Research Article

TumorBiology

Tumor Biol 2004;25:282–289
DOI: 10.1159/000081393

Received: July 3, 2004
Accepted after revision: July 13, 2004

# In silico Analysis of the *Human Kallikrein Gene 6*

George M. Yousef[a] Carla A. Borgoño[b,c] Nicole M.A. White[a]
John-Desmond Robb[a] Iacovos P. Michael[b,c] Katerina Oikonomopoulou[b,c]
Saba Khan[b] Eleftherios P. Diamandis[b,c]

[a]Discipline of Pathology, Memorial University, St. John's, [b]Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, and [c]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada

**Key Words**
Serine proteases · *Human kallikrein gene 6* · Zyme · Protease M · Neurosin · Breast cancer · Cancer genes · Tumor markers · Prognostic/predictive factors · Gene expression

**Abstract**
Kallikreins are a family of 15 serine proteases clustered together on the long arm of chromosome 19. Recent reports have linked kallikreins to malignancy. The *human kallikrein gene 6 (KLK6)* is a newly characterized member of the human kallikrein gene family. Recent work has focused on the possible role of this gene and its protein product as a tumor marker and its involvement in diseases of the central nervous system. In this study, we performed extensive in silico analyses of *KLK6* expression from different databases using various bioinformatic tools. These data enabled us to construct and verify the longest transcript for this kallikrein, to identify several polymorphisms among published sequences and to summarize the 21 single-nucleotide polymorphisms of the gene. Our expressed sequence tag (EST) analyses suggest the existence of seven new splice variants of the gene, in addition to the already reported ones. Most of these variants were identified in libraries from cancerous tissues. *KLK6* orthologues were identified from three other species with approximately 86% overall homology with rat and mouse orthologues. We also utilized several databases to compare *KLK6* gene expression in normal and cancerous tissues. The serial analysis of gene expression and EST expression profiles showed upregulation of the gene in female genital (ovarian and uterine) and gastrointestinal (gastric, colon, esophageal and pancreatic) cancers. Significant downregulation was observed in breast cancers and brain tumors, in relation to their normal counterparts.



## Introduction

The new millennium is considered by many researchers as the era of bioinformatics, or in silico (computer-based) analysis. With the completion of the Human Genome Project, an enormous amount of information has been created and made available to the research community through several databases. In addition, several analytical tools became available through many institutions. One such major search engine is the National Cen-

KARGER

Fax + 41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com


1010–4283/04/0256–0282$21.00/0

Accessible online at:
www.karger.com/tbi

Dr. E.P. Diamandis
Mount Sinai Hospital, Department of Pathology and Laboratory Medicine
600 University Avenue
Toronto, Ontario M5G 1X5 (Canada)
Tel. +1 416 586 8443, Fax +1 416 586 8628, E-Mail ediamandis@mtsinai.on.ca

ter for Biotechnology Information (NCBI) with its GenBank and various other sections including the Expressed Sequence Tag (EST) databases and the UniGene clusters, among others. Another major source is the Cancer Genome Anatomy Project (CGAP), which aims to catalog all genes expressed in normal and cancerous conditions, and to develop the necessary tools for their analysis, in order to facilitate our understanding of the pathogenesis of cancer and to identify new cancer biomarkers [1]. CGAP libraries have been developed using two approaches, i.e. EST analysis [2] and the serial analysis of gene expression (SAGE) [3]. The SAGE method is similar to the EST approach, in that it provides a sequence tag for a portion of a cDNA. However, in this approach, the tagged sequences are generally quite short (10–14 bp), and the individual tags are annealed to generate chimeric DNA molecules carrying many individual tags. This method has been shown to more accurately quantify transcript expression levels and to increase the chances of identifying rarely expressed transcripts.

Our current study represents an in silico approach that provides a comprehensive analysis of all reported information on the *human kallikrein gene 6 (KLK6)* using various search engines. This approach has been used successfully and has proven to be a valuable tool that is directing further research experiments, and its results were shown experimentally to be accurate [4, 5]. In this paper, we focus on the comprehensive in silico analysis of *KLK6* gene structure, including differential splicing and single-nucleotide polymorphisms (SNPs), expression, including differential expression between normal and cancerous tissues, and identification of orthologues in other species.

## Materials and Methods

*KLK6* ESTs and mRNA clones were obtained from the non-redundant and EST databases of the NCBI (http://www.ncbi.nlm.nih.gov/) and the University of California at Santa Cruz (UCSC) (http://www.ucsc.edu/).

Information about the *KLK6* gene was obtained from the following databases and web sites:

- The UniGene clusters (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene),
- The Online Mendelian Inheritance in Man (OMIM) databases (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM),
- The Swiss-Prot protein knowledgebase (http://us.expasy.org/sprot/),
- The MEROPS databases (http://merops.sanger.ac.uk/),
- The GeneCards™ web site (http://genecards.bcgsc.ca/cgi-bin/randomize.pl),
- The Human Gene Nomenclature database (Genew) (http://www.gene.ucl.ac.uk/cgi-bin/nomenclature/searchgenes.pl),
- The CleanEx database of gene expression profiles (http://www.cleanex.isb-sib.ch/),
- The GenAtlas human gene database (http://www.dsi.univ-paris5.fr/genatlas/),
- GeneLynx, a portal to the human and mouse genomes (http://www.genelynx.org/),
- The Stanford Online Universal Resource for Clones and ESTs (SOURCE) (http://genome-www5.stanford.edu/cgi-bin/SMD/source/sourceSearch),
- The Gene Ontology (GO) database (http://www.geneontology.org/),
- BLOCKS, a protein domain database (http://blocks.fhcrc.org/), and
- Protein Data Bank, a repository for three-dimensional biological macromolecular structure data (http://www.rcsb.org/pdb/).

Multiple alignments were performed using the 'ClustalW' software package [6] and the BLAST programs of the NCBI, and manually edited. Splice variants were identified using multiple alignments of sequences obtained from the GenBank, UniGene and UCSC databases. EST and mRNA sequences were compared to the reference genomic sequence of *KLK6* (GenBank accession No. AF149289). Alignment viewings were done using the 'boxshade' (www.ch.embnet.org/software/BOX_form.html) and 'chroma' (www.lg.ndirect.co.uk/chroma/) programs. Taxonomy information was obtained from the HomoloGene and taxonomy browsers of the NCBI (http://www.ncbi.nlm.nih.gov).

Analysis of *KLK6* gene expression in cancer was performed through the CGAP databases. EST information was obtained from the GenBank and CGAP web servers. The mRNA sequence of the gene was used to identify unique sequence tags of UniGene clusters, and two restriction digestion enzymes (*Nla*III and *Sau*3A) were used as anchoring enzymes. These sequence tags were then used to determine the levels of expression in the SAGE libraries. Detailed information for these libraries is available from the CGAP web site (http://www.ncbi.nlm.nih.gov/ncicgap/). Analyses were performed by comparing the proportion of libraries of each type (cancer vs. normal) that show expression of each tag, in addition to the average expression densities in these libraries. If more than one tag of the same gene appeared in the same library, we only included the one with the peak level of expression [maximum tags per million (t.p.m.)]; the other tags were excluded to avoid inaccurate estimation of expression.

## Results

### *Analysis of the* KLK6 *mRNA Sequence*

Access information and parameters of the *KLK6* gene and its protein product obtained from different databases are presented in table 1. There are six published mRNA clones for *KLK6,* shown in table 2. Using multiple alignments of these clones, we constructed the longest available clone (by overlapping clones AF013988 and U62801). This clone is formed of 1,512 nucleotides. The 3′ end of this clone was verified by the presence of a poly A tail of 28 nucleotides. No other mRNA or EST clones with a longer 5′ extension were identified. The exact transcription start site could not, however, be verified, and the

**Table 1.** Access information of the *KLK6* gene and parameters of its protein product as obtained from different databases[1]

| *KLK6* gene | |
|---|---|
| Gene name | human tissue kallikrein gene 6 |
| Official symbol | *KLK6* (gene), hK6 (protein) |
| Synonyms | zyme, protease M, neurosin, PRSS9 |
| GenBank accession No. | NM_002774 (mRNA)<br>AF149289 (full genomic structure) |
| UniGene Cluster | Hs. 79361 |
| UCSC | FLJ20080 |
| OMIM | 602652 |
| GeneCards ID | GC19M056137 |
| Genew ID | HGNC:6367; KLK6 |
| CleanEx ID | HGNC:6367; KLK6 |
| GeneLynx ID | KLK6; *Homo sapiens* |
| GenAtlas ID | KLK6 |
| GO ID | GO:0008236 |
| SOURCE ID | KLK6; *Homo sapiens* |
| BLOCKS ID | Q92876 |
| Protein Data Bank ID | 1LO6 |
| *hK6 protein* | |
| Swiss-Prot | Q92876 |
| MEROPS ID | S01.236 |
| Isoelectric point | 7.15 |
| Molecular weight[2], kD | 26.8 |
| Signal peptide, aa | 16 |
| Activating peptide, aa | 5 |
| Mature protein, aa | 223 |
| Cysteine residues in the mature enzyme, aa | 12 |
| Substrate specificity | trypsin-like |

aa = Amino acids.
[1] For web site addresses, please see text.
[2] Excluding any posttranslational modifications.

**Table 2.** Reported *KLK6* mRNA sequences

| GenBank accession No. | Length, bp | Reference |
|---|---|---|
| NM_002774 | 1,512 | 7, 10 |
| U62801 | 1,506 | 7 |
| AF013988 | 1,451 | 11 |
| D78203 | 1,419 | 11 |
| AF149289 | 1,506 | 22 |
| BT006852 | 735 | unpubl. |

possibility of a 5′ extension still exists. Multiple alignments of all published mRNA sequences of *KLK6* identified six polymorphic sites, in comparison to the sequence submitted by Anisowicz et al. [7] (GenBank accession No. U62801). These are: A to G in position 12, A to C in position 1412, G to A in position 1440, A to G in position 1464, A to G in position 1479, and A to T in position 1501 (numbers refer to the reference sequence, GenBank accession No. NM_002774). However, these polymorphisms do not affect the encoded protein sequence as they occur in the 5′ and 3′ untranslated regions of the mRNA.

*Single Nucleotide Polymorphisms*

As shown in figure 1 and table 3, there are 22 reported SNPs in the *KLK6* genomic sequence. Eleven are located in the area flanking the 5′ end of the mRNA sequence. There are nine intronic SNPs and two that fall within exons. One SNP is located in each of exon 1 and exon 6. These SNPs are synonymous and do not elicit change in the coding sequence.

*Splice Variants of* KLK6

As shown in table 4 and figure 2, there are five published splice variants of *KLK6,* in addition to the classic form. Two of these variants were given the same designation (splice variant-1). Sequence analysis showed that clone AY279383 is identical to splice variant-3 (accession No. AY318868). A revised nomenclature of all kallikrein splice variants is needed and a uniform abbreviation system should be developed.

Comprehensive analysis of 185 EST clones from different databases enabled us to identify the following additional splice variants of the gene (fig. 1):

(1) One EST clone (GenBank accession No. BG 720964) that partially matches both *KLK6* and *KLK5.* This clone was isolated from a testicular tissue library. The first exon of the transcript starts in the *KLK6* gene with a 90-nucleotide 5′ extension of exon 5, with a conserved AG splice site, and ends at the normal donor site. The second exon starts with a downstream acceptor site within exon 7 of *KLK6* with a nonconserved acceptor site. The third exon represents exon 1 of the *KLK5* gene with 182 nucleotides of upstream extension, with a nonconserved splice site, and the fourth exon consists of 120 nucleotides located in intron II of *KLK5* with conserved splice sites (not shown).

(2) Fifty-three EST clones were found to match the *KLK6* splice variant-1 (GenBank accession No. AY318867) (upstream 5′ splice acceptor site of exon 2). Interestingly, twenty of these clones were isolated from a gastric carcinoma cell line and the remaining six clones (accession No. BG821445, BG468256, BG469139, BE870700, BG469249 and BG824290) were isolated


**Fig. 1.** Schematic of *KLK6* showing the locations of all SNPs. Exons are shown as boxes, with the exon number indicated above each exon and the length (in base pairs) indicated inside. Introns are represented as lines connecting each box to the next. The positions of the start and stop codons are indicated by stars. SNP positions are indicated by arrows. See also table 3 for exact locations.

from colon adenocarcinoma cell lines. Two of these EST clones, BG468256 and BG469249, have a new conserved splice acceptor site 212 nucleotides upstream of exon 2, which is longer than that of the longest reported clone of *KLK6* splice variant-1 (GenBank accession No. AY318867). This raises the possibility that this variant extends further upstream to the first exon.

(3) Another EST clone (BE717737) was found to have a partial match (the first 300 nucleotides) with the *KLK6* mRNA, but the rest of its sequences (the last 200 nucleotides) match 99% with genomic sequences from chromosome 11. No other ESTs were found to be similar to this clone.

(4) Two EST groups were found to be similar to *KLK6* splice variant-2, which has a 5′ extension of exon 2 and is missing the third exon. The first group (BM 760724, BM760718, BM845548) is from the gastric carcinoma cell line SNU-216 and the second group (BM847296, BM848065) was isolated from the KMS-5 myeloma cell line. An important observation is that most, if not all, of the splice variants are isolated from cancer cells.

(5) Another EST clone (BE717114), isolated from a head and neck library, starts at the middle of exon 5 and ends in intron 5, 327 nucleotides downstream of exon 5.

(6) An EST clone from fetal tissue (AA444372) has an alternative splice acceptor site 42 nucleotides downstream in exon five with a conserved AG acceptor site.

(7) An interesting clone is the EST (BQ361291) isolated from an unspecified ovarian tissue library. This clone is formed of three exons; the last two of them match the forward strand of the last two exons (exons 6 and 7) of *KLK6.* The first exon, however, matches the reverse complementary strand of intron 1 and the beginning of exon 2 of the classic form of the gene. Attempts to translate this clone in all frames showed a frame which will give rise to an uninterrupted 169-amino acid polypeptide that partially matches with the human kallikrein 6 (hK6) protein near its end, but with no homology with any known protein.

**Table 3.** Summary of reported SNPs of the *KLK6* sequence[1]

| SNP position | refSNP ID | Contiguous position[2] | Nucleotide change[3] |
|---|---|---|---|
| 5′ end | rs268890 | 23742864 | A→G |
| | rs1722542 | 23742775 | C→T |
| | rs1701951 | 23742733 | A→G |
| | rs1707952 | 23742715 | C→T |
| | rs1654533 | 23742616 | A→G |
| | rs1654534 | 23742610 | C→T |
| | rs1722543 | 23742606 | C→T |
| | rs1722544 | 23742310 | A→T |
| | rs3810287 | 23742171 | G→A |
| | rs1654535 | 23742137 | C→T |
| | rs10409590 | 23741824 | C→T |
| Exon 1 | rs2659094 | 23741109 | C→T |
| Intron 1 | rs2569520 | 23740740 | C→T |
| | rs268891 | 23740173 | A→C |
| | rs268892 | 23740128 | C→T |
| Intron 2 | rs2736714 | 23739872 | A→C |
| | rs1722537 | 23739577 | C→T |
| Intron3 | rs2280811 | 23738906 | G→T |
| Intron 5 | rs1654537 | 23734706 | A→G |
| | rs2569521 | 23734688 | C→T |
| Exon 6 | rs1701950 | 23733286 | G→C |
| Intron 6 | rs2659080 | 23731979 | A→G |

[1] All nucleotide changes were found in the reverse complementary strand of the mRNA.
[2] Numbers refer to GenBank accession No. NT_011109.
[3] Nucleotide change according to the mRNA forward strand.

(8) Another splice variant (GenBank accession No. BM847505, from a myeloma library) was found to have an alternative splice acceptor site 105 nucleotides upstream of exon 3 with a conserved AG splice site. Figure 2 shows a diagrammatic presentation of all reported splice

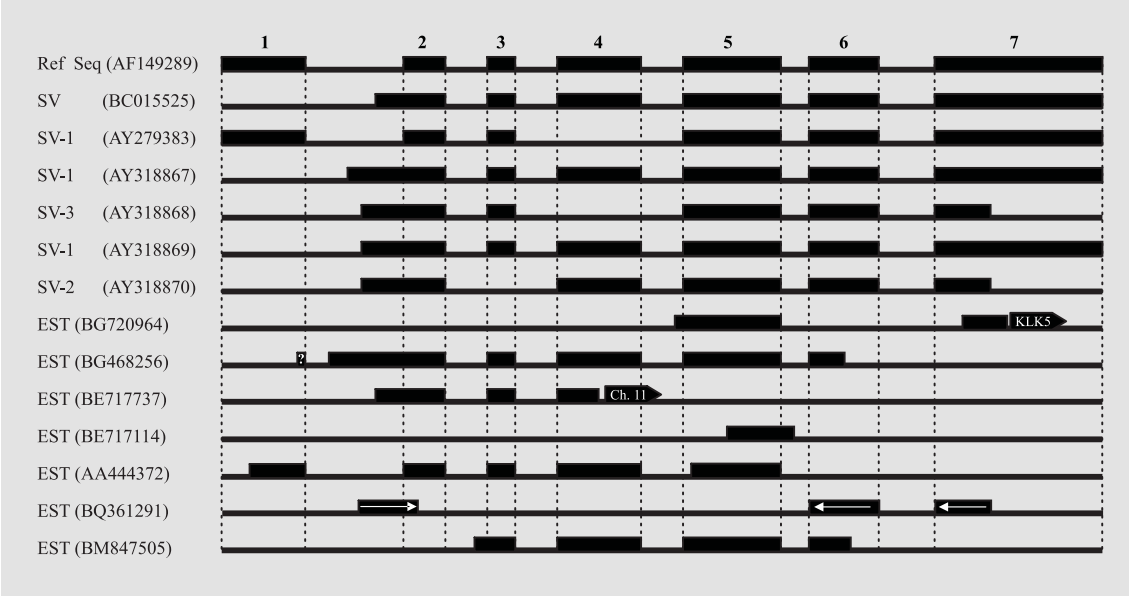


**Fig. 2.** Diagrammatic representation of *KLK6* splice variants. Solid boxes represent exons and connecting lines are introns. White arrows inside boxes indicate exons that are reverse complementary to the reference genomic sequence of *KLK6* (Ref Seq). SV = Reported splice variant. As discussed in the text, the same name was applied to different variants. EST = New variants obtained from EST analysis. Please refer to the text and table 4 for detailed information and full names. Ch. = Chromosome.

**Table 4.** Reported *KLK6* splice variant sequences

| GenBank accession No. | Clone ID | Length, bp | Comments |
|---|---|---|---|
| BC015525 | ? | 1,512 | 165-bp 5′ extension of exon 2 similar to splice variant-1 |
| AY279383 | KLK6 splice variant-1[1] | 1,355 | skips exon 4 |
| AY318867 | KLK6 splice variant-1[1] | 1,517 | 194-bp 5′ extension of exon 2 |
| AY318868 | KLK6 splice variant-3 | 820 | 180-bp 5′ extension of exon 2 skips exon 4 |
| AY318869 | KLK6 splice variant-1 | 1,503 | 180-bp 5′ extension of exon 2 |
| AY318870 | KLK6 splice variant-2 | 929 | 180-bp 5′ extension of exon 2 skips exon 3 |

[1] These two clones have the same name although they represent different variants. Clone AY279383 is actually similar to splice variant-3. For more discussion, please see text.

variants in addition to those identified in the current study.

*In silico SAGE Expression Profile*

We analyzed the expression pattern of *KLK6* in normal and cancer libraries of the SAGE databases of the CGAP. The results are summarized in table 5. Our results clearly indicate that *KLK6* is downregulated in breast cancer tissues. *KLK6*-specific tags are detected in only 8% of cancer libraries compared to 38% of normal breast libraries. The average expression density (estimated as t.p.m.) is also significantly higher in normal than cancer tissue (270 vs. 15 t.p.m., respectively). These results agree with previously published reports. The *KLK6* gene was originally

**Table 5.** In silico analysis of *KLK6* gene expression using SAGE database

| Tissue | Library type | Positivity[1] | Average density, t.p.m. |
|---|---|---|---|
| Breast | normal | 3/8 (38) | 270 |
| | cancer | 2/24 (8) | 15 |
| Ovary | normal | 0/2 (0) | 0 |
| | cancer | 5/10 (50) | 262 |
| Prostate | normal | 1/4 (25) | 16 |
| | cancer | 3/12 (25) | 38 |
| Colon | normal | 0/2 (0) | 0 |
| | cancer | 2/6 (33) | 24 |
| Pancreas | normal | 1/2 (50) | 31 |
| | cancer | 5/6 (83) | 155 |
| Stomach | normal | 0/1 (0) | 0 |
| | cancer | 2/3 (67) | 15 |
| Brain | normal | 7/8 (88) | 39 |
| | cancer | 3/23 (13) | 18 |

Figures in parentheses represent percentages.

[1] Defined as the number of libraries with positive gene-specific tags out of the total number of libraries screened.

**Table 6.** Analysis of *KLK6* expression in the EST databases

| Tissue | Library type | Matching clones | Positive libraries |
|---|---|---|---|
| Stomach | carcinoma | 64 | 5 |
| Colon | adenocarcinoma | 16 | 9 |
| Esophagus | squamous cell carcinoma | 2 | 1 |
| Pancreas | carcinoma | 1 | 1 |
| Brain | normal | 22 | 8 |
| | glioblastoma | 2 | 1 |
| | multiple sclerosis | 2 | 1 |
| Myeloma | | 12 | 2 |
| Head and neck | carcinoma | 15 | 6 |
| Breast | normal | 13 | 5 |
| | adenocarcinoma | 1 | 1 |
| Uterus | carcinoma | 5 | 2 |
| Ovary | carcinoma | 20 | 2 |
| Kidney | normal | 1 | 1 |
| | Wilms' tumor | 2 | 2 |

Only those tissues with differential expression between cancer and normal are presented in this table.

cloned as a gene that is differentially downregulated in metastatic breast cancer [7]. Nacht et al. [8] recently identified *KLK6* as one of the genes differentially expressed in breast cancer.

Lercher et al. [9], during global analysis of over 11,000 genes in 14 different tissues, set cutoff values for low ($\leq$ 37 t.p.m.) and intermediate or high ($\geq$ 134 t.p.m.) gene expression. If these cutoffs are to be applied to our data, we can conclude that while *KLK6* is 'low' in breast cancer, expression levels can be classified as 'high' in normal breast tissue.

Our results show a significant upregulation of *KLK6* in ovarian cancer compared to normal ovarian tissues. While no expression is detectable in the normal ovary, 50% of ovarian cancer libraries show high levels of expression of the *KLK6*-specific tags (average 262 t.p.m.).

As shown in table 5, upregulation of the gene is also found in colon and pancreatic cancer tissues compared to their normal counterparts. Gene-specific tags are also detectable in 67% of stomach cancer libraries, compared to no expression in normal gastric tissues. The average expression density in pancreatic cancer is five times that of normal tissue. While expression is undetectable in normal colon, 33% of colon cancer libraries are positive for the gene. Expression rates are, on the other hand, significantly lower in brain tumors (13%, with an average density of 18 t.p.m.) compared to normal brain (88%, with an average density of 39 t.p.m.). No significant variation is found in expression levels between normal and cancerous prostate. Further experimental verification of these results is warranted.

*The EST Expression Profile*

The EST expression profile of *KLK6*, analyzed through different databases, is summarized in table 6. Our EST screening shows that *KLK6* is upregulated in gastrointestinal malignancies, including stomach, colon, esophagus and pancreatic cancers. Sixty-four clones were isolated from 5 stomach cancer libraries, and 16 clones from 9 colon cancer libraries. These results support those of the independent SAGE databases discussed earlier. This upregulation is also evident in tumors of female genital (ovarian and uterine) malignancies (20 and 5 clones, respectively).

Significant numbers of clones were also identified from head and neck cancers (15 clones) and myeloma (12 clones) libraries. Breast cancer showed opposite findings with a higher number of clones isolated from normal tissues compared to adenocarcinoma of the breast. Also, while 22 clones were isolated from 8 normal brain librar-

**Table 7.** Summary of the *KLK6* taxonomy results

| Species | Name | GenBank accession No. | Length, aa | Degree of homology[1], % | Chromosomal location |
|---|---|---|---|---|---|
| *Homo sapiens* | KLK6 | Q92876 | 244 | 100 | 19q13.4 |
| *Mus musculus* | PRSS18, protease, serine, 18 | NP_035307 | 246 | 68 | 7B3 |
| *Rattus norvegicus* | kallikrein 6 (neurosin, zyme); protease, serine, 9 | NP_062048 | 250 | 65 | 1q22 |
| *Drosophila melanogaster* | T13596 trypsin homolog | T13596 | 247 | 32 | X |

aa = Amino acids.
[1] Homology compared to *Homo sapiens.*

ies, only 2 clones were identified from a glioblastoma library.

*Taxonomy Analysis*

*KLK6* orthologues were identified in three other species in addition to *Homo sapiens* (table 7). Interestingly, the length of the hK6 polypeptide is quite comparable in all species, pointing to a possible preserved function among species. The mouse and rat orthologues showed the highest degree of sequence homology to humans (68 and 65%, respectively).

## Discussion

The *KLK6* gene was recently cloned by three independent groups. The cDNA of the gene was first isolated using a differential display technique from primary and metastatic breast cancer cell lines, and it was named protease M [7]. The same gene was also cloned from a cDNA library prepared from a human colon adenocarcinoma cell line and was named neurosin [10]. Finally, the same mRNA was cloned from Alzheimer's disease brain by polymerase chain reaction amplification, and was named zyme [11].

Our database mining enabled us to construct the longest available clone of *KLK6.* We also identified new splice variants of the *KLK6* gene. Most of these variants were isolated from cancer libraries of different sources. The possible presence of cancer-specific variants of *KLK6* and their role in cancer development and/or progression needs to be further experimentally analyzed. It is also important that these variants be identified, since they may present a potential source of error in measuring expression of the *KLK6* active form in different tissues.

The pattern of expression of the gene, analyzed through SAGE and EST databases, is consistent with previously published experimental evidence that showed higher levels of expression in the central nervous system (CNS), mammary gland, stomach, testis, kidney, spleen and ovary. This further verifies the accuracy and reliability of our in silico analysis. Our analyses clearly indicate differential expression of *KLK6,* at the mRNA level, in various malignancies, verified by two independent databases. The gene is upregulated in female genital (ovarian and uterine) and gastrointestinal (pancreas, esophagus, stomach and colon) cancers. Expression is also elevated in head and neck cancers and myeloma. It is, however, downregulated in breast and brain tumors. These findings are not surprising. There is significant experimental evidence that hK6 is involved in the progression of various types of cancer. hK6 was found to be elevated in the serum of ovarian cancer patients [12, 13]. Tanimoto et al. [14] and Hoffman et al. [15] also showed that overexpression of hK6 tends to be greater in late stages of disease. In breast cancer patients, *KLK6* gene expression was reported to be decreased, and this downregulation is related to the stage of the disease [7]. It will be interesting to examine the possible utility of the gene as a cancer biomarker in the remaining types of cancer identified in the current analysis, especially because most of them are lacking sensitive diagnostic and/or prognostic markers.

We also performed a preliminary analysis of the *KLK6* transcript in different species. This should be the basis for further detailed phylogenetic analysis. The observed high degree of homology among species might also reflect a conserved function. Human hK6 was found to be highly expressed in the CNS. The mouse homolog of *KLK6,* named myelencephalon-specific protease or MSP [16, 17], has also been shown to play an important role in the regulation of myelin turnover and the demyelination of the CNS [18, 19], possibly leading to diseases such as multiple sclerosis [20] and Alzheimer's disease [21].

It should be noted that our results are based on database analyses. This in silico approach has been increasingly and successfully used in recent years. However, these results should be experimentally verified. The main objective of this analysis is to guide *KLK6* research to new directions and to investigations that will lead to a better understanding of the role of *KLK6* in normal physiology and physiopathology, as well as its clinical utility [22].

## References


1 Strausberg RL: The Cancer Genome Anatomy Project: New resources for reading the molecular signatures of cancer. J Pathol 2001;195:31–40.
2 Adams MD, Soares MB, Kerlavage AR, Fields C, Venter JC: Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. Nat Genet 1993;4:373–380.
3 Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: Serial analysis of gene expression. Science 1995;270:484–487.
4 Argani P, Rosty C, Reiter RE, Wilentz RE, Murugesan SR, Leach SD, Ryu B, Skinner HG, Goggins M, Jaffee EM, Yeo CJ, Cameron JL, Kern S, Hruban RH: Discovery of new markers of cancer through serial analysis of gene expression: Prostate stem cell antigen is overexpressed in pancreatic adenocarcinoma. Cancer Res 2001;61:4320–4324.
5 Welsh JB, Zarrinkar P, Sapinoso LM, Kern SG, Behling CA, Monk BJ, Lockhart DJ, Burger RA, Hampton GM: Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. Proc Natl Acad Sci USA 2001;98:1176–1181.
6 Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–4680.
7 Anisowicz A, Sotiropoulou G, Stenman G, Mok SC, Sager R: A novel protease homolog differentially expressed in breast and ovarian cancer. Mol Med 1996;2:624–636.
8 Nacht M, Ferguson A, Zhang W, Petroziello JM, Cook BP, Gao YH, Maguire S, Riley D, Coppola G, Landes GM, Madden SL, Sukumar S: Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer. Cancer Res 1999;59:5464–5470.
9 Lercher MJ, Urrutia AO, Hurst LD: Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat Genet 2002;31:180–183.
10 Yamashiro K, Tsuruoka N, Kodama S, Tsujimoto M, Yamamura Y, Tanaka T, Nakazato H, Yamaguchi N: Molecular cloning of a novel trypsin-like serine protease (neurosin) preferentially expressed in brain. Biochim Biophys Acta 1997;350:11–14.
11 Little SP, Dixon EP, Norris F, Buckley W, Becker GW, Johnson M, Dobbins JR, Wyrick T, Miller JR, MacKellar W, Hepburn D, Corvalan J, McClure D, Liu X, Stephenson D, Clemens J, Johnstone EM: Zyme, a novel and potentially amyloidogenic enzyme cDNA isolated from Alzheimer's disease brain. J Biol Chem 1997;272:25135–25142.
12 Diamandis EP, Yousef GM, Soosaipillai AR, Bunting P: Human kallikrein 6 (zyme/protease M/neurosin): A new serum biomarker of ovarian carcinoma. Clin Biochem 2000;33:579–583.
13 Diamandis EP, Scorilas A, Fracchioli S, Van Gramberen M, De Bruijn H, Henrik A, Soosaipillai A, Grass L, Yousef GM, Stenman UH, Massobrio M, Van Der Zee AG, Vergote I, Katsaros D: Human kallikrein 6 (hK6): A new potential serum biomarker for diagnosis and prognosis of ovarian carcinoma. J Clin Oncol 2003;21:1035–1043.
14 Tanimoto H, Underwood LJ, Shigemasa K, Parmley TH, O'Brien TJ: Increased expression of protease M in ovarian tumors. Tumour Biol 2001;22:11–18.
15 Hoffman BR, Katsaros D, Scorilas A, Diamandis P, Fracchioli S, Rigault De La Longrais IA, Colgan T, Puopolo M, Giardina G, Massobrio M, Diamandis EP: Immunofluorometric quantitation and histochemical localisation of kallikrein 6 protein in ovarian cancer tissue: A new independent unfavourable prognostic biomarker. Br J Cancer 2002;87:763–771.
16 Scarisbrick IA, Towner MD, Isackson PJ: Nervous system-specific expression of a novel serine protease: Regulation in the adult rat spinal cord by excitotoxic injury. J Neurosci 1997;17:8156–8168.
17 Scarisbrick IA, Blaber SI, Lucchinetti CF, Genain CP, Blaber M, Rodriguez M: Activity of a newly identified serine protease in CNS demyelination. Brain 2002;125:1283–1296.
18 Yamanaka H, He X, Matsumoto K, Shiosaka S, Yoshida S: Protease M/neurosin mRNA is expressed in mature oligodendrocytes. Brain Res Mol Brain Res 1999;71:217–224.
19 Scarisbrick IA, Isackson PJ, Ciric B, Windebank AJ, Rodriguez M: MSP, a trypsin-like serine protease, is abundantly expressed in the human nervous system. J Comp Neurol 2001;431:347–361.
20 Scarisbrick IA, Asakura K, Blaber S, Blaber M, Isackson PJ, Bieto T, Rodriguez M, Windebank AJ: Preferential expression of myelencephalon-specific protease by oligodendrocytes of the adult rat spinal cord white matter. Glia 2000;30:219–230.
21 Blaber SI, Scarisbrick IA, Bernett MJ, Dhanarajan P, Seavy MA, Jin Y, Schwartz MA, Rodriguez M, Blaber M: Enzymatic properties of rat myelencephalon-specific protease. Biochemistry 2002;41:1165–1173.
22 Yousef GM, Luo LY, Scherer SW, Sotiropoulou G, Diamandis EP: Molecular characterization of zyme/protease M/neurosin (PRSS9), a hormonally regulated kallikrein-like serine protease. Genomics 1999;62:251–259.