Am. J. Hum. Genet. 52:124–134, 1993

A Duplicated Region Is Responsible for the Poly(ADP-ribose) Polymerase Polymorphism, on Chromosome 13, Associated with a Predisposition to Cancer

Deborah Lyn,* Barry W. Cherney,*[,1] Marc Lalande,† James R. Berenson,‡ Alan Lichtenstein,‡ Shelley Lupold,* Kishor G. Bhatia,§ and Mark Smulson*

*Department of Biochemistry and Molecular Biology, Georgetown University School of Medicine, Washington, DC; †Genetics Division, Children's Hospital, and Department of Pediatrics, Harvard Medical School, Boston; ‡Department of Medicine, UCLA School of Medicine, and Veterans Administration Medical Center, Los Angeles; and §National Cancer Institute, Bethesda

Summary

The poly(ADP-ribose) polymerase (PADPRP) gene (13q33-qter) depicts a two-allele (A/B) polymorphism. In the noncancer population, the frequency of the B allele is higher among blacks than among whites. Since the incidence of multiple myeloma and prostate and lung cancer is higher in the U.S. black population, we have analyzed the B-allele frequency in germ-line DNA to determine whether the PADPRP gene correlates with a polymorphic susceptibility to these diseases. For multiple myeloma and prostate cancer, an increased frequency of the B allele appeared to be striking only in black patients. In contrast, the distribution of the B allele in germ-line DNA did not differ among white patients with these diseases, when compared with the control group. An elevated B-allele frequency was also found in germ-line DNA in blacks with colon cancer. These observations suggest that the PADPRP polymorphism may provide a valid marker for a predisposition to these cancers in black individuals. To determine the genomic structure of the polymorphic PADPRP sequences, a 2.68-kb *Hin*dIII clone was isolated and sequenced from a chromosome 13–enriched library. Sequence analysis of this clone (A allele) revealed a close sequence similarity (91.8%) to PADPRP cDNA (1q42) and an absence of introns, suggesting that the gene on 13q exists as a processed pseudogene. A 193-bp conserved duplicated region within the A allele was identified as the source of the polymorphism. The nucleotide differences between the PADPRP gene on chromosome 13 and related PADPRP genes were exploited to develop oligonucleotides that can detect the difference between the A/B genotypes in a PCR. This PCR assay offers the opportunity for analyzing additional black cancer patients, to determine how the PADPRP processed pseudogene or an unidentified gene that cosegregates with the PADPRP gene might be involved with the development of malignancy.

Introduction

Poly(ADP-ribose) polymerase (PADPRP) (E.C.2.4.2.30) is a DNA-binding protein that modulates chromatin structure adjacent to regions of DNA replication, recombination, and repair (Ueda and Hayaishi 1985). In the course of analyzing potential rearrangements within this gene in tumor cells, a simple two-allele (A/B) polymorphism localized to chromosome 13q33-qter was observed. A twofold increase in the frequency of the B allele (statistically significant) was also noted in tumor DNA from patients with Burkitt lymphoma, B-cell follicular lymphoma, and lung carcinoma, when compared with germ-line DNA in a noncancer population (Bhatia et al. 1990).

The RFLP was identified as a 2.7-kb or 2.5-kb *Hin*dIII fragment after hybridization to full-length human PADPRP cDNA. These fragments were thought to originate either from a PADPRP processed pseudogene or from a gene with extensive identity to

Received July 13, 1992; revision received September 15, 1992.

Address for correspondence and reprints: Mark Smulson, Department of Biochemistry and Molecular Biology, Georgetown University School of Medicine, 3900 Reservoir Road N.W., Washington DC 20007.

[1] Present address: Food and Drug Administration, Bethesda, MD.

0002-9297/93/5201-0015$02.00

PADPRP, but they did not reflect the gene encoding the authentic PADPRP protein on chromosome 1q or the pseudogene on chromosome 14 (Bhatia et al. 1990). A preliminary characterization of the two-allele polymorphism showed that a number of restriction enzymes, including *Kpn*I, *Eco*RI, *Bgl*II, *Rsa*I, and *Msp*I, also identified this polymorphism, which always cosegregated together and differed by 200 bp between the respective A and B alleles. Collectively, these RFLPs suggested a deletion or insertion of DNA of at least 200 bp adjacent to or within PADPRP-like sequences.

Initial analysis of DNA derived from tumor and normal tissue of the same individual revealed that the 200 bp difference did not occur as a somatic event, and the predominant source of the B allele was germ line, although a tumor-derived loss of heterozygosity was found in 5% of the matched samples. In the noncancer population, a marked difference in the frequency of the B allele was also observed in germ-line DNA from black (.35) and white (.14) individuals (Bhatia et al. 1990). An important aspect of the earlier study was the observation that at least one copy of the B allele was always present in tumor tissue from patients with endemic Burkitt lymphoma, and the frequency of the B allele was at least twofold higher than in the black, noncancer population. A limited survey of germ-line DNA from black patients having various cancers showed a 1.7-fold increase in the B-allele frequency, which compared favorably with the data collected for endemic Burkitt lymphoma (Bhatia et al. 1990). It has been suggested that racial differences in the cancer incidence rates are mainly attributable to socioeconomic or life-style factors. However, it is not well understood whether a genetic basis contributes to the observed epidemiological data. We have used the polymorphic PADPRP DNA marker on chromosome 13, associated with a possible predisposition to cancer, to gain insights into understanding how a genetic basis may account for a high occurrence of certain cancers in the black population.

Thus, we have extended the previous analysis to include a new group of patients with cancers that occur more frequently in the black population (multiple myeloma and prostate and lung cancer). To provide insight into the association between the PADPRP polymorphism on chromosome 13 and a predisposition to cancer, we have also characterized the genomic structure of the polymorphic PADPRP sequences. The present report presents data that indicate that the polymorphism reflects a 193-bp duplication of PADPRP processed–pseudogene sequences and that the absence of this duplicated region is often present in individuals with certain types of cancer. In addition, a strategy was developed to analyze the PADPRP genotype of patients, by using the PCR, in which the DNA sequences responsible for the A/B polymorphism could be selectively amplified.

Subjects, Material, and Methods

Cancer Patients

Patients with multiple myeloma were studied at both UCLA and the Veterans Administration Medical Center in West Los Angeles. All other cancer patients were under treatment at either Georgetown University Hospital or Howard University Cancer Center, Washington, DC. Data on the noncancer groups were taken from Bhatia et al. (1990).

Isolation of a Chromosome 13 PADPRP Genomic Clone

A recombinant DNA library containing *Hin*dIII inserts was prepared from flow-sorted chromosome 13, and the phage vector Charon 21A was screened using full-length PADPRP cDNA. This library has successfully been used to isolate DNA markers for the 13q14 region (Lalande et al. 1984). Approximately 5×10^5 phage were screened under stringent hybridization and wash conditions.

DNA Subcloning and Sequencing

DNA was isolated from one of the positive genomic clones, digested with *Hin*dIII, and subcloned into dephosphorylated, *Hin*dIII-digested pBluescript (Stratagene), by standard techniques (Sambrook et al. 1989, pp. 1.53–1.73). This clone was referred to as "pH2.68BT." The DNA was sequenced by the dideoxynucleotide chain-terminating method (Sanger et al. 1977) using Sequenase (U.S. Biochemical). Initially, the 5′ and 3′ termini of the clone were determined using primers to the T3 and T7 promoter regions. The entire sequence was determined on both strands by using a series of oligonucleotides obtained from sequential reactions. The products of the PCR were gel isolated and sequenced on both strands from independent reactions using nested primers generated from the sequence in figure 1. Sequence data were analyzed using the Genetics Computer Group program (Madison, WI).

```
             .         .         .         .         .
    1   AAGCTTGAAAAGGCCCTAAAGGCCCAGAACGACCTGATCTGGAACATCAA   A allele
        ||||||||||| ||||||||||| ||||||||||||||||||||||||||
  887   AAGCTTGAAAAAGCCCTAAAGGCTCAGAACGACCTGATCTGGAACATCAA   cDNA

             .         .         .         .         .
   51   GGACGAGCTAATGAAAGTGTGCTCTATTAATGACCTGAAGGAGATGCTCA   A allele
        ||||||||||| ||||||||| || | |||||||||||||||| | ||||
  937   GGACGAGCTAAAGAAAGTGTGTTCAACTAATGACCTGAAGGAGCTACTCA   cDNA

             .         .         .         .         .
  101   TCTTCAACAGGCAGCAGGTGCCTTCCGGGGAGTCGGCGATCTTGGACCGA   A allele
        ||||||||| |||||| |||||||| ||||||||||||||||||||||||
   98   TCTTCAACAAGCAGCAAGTGCCTTCTGGGGAGTCGGCGATCTTGGACCGA   cDNA

             .         .         .         .         .
  151   GTAGCTGACAGCATGACGTTCGGTGCCCTCCTTCCCTGTGAGGAATGCTC   A allele
        ||||||||  |||||  ||||||||||||||||||||| |||||||||||
 1037   GTAGCTGATGGCATGGTGTTCGGTGCCCTCCTTCCCTGCGAGGAATGCTC   cDNA

             .         .         .         .         .
  201   AGGTCAGCTGGTCTTCAAGAGCGACACTTATTACTGCACCGGGGACGTCA   A allele
         |||||||||||||||||||||||  | ||||||||||| ||||||||||
 1087   GGGTCAGCTGGTCTTCAAGAGCGATGCCTATTACTGCACTGGGGACGTCA   cDNA

             .         .         .         .         .
  251   CGGCCTGGACCAAGTGTATGGTCAAGACATGGACACCCAACCAGAAGGAA   A allele
        | |||||||||||||||||||||||||||  ||||||||||| ||||||
 1137   CTGCCTGGACCAAGTGTATGGTCAAGACACAGACACCCAACCGGAAGGAG   cDNA

             .         .         .         .         .
  301   TAGGTGATCCCA........CCGAGAAATCTCTTAACTCAAGAAATTTAA   A allele
        | ||| | ||||        ||||||||||||||| ||||||||||| ||
 1187   TGGGTAACCCCAAAGGAATTCCGAGAAATCTCTTACCTCAAGAAATTGAA   cDNA

             .         .         .         .         .
  343   GGTTAAAAAGCAGGACCGTATATTTCCCCCAGAAACCAGTGACCCAGTGG   A allele
        |||||||||||||||||||||||| |||||||||||||| | | | ||||
 1237   GGTTAAAAAGCAGGACCGTATATTCCCCCCAGAAACCAGCGCCTCCGTGG   cDNA

             .         .         .         .         .
  393   CGGCCATGCCCCCACCCTCCACAGCCTCCGAGCCTGCTGCCATGAACGCC   A allele
        |||||| ||| || |||||||||||||| |  ||||||||  ||||| ||
 1287   CGGCCACGCCTCCGCCCTCCACAGCCTCGGCTCCTGCTGCTGTGAACTCC   cDNA

             .         .         .         .         .
  443   TCTGCTCCAGCAGATAAGCCGTTATCCAACATGAAGATCCTTACTCTTGG   A allele
        |||||| ||||||||||||| |||||||||||||||||||| ||||| ||
 1337   TCTGCTTCAGCAGATAAGCCATTATCCAACATGAAGATCCTGACTCTCGG   cDNA

             .         .         .         .         .
  493   GAAGCTCTCCCGGAACAAGGATGAAGTGAAGGCCACGATTGAGAAACTCG   A allele
        |||||| |||||||||||||||||||||||||||| ||||||||||||||
 1387   GAAGCTGTCCCGGAACAAGGATGAAGTGAAGGCCATGATTGAGAAACTCG   cDNA

             .         .         .         .         .
  543   GGGGAAAGTTGATGGGGACGGCCAACAAGGCTTCTCCGTGCATCAGCACT   A allele
        |||| ||||||| ||||||||||||||||||||| | ||||||||||||
 1437   GGGGGAAGTTGACGGGGACGGCCAACAAGGCTTCCCTGTGCATCAGCACC   cDNA

             .         .         .         .         .
  593   AAAGAGGAGGTGGAAAAGAAGAAGAAGAAGATGGAGGAAGTGAAAGAAGC   A allele
        ||| ||||||||||||||| ||| ||||||||||||||||| || |||||
 1487   AAAAAGGAGGTGGAAAAGATGAATAAGAAGATGGAGGAAGTAAAGGAAGC   cDNA

                       .         .         .         .
  643   CAACATCTGAGCTGTGTCTGAGGACTTCCTCCAGGACTTCTCTGCCTCCA   A allele
        ||||||| ||| ||||||||||||||||||||||||| |||| |||||||
 1537   CAACATCCGAGTTGTGTCTGAGGACTTCCTCCAGGACGTCTCCGCCTCCA   cDNA

             .         .         .         .         .
  693   CCAAGGGGCTCCAGGAGTTGTTCTCAGCGCACATATTGACCCCCTGGGGG   A allele
        ||||| | || ||||||||||||| ||||||||| ||| |||| ||||||
 1587   CCAAGAGCCTTCAGGAGTTGTTCTTAGCGCACATCTTGTCCCCTTGGGGG   cDNA
```

```
            +++++++++      .         .         .         .
  743   GCAGAGGTGAAGGCAGAGCCTGTTGAAGTCGTAGCCCCAAGAGGGAAGTC   A allele
        ||||||||||||||||||||||||||||| || |||||||||||||||||
 1637   GCAGAGGTGAAGGCAGAGCCTGTTGAAGTTGTGGCCCCAAGAGGGAAGTC   cDNA

             .         .         .         .         .
  793   AGGAGCTGTGCTCTCCAAAAAAAGCAAGGGCCAGGTCAAGGAGGAAGGTA   A allele
        ||| |||| |||||||||||||||||||||||||||||||||||||||||
 1687   AGGGGCTGCGCTCTCCAAAAAAAGCAAGGGCCAGGTCAAGGAGGAAGGTA   cDNA

             .         .         .         .         .
  843   TCAACAAATCTGAAAAGAGAATGAAATTAACTCTTAAAGGAGGAGCAGCT   A allele
        ||||||||||||||||||||||||||||||||||||||||||||||||||
 1737   TCAACAAATCTGAAAAGAGAATGAAATTAACTCTTAAAGGAGGAGCAGCT   cDNA

             .         .         .         .         .
  893   GTGGATCCTGACTCTGGTCTGGAACACTCTGCGCATGTCCTGGAGAAAGG   A allele
        ||||||||||| ||||| |||||||||||||||||||||||||||||
 1787   GTGGATCCTGATTCTGGACTGGAACACTCTGCGCATGTCCTGGAGAA...   cDNA

             .         .         .         .         .
  943   TGAAGGCAGAGCCTGTTGAAGTCGTAGCCCCAAGAGGGAAGTCAGGAGCT   A allele

        ..................................................

             .         .         .         .         .
  993   GTGCTCTCCAAAAAAAGCAAGGGCCAGGTCAAGGAGGAAGGTATCAACAA   A allele

        ..................................................

             .         .         .         .         .
 1043   ATCTGAAAAGAGAATGAAATTAACTCTTAAAGGAGGAGCAGCTGTGGATC   A allele

        ..................................................

             .         .         .         .+++++  +++
 1093   CTGACTCTGGTCTGGAACACTCTGCACATGTTCTGGAGAAAGGTGGAAG    A allele
                                                ||||||||||
 1826   ........................................AGGTGGAAG    cDNA

        +        .         .         .         .         .
 1143   GTCTTCAGTGCCACCCTCAGCCTGGTGGACGTCGTTAAAGGAACCAACTC   A allele
        |||||||||||||||||  ||||||||||| |||||||||||||||||||
 1844   GTCTTCAGTGCCACCCTTGGCCTGGTGGACATCGTTAAAGGAACCAACTC   cDNA

             .                             .         .
 1193   CTATTACAAGCTGAAGTTGCTGAAGGATGACAAGGAAAGCAGGCATTGGA   A allele
        ||| ||||||||| || | ||| |||| |||||||||| |||| ||||||
 1894   CTACTACAAGCTGCAGCTTCTGGAGGACGACAAGGAAAACAGGTATTGGA   cDNA

             .         .         .         .         .
 1243   TATTCAAGTCCTGGGACCGTGTGGGCACGGTGATCGGTAGCAACAAACTG   A allele
        |||||| |||||||| ||||||||| ||||||||||||||||||||||||
 1944   TATTCAGGTCCTGGGGCCGTGTGGGTACGGTGATCGGTAGCAACAAACTG   cDNA

             .         .         .         .         .
 1293   GAACAGATGCTGTCCAAGGAGGACACCATTGAACACTTCATGAAATTATA   A allele
        |||||||||| ||||||||||||  ||||||| || ||||||||||||||
 1994   GAACAGATGCCGTCCAAGGAGGATGCCATTGAGCAGTTCATGAAATTATA   cDNA

             .         .         .         .         .
 1343   TGAAG.AAAAACTAGGAATGCTTGGCACTCCAAAAA.TTCACAAAGTATC   A allele
        ||||| ||||||  |||| ||||||||||||||||| ||||| |||||||
 2044   TGAAGAAAAAACCGGGAACGCTTGGCACTCCAAAAATTTCACGAAGTATC   cDNA

             .         .         .         .         .
 1391   CCAAAAAGTTCTACCCCCTGGAGATTGACTACGGCCAGGACAAAGAGGCG   A allele
        |||||||||| |||||||||||||||||||| ||||||||  ||||||| 
 2094   CCAAAAAGTTTTACCCCCTGGAGATTGACTATGGCCAGGATGAAGAGGCA   cDNA

             .         .         .         .         .
 1441   GTGAAGAAGCCGACAGTAAATCCTGGCACCAAGTCCATGCTCCCCAAGCC   A allele
        ||||||||||  ||||||||||||||||||||||||| ||||||||||||
 2144   GTGAAGAAGCTCACAGTAAATCCTGGCACCAAGTCCAAGCTCCCCAAGCC   cDNA
```

```
               .         .         .         .         .
1491  AGTTCAGGACCTCATCAAGATGATCTTTGATGTGGAAAGTATGAGCAAAG   A allele
      ||||||||||||||||||||||||||||||||||||||||||||  ||||
2194  AGTTCAGGACCTCATCAAGATGATCTTTGATGTGGAAAGTATGAAGAAAG   cDNA
               .         .         .         .         .
1541  CCATGGTGGGGTGTGAGATCAACCTTC...AGATGCCCTTGGGGAAGCTG   A allele
      ||||||||| || ||||||| ||||||   ||||||||||||||||||||
2244  CCATGGTGGAGTATGAGATCGACCTTCAGAAGATGCCCTTGGGGAAGCTG   cDNA
               .         .         .         .         .
1588  AGCAAAAGGCAAATCCAGGCCGCGTACTCCATCCTC....AGGTCCAGCA   A allele
      ||||||||||| ||||||||||| ||||||||||||    ||||||||||
2294  AGCAAAAGGCAGATCCAGGCCGCATACTCCATCCTCAGTGAGGTCCAGCA   cDNA
               .         .         .         .         .
1634  GGTGGTGTCCCAGGGCAGCAGCGACTCTCAGATCCTGGATCTCTCAAATC   A allele
      || |||||| ||||||||||||||||||||||||||||||||||||||||
2344  GGCGGTGTCTCAGGGCAGCAGCGACTCTCAGATCCTGGATCTCTCAAATC   cDNA
               .         .         .         .         .
1684  GCTTTTACATCCTGATCCCCCACGACTTTGGGATGAAGGATCCTCTGCTC   A allele
      ||||||||| |||||||||||||||||||||||||||| | |||| ||||
2394  GCTTTTACACCCTGATCCCCCACGACTTTGGGATGAAGAAGCCTCCGCTC   cDNA
               .         .         .         .         .
1734  CTGAACAATGCAGACAGTGTGCAGGCCAAGGTAGAAATGCTGGACAACCT   A allele
      |||||||||||||||||||||||||||||||| |||||||| ||||||||
2444  CTGAACAATGCAGACAGTGTGCAGGCCAAGGTGGAAATGCTTGACAACCT   cDNA
               .         .         .         .         .
1784  GCTGGACATTGAGGTAGCCTACGGTCTGCTCAGGGGAGGGTCTCACGATA   A allele
      ||||||||| ||||| |||||| |||||||||||||||||||| | ||||
2494  GCTGGACATCGAGGTGGCCTACAGTCTGCTCAGGGGAGGGTCTGATGATA   cDNA
               .         .         .         .         .
1834  GCAGGAAGGACTCCATCGATGTCAACTATGAGAAGCTCAAAACTGACATT   A allele
      |||| |||||  ||||||||||||||||||||||||||||||||||||||
2544  GCAGCAAGGATCCCATCGATGTCAACTATGAGAAGCTCAAAACTGACATT   cDNA
               .         .         .         .         .
1884  AAGGTGGTTGACAGAGATTCTGAAGAAGCTGAGATCATCAGGAAGTATGT   A allele
      ||||||||||||||||||||||||||||| ||||||||||||||||||||
2594  AAGGTGGTTGACAGAGATTCTGAAGAAGCCGAGATCATCAGGAAGTATGT   cDNA
               .         .         .         .         .
1934  TAAGAACACTCATGCAACCAACCACACACGATGCATATGACTTGGAAGTC   A allele
      ||||||||||||||    |||||||||||  ||| |||||||||||||||
2644  TAAGAACACTCATG....CAACCACACACAGTGCGTATGACTTGGAAGTC   cDNA
               .         .         .         .         .
1984  ATTGATAGCTTTAAGATAGAGTGTGAAGAGGAGTGCCAGCACTACAAGCC   A allele
      || |||| ||||||||||||| ||||||  || |||||||  ||||||||
2690  ATCGATATCTTTAAGATAGAGCGTGAAGGCGAATGCCAGCGTTACAAGCC   cDNA
               .         .         .         .         .
2034  CTTTAAGCAGCTTCATAACTGAAGGTTGCTGTGGCATGGGTCCAGGACCA   A allele
      ||||||||||||||||||| |||| ||||||||||| |||||||||||||
2740  CTTTAAGCAGCTTCATAACCGAAGATTGCTGTGGCACGGGTCCAGGACCA   cDNA
               .         .         .         .         .
2084  CCAACTTTGCTGGGATCCTGTCCCTGGGTCTTTGGATAGCCCTGCCTGAA   A allele
      |||||||||||||||||||||||| ||||||| ||||||||| |||||||
2790  CCAACTTTGCTGGGATCCTGTCCCAGGGTCTTCGGATAGCCCCGCCTGAA   cDNA
               .         .         .         .         .
2134  GCACCTGTGATGGGCTACATGTTTGGTAAAGTGATCTATTTCGCTGATCT   A allele
      || || ||||  ||||||||||||||||||| |||||||||||||||  |
2840  GCGCCCGTGACAGGCTACATGTTTGGTAAAGGGATCTATTTCGCTGACAT   cDNA
               .         .         .         .         .
2184  TGTCTCCAAGAGTGCCAACGACTGCCATACATCTTAGGAAGACCCAATAG   A allele
       |||||||||||||||||| ||| |||||| ||| ||| |||||||||||
2890  GGTCTCCAAGAGTGCCAACTACTACCATACGTCTCAGGGAGACCCAATAG   cDNA
               .         .         .         .         .
2234  GGTTAATCCTGTCGGAAGAAGTTGCCCTTGGAAACGTGTGTGAACTGAAG   A allele
      | |||||||||| || ||||||||||||||||||| ||| ||||||||||
2940  GCTTAATCCTGTTGGGAGAAGTTGCCCTTGGAAACATGTATGAACTGAAG   cDNA
               .         .         .         .         .
2284  CATGCTTCACATATCAGCAAGTTACCCAAGGGCAAGCACAGTGTCAAAGG   A allele
      || |||||||||||||||| ||||||||||||||||||||||||||||||
2990  CACGCTTCACATATCAGCAGGTTACCCAAGGGCAAGCACAGTGTCAAAGG   cDNA
               .         .         .         .         .
2334  TTTGGGCAAAACTACTCCTGACCTTTCAGCTAGTATCCCACTGGATGGTG   A allele
      ||||||||||||||| ||||| | ||||||||  ||    ||||||||||
3040  TTTGGGCAAAACTACCCCTGATCCTTCAGCTAACATTAGTCTGGATGGTG   cDNA
               .         .         .         .         .
2384  TAGAGGTTCCTCTTGGGACCAGGGTTTCATCTGGTGTGAATGACACCTGT   A allele
      |||| ||||||||||||||| || |||||||||||||||||||||||| |
3090  TAGACGTTCCTCTTGGGACCGGGATTTCATCTGGTGTGAATGACACCTCT   cDNA
               .         .         .         .         .
2434  CTACTGTATAATGAGTACATTGTCTATGATATTGCTCAGGTAAATCTGAA   A allele
      ||||| ||||| ||||||||||||||||||||||||||||||||||||||
3140  CTACTATATAACGAGTACATTGTCTATGATATTGCTCAGGTAAATCTGAA   cDNA
               .         .         .         .         .
2484  ATATCTGCTGAAACTGAAATTCAATTTTAAGACCTCCTTGTGGTAATTGG   A allele
       |||||||||||||||||||||||||||||||||||| ||||||||||||
3190  GTATCTGCTGAAACTGAAATTCAATTTTAAGACCTCCCTGTGGTAATTGG   cDNA
                                                 ***
               .         .         .         .         .
2534  GAGAGGTGGCTGAGTCACACACGGTGACTCTGGTATTAATTCACCCTAAG   A allele
      ||||||| || ||||||||| ||||| || |||||| ||||||||| |||
3240  GAGAGGTAGCCGAGTCACACCCGGTGGCTGTGGTATGAATTCACCCGAAG   cDNA
               .         .         .         .         .
2584  CGCTTCTGCACCAACTCACCTGGCTGGCTAAGTTGCTGGTGGGTAGTACC   A allele
      ||||||||||||||||||||||||  |||||||||||| |||||||||||
3290  CGCTTCTGCACCAACTCACCTGGC.CGCTAAGTTGCTGATGGGTAGTACC   cDNA
               .         .         .         .
2634  TGTACTAAACCTCCTCAGAAAGGATTTTGCAGAAATGCATTAGAAGCTT    A allele
      ||||||||||| |||||||||||||||| |||||| |  ||| | | ||
3339  TGTACTAAACCACCTCAGAAAGGATTTTACAGAAACGTGTTAAAGGTTT    cDNA
```

Figure 1 Alignment of the PADPRP *Hin*dIII clone (A allele) to the cDNA. The upper sequence represents the complete sequence of the isolated A allele (chromosome 13), while the lower sequence is the cDNA (derived from chromosome 1) as reported by Cherney et al. (1987). Sequence matches are shown by vertical lines. The boxed areas represent the 193-bp duplicated regions. The overlined region corresponds to the oligonucleotide primers used in the PCR; the area overlined by the plus sign (+) represents sequences at the start and flanking the 3′ end of the duplicated region, while triple asterisks (***) indicate the protein-termination codon of the cDNA that encodes the authentic PADPRP protein.

PCR

Genomic DNA (100–400 ng) or plasmid DNA (1–2 ng) was amplified using the primers 5′-AAGAAGCCAACATCTGAGCT-3′ and 5′-TTTCCTTGTCATCCTTCAGC-3′ for 30 cycles of the following: 45 s at 94°C, 1 min at 62°C, and 2 min at 72°C. The reaction was carried out in a 100-μl volume containing 2.5 units of AmpliTaq DNA polymerase (Perkin Elmer Cetus), 10 mM Tris-HCl pH 8.3, 50 mM KCl, 2.5 mM $MgCl_2$, 0.001% (w/v) gelatin, and 0.2 mM of each deoxynucleoside triphosphate.

Other Methods

Southern blotting, hybridization conditions, and preparation of peripheral blood lymphocytes were performed according to methods described elsewhere (Bhatia et al. 1990). Germ-line DNA was obtained from peripheral blood lymphocytes.

Results

Frequency of the PADPRP B Allele in Germ Line–derived DNA from Black Cancer Patients

Since the increased frequency of the B allele in patients with cancer was rarely a result of tumor-associated loss of heterozygosity, normal DNA from peripheral blood lymphocytes was used to screen a new group of cancer patients for the PADPRP polymorphism. To determine whether the increased frequency of the B allele observed in patients with endemic Burkitt lymphoma was also found in a related B-cell malignancy, we analyzed germ-line DNA from 68 patients who had multiple myeloma. Multiple myeloma is the only hemopoietic malignancy for which the incidence in U.S. blacks is twice that observed among whites (Riedel et al. 1991). It has been postulated that immunological determinants such as the human leukocyte antigens may contribute to the racial difference in incidence rates (Pottern et al. 1992). In 31 black patients with multiple myeloma, the frequency of the PADPRP B allele was .66, nearly double that observed in the noncancer population (table 1). In contrast, the frequency of the B genotype among the white patients was no different from that of the appropriate control group.

Prostate cancer is another cancer of high occurrence among U.S. blacks and is at least 50% higher than in the white population (Gloeckler Ries et al. 1990). Since the incidence of prostate cancer is higher among blacks in the United States than among those in Africa, an interaction between genetic and environmental factors is clearly involved. However, the exact environmental factors that contribute to the etiology of this disease are still unknown. The frequency of the B allele was .72 among the black patients, which was twice as high as that observed in the noncancer population (table 1). On the other hand, none of the 16 patients in this cohort of white patients showed the homozygous BB genotype, and the B-allele frequency (.19) did not differ from the frequency in the control, noncancer group (.14).

We observed a 1.7-fold and 2.4-fold increase in the frequency of the B allele in patients with lung cancer, from black and white individuals, respectively, compared with the control group (table 1). However, the increase in B-allele frequency was not statistically significant compared with the appropriate noncancer group ($P > .05$). Germ-line DNA from two other groups of black patients having either colon or breast cancer was also analyzed for the PADPRP genotype. It is notable that all 11 black patients with colon cancer had at least one copy of the B allele. In an earlier study (Bhatia et al. 1990), 62 matched samples of unknown racial distribution from the Vogelstein laboratory (Johns Hopkins University Hospital) showed a 1.6-fold higher frequency of the B allele over that of the white control group, for the PADPRP genotype. In contrast, the homozygous A allele was predominant in germ-line DNA from 21 black women with breast cancer, and, in our experience, this was one of the few cancers thus far studied that did not appear to be correlated with an increased frequency of the PADPRP B allele.

To summarize, for multiple myeloma and for prostate and colon cancer, an increased frequency of the B allele appeared to be striking in germ-line DNA from black patients. On the other hand, the distribution of the B genotype did not differ significantly among the white cancer patients compared with the noncancer population.

Analysis of the PADPRP Sequences on Chromosome 13

Previously, we used the full-length PADPRP cDNA to analyze the RFLPs associated with the PADPRP sequences on chromosome 13 (Bhatia et al. 1990; present study). Hybridization under stringent conditions showed a strong signal intensity of the 2.5-kb and 2.7-kb *Hin*dIII fragment in a Southern blot of genomic DNA. Preliminary data using different regions of the PADPRP cDNA as a probe to *Hin*dIII-restricted genomic DNA of the A and B genotypes

Table 1

Frequency of the B Allele in Germ-Line DNA from Patients with Various Cancers

	No. of Individuals with Genotype[a]					
Cancer Type and Population	AA	AB	BB	Total	B-Allele Frequency	P^b
Multiple myeloma:						
Black	5	11	15	31	.66	<.003
White	28	7	2	37	.15	.968
Prostate:						
Black	1	3	5	9	.72	.010
White	10	6	0	16	.19	.653
Lung:						
Black	1	5	3	9	.61	.080
White[c]	4	4	1	9	.33	.075
Colon:						
Black	0	8	3	11	.64	.033
Breast:						
Black	12	7	2	21	.26	.430
Control (noncancer population):						
Black[c]	15	18	4	37	.35	
White[c]	45	12	2	59	.14	

[a] Designation of the genotypes was as described in the study by Bhatia et al. (1990), in which germ-line DNA was restricted with *Hin*dIII and hybridized to full-length PADPRP cDNA after Southern blotting.

[b] *P* calculations compare the B-allele frequency in cancer patients with that in the noncancer population of the same racial group.

[c] Distribution of genotypes was taken from Bhatia et al. (1990).

indicated that the sequences on chromosome 13 were colinear with respect to the cDNA and that they probably represented a processed pseudogene (Cherney et al. 1987; B. W. Cherney and K. G. Bhatia, unpublished observations).

A chromosome 13-enriched library of *Hin*dIII fragments (see Subjects, Material, and Methods) was therefore screened using full-length cDNA to human PADPRP. The entire sequence of one of the isolated genomic clones was determined and was found to be the expected *Hin*dIII fragment of 2,682 bases. An earlier analysis of the polymorphic *Hin*dIII alleles had estimated the sizes to be 2.8 kb and 2.6 kb for the A and B genotypes, respectively (Bhatia et al. 1990). Genomic DNA from cell lines previously identified as representing either the A allele (fig. 2, lane 1) or B allele (fig. 2, lane 2) was restricted with *Hin*dIII and was compared with the isolated clone (fig. 2, lane 5) after Southern blotting. A direct size comparison between the cloned genomic fragment and the A and B genotypes showed that the isolated clone represented the *Hin*dIII A allele. We therefore correct the original *Hin*dIII allelic combination (Bhatia et al. 1990) and report them as 2.7 kb (A allele) and 2.5 kb (B allele). The chromosome 13 origin of the isolated *Hin*dIII clone was verified by hybridization to a human–mouse cell hybrid PGMEI (Cowell and Mitchell 1989) con-

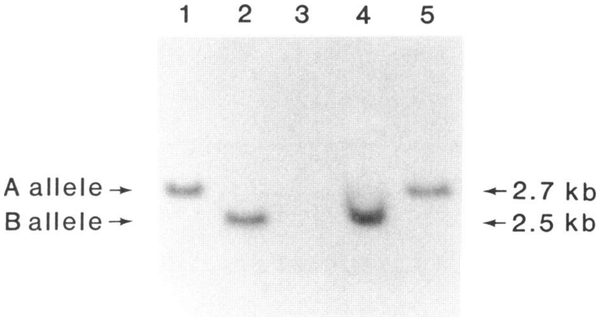

Figure 2 Southern blot of genomic DNA and the genomic clone restricted with *Hin*dIII and probed with the 2.68-kb *Hin*dIII fragment. Lanes 1–4, Genomic DNA (5 μg) from a keratinocyte cell line (A allele), HeLa (B allele), mouse spleen, and somatic cell hybrid PGMEI (7 μg), respectively. Lane 5, (0.008 ng) *Hin*dIII 2.68-kb clone in pBluescript.

taining an intact chromosome 13 (fig. 2, lane 4). This cell hybrid was also found to have the PADPRP B genotype. The hybridization signal did not represent the endogenous mouse PADPRP gene, as there was no hybridization to mouse spleen DNA (fig. 2, lane 3).

To determine the relationship between the *Hin*dIII clone (A allele) and the cDNA encoding the authentic PADPRP protein (Cherney et al. 1987), the nucleotide sequences were aligned as shown in figure 1. Sequence analysis revealed an overall shared identity of 91.8% between the PADPRP A allele and the cDNA (fig. 1). The genomic clone retained the *Hin*dIII site found on the cDNA sequence (fig. 1, base 887 of the lower sequence) and also encompassed a region that extended into the corresponding 33 noncoding portion of the cDNA (PADPRP protein terminates at TAA bases 3233–3235 [fig. 1]). The resemblance of the cloned fragment to an intronless cDNA copy, as well as the high sequence identity to the cDNA, suggest that the PADPRP sequences on chromosome 13 exist as a processed pseudogene (Weiner et al. 1986). A schematic representation of the *Hin*dIII genomic clone (A allele) and its relationship to the functional domains of the PADPRP protein is depicted in figure 3.

PADPRP Gene on Chromosome 13 Has a Duplicated 193-bp Region

Alignment of the *Hin*dIII clone to the cDNA revealed eight gaps, of which the most significant was a 193-bp gap in the cDNA sequence commencing at bp 1825 (fig. 1). This gap reflected a duplication of a 193-bp sequence in the A allele, which was not found in the cDNA. Restriction enzymes (such as *Eco*RI and *Msp*I) that were earlier found to be informative for detecting the PADPRP polymorphism, therefore, have sites outside the duplicated region. Inspection of the sequences around the duplicated region of the processed pseudogene showed a 9-bp repeat at the beginning and flanking the 3′ end (with two gaps) of the duplicated region (fig. 1, bases 747–755 and 1133–1143). It was therefore possible that an alteration in the processed pseudogene structure that forms the A allele occurred as a result of DNA looping during replication.

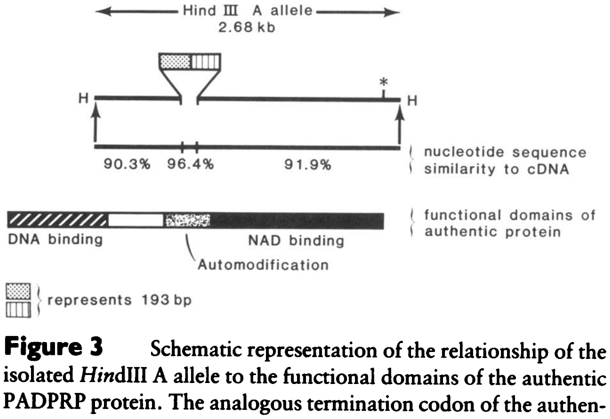

Figure 3 Schematic representation of the relationship of the isolated *Hin*dIII A allele to the functional domains of the authentic PADPRP protein. The analogous termination codon of the authentic PADPRP protein on the A allele is indicated by an asterisk (*).

To provide insight into the evolutionary relatedness between the duplicated regions, the sequences of each 193-bp region of the A allele (bases 747–939 and 940–1132) were compared with the homologous sequences of the PADPRP cDNA, as illustrated in figure 4. This comparison showed that, among the duplicated 193-bp regions, there were only two nucleotide alterations (1.0% difference), while we observed six and eight nucleotide changes, respectively, between each 193-bp region and the cDNA (an average of 3.6% nucleotide difference). As derived from figure 1, there was an overall 8.2% nucleotide difference between the PADPRP processed pseudogene on chromosome 13 and the cDNA, whereas there was only a 1.0% divergence with respect to the two duplicated regions (from fig. 4). Thus, it appears as if the duplication of PADPRP-related sequences on chromosome 13 was a recent occurrence, compared with the integration and formation of the PADPRP processed pseudogene on the q arm of chromosome 13. No other duplicated or inverted sequences were observed in the cloned *Hin*dIII fragment.

Verification of the PADPRP B Allele

Since size analysis in agarose gels, after restriction with various enzymes, produced an approximately 200-bp difference between the A and B alleles and since the duplicated region in the sequenced *Hin*dIII A allele was determined to be 193 bp (fig. 1 and Bhatia et al. 1990), it seemed reasonable to assume that the B allele reflected the nonduplicated version of the PADPRP sequences on chromosome 13. To verify whether this hypothesis was correct, oligonucleotide primers that flank the duplicated area were utilized in a PCR (containing bases 636–1230 of the *Hin*dIII clone) to amplify genomic DNA. Synthetic oligonucleotides (overlined in fig. 1) were designed to exploit the sequence differences between the PADPRP processed pseudogene (13q33-qter) and the PADPRP active gene (1q42). In addition, the primers selected for the PCR amplified a corresponding region of the PADPRP gene (chromosome 1), which encompassed an intron (Auer

```
 747  AGGTGAAGGC AGAGCCTGTT GAAGTCGTAG CCCCAAGAGG GAAGTCAGGA  Region 1 A allele
 940  .......... .......... .......... .......... ..........  Region 2 A allele
   1  .A........ .......... .......... .......... ..........  B allele
1641  .......... .......... .....T..G. .......... .........G  cDNA

 797  GCTGTGCTCT CCAAAAAAAG CAAGGGCCAG GTCAAGGAGG AAGGTATCAA  Region 1 A allele
 990  .......... .......... .......... .......... ..........  Region 2 A allele
  51  .......... .......... .......... .......... ..........  B allele
1691  ....C..... .......... .......... .......... ..........  cDNA

 847  CAAATCTGAA AAGAGAATGA AATTAACTCT TAAAGGAGGA GCAGCTGTGG  Region 1 A allele
1040  .......... .......... .......... .......... ..........  Region 2 A allele
 101  .......... .......... .......... .......... ..........  B allele
1741  .......... .......... .......... .......... ..........  cDNA

 897  ATCCTGACTC TGGTCTGGAA CACTCTGCGC ATGTCCTGGA GAA         Region 1 A allele
1090  .......... .......... ........A. ....T..... ...         Region 2 A allele
 151  .......... .......... ........A. ....T..... ...         B allele
1791  .......T.. ...A...... .......... .......... ...         cDNA
```

Figure 4 Comparison of the duplicated PADPRP sequences of the A allele to the corresponding region on the cDNA. The 193-bp duplicated sequences from the A allele were taken from fig. 1 (region 1 was from bases 747–939, and region 2 was from bases 940–1132), while the cDNA was from Cherney et al. (1987). The sequence of the PCR product (B allele) was obtained from the reaction shown in lane 3 of fig. 5. Only nonidentical nucleotides are indicated.

et al. 1989) and which could be distinguished from the processed pseudogene (chromosome 13) by a difference in DNA size.

Germ-line DNA from individuals with prostate cancer whose PADPRP genotype had been previously determined by Southern blotting (table 1) was used in the PCR. Figure 5 shows that the amplified DNA yielded only the expected size for patients who had a PADPRP AA genotype (595 bp; lane 2), BB genotype (402 bp; lane 3) or AB genotype (595 bp and 402 bp; lane 4). Lane 5 represents the control reaction in which the PADPRP cDNA encoding the authentic protein is used as the template and indicates that the designed primers did not amplify the PADPRP sequences from chromosome 1. Amplified DNA from the *Hin*dIII 2.68-kb clone in pBluescript is shown in lane 6. Furthermore, these primers did not amplify the PADPRP gene on chromosome 14, as a common DNA fragment, independent of the PADPRP genotype (fig. 5), would be observed in the PCR. These data indicated that mismatches at the 3′ end of the primers, as well as the internal mismatches, were sufficient to prevent amplification of related PADPRP sequences.

The 402-bp (B allele) PCR product (fig. 3) was sequenced and was found to contain the nonduplicated version of the PADPRP sequences. Alignment of this sequence with the cloned *Hin*dIII A allele uncovered one nucleotide difference (G replaced by A) between the amplified DNA and the second 193-bp duplicated region (fig. 4). This nucleotide difference was unlikely to be the consequence of an error in AmpliTaq polymerase, since identical sequences were also found in amplified germ-line DNA from two other individuals having the B allele. Thus, the observed polymorphism associated with the PADPRP sequences on chromosome 13 was attributed to a duplicated 193-bp sequence within the processed pseudogene.

Discussion

We were encouraged to extend the previous study on the germ line–derived PADPRP polymorphism in black cancer patients, as a twofold increase in the B allele frequency suggested that the PADPRP-like sequences on chromosome 13 were associated with a predisposition to cancer (Bhatia et al. 1990). Our analysis included cancers (multiple myeloma and prostate and lung cancer) that have a higher incidence in the U.S. black population compared with the white population. The increased frequency of the B allele observed in black patients with multiple myeloma and prostate and colon cancer, compared with that in the appropriate racial controls, suggests that the PADPRP polymorphism may represent a valid marker for a pre-

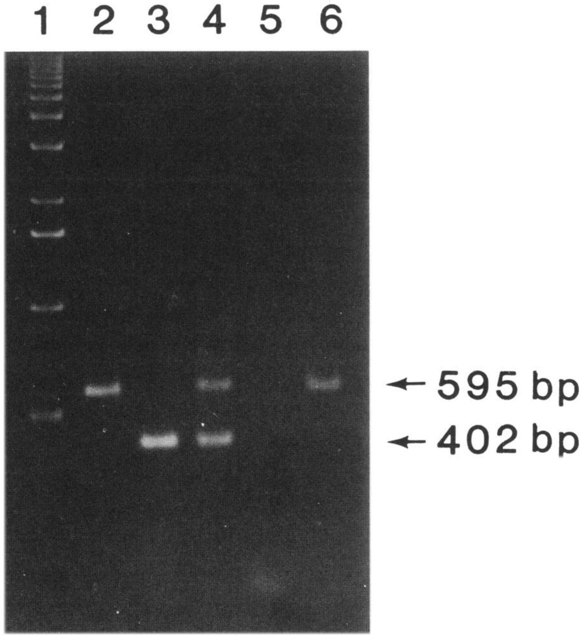

Figure 5 Ethidium bromide stain of a 1.5% agarose gel of the amplified PCR products. Lane 1, 1-kb ladder (from BRL). The amplified DNAs depicted are from individuals of the PADPRP A genotype (lane 2), B genotype (lane 3), or AB genotype (lane 4). Lane 5, Control reaction using full-length cDNA (pCD12) as described by Cherney et al. (1987). Lane 6, Amplified DNA from the *Hin*dIII 2.68-kb clone in pBluescript.

disposition (along with other genetic markers) to these cancers. However, for black patients with lung and breast cancer, the frequency of the B allele was not statistically different from that in the noncancer population. The lower B-allele frequency observed in this cohort of patients with breast cancer (.26) compared with that in the control group (.35), may be explained by the fact that the U.S. black population represents a genetically heterogeneous group (Chakraborty et al. 1992). No information was available concerning whether these individuals were American or included blacks from Africa. However, the B-allele frequency (.36) in a noncancer group of black Africans was no different from that noted in U.S. blacks (Bhatia et al. 1990). Thus, whether we consider a B-allele frequency of .35 or .26 as an appropriate control for the noncancer black population, the overall conclusions regarding the increased frequency of the B allele in multiple myeloma and prostate and colon cancer remain the same. On the other hand, the PADPRP polymorphism did not correlate with a germ line–derived increase of the B genotype among white patients with multiple myeloma or prostate or lung cancer. Our observations contrast with those of the previous study (Bhatia et al. 1990), in which an increase in the B-allele frequency (greater than twofold) was found in tumor DNA from B-cell lymphoma and lung carcinoma but not in myeloid leukemia of white patients. These data imply that the PADPRP polymorphism associated with a predisposition to cancer may be confined to specific diseases that are dependent on the racial population under study.

We were thus prompted to clone and characterize the PADPRP-like sequences on chromosome 13 that are associated with a predisposition to cancer. A 2.68-kb genomic clone was isolated and sequenced, which represented the *Hin*dIII A allele of the PADPRP processed pseudogene (fig. 2). This is the predominant PADPRP genotype in noncancer individuals (table 1). The characteristics of a processed pseudogene include features such as a lack of introns and the presence of a poly(dA) tail at the 3′ end (Weiner et al. 1986). However, the absence of a poly(dA) tail in the *Hin*dIII fragment suggests that the genomic clone isolated in the present study does not represent the entire PADPRP processed pseudogene on chromosome 13.

The most unexpected observation was the identification of a 193-bp duplication within the PADPRP sequences, which was responsible for the differences between the A/B polymorphism. The data relating the RFLP A/B polymorphism with the newly detected 193-bp duplication were confirmed in a PCR assay in which primers were developed to discriminate the processed pseudogene from other PADPRP-related sequences. The PCR also provides a relatively easy method for screening the genotype of a large number of individuals at high risk for certain cancers, over Southern blotting. Thus, we have identified the source of the RFLP associated with a predisposition to cancer in a selected population.

The 8.2% divergence in nucleotide sequence between the PADPRP (A allele) and the cDNA (derived from fig. 1) suggests that the integration of the processed pseudogene into chromosome 13 was not a recent event. In this context, a comparison of the noncoding sequences of the β-globin gene of human and chimpanzee suggests that nucleotide changes accumulate at an approximate rate of 0.3% per million years (Maeda et al. 1983). This allowed us to estimate that the processed pseudogene is probably at least 27 million years old. Other studies show that the noncoding 5′ and 3′ flanking, as well as other polymorphic regions of the very conserved ψη-globin pseudogene sequences, differ by 1.61%–1.84% between humans and African apes. This latter observation placed a species divergence as occurring 5–7 million years ago (Miyamoto et al. 1988). Taken together, our data suggest that the integration of the PADPRP (cDNA-like) sequences at the 13q33-qter locus occurred before the human/great ape divergence, and we favor the possibility that a PADPRP processed pseudogene may be present in African apes (supported by recent preliminary data [D. Lyn, unpublished observation]). In contrast, the duplicated PADPRP sequences on the A allele showed a lower degree of overall nucleotide divergence (1.0%, derived from fig. 4). Thus, the B allele probably represents the primordial gene, and the duplication of this region probably occurred considerably after the initial integration on chromosome 13. Since the nucleotide variation of a gene locus is estimated to be 0.5%–1.0% (Cooper et al. 1985), our observations are consistent with the suggestion that the A allele only occurs in humans.

To address how the polymorphic PADPRP sequences on chromosome 13 may be linked to a predisposition to certain cancers, two possible explanations emerge from an analysis of the processed pseudogene. Of potential significance, the duplicated sequences represented the most conserved area (96.4% average sequence identity, derived from fig. 1) between the PADPRP processed pseudogene and the cDNA. Figure 6 shows the predicted amino acid sequence of the only long open reading frame in the isolated *Hin*dIII clone from chromosome 13. In the A allele, the duplicated region would allow the introduction of a unique methionine residue, which is absent in the B allele. The authentic PADPRP protein has been implicated to play a role in DNA recombination, replication, and repair—processes that are involved in tumorigenesis. In this regard, if the A allele were to encode a functional protein, it would correspond to the C-terminal portion of the automodification domain of the PADPRP gene and, as such, would be highly homologous over a 65-amino-acid region (underlined in fig. 6), with only one conservative valine-for-alanine replacement. In addition, this protein would contain a homologous region of approximately 70 amino acid residues whose biochemical function with respect to the catalytic activity of PADPRP is unknown. The close sequence similarity between this potential protein from the A allele and the automodification domain of the authentic PADPRP protein may suggest an intolerance for amino acid change and could be an indication of functional similarity. This putative protein might therefore compete with the active polymerase for ribosylation sites at DNA strand breaks.

```
925
GCATGTCCTGGAGAAAGGTGAAGGCAGAGCCTGTTGAAGTCGTAGCCCCAAGAGGGAAGT
  M  S  W  R  K  V  K  A  E  P  V  E  V  V  A  P  R  G  K  S

985
CAGGAGCTGTGCTCTCCAAAAAAAGCAAGGGCCAGGTCAAGGAGGAAGGTATCAACAAAT
  G  A  V  L  S  K  K  S  K  G  O  V  K  E  E  G  I  N  K  S

1045
CTGAAAAGAGAATGAAATTAACTCTTAAAGGAGGAGCAGCTGTGGATCCTGACTCTGGTC
  E  K  R  M  K  L  T  L  K  G  G  A  A  V  D  P  D  S  G  L

1105
TGGAACACTCTGCACATGTTCTGGAGAAAGGTGGGAAGGTCTTCAGTGCCACCCTCAGCC
 E  H  S  A  H  V  L  E  K  G  G  K  V  F  S  A  T  L  S  L

1165
TGGTGGACGTCGTTAAAGGAACCAACTCCTATTACAAGCTGAAGTTGCTGAAGGATGACA
 V  D  V  V  K  G  T  N  S  Y  Y  K  L  K  L  L  K  D  D  K

1225
AGGAAAGCAGGCATTGGATATTCAAGTCCTGGGACCGTGTGGGCACGGTGATCGGTAGCA
 E  S  R  H  W  I  F  K  S  W  D  R  V  G  T  V  I  G  S  N

1285
ACAAACTGGAACAGATGCTGTCCAAGGAGGACACCATTGAACACTTCATGAAATTATATG
  K  L  E  Q  M  L  S  K  E  D  T  I  E  H  F  M  K  L  Y  E

1345
AAGAAAAACTAGGAATGCTTGGCACTCCAAAAATTCACAAAGTATCCCAAAAAGTTCTAC
 π E  K  L  G  M  L  G  T  P  K  I  H  K  V  S  Q  K  V  L  P

1405
CCCCTGGAGATTGA
  P  G  D  *
```

Figure 6 Predicted amino acid sequence of the potential open reading frame in the cloned *Hin*dIII A allele. The nucleotide sequence shown includes bases 925–1418 from fig. 1. The underlined amino acids are homologous to the C-terminal portion of the automodification domain in the authentic PADPRP protein (Cherney et al. 1987).

Another possible consideration is that the B allele may cosegregate with another gene whose function, or lack thereof, contributes to the development of malignancy. Since the integration of PADPRP-related sequences on chromosome 13 occurred millions of years ago, this linkage must be close to the processed pseudogene, otherwise a crossover event would have occurred to separate distally linked genes. In this regard, we previously identified a *Pst*I RFLP (7.2/5.3-kb allelic fragments) that cosegregated with the *Hin*dIII RFLP and was initially thought to reflect the same duplication detected by *Hin*dIII (Bhatia et al. 1990). Sequence analysis reveals no *Pst*I sites within the *Hin*dIII 2.7/2.5 alleles, demonstrating that the *Pst*I polymorphism is independent of the RFLP detected by *Hin*dIII. We are now analyzing whether the genetic change detected by *Pst*I is a better predictor of a predisposition to cancer.

In summary, a genomic *Hin*dIII 2.68-kb clone was successfully isolated and found to encompass the PADPRP polymorphic sequences at the q33-qter region on chromosome 13. These sequences provide an anchor point for elucidating the genomic structure of the distal end of 13q.

Acknowledgments

This work was supported by grant CA 13195 and by supplemental support from the National Cancer Institute. Blood samples of patients were provided by John H. Lynch, M.D., of the Surgery/Urology Division of Georgetown University Hospital and by Barry Gause, M.D., of Howard University, Washington, DC. We express our gratitude to Nancy Istock for providing skillful technical assistance and to Dawn Heyn for carrying out the statistical analysis.

References

Auer B, Nagl U, Herzog H, Schneider R, Schweiger M (1989) Human nuclear NAD$^+$ ADP-ribosyltransferase (polymerizing): organization of the gene. DNA:575–580

Bhatia KG, Cherney BW, Huppi K, Magrath IT, Cossman J, Sausville E, Barriga F, et al (1990) A deletion linked to a poly(ADP-ribose) polymerase gene on chromosome 13q33-qter occurs frequently in the normal black population as well as in multiple tumor DNA. Cancer Res 50: 5406–5413

Chakraborty R, Kamboh MI, Nwankwo M, Ferrell RE (1992) Caucasian genes in American blacks: new data. Am J Hum Genet 50:145–155

Cherney BW, McBride OW, Chen D, Alkhatib H, Bhatia K, Henseley P, Smulson ME (1987) cDNA sequence, protein structure, and chromosomal location of the human gene for poly (ADP-ribose) polymerase. Proc Natl Acad Sci USA 84:8370–8374

Cooper DN, Smith BA, Booke HJ, Niemann S, Schmidtke J (1985) An estimate of unique DNA sequence heterozygosity in the human genome. Hum Genet 69:201–210

Cowell JK, Mitchell CD (1989) A somatic cell hybrid mapping panel for regional assignment of human chromosome 13 DNA sequences. Cytogenet Cell Genet 52:1–6

Gloeckler Ries LA, Hankey BF, Edwards BK (1990) Cancer statistics review 1973–1987. U.S. Department of Health and Human Services, Bethesda, MD

Lalande M, Dryja TP, Schreck RR, Shipley J, Flint A, Latt S (1984) Isolation of human chromosome 13–specific DNA sequences cloned from flow sorted chromosomes and potentially linked to the retinoblastoma locus. Cancer Genet Cytogenet 13:283–295

Maeda N, Bliska JB, Smithies O (1983) Recombination and balanced chromosome polymorphism suggested by DNA sequence 5′ to the human δ-globin gene. Proc Natl Acad Sci USA 80:5012–5016

Miyamoto MM, Koop BF, Slightom JL, Goodman M, Tennant MR (1988) Molecular systematics of higher primates: genealogical relations and classification. Proc Natl Acad Sci USA 85:7627–7631

Pottern LM, Gart JJ, Nam J, Dunston G, Wilson J, Greenberg R, Schoenberg J, et al (1992) HLA and multiple myeloma among black and white men: evidence of a genetic association. Cancer Epidemiol Biol Prev 1:177–182

Riedel D, Pottern LM, Blattner WA (1991) Etiology and epidemiology of multiple myeloma. In: Wiernik PH, Canellos G, Kyle RA, Schiffer CA (eds) Neoplastic diseases of the blood and blood forming organs. Churchill-Livingston, New York, pp 347–372

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2d ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Ueda K, Hayaishi O (1985) ADP-ribosylation. Annu Rev Biochem 54:73–98

Weiner AM, Deininger PL, Efstratiadis A (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu Rev Biochem 55:631–661