
CCC AGG CGC CGC TTC CTC TGC GGA GGC GCC CTG CTT TCA GGC CAG TGG
 P   R   R   F   L   C   G   L   A   G   L   L   S   G   Q   W
GTC ATC ACT GCT GCT CAC TGC GGC CGC CC gtaagtgac..............
 V   I   T   A   A  ⊞H   C   G   R   P
................Intron IV...............gccag G ATC CTT CAG GTT GCC CTG GGC
                                              I   L   Q   V   A   L   G
AAG CAC AAC CTG AGG AGG TGG GAG GCC ACC CAG CAG GTG CTG CGC GTG
 K   H   N   L   R   R   W   E   A   T   Q   Q   V   L   R   V
GTT CGT CAG GTG ACG CAC CCC AAC TAC AAC TCC CGG ACC CAC GAC AAC
 V   R   Q   V   T   H   P   N   Y   N   S   R   T   H   D   N
GAC CTC ATG CTG CTG CAG CTA CAG CAG CCC GCA CGG ATC GGG AGG GCA
⊞D   L   M   L   L   Q   L   Q   Q   P   A   R   I   G   R   A
GTC AGG CCC ATT GAG GTC ACC CAG GCC TGT GCC AGC CCC GGG ACC TCC
 V   R   P   I   E   V   T   Q   A   C   A   S   P   G   T   S
TGC CGA GTG TCA GGC TGG GGA ACT ATA TCC AGC CCC ATC G gtgaggacc
 C   R   V   S   G   W   G   T   I   S   S   P   I
...........Intron V............................ccccag CC
                                                        A
AGG TAC CCC GCC TCT CTG CAA TGC GTG AAC ATC AAC ATC TCC CCG GAT
 R   Y   P   A   S   L   Q   C   V   N   I   N   I   S   P   D
GAG GTG TGC CAG AAG GCC TAT CCT AGA ACC ATC ACG CCT GGC ATG GTC
 E   V   C   Q   K   A   Y   P   R   T   I   T   P   G   M   V
TGT GCA GGA GTT CCC CAG GGC GGG AAG GAC TCT TGT CAG gtaaggcc.........
 C   A   G   V   P   Q   G   G   K   D   S   C   Q
.............Intron VI...............ccctcag GGT GAC ⊞TCT GGG GGA CCC
                                             G   D   S   G   G   P
CTG GTG TGC AGA GGA CAG CTC CAG GGC CTC GTG TCT TGG GGA ATG GAG
 L   V   C   R   G   Q   L   Q   G   L   V   S   W   G   M   E
CGC GCC CTG CCT GGC TAC CCC GGT GTC TAC ACC AAC CTG TGC AAG
 R   A   L   P   G   Y   P   G   V   Y   T   N   L   C   K
TAC AGA AGC TGG ATT GAG GAA ACG ATG CGG GAC AAA ⟨TGA⟩ TGGTCTTCACG
 Y   R   S   W   I   E   E   T   M   R   D   K
GTGGGATGGACCTCGTCAGCTGCCCAGGCCCTCCTCTCTCTACTCAGGACCAGGAGTCCAGGCCCCAGCCCCTCCTCC
CTCAGACCCAGGAGTCCAGGCCCCCAGCCCCTCCTCCTCAGACCCGGGAGTCCAGGCCCCAGCCCCTCCTCCCTCA
GACCCAGGAGTCCAGGCCCCAGCCCCTCCTCCCTCAGACCCGGGAGTCCAGGCCCCAGCCCCTCCTCCCTCAGACCC
AGGAGTCCAGGCCCCAGTCCCTCCTCCCTCAGACCCCAGGAGTCCAGGCCCCCAGCCCCTCCTCCCTCAGACCCAGGAA
TCCAGGCCCAGCCCCTCCTCCCTCAGACCCAGGAGCCCCAGTCCCCCAGCCCCTCCTCCTTGAGACCCAGGAGTCCAG
GCCCAGCCCCTCCTCCCTCAGACCCAGGAGCCCCAGTCCCCAGCATCCTGATCTTTACTCCGGCTCTGATCTCTCCTT
TCCCAGAGCAGTTGCTTCAGGCGTTTTCTCCCCACCAAGCCCCCACCCTTGCTGTGTCACCATCACTACTCAAGACCG
GAGGCACAGAGGGCAGGAGCACAGACCCCTTAAACCGGCATTGTATTCCAAAGACGACAATTTTTAACACGCTTAGTG
TCTCTAAAAACCGAATAAATAATGACAATAAAAA
```
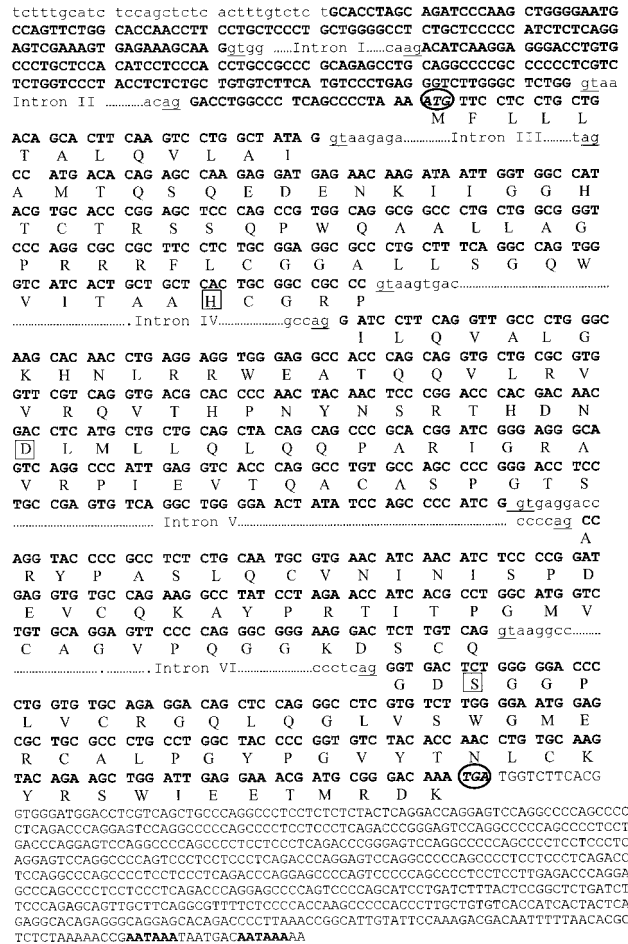
Fig. 1. Genomic organization and partial genomic sequence of the *KLK14* gene. Intronic sequences are not shown except for short sequences around the splice junctions. Introns are shown with *lowercase letters,* and exons are shown with *bold capital letters*. The 3′-untranslated region is shown in *capital non-bold letters.* For full sequence, see GenBank accession number AF161221. The start and stop codons are *encircled,* and the exon-intron junctions are *underlined.* The translated amino acids of the coding region are shown *underneath* by a *single-letter abbreviation.* The catalytic residues are *boxed.* Putative polyadenylation signals are shown in *bold.* The 3′-end of the gene has been verified for only 110 bases after the stop codon.

Structure Analysis. Multiple alignment was performed using the Clustal X software package[5] and the multiple alignment program[6] available from the Baylor College of Medicine (Houston, TX). Phylogenetic studies were performed using the Phylip software package.[7] Distance matrix analysis was performed using the Neighbor-joining/UPGMA program, and parsimony analysis was done using the Protpars program. The hydrophobicity study was performed using the Baylor College of Medicine **s**earch launcher programs.[8] Signal peptide was predicted using the SignalP server.[9] Protein structure analysis was performed by SAPS (structural analysis of protein sequence) program.[10]

RESULTS

Identification of the *KLK14* Gene. The genomic area encompassing the putative human KLK gene locus was subjected to different gene prediction programs as described above. We predicted a new gene formed of at least five exons. Screening of the human EST

[5] ftp://ftp.ebi.ac.uk/pub/software/dos/clustalw/clustalx/(clustalx1.64b.msw.exe).
[6] kiwi.imgen.bcm.tmc. edu:8808/search-launcher/launcher/html.
[7] http://evolution.genetics.washington.edu/phylip/getme.html.
[8] http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html.
[9] http://www.cbs.dtu.dk/services/signal.
[10] http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html.
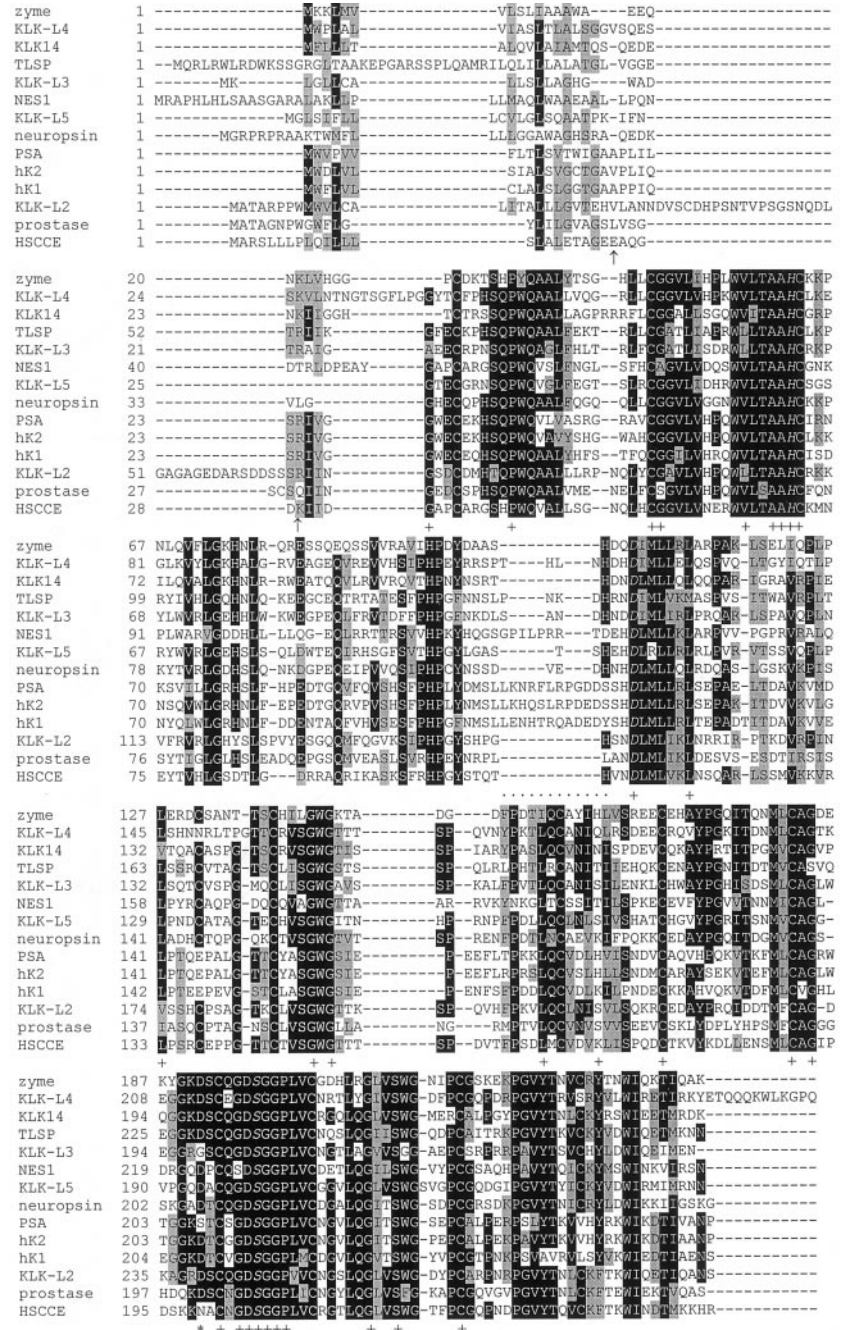
Fig. 2. Alignment of the deduced amino acid sequence of hK14 with other members of the KLK multigene family. For full gene names, see the text. *Dashes* represent gaps to bring the sequences to better alignment. The residues of the catalytic triad are represented in *italic,* and the 29 invariant serine protease residues are marked with +. Identical residues are *highlighted in black,* and similar residues are *highlighted in gray.* The predicted cleavage sites of signal and activation peptides of hK14 are indicated by *vertical arrows.* The *dotted area* represents the KLK loop sequence. The amino acid of the trypsin-like specificity (D) is marked with an *asterisk.*

database dbEST revealed an EST clone (GenBank accession number AI935298) with 99% identity with our predicted exons. The full-length sequence of the EST was compared with the genomic sequence of the putative gene and showed 100% identity with three areas (exons 5–7, according to our GenBank submission number AF161221; the new gene was originally designated *KLK-L6* in our GenBank submission). The three exons were separated by introns of variable length. The presence of two conserved amino acids of the catalytic triad of serine proteases suggested that this new gene encodes for a novel serine protease. Sequence homology and phylogenetic analysis further revealed that this new gene is closely related to the KLK subfamily of serine proteases (as discussed below). These three exons were further experimentally verified by RT-PCR analysis using primers L6-F1 and L6-R1 and brain cDNA as template (Table 1). Exon 4 was identified based on: (*a*) a consensus exon/intron splice site (AG . . . . GT); (*b*)

preservation of intron phases of the KLK genes (see "Discussion" for details); (*c*) the presence of the histidine residue of the catalytic triad (His[67]) surrounded by a well-conserved peptide motif just before the end of this exon (Fig. 1); and (*d*) comparable exon length to other KLK genes. Furthermore, exon 4 was experimentally verified by RT-PCR of brain cDNA, followed by sequencing of the PCR product.

To obtain the full structure of the gene, several RT-PCR reactions were performed with cDNA from different tissues, under different reaction conditions, using the reverse primer (L6-R1) together with different forward primers located in upstream putative exons predicted by the computer programs. Only one of these primers (L6-F2; Table 1) was able to produce a PCR product of ~1.1 kb. This PCR product was fully sequenced. By aligning the sequence obtained from the PCR reaction with the genomic sequence of the putative gene, we were able to characterize the full genomic structure of the gene.
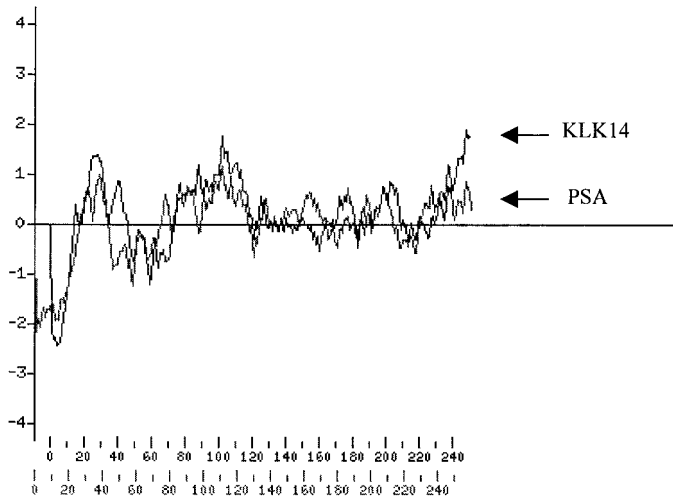
Fig. 3. Plot of hydrophobicity and hydrophilicity of hK14 protein, compared to PSA. The hydrophobic NH$_2$ terminus may harbor a signal and activation peptide. For details, see the text.
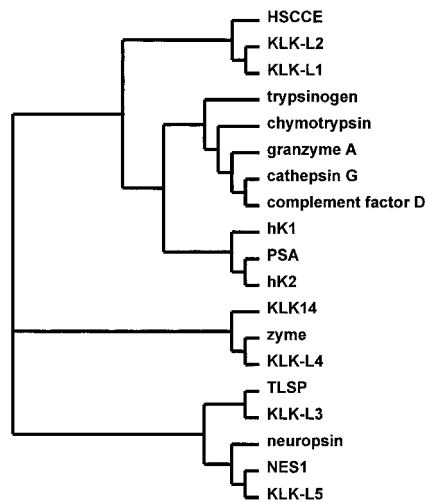
an intron phase of I, in all known KLK genes. A stop codon was found 153 bp downstream from the start of the last coding exon that harbors the serine residue of the catalytic triad, at a distance that is almost identical in all KLK genes (Fig. 1).

Structural Characterization of the *KLK14* Gene. The *KLK14* gene, as presented in Fig. 1, is formed of seven exons and six intervening introns, spanning an area of 6349 bp of genomic sequence on chromosome 19q13.3–q13.4. The lengths of the exons are 110, 135, 62, 172, 254, 137, and 263 bp, respectively (the 3′-untranslated region may be longer). The intron/exon splice sites (mGT . . . . AGm) and their flanking sequences are closely related to the consensus splicing sites (-mGTAAGT . . . CAGm-) (21).

As is the case with some other KLK genes, the first two exons are untranslated (22, 23), and the start codon is located at a distance of 22 bp from the start of the first coding exon. The predicted protein-coding region of the gene is formed of 756 bp encoding a deduced 251-amino acid polypeptide chain with a predicted molecular weight of 27,500. The 3′-end of the gene is not yet fully verified; however,

Fig. 4. Dendrogram of the predicted phylogenetic tree for some serine proteases and KLK proteins using the Neighbor-joining/UPGMA method. The tree grouped the classical KLKs (hK1, hK2, and PSA) together and aligned the hK14 protein in one group with zyme and KLK-L4. Other serine proteases were aligned in different groups.

Because five other new KLK genes were previously identified in this region (6, 14, 18, 19), this novel gene was named *KLK-L6*. With the new human KLK gene nomenclature, this gene was designated the official symbol *KLK14*. Translation of the mRNA of the gene in all reading frames indicated that there is only one reading frame that produces a full-length polypeptide chain without interrupting in-frame stop codons. Further support was obtained by the observation that this was the only frame that will preserve the three amino acid residues essential for serine protease activity (the catalytic triad) and the conserved motifs surrounding them (Fig. 1). Two putative in-frame methionine start codons were found in positions 2803 and 3716 of the genomic sequences (numbers refer to our GenBank submission number AF161221). The first is most likely the actual start codon of the gene because (*a*) it has an adenine nucleotide in position −3, an observation that was found to be conserved in 97% of vertebrate mRNAs (20); (*b*) in all other known KLKs, the start codon is located in the exon preceding the exon containing the histidine residue of the catalytic triad (6); and (*c*) this start codon is located at a comparable distance from the end of the first coding exon, which always ends with
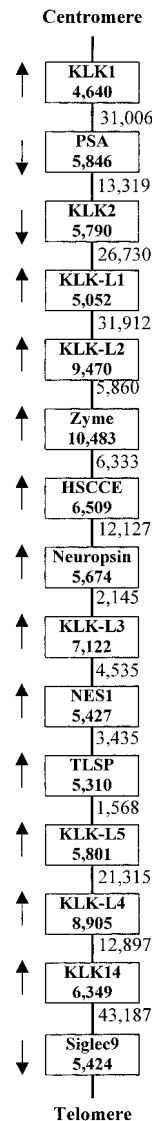

Fig. 5. An approximate 300-kb region of contiguous genomic sequence around chromosome 19q13.3–q13.4. The direction of transcription is shown by the *arrows*. *Boxes* represent genes, and the genomic lengths are shown in bp. Distances between genes in bp are also shown next to the connecting lines. For full gene names, see the text. The figure is not drawn to scale. The Siglec9 gene does not belong to the KLK gene family and is shown to mark the telomeric end of the KLK gene family.

3428

Fig. 6. Tissue expression of the *KLK14* gene as determined by RT-PCR (*top panel*). Actin (*bottom panel*) is a control gene. *KLK14* is expressed at the highest levels in the CNS. For further discussion, see the text.
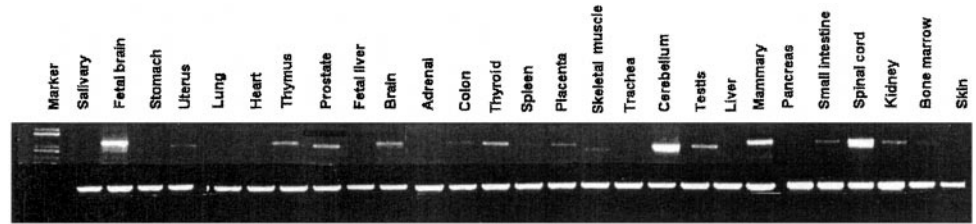
two putative polyadenylation signals (AATAAA) are located at positions 8060 and 8073 of the genomic sequence, at a comparable distance from the stop codon to that of the PSA and *KLK2* genes. The presence of aspartate (D) at position 198 suggests that *KLK14* possesses a trypsin-like cleavage pattern similar to that in most of the other KLKs but different from that in PSA, which has a serine (S) residue in the corresponding position, and is known to have chymotrypsin like activity (Fig. 2; Ref. 6).

Fig. 3 shows a comparative hydrophobicity analysis of the hK14 and PSA proteins. The NH_2-terminal regions of both proteins are quite hydrophobic, indicating that this region of hK14 possibly harbors a signal peptide, analogous to PSA. Fig. 3 also shows the presence of several evenly distributed hydrophobic regions throughout the hK14 polypeptide, which are consistent with a globular protein, similar to other serine proteases (24). Computer analysis of the amino acid sequence of hK14 predicted a cleavage site between amino acids 18 and 19 (TQS-QE). Sequence alignment (Fig. 2) also revealed a potential cleavage site (Lys[24]) at a site homologous to other serine proteases (lysine or arginine are present in most cases; Ref. 24). Thus, as is the case with PSA, and although direct experimental evidence is lacking, *KLK14* is presumably translated as an inactive 251-amino acid preproenzyme precursor. Prepro-KLK14 has 24 additional residues that constitute the preregion (18 residues), which is the signal peptide, and the propeptide (6 residues).

hK14 Homology with Other Members of the KLK Gene Family. Comparative analysis revealed that hK14 is highly homologous to other members of the KLK multigene family. It shows 48% protein identity with zyme/protease M/neurosin (hK6) and 47% identity with the neuropsin (hK8), human stratum corneum chymotryptic enzyme (HSSCE; hK7), and trypsin-like serine protease (hK11). Multiple alignment shows that the typical catalytic triad of serine proteases is conserved in the hK14 protein (His[67], Asp[111], and Ser[204]; Fig. 2). In addition, a well-conserved peptide motif is found around the amino acid residues of the catalytic triad, as is the case with other

serine proteases [*i.e.,* histidine (VITAA*H*C), serine (GD*S*GGP), and aspartate (*D*LMLL); Refs. 25 and 26; Fig. 2]. Fig. 2 also shows other amino acid residues that are completely conserved between KLKs and KLK-L proteins and shows an 11-amino acid loop characteristic of the classical KLKs (PSA, hK1, and hK2) that is not found in hK14 or other members of the KLK multigene family (18, 19, 22–24).

Sequence analysis of eukaryotic serine proteases indicates the presence of 29 invariant amino acids (27). Twenty-seven of them are conserved in the hK14 polypeptide. One of the nonconserved amino acids (Val[132] instead of Leu) is also found in the KLK-L2 (hK5) protein, and according to protein evolution studies, this amino acid change represents a conserved evolutionary substitution to a protein of the same group (27, 28). The remaining amino acid (Asn[156] instead of Pro) is not conserved among all other human KLKs (Fig. 2). Twelve cysteine residues are present in the mature hK14 protein; 10 of them are conserved in all KLKs, and the remaining 2 (Cys[136] and Cys[238]) are present in most of the other KLK-L proteins but not in the classical KLKs, and they are expected to form an additional disulfide bridge (Fig. 2).

To predict the phylogenetic relatedness of the hK14 protein, the amino acid sequences of the KLK proteins and some other serine proteases were aligned together using the Clustal X multiple alignment program, and a distance matrix tree was predicted using the Neighbor-joining/UPGMA and Protpars parsimony methods. Fig. 4 shows the phylogenetic analysis that separated the classical KLKs (hK1, hK2, and PSA) and clustered hK14 with zyme and KLK-L4 proteins in a separate group away from other serine proteases, consistent with previously published studies (19, 29) and indicating that this group of genes probably arose from a common ancestral gene by gene duplication.

Mapping and Chromosomal Localization of the *KLK14* Gene. Alignment of the *KLK14* gene and the sequences of other known KLK genes within the 300-kb area of interest enabled us to precisely localize all genes and to determine the direction of transcription, as shown in Fig. 5. The *KLK1* gene is the most centromeric, transcribing from telomere to centromere, followed by *PSA* and then by *KLK2,* which transcribe in the opposite direction. All other KLK-L genes transcribe in the same direction as *KLK1*. *KLK14* resides 12,897 bp telomeric to *KLK-L4,* and 43,187 bp centromeric to a novel Siglec gene (GenBank submission number AF135027). The Siglec gene is not a serine protease and has no homology to any of the KLK gene family members.

Tissue Expression of the *KLK14* Gene. As shown in Fig. 6, the *KLK14* gene is primarily expressed in tissues of the CNS, including the brain, cerebellum, and spinal cord. As is the case with other KLKs, lower levels of expression are found in many other tissues including breast, thyroid, kidney, colon, spleen, bone marrow, uterus, small intestine, and placenta. To verify the RT-PCR specificity, the PCR products were cloned and sequenced.

Expression of *KLK14* in Normal and Malignant Tissues. We studied the differential expression of *KLK14* in different normal and malignant tissues. To characterize the extent and frequency of expression of the *KLK14* gene in breast tumors compared to normal breast
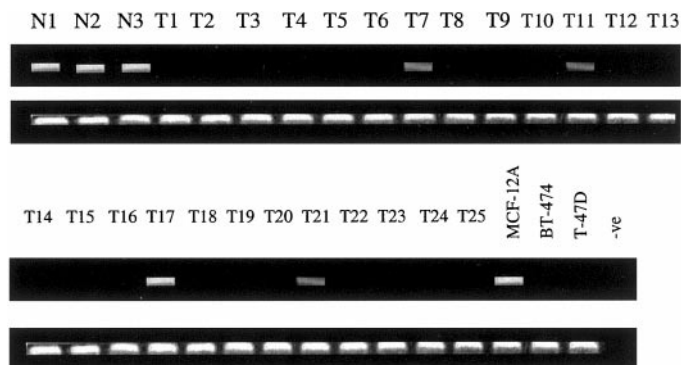

Fig. 7. Expression of the *KLK14* gene in normal and cancerous breast tissues and cell lines. The gene is expressed in all 3 normal tissues (*N1, N2,* and *N3*) but is not expressed at all in 21 tumors, has lower than normal expression in 3 tumors, and has an expression level comparable to that in normal tissue in 1 tumor (*T,* tumor). Whereas the gene was expressed in the normal breast cell line (MCF-12A), it was not detected in the two breast cancer cell lines examined (BT-474 and T-47D). *–ve,* negative PCR control. For data interpretation, see the text. The actin gene (*bottom panel*) was used as a control.
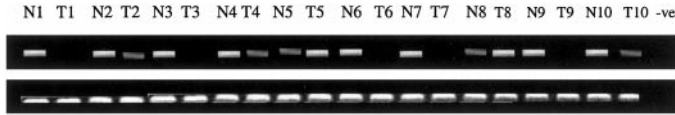
Fig. 8. Expression of the *KLK14* gene in 10 matched pairs of normal and malignant prostate tissues. The gene is down-regulated in 8 of 10 tumor tissues examined. *N,* normal; *T,* tumor. *–ve,* negative PCR control. The *bottom panel* represents expression of the actin gene (control).

tissues, we used cDNA derived from 3 normal and 25 malignant breast tissues. *KLK14* gene expression was detectable in all three normal tissues. Of the 25 tumors, *KLK14* gene expression was undetectable in 21 tumors, lower than normal in 3 tumors, and comparable to that of the normal tissues in 1 tumor (Fig. 7). We also compared *KLK14* gene expression in the MCF-12A normal breast cell line and two breast cancer cell lines (BT-474 and T-47D). *KLK14* was highly expressed in the normal breast cell line but was undetectable in both breast cancer cell lines examined (Fig. 7).

Matched pairs of normal and malignant tissues from 10 patients were also examined to study the differential expression of *KLK14* in both prostate and testicular cancer. In prostatic tissues, the expression of the gene was absent or lower in eight tumor tissues compared with their normal counterparts and was higher than normal in only two tumor tissues (Fig. 8). *KLK14* was also found to have lower or no expression in 7 of 10 testicular tumor tissues examined (compared with their normal counterparts), a comparable level of expression in 1 tumor, and higher than normal expression in 2 tumors. Whereas *KLK14* expression was easily detectable in the normal ovarian cDNA examined, it was expressed in only 1 of the 15 ovarian cancer tissues examined. In all cases, the actin gene was used as a control gene to check for the quality and quantity of the cDNA used. These preliminary results suggest that this gene is down-regulated, at the mRNA level, in diverse malignancies.

DISCUSSION

KLKs are a subgroup of serine proteases that play important roles in diverse physiological processes (2, 7). A number of structural and localization criteria have been proposed to classify putative KLKs into

this multigene family (30, 31) As shown in Fig. 9, KLKs are characterized by the following common structural features: (*a*) all genes are formed of five coding exons and four intervening introns [some genes have extra 5′-untranslated exon(s) (22, 23, 32)]; (*b*) the coding exon lengths are usually comparable or identical between family members; (*c*) the intron phases are always conserved (I-II-I-0; see the Fig. 9 legend for definition of intron phases); (*d*) these genes are clustered in the same chromosomal region, without any intervening non-KLK-L genes (Fig. 5); and (*e*) the histidine residue of the catalytic triad of serine proteases is located near the end of the second coding exon, the aspartate residue is located in the middle of the third coding exon, and the serine is located at the beginning of the fifth coding exon. As shown in Fig. 9, all these criteria apply to the newly identified *KLK14* gene. Thus, *KLK14* should be considered as a new member of the KLK multigene family.

The wide range of tissue expression of *KLK14* should not be surprising because by using the more sensitive RT-PCR technique instead of Northern blot analysis, many KLK genes were found to be expressed in a wide variety of tissues including the salivary glands, kidney, pancreas, brain, and tissues of the reproductive system (uterus, mammary gland, ovary, and testis; Ref. 2).

Our preliminary finding that the *KLK14* gene is down-regulated at the mRNA level in different malignancies is in accord with findings for other KLKs [*e.g., PSA, NES1,* and *protease M* (6)]. There is now growing evidence that many of the KLKs and KLK-L genes are related to malignancy. PSA is a useful biomarker for prostate cancer (8). A recent report provided evidence that PSA has antiangiogenic activity and that this activity is due to its action as a serine protease (33). This study also suggested that other serine proteases, including new members of the KLK multigene family of enzymes, should also be evaluated for potential antiangiogenic action. Recent reports suggest that hK2 could be another useful diagnostic marker for prostate cancer (9). *NES1* appears to be a tumor suppressor gene (10). The *zyme/protease M/neurosin* gene is differentially expressed in primary breast and ovarian tumors (11), and the human stratum corneum chymotryptic enzyme (*HSCCE*) has been shown to be expressed at abnormally high levels in a subset of ovarian cancers (13). Another recently identified KLK-L gene, tumor-associated differentially ex-

Fig. 9. Schematic diagram showing the comparison of the genomic structure of *KLK14* with all other known genes of the human KLK family. Exons are shown by *boxes,* and introns are shown by the *connecting lines. Arrowhead* marks the site of the start codon, and the *arrow* represents the stop codon. *Letters above boxes* indicate relative positions of the amino acids of the catalytic triad. *H,* histidine; *D,* aspartic acid; *S,* serine. *Roman numerals* indicate intron phases. The intron phase refers to the location of the intron within the codon; *I* denotes that the intron occurs after the first nucleotide of the codon; *II* denotes that the intron occurs after the second nucleotide; *0* denotes that the intron occurs between codons. The intron phases were found to be identical in all genes. *Numbers inside boxes* indicate exon lengths in bp, and *numbers below the connecting lines* show the intron lengths. *Question marks* indicate that exact length is not accurately known. The figure is not drawn to scale. We indicate the official symbol for each gene, as defined by the Human Gene Nomenclature Committee, in *parentheses.*
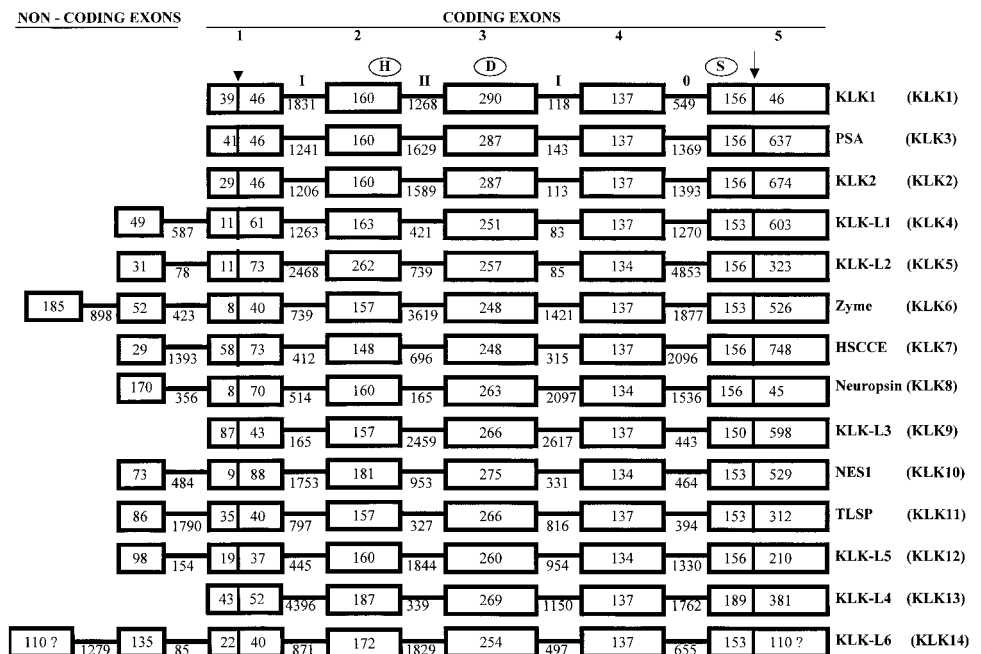
pressed gene 14 (*TADG14/neuropsin*), was found to be overexpressed in about 60% of ovarian cancer tissues (12). *Prostase/KLK-L1,* another newly discovered KLK-L gene, is speculated to be linked to prostate cancer (29). A newly discovered KLK, *KLK-L4,* was found to be down-regulated in breast cancer (34). Thus, extensive recent literature suggests multiple connections of many KLK genes to various forms of human cancer (reviewed in Ref. 6).

Many KLKs seem to play important physiological roles in the CNS. In mouse, neuropsin appears to have an important role in neural plasticity, and the amount of neuropsin mRNA is related to memory retention after a chemically induced ischemic insult (35). The human neuropsin gene was first isolated from the hippocampus (23). Another KLK, the *zyme/protease M/neurosin* gene, was isolated from Alzheimer's disease brain and shown to have amyloidogenic activity (25, 26). Zyme was also found to be localized in perivascular cells and microglial cells in human Alzheimer's disease brain (25). Trypsin-like serine protease (*TLSP*), another newly discovered KLK gene, was isolated from brain hippocampus cDNA and is thought to play a role in brain plasticity (36). Finally, the *KLK-L2* gene is also expressed at high levels in the brain (19). Thus, we speculate that the *KLK14* gene, being expressed at high levels in all examined CNS tissues, might have a role in brain physiology and pathobiology.

In conclusion, we have cloned a new KLK gene and describe its genomic structure, chromosomal localization, and tissue expression pattern. Our preliminary results suggest that the gene is down-regulated in diverse cancer tissues. This gene should be examined as a potential tumor marker because it is predicted to be secreted and is likely present in the circulation. Also, the high level of expression of the *KLK14* gene in the CNS suggests that this gene may have an important function in brain physiology.

REFERENCES

1. Tilghman, S. M. Lessons learned, promises kept: a biologist's eye view of the Genome Project. Genome Res., *6:* 773–780, 1996.
2. Clements, J. The molecular biology of the kallikreins and their roles in inflammation. *In:* S. Farmer (ed.), The Kinin System, pp. 71–97. New York: Academic Press, 1997.
3. Evans, B. A., Yun, Z. X., Close, J. A., Tregear, G. W., Kitamura, N., Nakanishi, S., Callen, D. F., Baker, E., Hyland, V. J., Sutherland, G. R., *et al.* Structure and chromosomal localization of the human renal kallikrein gene. Biochemistry, *27:* 3124–3129, 1988.
4. Schedlich, L. J., Bennetts, B. H., and Morris, B. J. Primary structure of a human glandular kallikrein gene. DNA, *6:* 429–437, 1987.
5. Riegman, P. H., Vlietstra, R. J., van der Korput, J. A., Romijn, J. C., and Trapman, J. Characterization of the prostate-specific antigen gene: a novel human kallikrein-like gene. Biochem. Biophys. Res. Commun., *159:* 95–102, 1989.
6. Diamandis, E. P., Yousef, G. M., Luo, L. Y., Magklara, A., and Obiezu, C. V. The new human kallikrein gene family: implications in carcinogenesis. Trends Endocrinol. Metab., *11:* 54–60, 2000.
7. Margolius, H. S. Tissue kallikreins: structure, regulation, and participation in mammalian physiology and disease. Clin. Rev. Allergy Immunol., *16:* 337–349, 1998.
8. Diamandis, E. P. Prostate-specific antigen: its usefulness in clinical medicine. Trends Endocrinol. Metab., *9:* 310–316, 1998.
9. Stenman, U. H. New ultrasensitive assays facilitate studies on the role of human glandular kallikrein (hK2) as a marker for prostatic disease. Clin. Chem., *45:* 753–754, 1999.
10. Goyal, J., Smith, K. M., Cowan, J. M., Wazer, D. E., Lee, S. W., and Band, V. The role for NES1 serine protease as a novel tumor suppressor. Cancer Res., *58:* 4782–4786, 1998.
11. Anisowicz, A., Sotiropoulou, G., Stenman, G., Mok, S. C., and Sager, R. A novel protease homolog differentially expressed in breast and ovarian cancer. Mol. Med., *2:* 624–636, 1996.
12. Underwood, L. J., Tanimoto, H., Wang, Y., Shigemasa, K., Parmley, T. H., and O'Brien, T. J. Cloning of tumor-associated differentially expressed gene-14, a novel

13. serine protease overexpressed by ovarian carcinoma. Cancer Res., *59:* 4435–4439, 1999.
13. Tanimoto, H., Underwood, L. J., Shigemasa, K., Yan Yan, M. S., Clarke, J., Parmley, T. H., and O'Brien, T. J. The stratum corneum chymotryptic enzyme that mediates shedding and desquamation of skin cells is highly overexpressed in ovarian tumor cells. Cancer (Phila.), *86:* 2074–2082, 1999.
14. Yousef, G. M., Luo, L. Y., and Diamandis, E. P. Identification of novel human kallikrein-like genes on chromosome 19q13.3-q13.4. Anticancer Res., *19:* 2843–2852, 1999.
15. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res., *25:* 3389–3402, 1997.
16. Lennon, G., Auffray, C., Polymeropoulos, M., and Soares, M. B. The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. Genomics, *33:* 151–152, 1996.
17. Meyer, A., Jung, K., Lein, M., Rudolph, B., Schnorr, D., and Loening, S. A. Factors influencing the ratio of free to total prostate-specific antigen in serum. Int. J. Cancer, *74:* 630–636, 1997.
18. Yousef, G. M., Obiezu, C. V., Luo, L. Y., Black, M. H., and Diamandis, E. P. 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, 1999.
19. Yousef, G. M., and Diamandis, E. P. The new kallikrein-like gene, *KLK-L2*. Molecular characterization, mapping, tissue expression, and hormonal regulation. J. Biol. Chem., *274:* 37511–37516, 1999.
20. Kozak, M. An analysis of vertebrate mRNA sequences: intimations of translational control. J. Cell Biol., *115:* 887–903, 1991.
21. Iida, Y. Quantification analysis of 5′-splice signal sequences in mRNA precursors. Mutations in 5′-splice signal sequence of human β-globin gene and β-thalassemia. J. Theor. Biol., *145:* 523–533, 1990.
22. Yousef, G. M., Luo, L. Y., Scherer, S. W., Sotiropoulou, G., and Diamandis, E. P. Molecular characterization of zyme/protease M/neurosin (PRSS9), a hormonally regulated kallikrein-like serine protease. Genomics, *62:* 251–259, 1999.
23. Yoshida, S., Taniguchi, M., Hirata, A., and Shiosaka, S. Sequence analysis and expression of human neuropsin cDNA and gene. Gene (Amst.), *213:* 9–16, 1998.
24. Liu, X. L., Wazer, D. E., Watanabe, K., and Band, V. Identification of a novel serine protease-like gene, the expression of which is down-regulated during breast cancer progression. Cancer Res., *56:* 3371–3379, 1996.
25. Little, S. P., Dixon, E. P., Norris, F., Buckley, W., Becker, G. W., Johnson, M., Dobbins, J. R., Wyrick, T., Miller, J. R., MacKellar, W., Hepburn, D., Corvalan, J., McClure, D., Liu, X., Stephenson, D., Clemens, J., and Johnstone, E. M. Zyme, a novel and potentially amyloidogenic enzyme cDNA isolated from Alzheimer's disease brain. J. Biol. Chem., *272:* 25135–25142, 1997.
26. Yamashiro, K., Tsuruoka, N., Kodama, S., Tsujimoto, M., Yamamura, Y., Tanaka, T., Nakazato, H., and Yamaguchi, N. Molecular cloning of a novel trypsin-like serine protease (neurosin) preferentially expressed in brain. Biochim. Biophys. Acta, *1350:* 11–14, 1997.
27. Dayhoff, M. O. The origin and evolution of protein superfamilies. Fed Proc., *35:* 2132–2138, 1976.
28. Miyata, T., Miyazawa, S., and Yasunaga, T. Two types of amino acid substitutions in protein evolution. J. Mol. Evol., *12:* 219–236, 1979.
29. Nelson, P. S., Gan, L., Ferguson, C., Moss, P., Gelinas, R., Hood, L., and Wang, K. Molecular cloning and characterization of prostase, an androgen-regulated serine protease with prostate-restricted expression. Proc. Natl. Acad. Sci. USA, *96:* 3114–3119, 1999.
30. Yousef, G. M., and Diamandis, E. P. The expanded human kallikrein gene family: locus characterization and molecular cloning of a new member, KLK-L3 (KLK9). Genomics, *65:* 184–194, 2000.
31. Irwin, D. M., Robertson, K. A., and MacGillivray, R. T. Structure and evolution of the bovine prothrombin gene. J. Mol. Biol., *200:* 31–45, 1988.
32. Luo, L., Herbrick, J. A., Scherer, S. W., Beatty, B., Squire, J., and Diamandis, E. P. Structural characterization and mapping of the normal epithelial cell-specific 1 gene. Biochem. Biophys. Res. Commun., *247:* 580–586, 1998.
33. Fortier, A. H., Nelson, B. J., Grella, D. K., and Holaday, J. W. Antiangiogenic activity of prostate-specific antigen. J. Natl. Cancer Inst., *91:* 1635–1640, 1999.
34. Yousef, G. M., Chang, A., and Diamandis, E. P. Identification and characterization of *KLK-L4*, a new kallikrein-like gene that appears to be down-regulated in breast cancer tissues. J. Biol. Chem. *275:* 11891–11898, 2000.
35. Akita, H., Matsuyama, T., Iso, H., Sugita, M., and Yoshida, S. Effects of oxidative stress on the expression of limbic-specific protease neuropsin and avoidance learning in mice. Brain Res., *769:* 86–96, 1997.
36. Yoshida, S., Taniguchi, M., Suemoto, T., Oka, T., He, X., and Shiosaka, S. cDNA cloning and expression of a novel serine protease, TLSP. Biochim. Biophys. Acta, *1399:* 225–228, 1998.