Biochemical and Biophysical Research Communications **285,** 1321–1329 (2001)
doi:10.1006/bbrc.2001.5321, available online at http://www.idealibrary.com on IDEAL®

Sequence Analysis of the Human Kallikrein Gene Locus Identifies a Unique Polymorphic Minisatellite Element

George M. Yousef,*,† Bhupinder S. Bharaj,*,† He Yu,‡
John Poulopoulos,* and Eleftherios P. Diamandis*,†,[1]

*Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada;
†Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada; and
‡Department of Medicine, Louisiana State University Medical Center, Shreveport, Louisiana

Received July 2, 2001

Minisatellites are repetitive sequences of DNA that are present throughout the genome. Although the origin and function of these minisatellites is still unknown, they found clinical applications as markers of many diseases, including cancer. Also, they are useful tools for DNA fingerprinting and linkage analysis. Kallikreins are serine proteases that appear to be involved in many diseases including brain disorders and malignancy. We have recently characterized the human kallikrein gene locus on chromosome 19q13.4, which includes 15 kallikrein genes. In this study, we examined the kallikrein locus (~300 Kb) for all known repeat elements. About 50% of this genomic area is occupied by different repeat elements. We also identified unique minisatellite elements that are restricted to chromosome 19q13. Ten clusters of these minisatellites are distributed along the locus on either DNA strand. The clusters are located in the promoters and enhancers of genes, in introns, and in untranslated regions of the mRNA. Analysis of these elements indicates that they are polymorphic, thus they can be useful in linkage analysis and DNA fingerprinting. Our preliminary results indicate also that the distribution of the different alleles of these minisatellites might be associated with malignancy.

Key Words: **kallikreins; minisatellites; human kallikrein locus; repeat elements; chromosome 19; polymorphism; hypervariable minisatellites; repeat elements; malignancy.**

An intriguing feature of the human genome is the accumulation of distinctive classes of repetitive DNA. These tandemly repetitive sequences were originally identified as satellite bands in ultracentrifuge density gradients of complex eukaryotic genomes, but now the term "satellite" is applied to any tandemly repetitive sequence (1). Satellite DNA units can be divided into two classes: short sequence repeats consisting of 2–20 bp units, and more complex repeat units, or minisatellites (1). One family of the latter class, designated "hypervariable minisatellites" arises from variable number of tandem repeats of short sequence motifs 10- to 100-bp long (2). These arrays are frequently unstable and produce highly polymorphic loci.

Despite the debate about the origin and function of the minisatellite repeats, they found application in many clinical aspects. A probe based on the core sequence of a tandem repeat was able to detect many highly variable loci simultaneously and provided an individual-specific DNA "fingerprint" for general use in human genetic linkage analysis and medicolegal aspects (2, 3). Many satellites are also linked to different diseases. In progressive myoclonus epilepsy, the elongation of a minisatellite repeat seems to be responsible for the disease (4) and a subgroup of alleles at the minisatellite locus for the insulin gene has been associated with type I diabetes (5). Furthermore, rare mutant alleles of the HRAS1 minisatellite were found to occur three times more frequently in patients with cancer and to represent a major risk factor for many types of cancer (6, 7).

One unique feature of chromosome 19 is that it has the highest repeat density in the whole genome (57%), as well as the highest gene density (8). In our previous

Abbreviations used: KLK, kallikrein; PCR, polymerase chain reaction; dNTPs, deoxynucleoside triphosphates; RT, reverse transcription; TLSP, trypsin-like serine protease; HSCCE, human stratum corneum chymotryptic enzyme; NES1, normal epithelial cell-specific 1 gene; kb, kilobase; BAC, bacterial artificial chromosome; SINE, short interspersed nuclear element; LINE, long interspersed nuclear element; LTR, long terminal repeat.

[1] To whom correspondence should be addressed at Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. Fax: 416-586-8628. E-mail: ediamandis@mtsinai.on.ca.

0006-291X/01 $35.00

TABLE 1

Primers Used for Polymerase Chain Reaction (PCR) Amplification of the Minisatellite Clusters

Cluster ID	Primer name	Sequence[1]	Product size[2] (base pairs)
C3	G2F	CACCAACCTCTGCAAATTCA	
	G2RC	GGATCTAGGGGAAAGGGACA	546
C9	G1F	GAAGCTGGATTGAGGAAACG	600
	G1RC	GTGCCTCCGGTCTTGAGTAG	

[1] All nucleotide sequence are given in the 5′ → 3′ orientation.
[2] Product sizes were calculated according to the human genome project sequences available at the Lawrence Livermore National Laboratory web site.

work, we characterized the human kallikrein gene locus on chromosome 19q13.3–q13.4 (9, 10). In this paper, we describe sequence analysis of the different microsatellite and minisatellite repeats of this locus. We also identify a unique 37-bp minisatellite element which is specific to this chromosomal region and is heavily distributed in the locus. We further provide evidence that this minisatellite exhibits a polymorphic pattern that may have clinical and medicolegal applications. Our preliminary results indicate that these minisatellites show different frequency patterns in breast and prostate cancer patients.

MATERIALS AND METHODS

DNA extraction and PCR amplification. DNA was extracted from whole blood and tissues using the Qiagen QIAamp blood kit, following the manufacturer's recommendations (Qiagen, Chatsworth, CA). A PCR-based, fragment analysis method was employed as described in detail elsewhere (11). Specific pairs of primers were used for PCR amplification of each repeat cluster, as shown in Table 1. In summary, PCR was carried out in a reaction mixture containing 1 μl of genomic DNA (approximately 1 μg/reaction), 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM dNTPs (deoxynucleoside triphosphates), 150 ng of primers, and 2.5 units of HotStar DNA polymerase (Qiagen Inc., Valencia, CA) on a Perkin-Elmer 9600 thermal cycler. The cycling conditions were 95°C for 15 min to activate the enzyme, followed by 35 cycles of 94°C for 30 s, 64°C for repeat cluster C3 (64°C for C9) for 30 s, 72°C for 1 min, and a final extension step at 72°C for 10 min. One of the PCR primers was labeled with the fluorescent dye Cy5.5 in order to facilitate detection of the PCR product on the automated DNA sequencer (Visible Genetics Inc., Toronto, On, Canada).

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

Identification of the PCR product. Two microliters of the PCR product were mixed with 4 μL of a gel loading buffer (Visible Genetics) to which 2 μL each of the two molecular weight markers were added. This mixture was denatured at 95°C for 2 min, placed on ice and then loaded onto the sequencing gel (see below).

The Long Read Tower (Visible Genetics Inc.) automated sequencer was used for fragment analysis. Once the gel is loaded, it is positioned into the sequencer and the buffer chambers filled with 150 mL of 1 × TBE (Tris-Borate-EDTA) buffer, pH 8.3. A 5 min prerun is initiated to bring the gel temperature to the set values. Each lane is loaded with 2 μL of denatured mixture of unknown DNA, size standard marker, and loading buffer. For quality control purposes, every fourth lane is loaded with size standard marker only. Laser intensity was adjusted to bring the signals within the dynamic range of the detectors. At the end of the run (60 min), the data was analyzed using the Fragment Tool of the Gene Objects software (Visible Genetics). The unknown PCR product in the three lanes was aligned with the fourth lane which contains the size standards only.

Prostate and breast cancer tissues and control samples. Prostate tissue samples were obtained from nine patients who had undergone radical retropubic prostatectomy for prostatic adenocarcinoma at Mount Sinai Hospital. The patients did not receive any hormonal therapy before surgery. Fresh tissue samples were obtained from the cancerous and noncancerous parts of the same prostates that had been removed. Histologically confirmed tissue specimens were obtained from breast cancer patients who underwent surgical treatment for primary breast carcinoma at Mount Sinai Hospital. The use of these tissues for research purposes was approved by the Ethics Committee of Mount Sinai Hospital. Small pieces of tissue were dissected immediately and stored in liquid nitrogen until analysis. Histological analysis from all the tissue pieces was performed to ensure that the tissue was either malignant or benign. The tissues were pulverized and DNA extracted as described above. Whole blood samples were obtained from age-matched, apparently healthy men (blood donors). Individuals with documented evidence or suspicion of malignancy were excluded.

RESULTS

Sequence Analysis of the Human Kallikrein Gene Locus

We have recently constructed the first detailed map of the human kallikrein gene locus on chromosome 19q13.3–q13.4 (10). The locus consists of 261,558 bp of genomic sequence that harbors 15 kallikrein genes on either DNA strand, with no intervening nonkallikrein genes. In order to further characterize the locus, we analyzed the whole sequence for existence of known repeat elements. The sequence has 49.59% GC content, which is comparable to other genomic regions. Approximately 52% of the region was found to be occupied by various repetitive elements (on either strand). Short interspersed nuclear elements (SINEs), like the ALU and MIR repeats, are the most abundant repeat elements in the region (22.53%), followed by the Long interspersed nuclear elements (LINEs) which represent 13.1%. Table 2 describes the repeat elements found in kallikrein locus.

Identification of a Unique Minisatellite Repeat

During structural analysis of the newly identified KLK14 gene (12), we found that the 3′ untranslated region of the gene is formed mainly by a tandem of ~36-bp of a unique minisatellite repeat. We examined the presence of this repeat element in the rest of the locus, and we were able to identify 10 clusters of this minisatellite in the kallikrein locus. Table 3 and Fig. 1 summarize the distribution of these repeats along the

TABLE 2

Summary of the Repeat Elements Found in the Human Kallikrein Gene Locus

	Number/ elements	Length (bp)	Percentage of sequence
SINEs:	269	66938	22.5%
ALUs	202	56668	19.1%
MIRs	67	10270	3.5%
LINEs:	102	38919	13.1%
LINE1	47	19831	6.7%
LINE2	53	18881	6.4%
L3/CR1	2	207	0.0%
LTR elements:	53	21711	7.3%
MaLRs	14	5356	1.8%
ERVL	22	8137	2.7%
ERV_classI	16	7787	2.6%
ERV_classII	1	431	0.2%
DNA elements:	33	9087	3.1%
MER1_type	28	6066	2.0%
MER2_type	5	1472	0.5%

kallikrein locus. The clusters were distributed on both strands of the DNA. As shown in Table 3, the number of minisatellites in each cluster varied widely from just three repeats between the KLK6 and KLK7 genes to 36 repeats in intron I of the KLK7 gene. Four clusters are located between genes, at distances ranging from 1–6 kb upstream of the transcription initiation site, two are located in the 3′ UTR of genes and three are inside introns. Interestingly, one of the clusters (C10) was found to extend from the last part of exon 3 of the KLK14 gene (Table 3). Genomic regions with clusters of repeats were always composed of either the forward or the reverse-complementary sequence of the minisatellite. Around the KLK14 locus, which had a cluster in each direction, the forward-direction cluster was on a different intron than the reverse-complementary one (Table 3, Fig. 1).

As shown in Fig. 2, multiple alignment of the minisatellite repeats show that they are highly conserved, with more than 80% degree of identity. We developed a consensus sequence (AGTCCAGGCCCCCAGCCCCT-CCTCCCTCAGACCCAGG) based on the frequency of occurrence of each base pair, and this sequence was "blasted" against the nonredundant GenBank database and the unfinished genomic sequences database. As shown in Table 4, high degrees of homology (>80%) were only found in genes located in the chromosomal region 19q13.2–q13.4, including the kallikrein genes (not shown in Table 4), apolipoprotein C-II, cardiac troponin I, and XRCC1 DNA repair genes. In genes other than the kallikreins, the repeat element occurs in variable copy numbers ranging from only one copy in the ATP1A1 gene, to 23 copies in the cardiac troponin I gene. These data indicate that this repeat element is specific to this chromosomal region. Sequence homology with the unfinished genomic sequences also revealed that all matches are located on chromosome 19-derived bacterial artificial chromosome (BAC) and cosmid clones (data not shown).

Polymorphism of the Minisatellite Repeats

We conducted a preliminary screening for two randomly selected clusters, C3 and C9 with DNA samples from healthy donors, using the PCR technique. PCR amplification of the DNA was followed by fragment length analysis using a sequencing gel, as described in the experimental section. Screening for C3 in 35 DNA samples from healthy individuals indicated the presence of three different size alleles for this cluster. In addition to the expected 546-bp fragment (the length of the expected PCR product was calculated according to the Human Genome Project published sequence of this region), two other fragments of 618 and 675 bp were obtained. The latter fragments represent an additional two and four minisatellite units, respectively. Figure 3

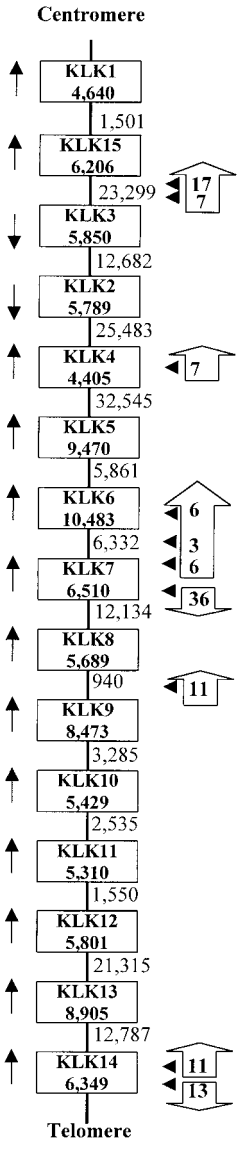
TABLE 3

Analysis of the Clusters of the Unique Minisatellite of the Kallikrein Locus

Cluster ID	Map location	Relative position to genes	No. of repeats	DNA strand[1]
C1	Between KLK3 & KLK15	KLK15 promoter	17	F
C2	Between KLK3 & KLK15	KLK15 promoter	7	F
C3	KLK4	3′ UTR[2]	7	F
C4	KLK6	Intron I	6	F
C5	Between KLK6 & KLK7	KLK6 promoter	3	F
C6	KLK7	Intron IV	6	F
C7	KLK7	Intron I	36	RC
C8	Between KLK8 & KLK9	KLK8 promoter	11	F
C9	KLK14	3′ UTR	11	F
C10	KLK14	Exon 5 Intron V	13	RC

[1] The forward (F) and reverse complementary (RC) strands are defined with respect to the consensus sequence, developed based on multiple alignments (see results).

[2] UTR; untranslated region.



Chromosomal locus 19q13.3-q13.4

FIG. 1. Schematic representation of the relative positions of the minisatellite repeats along the human kallikrein gene locus on chromosome 19q13.3–q13.4. Boxes represent kallikrein genes, and the genomic lengths are shown in base pairs. The direction of transcription is shown by the solid arrows and distances between genes in base pairs are also shown next to the connecting lines. The arrowheads point to the relative positions of the different clusters in relation to genes, and the numbers inside the hollowed arrows indicate the number of repeats present in each cluster. The direction of the hollowed arrows indicate the direction of the strand (whether centromeric or telomeric) of each cluster.

shows a representative gel of PCR products with the three alleles and Fig. 4 shows the lengths of these alleles as determined by fragment length analysis. The 546-bp allele was found to be the most frequent, occurring in 25 out of 35 samples (69%). The remaining two alleles (618 and 675 bp) occur at a frequency of 34 and 37%, respectively.

For the C9 cluster, three alleles were identified; the expected 600-bp allele, and in addition, a 528-bp (two one less minisatellite unit) and a 490-bp (three less units) fragment was observed (Fig. 5). The 600-bp allele was the most frequent (occurs in 96% of the samples), followed by the 528 allele (38%). The 490 bp-allele was found in one sample only (2%).

Screening of the Kallikrein Locus for Other Minisatellites

We also screened the kallikrein locus for the presence of other, previously reported satellites, but none of these elements were found to exist in our sequence, except for the GGAT microsatellite, reported by Weller *et al.* (13) upstream of the myoglobin gene. A tandem repetitive sequence based on $(GGAT)_{83}$ is located in intron I of the KLK13 gene. Jeffreys *et al.* reported the isolation of a heterogeneous group of minisatellites using a probe based on a 33-bp satellite repeat identified in the human myoglobin gene (2, 13). The kallikrein locus was devoid of these minisatellites.

```
AGTCCAGGCCCC*AGCCCCTCCTCCCTCAGACCCAGG
AGTCCAGGCCCCCAGCCCCTCCTCCCTCAGACCCGGG
AGTCCAGGCCCCCAGCCCCTCCTCCCTCAGACCCAGG
AGTCCAGGCCCC*AGCCCCTCCTCCCTCAGACCCGGG
AGTCCAGGCCCCCAGCCCCTCCTCCCTCAGACCCAGG
AGTCCAGGCCCC*AGTCCCTCCTCCCTCAGACCCAGG
AGTCCAGGCCCCCAGCCCCTCCTCCCTCAGACCCAGG
AATCCAGGCCC**AGCCCCTCCTCCCTCAGACCCAGG
AGTCCAGGCCC**AGCCCCTCCTCCCTCAGACCCAGG
AATCCAGGCCCCTAGCCCCTCCTCCCTCAGACCCAGG
AATCCAGGCCCCTAGCCCCTCCTCCCTCAGACCCAGG
AGTCCAGGCCCCCAGCCTCTCCTCCCTCAGACCCAGG
AGTCCAGGCCCCCAGCCCCTCCTCCCTCAGACCCAGG
AGTCCAGGCCCCCAGCCCCTCCTCCCTCAGACCCGGG
AGTCCAGACCCC*AGCCCCTCCTCCCTCAGACCCAGC
AGTCCAGGCCCCCAGCCCCTCCTTCCTCAGACCCAGG
AGTCCAGGCCCCCAGCCCCTCCTCCCTCAGACCCAGG
AGTCCAGGCCCCAAGCCCCTCCTCCCTCAGACGCAAG
*GTCCAGGCCCCCAGCCC*TCCTCCCTCAGACTCAGG
AGTCCAGGCCCCAAGCCCCTCCTCCCTCAGACCCAGG
AGTCCAGGCCCCCAGCCCCTCCTCCCTCAAACCAAGG
*GTACAGATCCCCAGCCCCTCCTCCCTCAGACCCAGG
GGTCCAGGCCCCCAACCCCTCCTCCCTCAGACTCAGA
GGTCCAGGTCCC*AGCCCCTCCTCCCTCAGACCCAG*
AGTCCAGGACCCCAGCCCCTCCTCCCTCAGACCCAGG
AGTCCAGGCACC*AGCTTCTCCTCCCTCAGACCCAGG
AGTGCAGGCCCTCAGCCCCACCTCCCTCAGACTCAGG
AGTCCGGGCCCTCAGTCCCACCTCCCTCAGACCCAGG
AGTCCAGACCCC*AGCCGCTCCTCCCTTGGACCCAAG
AGTCCCGGCCCCCAGCCTCTCCTTTCTCAGACCCAGG
AGTCCAGGCCCCCAGCCCCTCCTCCCTCAGACCCAGG
AGTCCAGACCCCCAGCCCCTCCTCCCTCAGACCCAGA
AGTCCAGACCCCCAGCCCCTCCTCCCTCAGACCCAGG
```

FIG. 2. A contiguous DNA sequence within intron I of the KLK7 gene, representing 33 of the 36 minisatellite units. The sequence was aligned to highlight the differences (in bold) between the repeats. Stars represent gaps to bring the sequences to better alignment.

TABLE 4

Genes that Show Regions of High Degree of Homology with the Consensus Minisatellite Sequence Found in the Kallikrein Gene Locus[1]

Gene name	Number of matches[2]	Chromosomal localization[3]	GenBank Accession No.
Apolipoprotein C-II	4	19q13.2	M10612
Cardiac troponin I	23	19q13.4	X90780
XRCC1 DNA repair gene	14	19q13.2	L34079
Alpha 1,2 fucosyltransferase	9	19q13.3	AB006136
LENG3[4]	7	19q13.4	AF211968
ATP1A1	1	1p13	M28284
Poliovirus related protein 2	2	19q13.2–q13.3	AF044964
Interleukin 11 (mRNA)	3	19q13.3–q13.4	M81890
Troponin T (TNNT1)	18	19q13.4	AJ011712
BCL3	7	19q13.1	U05681

[1] Kallikrein genes located in the kallikrein locus were excluded.
[2] Matches on both strands were considered and only matches with >80% homologies were included.
[3] According to GenBank report.
[4] Leucocyte receptor cluster encoded novel gene.

Association of the Unique Minisatellite Repeats with Malignancy

We performed a preliminary screening for the occurrence of the C3 minisatellites in nine pairs of normal and malignant prostatic tissue from the same patient. prostate tissues. Out of the nine pairs examined, three pairs showed heterogeneity in the distribution of the different alleles in the normal and cancerous tissues. Figure 4 shows the results of two of these samples. The prostate cancer sample (PC8) has the 546 and 618 alleles, while its normal counterpart (PN8) has the 546 allele only. The normal tissue sample (PN7) has two alleles, while its cancerous partner (PC7) has the 546 allele only. This points out to the possibility that any of these clusters might have somatic mutations associated with malignancy.

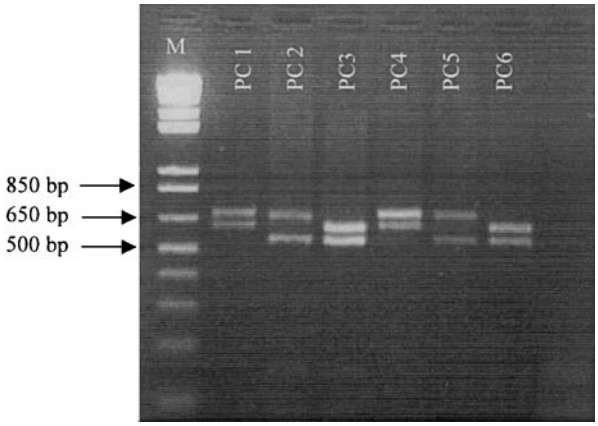


FIG. 3. A representative gel showing the three alleles that were amplified by PCR for the C3 cluster, from six prostate cancer (PC) tissues. The PC1 sample has the 675 and 618 bp bands, while the PC2 has 675 and 546 bp bands. Band identity was verified by sequencing. M, molecular weight marker, with lengths of markers shown on left.

We have also conducted preliminary screening for the frequency of the different alleles of these two clusters in breast and prostate cancer tissues and normal matched controls. Table 5 shows that while the frequency of the 600 and 490 alleles of the C9 cluster were almost the same in the cancer and control patients, the 528-bp allele was significantly more frequent in patients with histologically confirmed breast tumors compared with their normal matched controls ($P = 0.019$). For the C3 cluster, the frequency of the 546 allele was higher in patients with prostate tumors (histologically confirmed patients with high PSA values) compared to the control subjects (with normal PSA values and no evidence of malignancy by physical and rectal examination).

DISCUSSION

Kallikreins are serine proteases with diverse physiological functions (14). The unique minisatellites of the human kallikrein locus are located in exons, introns, and in the promoter/enhancer regions of genes. Other, previously reported minisatellites were found to be intimately associated with genes and gene clusters within chromosomes (15). Both the alpha-globin and immunoglobulin heavy chain loci contain several distinct minisatellites in both intergenic and intronic locations (16–18). Minisatellites appear within introns of the retinoblastoma, interleukin-6, myoglobin, and apolipoprotein C genes (19). Insulin, apolipoprotein B, and collagen type II genes bear minisatellites immediately upstream or downstream from coding sequences (20). Other genes, such as epithelial mucin, involucin, and proline-rich proteins contain transcribed and translated minisatellite arrays.

Four clusters of this minisatellite were found to be located in the promoter/enhancer region (1–6 Kb up-

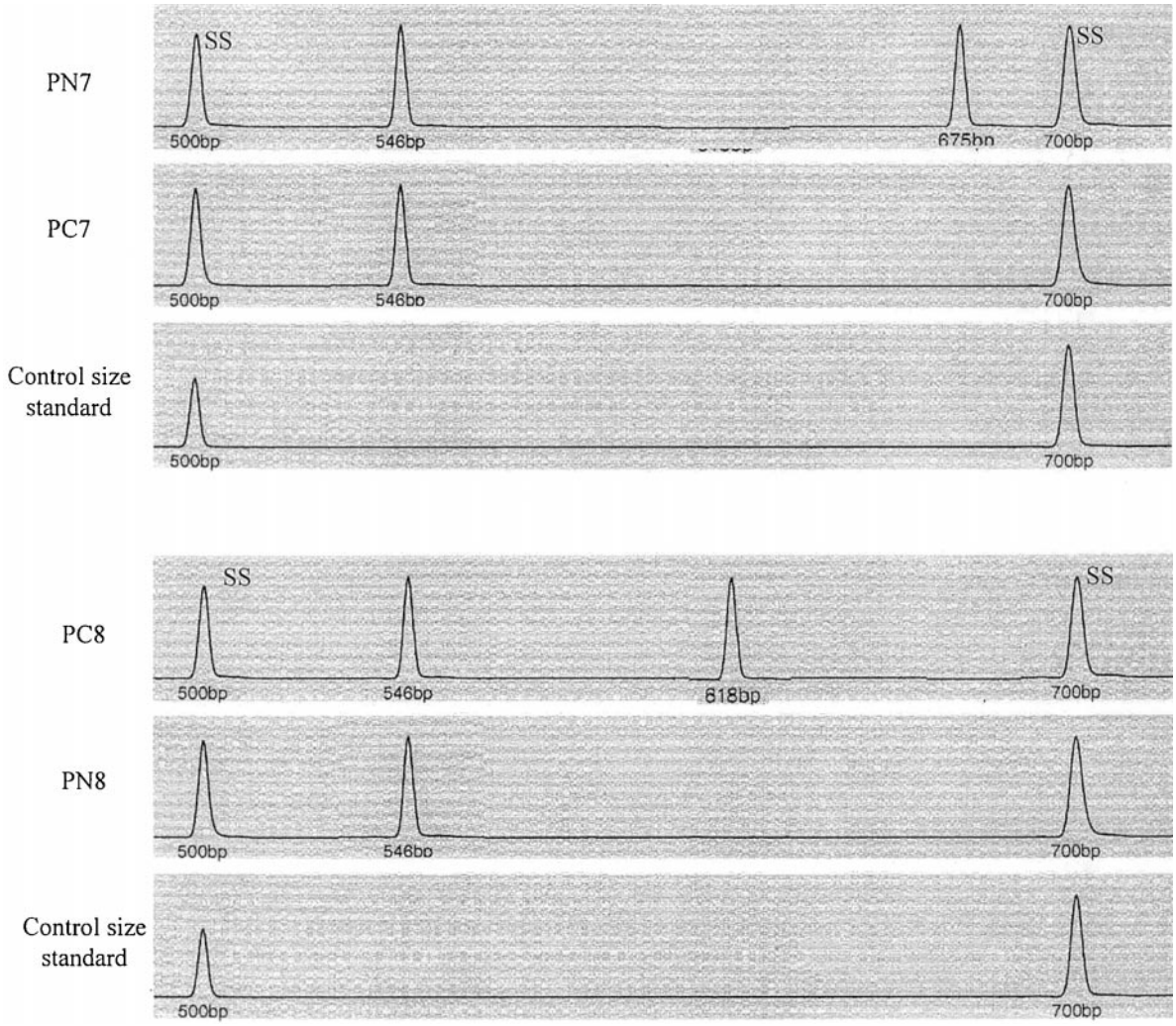


FIG. 4. A representative graph showing the three alleles of the C3 cluster of minisatellites. The numbers below the peaks represent lengths of the PCR products in base pairs. The upper panel shows a comparison between the cancerous tissue (PC7) and its normal counterpart (PN7), and the lower panel shows another pair of normal and cancerous tissues (PC8/PN8). For each lane, size standard markers (SS) were added to the sample, and the markers were run alone in the third and sixth lanes.

stream of the transcription initiation site) of some kallikrein genes, pointing out to the possibility that they might play a role in the regulation of gene expression. This idea is supported by comparing the consensus sequence with known transcription-factor binding sites where we found the presence of conserved AP-2, TFIID, and SP-1 binding sites in the consensus sequence (data not shown).

Our results indicate that this minisatellite repeat is specific to the chromosomal region 19q13.2–q13.4, consistent with an earlier report that points out to this possibility (21). Being chromosome and band-specific, these minisatellites represent a unique phenomenon. Yoshida *et al.* (22) reported the presence of two repeat sequences in the promoter region of the neuropsin gene. Careful examination of these repeats reveals that they are components of the main minisatellite repeat presented here.

The pattern of base changes between the repeat blocks is distinctly non-random, e.g., bases 2, 8, 10, 20, and 36 are always replaced by (A) and bases 12, 16, and 24 are always changed from C to T (Fig. 2). In addition, some of the mutated repeats are identical [see e.g., lines 2 and 4, and 10 and 11 in Fig. 2]. This observation was previously reported for other minisatellites (18) and could be explained by either multiple unequal crossing over events or multiple rounds of localized gene conversion.

An interesting feature of these minisatellites is that they exhibit polymorphism. This phenomenon can find applicability in DNA fingerprinting for medicolegal purposes and can also be utilized in linkage analysis. It is not obvious how such sequences came to reside in different locations. The sequences flanking the repeats are not related and show no evidence to support a transposition theory. Jeffreys *et al.* reported the pres-

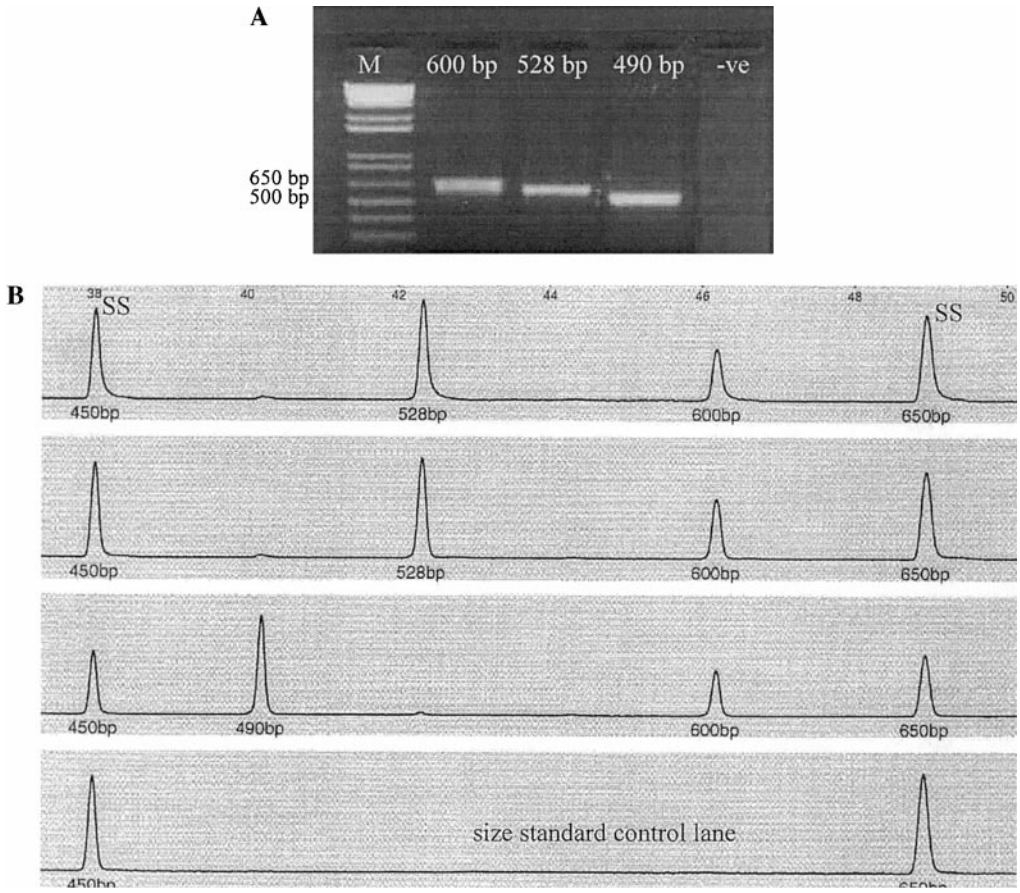


FIG. 5. A representative graph showing the three alleles of the C9 cluster. The numbers below the peaks represent lengths of the PCR products in base pairs. For each lane, two size standard (SS) markers (450 bp and 650 bp) were added to the sample, and the markers were run alone in the fourth lane. For more details, see text.

ence of a core region of an almost invariant sequence (GGGGCAGGAXG) in a heterogeneous group of minisatellites isolated by probing with the myoglobin gene minisatellites. Multiple alignment of our satellite with these repeats shows that a comparable sequence (GGGAGGAGG) exists in almost all of our minisatellites. This core region might help to generate minisatellites by promoting the initial tandem duplication of unique sequence DNA and/or by stimulating the subsequent unequal exchanges required to amplify the duplication into a minisatellite (2). Also, this core sequence is similar to the generalized recombination signal (X) of *E. coli* (23). As polymorphic minisatellites may also be recombination hotspots (2), it might be significant that the core sequence is similar in length and "G" content to the (X) sequence.

Our preliminary results indicate that at least some of these minisatellites occur in higher frequency in cancer. This finding is not unprecedented. Previous reports indicate that rare mutant alleles of the HRAS1 minisatellite were found to occur three times more frequently in patients with cancer and that they are a major risk factor for common types of cancer (6, 7). In three out of nine pairs (normal/cancer) of prostate tissues examined, different alleles were found when comparing the normal and malignant tissues from the same patient. There is now a debate regarding the association of minisatellite repeats and different diseases, including cancer. One theory hypothesizes that they demonstrate linkage disequlibrium with inherited pathogenic lesions, and they simply represent markers for the disease without being actually involved in the pathogenesis. However, other studies indicated that some of these minisatellites bind to transcriptional regulatory factors and can up- or down-regulate the transcription of nearby genes (24, 25).

We have evidence that our minisatellites posses binding sites for transcription factors. The intronic minisatellite of the interleukin-1α also has a binding site for SP-1 transcription factor (26). Krontiris *et al.* proposed a model explaining how could the minisatellite alterations affect transcription directly (7). In addition, it has been shown that many of the kallikrein genes are differentially expressed in different malignancies (12, 14, 27–33). The KLK3 (PSA) and KLK2 (human glandular kallikrein) already found clinical

TABLE 5

Frequency of the Different Alleles of the C3 and C9 Clusters in Breast and Prostate Cancer Patients and Controls

(A) C3 cluster

Type of patient	Allele frequency (%) 546	618	675
Breast cancer cases	29 (81)	9 (25)	16 (44)
Breast controls	19 (79)	7 (29)	9 (38)
P value	0.99	0.77	0.79
Prostate cancer cases	10 (83)	5 (42)	6 (50)
Prostate controls	5 (42)	6 (50)	5 (42)
P value	0.089	1.00	1.00

(B) C9 cluster

Type of patient	Allele frequency (%) 490	528	600
Breast cancer cases	0 (0)	18 (75)	23 (96)
Breast controls	0 (0)	7 (38)	22 (92)
P value	1.00	0.019	1.00
Prostate cancer cases	2 (8)	9 (38)	22 (92)
Prostate controls	1 (4)	9 (38)	23 (96)
P value	1.00	1.00	1.00

applicability as prostate cancer markers, and other kallikrein genes are recently emerging as potential biomarkers for different cancers (34–36).

ACKNOWLEDGMENTS

This work was supported by a University-Industry Collaboration Research and Development Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC), and Visisble Genetics Inc., Toronto, Canada.

REFERENCES

1. Csink, A. K., and Henikoff, S. (1998) Something from nothing: The evolution and utility of satellite repeats. *Trends Genet.* **14,** 200–204.
2. Jeffreys, A. J., Wilson, V., and Thein, S. L. (1985) Hypervariable "minisatellite" regions in human DNA. *Nature* **314,** 67–73.
3. Jeffreys, A. J., Wilson, V., and Thein, S. L. (1985) Individual-specific "fingerprints" of human DNA. *Nature* **316,** 76–79.
4. Lafreniere, R. G., Rochefort, D. L., Chretien, N., Rommens, J. M., Cochius, J. I., Kalviainen, R., Nousiainen, U., Patry, G., Farrell, K., Soderfeldt, B., Federico, A., Hale, B. R., Cossio, O. H., Sorensen, T., Pouliot, M. A., Kmiec, T., Uldall, P., Janszky, J., Pranzatelli, M. R., Andermann, F., Andermann, E., and Rouleau, G. A. (1997) Unstable insertion in the 5′ flanking region of the cystatin B gene is the most common mutation in progressive myoclonus epilepsy type 1, EPM1. *Nat. Genet.* **15,** 298–302.
5. Raffel, L. J., Vadheim, C. M., Roth, M. P., Klein, R., Moss, S. E., and Rotter, J. I. (1991) The 5′ insulin gene polymorphism and the genetics of vascular complications in type 1 (insulin-dependent) diabetes mellitus. *Diabetologia* **34,** 680–683.
6. Krontiris, T. G., DiMartino, N. A., Colb, M., and Parkinson, D. R. (1985) Unique allelic restriction fragments of the human Ha-ras locus in leukocyte and tumour DNAs of cancer patients. *Nature* **313,** 369–374.
7. Krontiris, T. G., Devlin, B., Karp, D. D., Robert, N. J., and Risch, N. (1993) An association between the risk of cancer and mutations in the HRAS1 minisatellite locus. *N. Engl. J. Med.* **329,** 517–523.
8. Venter, J. C., *et al.* (2001) The sequence of the human genome. *Science* **291,** 1304–1350.
9. Yousef, G. M., and Diamandis, E. P. (2000) The expanded human kallikrein gene family: Locus characterization and molecular cloning of a new member, KLK-L3 (KLK9). *Genomics* **65,** 184–194.
10. Yousef, G. M., Chang, A., Scorilas, A., and Diamandis, E. P. (2000) Genomic organization of the human kallikrein gene family on chromosome 19q13.3-q13.4 [In process citation]. *Biochem. Biophys. Res. Commun.* **276,** 125–133.
11. Bharaj, B. S., Vassilikos, E. J., and Diamandis, E. P. (1999) Rapid and accurate determination of (CAG)n repeats in the androgen receptor gene using polymerase chain reaction and automated fragment analysis. *Clin. Biochem.* **32,** 327–332.
12. Yousef, G. M., Magklara, A., Chang, A., Jung, K., Katsaros, D., and Diamandis, E. P. (2001) Cloning of a new member of the human kallikrein gene family, KLK14, which is down-regulated in different malignancies. *Cancer Res.* **61,** 3425–3431.
13. Weller, P., Jeffreys, A. J., Wilson, V., and Blanchetot, A. (1984) Organization of the human myoglobin gene. *EMBO J,* **3,** 439–446.
14. Yousef, G. M., and Diamandis, E. P. (2001) The new human tissue kallikrein gene family: Structure, function, and association to disease. *Endo. Rev.* **22,** 184–204.
15. Dover, G. A. (1989) DNA fingerprints: Victims or perpetrators of DNA turnover? *Nature* **342,** 347–348.
16. Higgs, D. R., Goodbourn, S. E., Wainscoat, J. S., Clegg, J. B., and Weatherall, D. J. (1981) Highly variable regions of DNA flank the human alpha globin genes. *Nucleic Acids Res.* **9,** 4213–4224.
17. Proudfoot, N. J., Gil, A., and Maniatis, T. (1982) The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. *Cell* **31,** 553–563.
18. Goodbourn, S. E., Higgs, D. R., Clegg, J. B., and Weatherall, D. J. (1983) Molecular basis of length polymorphism in the human zeta-globin gene complex. *Proc. Natl. Acad. Sci. USA* **80,** 5022–5026.
19. Wei, C. F., Tsao, Y. K., Robberson, D. L., Gotto, A. M., Brown, K., and Chan, L. (1985) The structure of the human apolipoprotein C-II gene. Electron microscopic analysis of RNA:DNA hybrids, complete nucleotide sequence, and identification of 5′ homologous sequences among apolipoprotein genes. *J. Biol. Chem.* **260,** 15211–15221.
20. Bell, G. I., Selby, M. J., and Rutter, W. J. (1982) The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. *Nature* **295,** 31–35.
21. Das, H. K., Jackson, C. L., Miller, D. A., Leff, T., and Breslow, J. L. (1987) The human apolipoprotein C-II gene sequence contains a novel chromosome 19-specific minisatellite in its third intron. *J. Biol. Chem.* **262,** 4787–4793.
22. Yoshida, S., Taniguchi, M., Hirata, A., and Shiosaka, S. (1998) Sequence analysis and expression of human neuropsin cDNA and gene. *Gene* **213,** 9–16.
23. Smith, G. R., Kunes, S. M., Schultz, D. W., Taylor, A., and Triman, K. L. (1981) Structure of chi hotspots of generalized recombination. *Cell* **24,** 429–436.
24. Trepicchio, W. L., and Krontiris, T. G. (1992) Members of the rel/NF-kappa B family of transcriptional regulatory proteins

bind the HRAS1 minisatellite DNA sequence. *Nucleic Acids Res.* **20,** 2427–2434.

25. Green, M., and Krontiris, T. G. (1993) Allelic variation of reporter gene activation by the HRAS1 minisatellite. *Genomics* **17,** 429–434.
26. Haugen, A., Mann, D., Murray, C., Weston, A., and Willey, J. C. (1989) Interleukin-1 alpha gene intron containing variable repeat region coding for the SP1 transcription factor recognition sequence is polymorphic. *Mol. Carcinog.* **2,** 68–71.
27. Diamandis, E. P., Yousef, G. M., Luo, L. Y., Magklara, A., and Obiezu, C. V. (2000) The new human kallikrein gene family: Implications in carcinogenesis. *Trends Endocrinol. Metab.* **11,** 54–60.
28. Yousef, G. M., Magklara, A., and Diamandis, E. P. (2000) KLK12 is a novel serine protease and a new member of the human kallikrein gene family-differential expression in breast cancer [In process citation]. *Genomics* **69,** 331–341.
29. Goyal, J., Smith, K. M., Cowan, J. M., Wazer, D. E., Lee, S. W., and Band, V. (1998) The role for NES1 serine protease as a novel tumor suppressor. *Cancer Res.* **58,** 4782–4786.
30. Anisowicz, A., Sotiropoulou, G., Stenman, G., Mok, S. C., and Sager, R. (1996) A novel protease homolog differentially expressed in breast and ovarian cancer. *Mol. Med.* **2,** 624–636.
31. Kim, H., Scorilas, A., Katsaros, D., Yousef, G. M., Massobrio, M., Fracchioli, S., Piccinno, R., Gordini, G., and Diamandis, E. P. (2001) Human kallikrein gene 5 (KLK5) expression is an indicator of poor prognosis in ovarian cancer. *Br. J. Cancer* **84,** 643–650.
32. Yousef, G. M., Scorilas, A., Jung, K., Ashworth, L. K., and Diamandis, E. P. (2001) Molecular cloning of the human kallikrein 15 gene (KLK15). Up- regulation in prostate cancer. *J. Biol. Chem.* **276,** 53–61.
33. Yousef, G. M., Chang, A., and Diamandis, E. P. (2000) 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.
34. Diamandis, E. P., Yousef, G. M., Soosaipillai, A. R., and Bunting, P. (2000) Human kallikrein 6 (zyme/protease M/neurosin): A new serum biomarker of ovarian carcinoma. *Clin. Biochem.* **33,** 579–583.
35. Diamandis, E. P., Yousef, G. M., Soosaipillai, A. R., Grass, L., Porter, A., Little, S., and Sotiropoulou, G. (2000) Immunofluorometric assay of human kallikrein 6 (zyme/protease M/neurosin) and preliminary clinical applications [In process citation]. *Clin. Biochem.* **33,** 369–375.
36. Luo, L. Y., Grass, L., Howarth, D. J., Thibault, P., Ong, H., and Diamandis, E. P. (2001) Immunofluorometric assay of human kallikrein 10 and its identification in biological fluids and tissues. *Clin. Chem.* **47,** 237–246.