Research Article

TumorBiology

Tumor Biol 2004;25:122–133
DOI: 10.1159/000079144

Received: March 24, 2004
Accepted after revision: March 29, 2004

Molecular Cloning of a New Gene Which Is Differentially Expressed in Breast and Prostate Cancers

George M. Yousef[a,b] Carla A. Borgoño[b,c] Iacovos P. Michael[b,c]
Christine Davidian[c] Carsten Stephan[d] Klaus Jung[d]
Eleftherios P. Diamandis[b,c]

[a]Department of Laboratory Medicine, Memorial University, St. John's, Newfoundland,
[b]Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario,
[c]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada,
[d]Department of Urology, University Hospital Charité, Humboldt University, Berlin, Germany

Key Words
Cancer-associated gene · Cancer genes · Prostate cancer · Cancer biomarkers · Breast cancer · Differentially expressed genes

Abstract
Objective: The chromosomal region 19q13 is non-randomly rearranged in many solid tumors. ***Methods:*** Using the positional candidate gene approach, we cloned a new gene, tentatively named cancer-associated gene *(CAG),* which is differentially expressed in breast and prostate cancers. ***Results:*** The gene is formed of 3 exons and 2 intervening introns. Its coding region is 1,047 bp in length and is predicted to encode a 348-amino-acid polypeptide. The new gene maps to chromosome 19q13.4 and is located 14 kb telomeric to the kallikrein gene locus (*KLK14* gene) and 17 kb centromeric from the Siglec family of genes *(Siglec-9).* The gene is expressed in a wide variety of tissues including the brain, colon, kidney and pancreas. The CAG protein shows a high degree of conservation among species and phylogenetically is most closely related to its mouse ortholog. In silico analysis indicates that this gene is differentially expressed in a variety of tumors including brain, colon, ovarian and prostate cancers. ***Conclusions:*** Our preliminary experimental data show that *CAG* is upregulated in prostate cancer tissues compared to normal prostatic tissues. *CAG* also appears to be downregulated in breast cancer tissues. The physiological function of the CAG protein is currently unknown.

Copyright © 2004 S. Karger AG, Basel

Genbank Submission: Sequences reported in this paper have been submitted to Genbank (accession No. AY279382).

Introduction

Cancer is fundamentally a disease of the genome, arising from inherited and/or somatically acquired mutations at various genetic loci [1]. A key component in understanding the genetic basis of cancer involves the identification of aberrantly and differentially expressed genes. Over the past two decades, extensive progress has been made in identifying such genes; their characterization can be utilized to better understand the pathogenesis of cancer, to

KARGER

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

© 2004 S. Karger AG, Basel
1010–4283/04/0253–0122$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

define molecular signatures and to develop biomarkers for early detection and targets for intervention [2].

Accumulating evidence suggests that the chromosomal locus 19q13 is implicated in cancer. For one, this region harbors a number of oncogenes, including *Bcl-3* [3], *RRAS*, *RAB14*, *AKT2* [4], *SEI-1* [5], tumor suppressor genes such as *p190-A* [6], the apoptosis regulator genes *BAX* [7] and *BCL2L12*, a newly identified homologue of *Bcl-2* [8], as well as a number of genes differentially expressed in cancer [9, 10]. Furthermore, this region is genetically altered in a variety of tumors. These alterations include gene amplification [11–14], loss of heterozygosity [15, 16] and chromosomal instability (aberrations and non-random translocations) [17–19].

We have previously characterized the human kallikrein gene locus on chromosome 19q13.4, a family of 15 hormonally regulated, structurally similar serine protease genes [20, 21]. Numerous reports indicate that many kallikreins are differentially expressed in hormone-dependent cancers at both the mRNA and/or protein levels and several possess clinical utility as cancer biomarkers [9, 22–24]. Considering that the chromosomal band 19q13 is known to be a hotspot for malignancy, we further examined a ~43.2-kb region separating the kallikrein (centromeric side) and Siglec gene family loci (telomeric side) [25] for the presence of other genes. Here we describe the identification of a new gene, its genomic and mRNA structure, its chromosomal location in relation to other known kallikrein and Siglec genes and its tissue expression pattern. Our preliminary in silico data suggest that this gene is differentially expressed in hormone-related malignancies. We also provide experimental evidence that this gene is upregulated in prostate cancer and downregulated in breast cancer.

Materials and Methods

Expressed Sequence Tag (EST) Searching

The predicted exons of the putative new gene were subjected to homology search using the BLASTN algorithm [26] on the National Center for Biotechnology Information (NCBI) web server (http://www ncbi.nlm.nih.gov/BLAST/) against the human EST database. Clones with >95% homology were obtained from the IMAGE consortium [27] through Research Genetics Inc, Huntsville, Ala., USA. The clones were propagated, purified as described elsewhere [28] and sequenced from both directions with an automated sequencer, using insert-flanking vector primers. The full-length mRNA sequence of the gene was compared against the human EST databases of the NCBI. Expression was calculated as the number of positive libraries in addition to the total number of clones detected in each.

Reverse Transcriptase Polymerase Chain Reaction

Total RNA was extracted from tissues using Trizol reagent (Gibco BRL, Bethesda, Md., USA) following the manufacturer's instructions. RNA concentration was determined spectrophotometrically. Two micrograms of total RNA was reverse-transcribed into first-strand cDNA using the Superscript™ pre-amplification system (Gibco BRL). The final volume was 20 µl. To confirm the efficiency of reverse transcriptase polymerase chain reaction (RT-PCR), 1 µl of cDNA was subsequently amplified by PCR with primers specific for actin, a housekeeping gene (ActinS: 5′ ACAATGAGCTGCGTGTGGCT, ActinAS: 5′ TCTCCTTAATGTCACGCACGA). Actin PCR products with an expected length of 372 bp were visualized on 2% agarose gels stained with ethidium bromide.

Two different sets of gene-specific primers were designed for PCR amplification of the putative gene (table 1). All primers spanned at least two exons to avoid contamination by genomic DNA during RT-PCR. Both sets of primers were used to determine expression in normal tissues and cancerous prostate and breast tissues.

PCR with primer set B-F1/B-R1 was carried out in a 25-µl reaction mixture containing 1 µl of cDNA, 10 m*M* Tris-HCl (pH 8.3), 50 m*M* KCl, 1.5 m*M* $MgCl_2$, 200 µ*M* dNTPs (deoxynucleoside triphosphates), 100 ng of each primer and 2.5 units of HotStarTaq™ DNA polymerase (Qiagen Inc., Valencia, Calif., USA) on an Eppendorf master cycler. The cycling conditions were 95 °C for 15 min, followed by 40 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 45 s and a final extension step of 10 min at 72 °C. PCR with primer set FTEFB/RTEFB was also carried out in a 25-µl reaction mixture containing 1 µl of cDNA, 10 m*M* Tris-HCl (pH 8.3), 50 m*M* KCl, 200 µ*M* dNTPs, 100 ng of each primer, 5 µl of Q-solution™ (Qiagen) and 2.5 units of HotStarTaq™ DNA polymerase (Qiagen) on an Eppendorf master cycler. The cycling conditions for this reaction were 95 °C for 15 min, followed by 35 cycles of 94 °C for 30 s, 65 °C for 30 s, 72 °C for 30 s and a final extension step of 10 min at 72 °C.

Primer set B-F1/B-R1 was additionally utilized with the Expand Long Template PCR System (Roche Diagnostics, Mannheim, Germany) in a 50 µl PCR reaction mixture containing 100 ng of genomic DNA, 10 m*M* Tris-HCl (pH 8.3), 50 m*M* KCl, 1.5 m*M* $MgCl_2$, 200 µ*M* dNTPs, 100 ng of each primer, and 2.6 units of Expand Long Template PCR polymerase mix, using an Eppendorf master cycler. The PCR conditions were 94 °C for 2 min, followed by 94 °C for 10 s, 52 °C for 30 s, 68 °C for 1 min for 40 cycles, and a final extension at 68 °C for 7 min. Equal amounts of PCR products were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining.

To verify the identity of the PCR products, they were cloned into the PCR 2.1-TOPO vector (Invitrogen, Carlsbad, Calif., USA) according to the manufacturer's instructions. The inserts were sequenced from both directions using vector-specific primers, with an automated DNA sequencer.

Tissue Expression

Total RNA isolated from 35 different human tissues was purchased from Clontech (Palo Alto, Calif., USA). cDNA was prepared as described above and amplified at various dilutions using two sets of gene-specific primers, B-F1/B-R1 and FTEFB/RTEFB (table 1).

Table 1. Primers used for RT-PCR analysis

Gene	Primer name	Primer sequence[1]	Product size, bp	Annealing temperature ℃
CAG	B-F1	GTG CTG CAC ACG GTG CTC	950	55
	B-R1	GAG GCC CGA AGT CGC TAG A		
	FTEFB	CAT CGG TGG CTA CCG GGA CG	497	65
	RTEFB	CCA GGC GCT TGA GCA GGT CC		
KLK14	L6-FG1	GGA AGG AAG TGC CCC GAC CT	437	67
	L6-R1	CCT GGA CTC CTG GGT CTG CA		
Siglec-9	BPL-FG1	CCT CTG CCT CCT AGG TTC AA	648	64
	BPL-R2	AAA AGG GAG GGC ACA GTG TG		

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

Prostate Cancer Tissues

Prostate tissue samples were obtained from 29 patients who had undergone radical retropubic prostatectomy for prostatic adenocarcinoma at the Charité University Hospital, Berlin, Germany. The patients had not received any hormonal therapy before surgery. The use of these tissues for research purposes was approved by the Ethics Committee of the Charité Hospital. Fresh prostate tissue samples were obtained from the cancerous and non-cancerous parts of the same prostates that had been removed. Small pieces of tissues were dissected immediately after removal of the prostate and stored in liquid nitrogen until analysis. Histological analysis of all tissue samples was performed as previously described [29], to ensure that the tissue was either malignant or benign. The tissues were pulverized with a hammer under liquid nitrogen and RNA was extracted as described above, using Trizol reagent.

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.

SAGE Analysis and Virtual Northern Blots

The mRNA sequences of the new gene and two restriction digestion anchoring enzymes (*Nla*III and *Sau*3A) were used to identify a unique sequence tag specific for this gene. This tag was then used to determine the levels of gene expression in normal and cancer SAGE libraries from various tissues. Detailed library descriptions are available from the website of the Cancer Genome Anatomy Project (CGAP) (http://www.ncbi.nlm.nih.gov/ncicgap/). Analyses were performed by comparing the proportion of libraries of each type (cancer vs. normal) expressing each tag, in addition to the average expression densities in these libraries. Expression levels were displayed as blots of different densities and corrected as tags per million (tpm) to facilitate comparison.

Structure Analysis

Multiple alignment was performed using the 'Clustal X' software package and the multiple alignment program available from the Baylor College of Medicine (Houston, Tex., USA). Phylogenetic studies were performed using the 'Phylip' software package. Distance matrix analysis was performed using the 'Neighbor-Joining/UPGMA' program and parsimony analysis was done using the 'Protpars' program. Hydrophobicity study was performed using the Baylor College of Medicine search launcher. Signal peptide was predicted using the 'SignalP' server. Protein structure analysis was performed by the 'SAPS' (structural analysis of protein sequence) program.

Results

Identification of the Cancer-Associated Gene

We have previously characterized the human kallikrein gene locus on chromosome 19q13.4 [20, 30], extending from the most centromeric kallikrein gene, *KLK1*, to the most telomeric gene, *KLK14* [31]. Telomerically, the kallikrein gene locus is flanked by the Siglec family of genes, which has recently been characterized in detail [25]. In the current study, we further analyzed the 43.2-kb distance separating these two families [25, 32]. Genomic sequences spanning this region were obtained from the Human Genome Project and subjected to sequence analysis using bioinformatics, EST databases, homology searches and PCR.

Our preliminary analysis suggested presence of a putative new gene in this region. Gene prediction programs and EST database search identified a potential novel gene formed of 3 exons with two intervening introns. This putative gene sequence was then blasted against the human EST database and several EST clones with >95% identity were found (table 2). The majority of these clones were members of the same Unigene cluster Hs.148425. These

Table 2. EST clones with >95% similarity to *CAG*

Organ	Library ID	Clones	Tissue type(s)
Blood	Hembase erythroid precursor cells	1	blood
Brain	NIH_MGC_98	4	astrocytoma
	NCI_CGAP_Brn25	8	anaplastic oligodendroglioma
	Fetal brain 00004	2	normal fetal brain
	NCI_CGAP_Brn64	2	glioblastoma
	NIH_MGC_119	1	normal medulla
	NIH_MGC_56	1	primitive neuroectoderm
	NIH_MGC_19	1	neuroblastoma
	LTI_NFL001_NBC4	1	
Breast	BT104	1	normal mammary gland
Chondrosarcoma	NCI_CGAP_FL1	1	chondrosarcoma
Colon	CI0092	1	normal colon
Kidney	NCI_CGAP_Kid11	4	normal kidney
	NCI_CGAP_Kid12	3	clear cell type renal cell carcinoma
Lung	UI-CF-DU1	1	epithelial cells
	NCI_CGAP_Lu28	1	squamous cell carcinoma
	NIH_MGC_7	1	small cell carcinoma
Lymph node	NIH_MGC_52	1	normal germinal center B cells
	NIH_MGC_85	1	lymphoma, cell line
Nervous	NN1006	1	normal nervous
	NN0245	1	normal nervous
Pancreas	Melton normalized human islet 4 N4-HIS1	6	islets of Langerhans
Placenta	NIH_MGC_79	2	placenta
	NIH_MGC_21	2	choriocarcinoma
Prostate	NCI_CGAP_Pr28	1	normal prostate
Skin	NIH_MGC_41	3	melanoma
	NIH_MGC_112	1	melanoma
Testis	NIH_MGC_92	1	embryonal carcinoma
Pooled	NIH_MGC_116	2	normal colon, kidney and stomach
	NIH_MGC_120	1	normal pancreas and spleen
	Soares fetal liver spleen 1NFLS	1	normal fetal liver and spleen
	Soares_NFL_T_GBC_S1	1	normal fetal lung, testis and B-cells
	NCI_CGAP_GC6	3	germ cell tumors

clones were obtained, propagated, purified and fully sequenced. Five of these clones were 99% identical to the last exon and the putative 3′ untranslated region of the gene, in addition to a stretch of adenine (A) nucleotides of variable lengths that were not found in the genomic sequence, thus verifying the 3′ end of the gene and the position of the poly A tail.

To verify its genomic organization, a gene-specific probe (based on the predicted structure of the gene) was used to screen a human BAC library and a positive clone (BC349142) was identified. PCR amplification using gene-specific primers was used to verify that the genomic sequence of this clone was positive for *KLK14* (the most telomeric kallikrein), *Siglec 9* (the most centromeric Siglec) and the putative new gene (table 1).

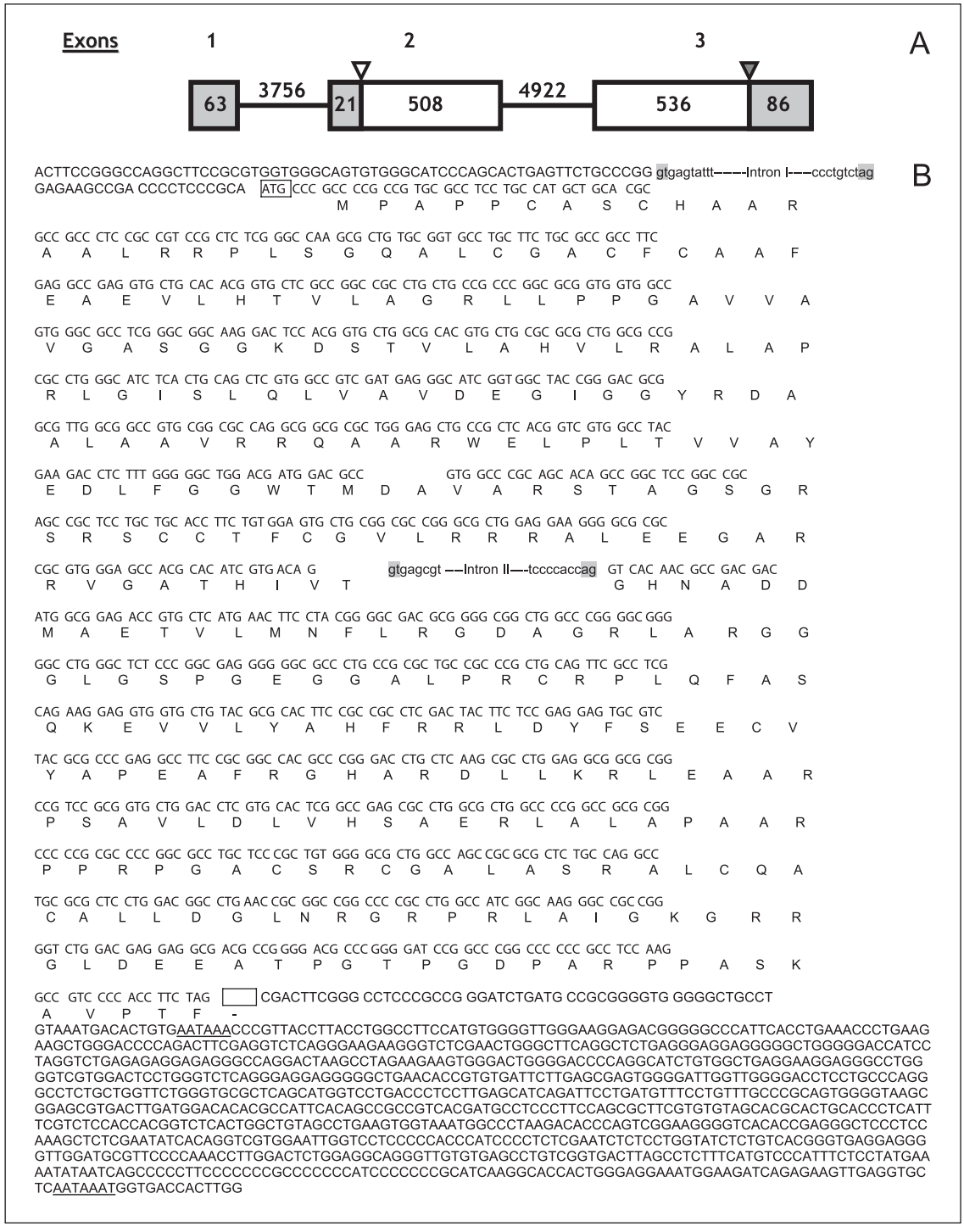

Fig. 1. Genomic organization and partial genomic sequence of *CAG*. **A** A schematic representation of the genomic structure of *CAG*. Exons are shown by boxes and introns by the connecting lines. The numbers inside the boxes indicate the exon lengths in base pairs, and numbers above the connecting lines show the intron lengths. Shaded boxes indicate untranslated sequences and white boxes represent translated regions. The white arrowhead denotes the location of the start codon, and the gray arrowhead represents the stop codon. Figure is not drawn to scale. **B** Partial genomic structure of *CAG*. Intron sequences are not shown except for splice junction areas. Exon-intron junctions are shaded. Introns boundaries are shown in lowercase letters and exons are represented by uppercase letters. The start and stop codons are boxed. The translated amino acids of the coding region are shown below in single-letter code. The putative polyadenylation signals are underlined. For more discussion, see the Results section.

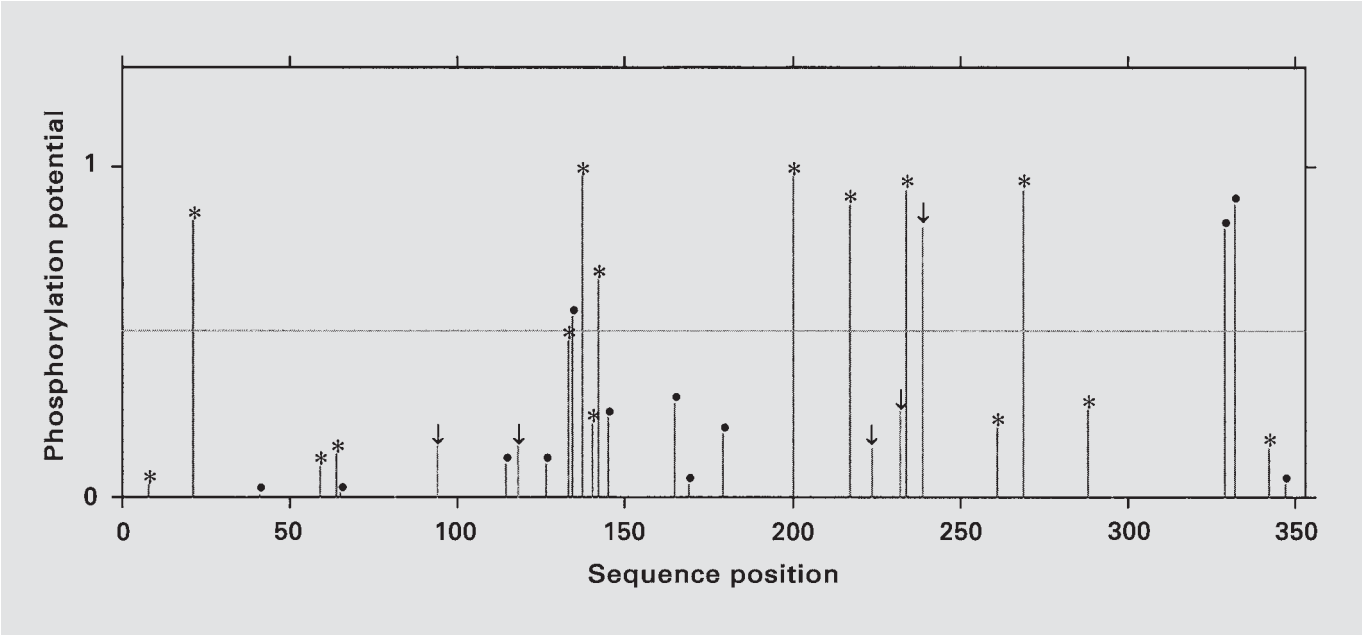

Fig. 2. Prediction of potential phosphorylation sites within the CAG protein. Serine phosphorylation sites are marked by *, tyrosine sites by ● and threonine sites by ↓.

To confirm the mRNA structure and exon/intron boundaries, PCR analysis was performed using primers that span the entire length of the gene and a panel of 35 human tissue cDNAs as templates. By aligning the sequence obtained from the PCR with the genomic sequence we were able to characterize its structure, consisting of 1 non-coding exon, 2 coding exons with 2 intervening introns and the splice junctions. Translation of the mRNA sequence in all possible reading frames revealed the presence of only one frame that gives an uninterrupted polypeptide chain, as discussed below.

Structural Characterization of the Cancer-Associated Gene

The cancer-associated gene (CAG) spans 9,892 bp of genomic sequence on chromosome 19q13.4. As shown in figure 1, this gene is formed of 3 exons (1 non-coding and 2 coding) and 2 intervening introns, although the presence of further upstream untranslated exon(s) could not be ruled out. All of the exon/intron splice sites (mGT.... AGm) conform to the consensus sequence for eukaryotic splice sites [33]. The first exon is untranslated and the predicted protein-coding region of the gene starts 21 nucleotides from the beginning of exon 2 and is formed of 1,044 bp, encoding a deduced 348 amino acid polypeptide with a predicted molecular weight of 36.52 kDa. The potential translation initiation codon is similar to the consensus Kozak sequence and a purine is present at position (–3), which occurs in 97% of vertebrate mRNAs [34]. However, as is the case with many members of the neighboring kallikrein gene family, *CAG* does not have the consensus G nucleotide at position (+4). Two well-conserved polyadenylation signals (AATAAA) are located 65 and 974 bp downstream from the stop codon. However, with the exception of 2 ESTs (Genbank accession No. BU632056 and BQ422774), the remainder utilize the former signal, located 12 bp from their poly A tails. These findings suggest that isoforms or alternatively spliced variants of *CAG* may exist. No other potential polyadenylation signals were discernible in the 3′ untranslated region.

Computer analysis of the predicted protein sequence indicated absence of a signal peptide, tyrosine sulfation or N-glycosylation sites. A mitochondrial targeting sequence was predicted with high score (0.975). In addition, multiple phosphorylation sites (fig. 2) were predicted. Hydrophobicity analyses, using different programs, indicated a soluble protein with no predicted transmembrane domains. Comparing the putative mRNA and protein sequences of the new gene against GenBank and other

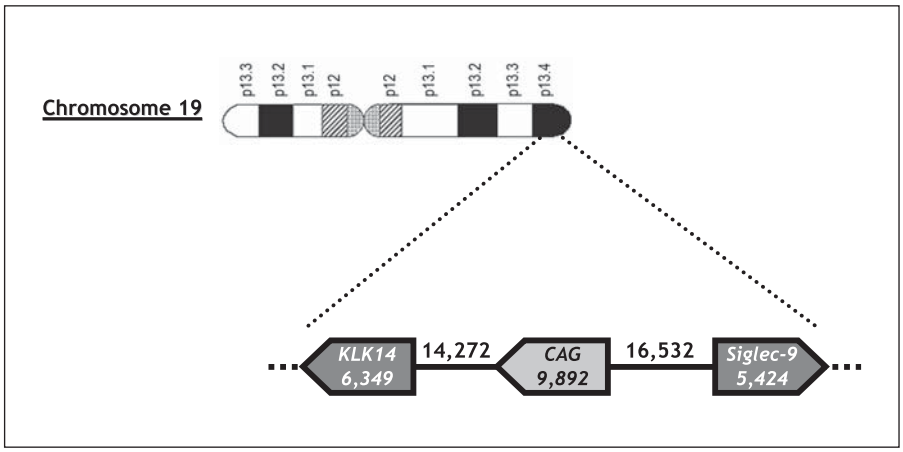

Fig. 3. Chromosomal localization of *CAG*. *CAG* is located in chromosomal region 19q13.4 and is flanked centromerically by *KLK14* and telomerically by *Siglec-9*. The block arrows illustrate genes and their direction of transcription. Gene names and genomic lengths, in base pairs, are shown within the block arrows. Intergenic distances are indicated above the solid lines in base pairs. Figure is not drawn to scale.

protein databases revealed 29% similarity with a 370-amino-acid region of the human collagen alpha-1protein.

Conserved domain search using different databases indicated the presence of a highly conserved motif (85%) from an uncharacterized gene family and a less conserved domain (65%) from a predicted ATPase of the PP-loop superfamily, implicated in cell cycle control (cell division and chromosome partitioning).

Chromosomal Localization of the Cancer-Associated Gene

Alignment of *CAG* with *KLK14* and *Siglec-9* gene sequences within and around the 42.3-kb area separating the kallikrein and Siglec gene families on chromosome 19q13.4 enabled us to precisely localize *CAG* and to determine its direction of transcription, as shown in figure 3. *CAG* resides 14,272 bp telomeric to *KLK14* and 16,532 bp centromeric to *Siglec-9*, and is transcribed from telomere to centromere, similar to *KLK14*. These results are consistent with previous reports that estimate the distance between *KLK14* and *Siglec-9* to be ~43.2 kb [20, 31].

Phylogenetic Analysis

To predict the evolutionary history of CAG, its protein sequence was subjected to homology search against the non-redundant databases of the NCBI. Taxonomy analysis indicated the presence of highly homologous orthologues in many other species. The CAG protein shares the highest degree of homology with the mouse orthologue (86%), followed by the fruit fly (51%). A representative multiple alignment graph of CAG orthologues from different species is presented in figure 4. Phylogenetic analyses were performed on the sequences from different species in addition to the protein sequence of the most related protein, human collagen alpha-1 chain precursor from different species, using the Unweighted Pair Group Method with Arithmetic mean (UPGMA), the Neighbor-Joining distance matrix methods, and the 'Protpars' parsimony method. A representative tree is shown in figure 5. The human CAG protein was grouped together with its mouse orthologue, separate from its homologues from 5 other species, suggesting divergence before separation of the human-rodent lineages. Furthermore, the potential CAG orthologues form a group separate from the human and rodent collagen alpha-1 proteins, which cluster together.

Tissue Expression

The tissue expression pattern of *CAG* was assessed both experimentally and in silico. We examined its expression by RT-PCR in a panel of 35 normal adult tissues. As shown in figure 6, *CAG* was found to be transcribed in many tissues with relatively higher expression levels observed in the adrenal gland, whole brain, cervix, colon, fetal brain, fetal liver, heart, kidney, lung, mammary gland, ovary, pancreas, placenta, prostate, skeletal muscle, spinal cord, spleen, small intestine, stomach, testis, thymus, thyroid and uterus tissues, with lower levels of expression in adipose, bone marrow, cerebellum, fallopian tube, hippocampus, liver, pituitary, salivary gland, trachea and vaginal tissues. Mining the human EST database of the NCBI revealed a number of EST clones for the new gene, from a range of libraries (table 2). ESTs were mainly identified in the brain (20 clones from 8 libraries), followed by the kidney (7 clones), pancreas (6

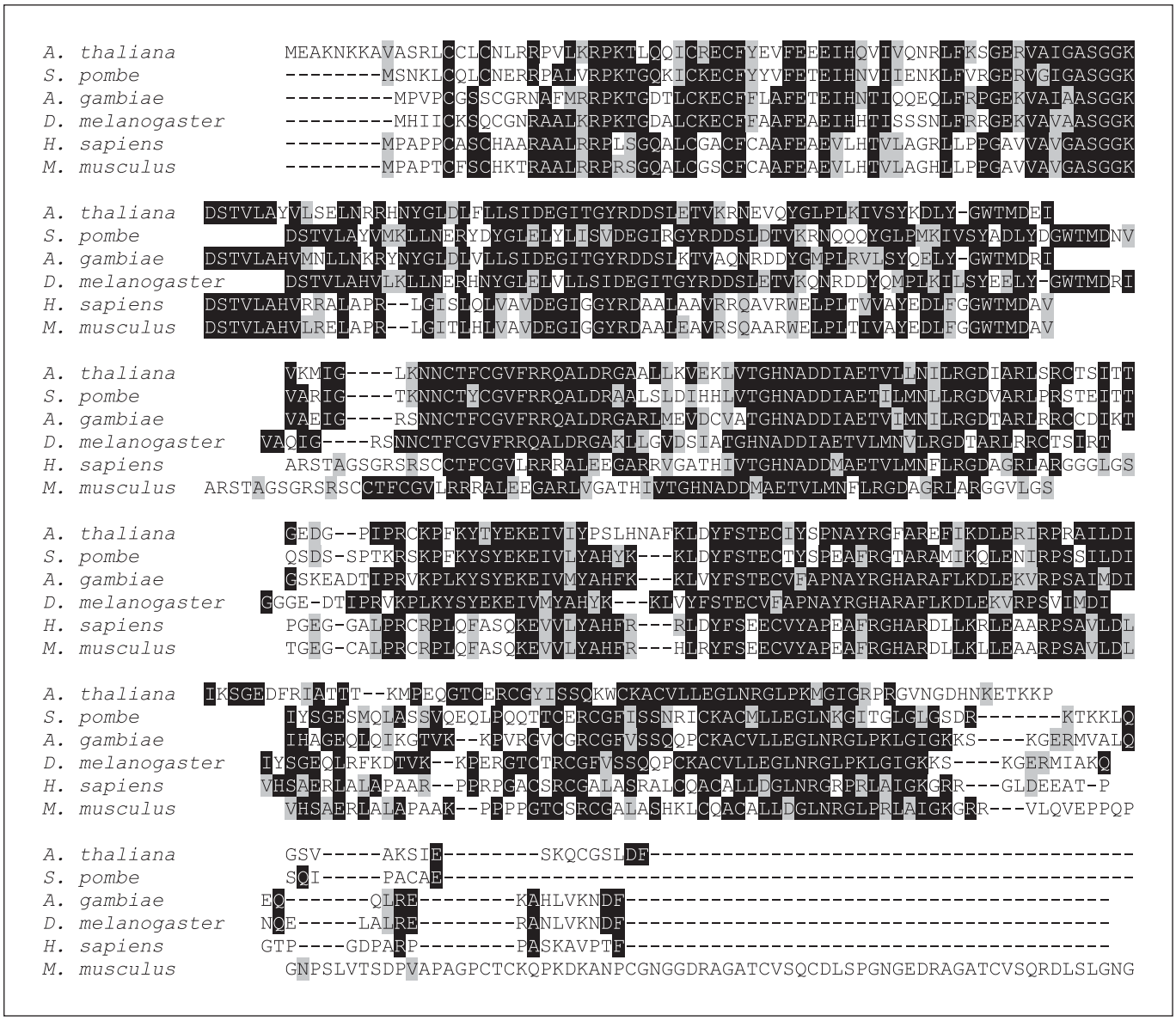

Fig. 4. Alignment of the deduced amino acid sequence of the human CAG protein with other CAG orthologues. Dashes represent gaps to bring the sequences to better alignment. Identical amino acids are highlighted in black and similar residues in gray.

clones), placenta (4 clones) and skin (4 clones). Serial analysis of gene expression (SAGE) analysis indicated low to moderate expression levels of the *CAG* mRNA in many tissues including the brain, colon, kidney, ovary, pancreas, prostate and skin (table 3). Our collective experimental and in silico data demonstrate that *CAG* is predominantly expressed in normal brain, colon, kidney and pancreas and to a lower extent in many other tissues.

Differential Expression of CAG *in Cancer*

In silico analysis indicated that *CAG* is differentially expressed in several cancers (table 3). While expression was detected in the normal brain (16 tpm in only 2 SAGE libraries), higher levels of expression were found in 8 malignant brain libraries (24 tpm on average). Only one normal prostate tissue library was positive for *CAG*, with weak expression levels (15 tpm), compared to 6 prostate cancer libraries displaying higher average expression lev-

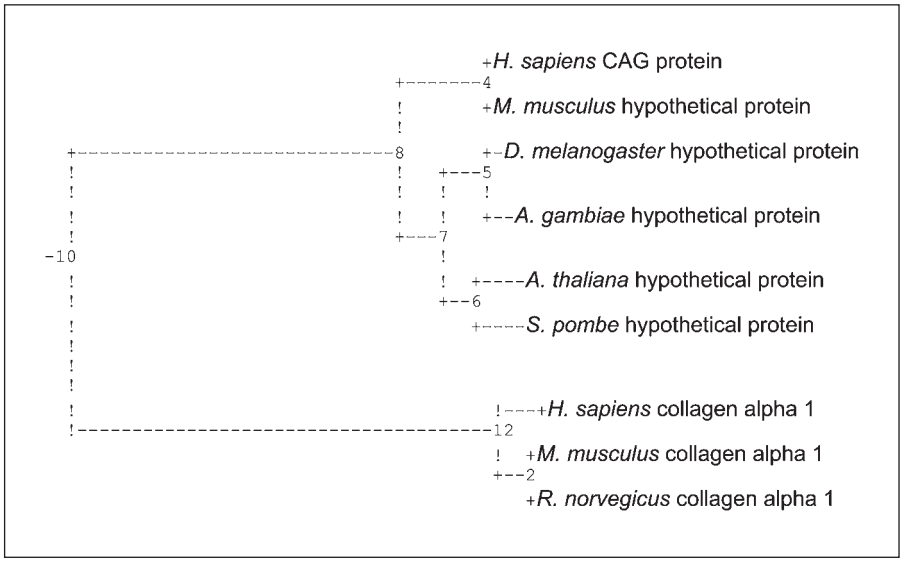

Fig. 5. Dendrogram of the predicted phylogenetic tree for the human CAG protein and other homologous proteins. The neighbor-joining method was used to align the human CAG protein with other homologous proteins in *Homo sapiens* and several other species. The tree grouped the human CAG protein together with 5 other hypothetical proteins in a group separate from the 3 collagen alpha-1 proteins. The human CAG and the homologous mouse hypothetical protein form a distinct group separate from the other CAG protein homologues.

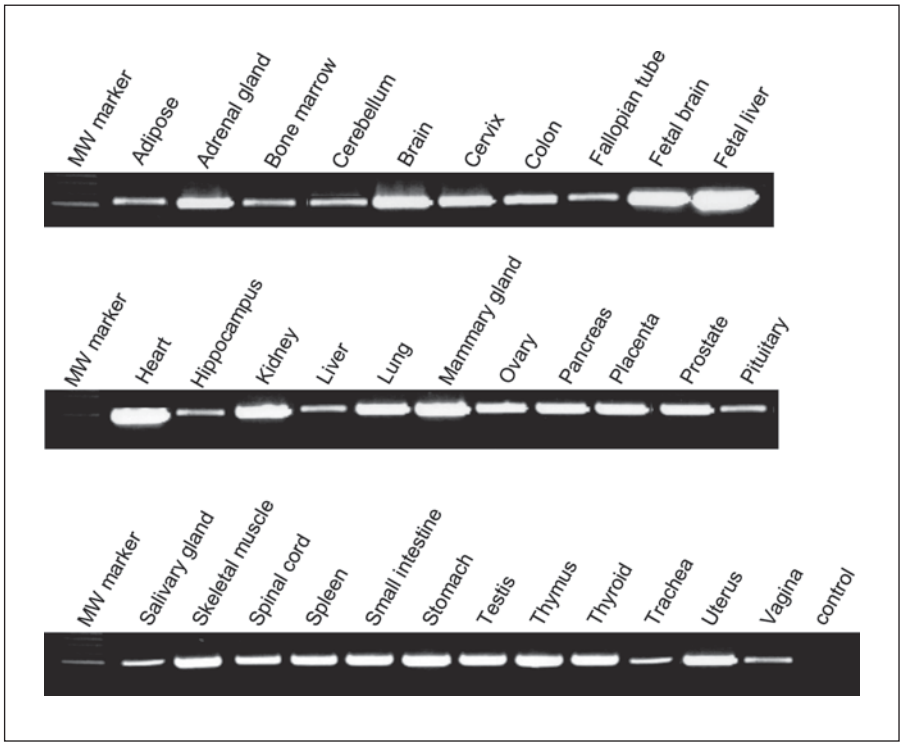

Fig. 6. Tissue expression of *CAG*, as determined by RT-PCR. *CAG* is expressed in a variety of normal human tissues, with highest levels in the brain, liver, heart, kidney and breast. M = Molecular weight marker. PCR was performed with primer set FTEFB/RTEFB (table 1).

els (30 tpm). Comparable results were obtained for colon, stomach and ovarian cancers, in which slightly higher *CAG* expression was apparent in cancerous SAGE libraries, compared to normal. In contrast, *CAG* seems to be expressed at lower levels in cancers of the kidney, pancreas and skin, compared to normal tissues. Our findings with respect to breast cancer were inconclusive.

We experimentally examined the expression of *CAG* in 15 pairs of matched normal and cancerous prostatic tissues. A representative example of 5 matched pairs is shown in figure 7. Expression levels were higher in cancerous prostatic tissues compared to their normal counterparts in 4 patients and comparable to normal in 1 pair. We also analyzed 20 breast cancer tissues and 5 normal

Table 3. In silico analysis of *CAG* gene expression in normal and cancerous tissues using SAGE databases

Organ	Tissue type	Average density (tpm*)
Brain	normal	16
	cancer	24
Colon	normal	20
	cancer	25
Kidney	normal	22
	cancer	0
Ovary	normal	20
	cancer	41
Pancreas	normal	31
	cancer	0
Prostate	normal	15
	cancer	30
Skin	normal	75
	cancer	0
Stomach	normal	0
	cancer	30
Vasculature	normal	19
	cancer	0

* tpm = Tag per million.

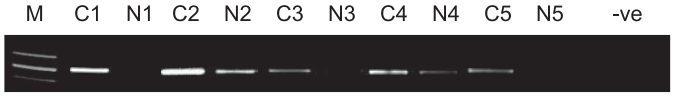

Fig. 7. Expression of *CAG* in matched normal and cancerous prostate tissues, as determined by RT-PCR. *CAG* is overexpressed in prostate cancer tissues compared to normal in 4 matched pairs (C1/N1, C2/N2, C3/N3 and C5/N5) and comparable to normal in 1 pair (C4/N4). M = Molecular weight marker. –ve, negative PCR control. PCR was performed with primer set FTEFB/RTEFB (table 1).

Fig. 8. Expression of *CAG* in normal and cancerous breast tissues, as determined by RT-PCR. *CAG* is expressed in the two normal breast tissues examined (N1 and N2), but transcribed at lower than normal levels in the cancerous tissues (C1–C5). –ve, negative PCR control. PCR was performed with primer set FTEFB/RTEFB (table 1).

breast tissues. Expression levels were higher in all normal tissues examined. Representative data are shown in figure 8.

Discussion

Molecular profiles of individual cancers based on patterns of gene expression often unveil important biological, diagnostic and prognostic information useful for the classification and individualized treatment of patients [35–37]. Recently, microarray analyses and bioinformatic approaches, such as SAGE, have been used to identify such genes differentially expressed in cancer [38, 39]. In this paper, we describe the identification and characterization of a novel gene which is differentially expressed in cancer.

The high degree of conservation of the CAG protein sequence among different species, together with its widespread pattern of expression, points to the possibility that it is a housekeeping gene. The parallel downregulation of *CAG*, together with 4 other adjacent kallikrein genes, namely *KLK10* [40], *KLK12* [41], *KLK13* [42], and *KLK14* [31] in breast cancer, could be due to a common mechanism controlling the expression of many genes in a cluster, e.g. a 'locus control region' [43]. Clustering of co-expressed non-homologous genes on chromosomes implies their co-regulation and has been reported in higher eukaryotes. Recent reports have shown clustering to occur at a rate much higher than that expected by chance and a widespread chromatin-mediated transcriptional co-regulation of such genes [44, 45].

Our in silico and experimental data suggest differential expression of *CAG* in several malignancies, including brain, breast, colon, ovarian and prostate cancer. The chromosomal band 19q13 is known to include many other genes that are differentially expressed in cancer [9, 10, 46, 47]. Since the expression of *CAG* seems to be altered in cancerous tissues, it may be a useful tumor marker for diagnosis and prognosis and/or a drug target. These suggestions warrant further investigation.The opposite pattern of regulation of *CAG* in breast and prostate cancers is not surprising and could be due to hormonal influences. Epidemiological and experimental evidence suggests that steroid hormones are implicated in the etiology of both breast and prostate carcinomas [48]. Breast cancer is an estrogen-dependent cancer and its pathogenesis involves unopposed estrogen exposure [49, 50]. Prostate cancer, on the other hand, is androgen-related, and estrogens have been successfully used for its treatment [51, 52].

CAG regulation by steroids, as is the case with the adjacent kallikrein genes [22], warrants investigation.

In conclusion, we characterized a new gene, *CAG*, differentially expressed in cancer and reported its genomic structure, chromosomal localization and tissue expression pattern. The physiological role of this gene needs further investigation.

Acknowledgements

This work was supported by a grant to E. P. D. from the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Genomic Projects Initiative.

References

1 Balmain A, Gray J, Ponder B: The genetics and genomics of cancer. Nat Genet 2003;33(suppl): 238–244.
2 Strausberg RL: The Cancer Genome Anatomy Project: New resources for reading the molecular signatures of cancer. J Pathol 2001;195: 31–40.
3 Ohno H, Takimoto G, McKeithan TW: The candidate proto-oncogene *bcl-3* is related to genes implicated in cell lineage determination and cell cycle control. Cell 1990;60:991–997.
4 Cheng JQ, Godwin AK, Bellacosa A, Taguchi T, Franke TF, Hamilton TC, Tsichlis PN, Testa JR: AKT2, a putative oncogene encoding a member of a subfamily of protein-serine/threonine kinases, is amplified in human ovarian carcinomas. Proc Natl Acad Sci USA 1992;89: 9267–9271.
5 Tang TC, Sham JS, Xie D, Fang Y, Huo KK, Wu QL, Guan XY: Identification of a candidate oncogene SEI-1 within a minimal amplified region at 19q13.1 in ovarian cancer cell lines. Cancer Res 2002;62:7157–7161.
6 Tikoo A, Czekay S, Viars C, White S, Heath JK, Arden K, Maruta H: p190-A, a human tumor suppressor gene, maps to the chromosomal region 19q13.3 that is reportedly deleted in some gliomas. Gene 2000;257:23–31.
7 Apte SS, Mattei MG, Olsen BR: Mapping of the human *BAX* gene to chromosome 19q13.3-q13.4 and isolation of a novel alternatively spliced transcript, *BAX delta.* Genomics 1995; 26:592–594.
8 Scorilas A, Kyriakopoulou L, Yousef GM, Ashworth LK, Kwamie A, Diamandis EP: Molecular cloning, physical mapping, and expression analysis of a novel gene, *BCL2L12,* encoding a proline-rich protein with a highly conserved BH2 domain of the *Bcl-2* family. Genomics 2001;72:217–221.
9 Diamandis EP, Yousef GM: Human tissue kallikreins: A family of new cancer biomarkers. Clin Chem 2002;48:1198–1205.
10 Yousef GM, Diamandis M, Jung K, Diamandis EP: Molecular cloning of a novel human acid phosphatase gene *(acpt)* that is highly expressed in the testis. Genomics 2001;74:385–395.
11 Muleris M, Almeida A, Gerbault-Seureau M, Malfoy B, Dutrillaux B: Identification of amplified DNA sequences in breast cancer and their organization within homogeneously staining regions. Genes Chromosomes Cancer 1995;14:155–163.
12 Curtis LJ, Li Y, Gerbault-Seureau M, Kuick R, Dutrillaux AM, Goubin G, Fawcett J, Cram S, Dutrillaux B, Hanash S, Muleris M: Amplification of DNA sequences from chromosome 19q13.1 in human pancreatic cell lines. Genomics 1998;53:42–55.
13 Huntsman DG, Chin SF, Muleris M, Batley SJ, CollinsVP, Wiedemann LM, Aparicio S, Caldas C: MLL2, the second human homolog of the *Drosophila trithorax* gene, maps to 19q13.1 and is amplified in solid tumor cell lines. Oncogene 1999;18:7975–7984.
14 Ferbus D, Flechon A, Muleris M, Li Y, Hanash S, Terris B, Hammel P, Pibouin L, Dutrillaux B, Goubin G: Amplification and over-expression of OZF, a gene encoding a zinc finger protein, in human pancreatic carcinomas. Int J Cancer 1999;80:369–372.
15 Bicher A, Ault K, Kimmelman A, Gershenson D, Reed E, Liang B: Loss of heterozygosity in human ovarian cancer on chromosome 19q. Gynecol Oncol 1997;66:36–40.
16 Dumur CI, Dechsukhum C, Ware JL, Cofield SS, Best AM, Wilkinson DS, Garrett CT, Ferreira-Gonzalez A: Genome-wide detection of LOH in prostate cancer using human SNP microarray technology. Genomics 2003;81:260–269.
17 Hoglund M, Gorunova L, Andren-Sandberg A, Dawiskiba S, Mitelman F, Johansson B: Cytogenetic and fluorescence in situ hybridization analyses of chromosome 19 aberrations in pancreatic carcinomas: Frequent loss of 19p13.3 and gain of 19q13.1–13.2. Genes Chromosomes Cancer 1998;21:8–16.
18 Smith JS, Alderete B, Minn Y, Borell TJ, Perry A, Mohapatra G, Hosek SM, Kimmel D, O'Fallon J, Yates A, Feuerstein BG, Burger PC, Scheithauer BW, Jenkins RB: Localization of common deletion regions on 1p and 19q in human gliomas and their association with histological subtype. Oncogene 1999;18:4144–4152.
19 Belge G, Rippe V, Meiboom M, Drieschner N, Garcia E, Bullerdiek J: Delineation of a 150-kb breakpoint cluster in benign thyroid tumors with 19q13.4 aberrations, Cytogenet Cell Genet 2001;93:48–51.
20 Yousef GM, Chang A, Scorilas A, Diamandis EP: Genomic organization of the human kallikrein gene family on chromosome 19q13.3–q13.4. Biochem Biophys Res Commun 2000; 276:125–133.
21 Yousef GM, Diamandis EP: Human kallikreins: Common structural features, sequence analysis and evolution. Current Genomics 2003;4:147–165.
22 Yousef GM, Diamandis EP: The new human tissue kallikrein gene family: Structure, function, and association to disease. Endocr Rev 2001;22:184–204.
23 Yousef GM, Diamandis EP: Kallikreins, steroid hormones and ovarian cancer: Is there a link? Minerva Endocrinol 2002;27:157–166.
24 Luo LY, Yousef G, Diamandis EP: Human tissue kallikreins and testicular cancer. APMIS 2003;111:225–232; discussion 232–233.
25 Yousef GM, Ordon MH, Foussias G, Diamandis EP: Genomic organization of the siglec gene locus on chromosome 19q13.4 and cloning of two new siglec pseudogenes. Gene 2002; 286:259–270.
26 Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
27 Lennon G, Auffray C, Polymeropoulos M, Soares MB: The IMAGE Consortium: An integrated molecular analysis of genomes and their expression. Genomics 1996; 33: 151–152.
28 Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual, ed 2. Cold Spring Harbor, Cold Sping Harbor Laboratory, 1989.
29 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;74:630–636.
30 Yousef GM, Diamandis EP: The expanded human kallikrein gene family: Locus characterization and molecular cloning of a new member, *KLK-L3 (KLK9)*. Genomics 2000;65: 184–194.

31 Yousef GM, Magklara A, Chang A, Jung K, Katsaros D, Diamandis EP: Cloning of a new member of the human kallikrein gene family, *KLK14*, which is down-regulated in different malignancies. Cancer Res 2001;61:3425–3431.
32 Foussias G, Yousef GM, Diamandis EP: Identification and molecular characterization of a novel member of the siglec family *(SIGLEC9).* Genomics 2000;67:171–178.
33 Iida Y: Quantification analysis of 5′-splice signal sequences in mRNA precursors. Mutations in 5′-splice signal sequence of human beta-globin gene and beta-thalassemia. J Theor Biol 1990;145:523–533.
34 Kozak M: An analysis of vertebrate mRNA sequences: Intimations of translational control. J Cell Biol 1991;115:887–903.
35 Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science 1999;286: 531–537.
36 Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C, Meltzer PS: Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nat Med 2001;7:673–679.
37 van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH: Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530–536.
38 Mohr S, Leikauf GD, Keith G, Rihn BH: Microarrays as cancer keys: An array of possibilities. J Clin Oncol 2002;20:3165–3175.
39 Polyak K, Riggins GJ: Gene discovery using the serial analysis of gene expression technique: Implications for cancer research. J Clin Oncol 2001;19:2948–2958.
40 Liu XL, Wazer DE, Watanabe K, Band V: Identification of a novel serine protease-like gene, the expression of which is down-regulated during breast cancer progression. Cancer Res 1996;56:3371–3379.
41 Yousef GM, Magklara A, Diamandis EP: KLK12 is a novel serine protease and a new member of the human kallikrein gene family-differential expression in breast cancer. Genomics 2000;69:331–341.
42 Yousef GM, Chang A, Diamandis EP: Identification and characterization of KLK-L4, a new kallikrein-like gene that appears to be down-regulated in breast cancer tissues. J Biol Chem 2000;275:11891–11898.
43 Li Q, Harju S, Peterson KR: Locus control regions: Coming of age at a decade plus. Trends Genet 1999;15:403–408.
44 Boutanaev AM, Kalmykova AI, Shevelyov YY, Nurminsky DI: Large clusters of co-expressed genes in the Drosophila genome. Nature 2002;420:666–669.
45 Megy K, Audic S, Claverie JM: Positional clustering of differentially expressed genes on human chromosomes 20, 21 and 22. Genome Biol 2003;4:P1.
46 Rondinelli RH, Tricoli JV: CLAR1, a novel gene that exhibits enhanced expression in advanced human prostate cancer. Clin Cancer Res 1999;5:1595–1602.
47 Benedit P, Paciucci R, Thomson TM, Valeri M, Nadal M, Caceres C, de Torres I, Estivill X, Lozano JJ, Morote J, Reventos J: PTOV1, a novel protein overexpressed in prostate cancer containing a new class of protein homology blocks. Oncogene 2001;20:1455–1464.
48 Henderson BE, Feigelson HS: Hormonal carcinogenesis. Carcinogenesis 2000;21:427–433.
49 Russo IH, Russo J: Role of hormones in mammary cancer initiation and progression. J Mammary Gland Biol Neoplasia 1998;3:49–61.
50 Persson I: Estrogens in the causation of breast, endometrial and ovarian cancers – evidence and hypotheses from epidemiological findings. J Steroid Biochem Mol Biol 2000;74:357–364.
51 Bosland MC: The role of steroid hormones in prostate carcinogenesis. J Natl Cancer Inst Monogr 2000;29:39–66.
52 Debes JD, Tindall DJ: The role of androgens and the androgen receptor in prostate cancer. Cancer Lett 2002;187:1–7.