AD_____

GRANT NUMBER DAMD17-94-J-4247

TITLE: Genetic Analysis of Human Breast Cancer

PRINCIPAL INVESTIGATOR: Michael H. Wigler, Ph.D.

CONTRACTING ORGANIZATION: Cold Spring Harbor Laboratory
Cold Spring Harbor, New York 11724

REPORT DATE: August 1998

TYPE OF REPORT: Final

PREPARED FOR: Commanding General
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 1998	Final (15 Jul 94 - 15 Jul 98)

4. TITLE AND SUBTITLE
Genetic Analysis of Human Breast Cancer

5. FUNDING NUMBERS
DAMD17-94-J-4247

6. AUTHOR(S)
Michael H. Wigler, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Cold Spring Harbor Laboratory
Cold Spring Harbor, New York 11724

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The genetic abnormalities that accumulate in tumor cells are potential clues to their evolution. To identify such abnormalities, we have developed RDA (representational difference analysis), a method that selects small restriction fragments present in only one of two genomic DNAs, or present in one in higher copy number. When nearly pure tumor DNA (driver) is used in excess to subtract normal DNA from the same individual (tester), RDA yields fragments deleted in the tumor. When tester and drivers are reversed, RDA yields fragments amplified in tumor DNAs.

DNA was extracted from the nuclei of about 20 breast cancer biopsies, sorted into diploid and aneuploid fractions by flow cytometry. From these pairs, we derived probes for about a half dozen loci that appear to undergo homozygous deletion. Some of the loci of deletion are also sites for high frequency loss-of-heterozygosity. The **PTEN** tumor suppressor gene was identified from one of these loci. P-TEN encodes a protein and phospholipid phosphatase.

14. SUBJECT TERMS
Breast Cancer

15. NUMBER OF PAGES
62

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

__✓__ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

__✓__ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

__✓__ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

M W. Jen 9/14/98

PI - Signature Date

TABLE OF CONTENTS

Page Number

INTRODUCTION

In this, the final report of our progress during the four year funding period, I will summarize our accomplishments. During this period we have sought to find tumor suppressor genes that are involved in breast cancer. Our method has been to search for homozygously deleted loci, and then to search for genes within these loci.

BODY

We applied RDA, a molecular tool we developed for difference analysis (3,4), to human breast cancer. We assembled a collection of greater than 500 biopsies, and sorted these into aneuploid and diploid nuclei, probably the largest collection of this type in the world. Using the aneuploid DNA as driver and diploid DNA as tester, we derived over 100 probes and analysed these for homozygous deletion in this collection. A total of six new loci were thereby discovered and confirmed.

One of these, on chromosome 10, was among the first to be analysed and led to the discovery of the PTEN tumor suppressor gene (2). PTEN is mutated in many glioblastomas, prostate and breast cancers (6,9). Germline mutation of PTEN occurs in patients with Cowden's disease, an inherited condition that predisposes to neoplasms. Most recently, germline knockouts of the PTEN gene in transgenic mice has led to direct confirmation that PTEN is a tumor suppressor.

PTEN encodes a protein phosphatase (8) that also possesses phospholipid phosphatase activity (7). In particular, PTEN can remove the 3'phosphate from PIP_3. Current evidence suggests that the lipid phosphatase activity is directly involved in PTEN's role as a tumor suppressor.

The remaining loci reside on chromosome 4, two on 8, one on 21 and one on 6. We have made progress defining the common boundaries of the homozygously deleted regions, a prerequisite for gene hunting. The region on chromosome 4 appears to be contained with a few hundred kilobases, those on 8 and 21 to within a megabase, and we have just begun to characterize the one on 6. To conduct these studies we have implemented some new technology (quantitative PCR) and invented some new technology, namely the immortalizing of small samples of tumor DNA by preparing high complexity representations (HCR), and wedded the two together so that we can now accurately measure gene copy number in minute biopsy samples (5). We expect this technique may be widely used in the molecular assessment of tumors by cancer biologists.

We also developed another tool for gene searching. Often genetic analysis, such as ours, defines a large region of the genome which may contain a gene of particular interest. Searching for the gene in this large region is often a rate limiting step in the research. Three basic techniques are employed: DNA sequence analysis, exon

trapping, and hybrid selection. Each method has advantages and pitfalls. We have developed a new method based on hybrid selection, called RICH (1), which adds to the armamentoria of search tools. RICH finds genes within a region that have overlap in sequence with collections of cDNAs.

CONCLUSION

In summary, we successfully applied a new tool (RDA) to the search for tumor suppressor genes in breast cancer, defined one, defined the function of the protein product, found and began the characterization of several other regions that harbor candidate tumor suppressor genes, and invented two new molecular tools: one for analysis the genomes of clinical cancer specimens, and one for the accelerated discovery of genes in large genomic regions.

Although RDA using tumor as driver and normal as tester is an effective way to find probes deleted in the tumor, these deletions are often large, and the search for genes within the deleted region is very time consuming. We have therefore begun new protocols, using tumor as tester and normal as driver to search for rearrangements, which may lead to the more rapid discovery of target genes.

REFERENCES

1. Hamaguchi, M., E.A. O'Connor, T. Chen, L. Parnell, R.W. McCombie, and M. Wigler. 1998. Rapid isolation of cDNA by hybridization. Proc.Natl.Acad.Sci. 95:3764-3769.

2. Li, J., C. Yen, D. Liaw, K. Podsypanina, S. Bose, S. Wang, J. Puc, C. Miliarcsis, L. Rodgers, R. McCombie, S.H. Bigner, B.C. Giovanella, M. Ittman, B. Tycko, H. Hibshoosh, M.H. Wigler, and R. Parsons. 1997. *PTEN* a putative protein tyrosine phosphatase gene mutated in human brain, breast and prostate cancer. Science 275:1943-1947.

3. Lisitsyn, N.A., N.M. Lisitsina, G. Dalbagni, P. Barker, C.A. Sanches, J. Gnarra, W.M. Linehan, B.J. Reid, and M. Wigler. 1995. Comparative genomic analysis of tumors: detection of DNA losses and amplification. Proc.Natl.Acad.Sci.USA 92:151-155.

4. Lisitsyn, N.A., N.M. Lisitsyn, and M. Wigler. 1993. Cloning the differences between two complex genomes. Science 259:946-951.

5. Lucito, R., M. Nakamura, J.A. West, Y. Han, K. Chin, K. Jensen, R. McCombie, J.W. Gray, and M. Wigler. 1998. Genetic analysis using genomic representations. Proc.Natl.Acad.Sci.USA 95:4487-4492.

6. Maehama, T. and J.E. Dixon. 1998. The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. J.Biol.Chem. 273:1-4.

7. Myers, M.P., I. Pass, I.H. Batty, J. Van der Kaay, J.P. Stolarov, B.A. Hemmings, M.H. Wigler, C.P. Downes, and N.K. Tonks. 1998. PTEN is a functional antagonist of PI 3-kinase. Proc.Natl.Acad.Sci. In press:

8. Myers, M.P., J.P. Stolarov, C. Eng, J. Li, S.I. Wang, M.H. Wigler, R. Parsons, and N.K. Tonks. 1997. P-TEN, the tumor suppressor from human chromosome 10q23 is a dual-specificity phosphatase. Proc.Natl.Acad.Sci.USA 94:9052-9052.

9. Tamura, M., J. Gu, K. Matsumoto, S.-I. Aota, R. Parsons, and K.M. Yamada. 1998. Inhibition of cell migration, spreading, and focal adhesions by tumor suppressor PTEN. Science 280:1614-1617.

PUBLICATIONS

109. Lisitsyn, N.A., Leach, F.S., Vogelstein, B. and Wigler, M. (1995) Detection of genetic loss in tumors by representational difference analysis. Cold Spring Harbor Symposium on Quantative Biology, Volume *59*: pp. 585-587.

116. Li, J., Yen, C., Liaw, D., Podsypanina,K., Bose, S., Wang, S., Puc, J., Miliarcsis, C., Rodgers, L., McCombie, R., Bigner, S.H., Giovanella, C., Ittman, M., Tycko, B., Hibshoosh, H., Wigler, M.H. and Parsons, R. (1997) *PTEN* a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science, *275*: 1943-1947.

117. Myers, M.P., Stolarov, J.P., Eng, C., Li, J., Wang, S.I., Wigler, M., Parsons, R. and Tonks, N.K. (1997) P-TEN, the tumor suppressor from human chromosome 10q23, is a dual-specificity phosphatase. Proc. Natl. Acad. Sci., USA *94*: 9052-9057.

119. Hamaguchi, M., O'Connor, E.A., Chen, T., Parnell, L., McCombie, R.W., and Wigler, M. (1998). Rapid isolation of cDNA by hybridization. Proc.Natl.Acad.Sci. *95*, 3764-3769

120. Lucito, R., Nakamura, M., West, J.A., Han, Y., Chin, K., Jensen, K., McCombie, R., Gray, J.W., and Wigler, M. (1998). Genetic analysis using genomic representations. Proc.Natl.Acad.Sci.USA *95*, 4487-4492.

122. Myers, M.P., Pass, I., Batty, I.H., Van der Kaay, J., Stolarov, J.P., Hemmings, B.A., Wigler, M.H. Downes, C.P. and Tonks, N.K. (1998). PTEN is a functional antagonist of PI 3-kinase. Proc. Natl. Acad. Sci. USA in press.

PERSONNEL

Year 1
Hao-Peng Xu, Linda Rodgers, Nikolai Listsyn, Kim Farina, Peter Barker

Year 2
Linda Rodgers, Nikolai Lisitsyn, Peter Barker, Clifford Yen, Michael Riggs, Jennifer Troge, Wei Wen, Xiaozhu Duan

Year 3
Linda Rodgers, Michael Riggs, Jennifer Troge, Wei Wen, Xiaozhu Duan, Clifford Yen, Masaaki Hamaguchi

Year 4
Linda Rodgers, Clifford Yen, Masaki Hamaguchi

Detection of Genetic Loss in Tumors by Representational Difference Analysis

N.A. LISITSYN,* F.S. LEACH,† B. VOGELSTEIN,† AND M.H. WIGLER*
*Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724;
†The Johns Hopkins Oncology Center, Baltimore, Maryland 21231

A variety of genetic lesions are found in tumors, including DNA losses, point mutations, gene amplifications, and rearrangements (Lasko et al. 1991; Salomon et al. 1991). Frequent losses of both alleles at a given locus or losses of one allele with functional inactivation of the other have been detected in many tumor types. These genetic lesions, manifesting themselves as loss of heterozygosity (LOH) and hemizygous and homozygous deletions, have been found to be the hallmarks of the presence of tumor suppressor genes. Many approaches have been taken in the past to identify these genes, but recently we have developed a new method that is both general and efficient (Lisitsyn et al. 1993, 1995). The method, called representational difference analysis, or RDA, is designed for analyzing the differences between complex but highly related genomes and combines three elements: representation, subtractive enrichment, and kinetic enrichment. The first stage of the procedure comprises the preparation of representations from the genomes, during which DNAs are cut with restriction endonuclease, ligated to oligonucleotide adapters, and amplified by the polymerase chain reaction (PCR). Since only small fragments (<1 kb in length), called ARFs, are efficiently amplified by standard PCR procedures, representations have at least tenfold lower complexity than initial DNAs. This enormously increases the efficiency of the second stage, comprising the reiterative hybridization/selection steps during which ARFs present in one sample, the tester, but not in the other, the driver, are selectively enriched. We describe here the application of RDA to discover sequences that are lost in tumors.

CLONING SEQUENCES LOST IN TUMORS

We performed RDA on 16 individual pairs of tumor DNA (used to derive driver) and matched normal DNA (used to derive tester) from the same patient. We isolated 15 DNAs from tumor cell lines (including 9 renal cell and 6 colon cancer cell lines) with normal DNAs derived from unaffected blood or tissue. In one case, we used a fluorescent activated cell sorter to fractionate nuclei from an esophageal cancer biopsy into aneuploid and diploid fractions used for preparation of driver and tester DNA, respectively.

In each application of RDA, difference products were cloned and analyzed by blot hybridization. The "informative" probes hybridized to DNA from the normal representation but not the tumor representation. Some of these probes mapped to the Y chromosome. Loss of Y chromosome information is frequently observed in renal cell carcinomas (Presti et al. 1991). Other probes detected binary polymorphisms at *Bgl*II sites and were presumed to reflect loss of heterozygosity in tumor. Finally, some probes did not hybridize at all to total genomic DNA from the tumor. Probes of this type were sequenced, and oligonucleotides were derived for use in PCR screening of genomic DNA from the tester and driver sources, and from panels of normal human and tumor cell lines. Occasionally, we found probes that did not hybridize to several normal human DNAs. We presume that these probes reflect hemizygous loss in the tumor of a deletion polymorphism common in the human population (see Table 1, footnote c). Table 1 summarizes the types of probes we obtained (for further details, see Lisitsyn et al. 1995).

HOMOZYGOUS LOSS ON CHROMOSOME 3p

From 16 comparisons, 6 pairs yielded probes that appear to detect homozygous loss in the tumor used as driver. Three probes were found to be homozygously deleted in other DNAs isolated from a collection of 100 tumor cell lines established from different types of cancer but not from normals (N. Lisitsyn and R. Lucito, unpubl.). Probe 758-6, derived from a patient with Barrett's esophagus, detected frequent losses in cancers of the digestive tract. Loss detected with this probe has also been observed in breast, bladder, and lung tumors. This probe has been analyzed in greatest detail.

758-6 was mapped to chromosomal region 3p by PCR analysis of monochromosomal human/rodent cell hybrids. The probe was used for screening a chromosome-3 cosmid library, and a cosmid contig was built by chromosome walking (see Fig. 1). Single-copy probes were derived from this contig and used to screen DNAs from a collection of human colon cancer cell lines and xenografts. Of 175 tumor DNAs, 20 (11%) lacked sequences from at least one of the probes from this region. In contrast, losses were observed in 1 of 122 lung cancer cell lines (S. Bader and J. Minna, unpubl.). Figure 1 shows the different patterns of loss that were observed in colorectal cancer cell lines and xenografts.

Additional probes derived from a cosmid contig were used in hybridizations to Southern blots containing DNAs harboring deletions. Several hybridizing restric-

Table 1. Analysis of RDA Probes Derived Using Tumor DNA as Driver

	Selected for initial characterization	Found to be informative[a]
Renal cell carcinoma cell lines		
UOK 112 (male)	13[b]	13 (0/13/0)
UOK 114 (female)	12[b]	4 (3/0/1)
UOK 124 (female)	12[b]	4 (4/0/0)
UOK 132 (male)	10[b]	9 (3/6/0)
UOK 108 (female)	2	2 (2/0/0)
UOK 111 (female)	5	5 (5/0/0)
UOK 127 (male)	3	3 (2/1[c]/0)
UOK 146 (female)	3	3 (1/1[c]/1)
UOK 154 (female)	5	1 (1/0/0)
Colon cancer cell lines		
VACO 429 (male)	2	1 (0/0/1)
VACO 441 (female)	3	3 (1/0/2)
VACO 432 (male)	2	1 (1/0/0)
VACO 456 (female)	2	1 (1/0/0)
VACO 576 (female)	2	2 (2/0/0)
RBX (male)	2	1 (1/0/0)
Barrett's esophagus		
BE 758[d] (male)	5	5 (0/4/1[e])
Total:	83	58 (27/25/6)

[a] Entries are a(b, c, d), where a is the total number of probes detecting DNA loss in tumors, judged to be: b, loss-of-heterozygosity; c, hemizygous loss; d, presumably homozygous loss (see Discussion). All but two probes judged to detect hemizygous loss were derived from the Y chromosome. The difference between quantities of initially selected probes (83) and informative probes (58) was due to the presence of the repeat sequences (9 cases), nonhuman DNA contaminating tester (5 cases), and single-copy sequences present in both tester and driver DNAs (11 cases).

[b] The difference products after two rounds of hybridization/selection were cloned; in all the rest of the experiments cloning was performed after three rounds.

[c] Probes 127-1 and 146-1 were found to be deletion polymorphisms, absent on both autosomes of 7 out of 35 and 3 out of 35 of normal humans, respectively.

[d] Nuclei from a biopsy were sorted by flow cytometry into aneuploid (tumor) and diploid (normal) fractions.

[e] This result is presumed, but was not confirmed because of the small amount of sorted tumor nuclei available.

tion fragments were observed in tumors that were absent in normals, presumably as a result of rearrangements occurring at the ends of some of the deletions. This observation rules out the possibility that probe 758-6 detects a deletion polymorphism and that the loss of sequences is caused by the same types of mechanisms that underlie loss of heterozygosity at polymorphic markers. Work is in progress to identify transcribed sequences from this region.

DISCUSSION

The RDA methodology may be successfully applied to detection of DNA losses in tumors, readily providing probes that detect homozygous deletions. As we were able to demonstrate, some of these deletions are relatively small (< 50 kb) and, thus, positional cloning of genes that must be inactivated in tumors becomes much more efficient, as compared to other techniques used for this purpose (e.g., allelotyping, linkage analysis of predispositions in families, and cytogenetic studies followed by microdissection). Since some of these probes are deleted in more than one DNA isolated from tumor cells, it is possible that the deleted locus contains a gene that is commonly inactivated in tumors.

It is well documented that some genes known to be disrupted by homozygous deletions in tumors regulate cellular growth, differentiation, and genomic stability. Some of these genes have strong tumor suppressor phenotypes after transfection into tumor cell lines. Our screening technique based on RDA methodology holds promise for the identification of new genes participating in these or other processes. We have taken a similar approach for the cloning of dominant oncogenes. These are frequently amplified in tumors, and probes detecting amplifications can be efficiently cloned by RDA when tumor DNA is used to derive tester (Lisitsyn et al. 1995). Although amplified regions are usually rather large, one can map candidate oncogenes more precisely by finding the minimal region common to all amplifications at a given locus. The use of RDA for the analysis of cancers thus opens up new avenues for understanding the etiology of the disease and for the development of prognostic and diagnostic markers.

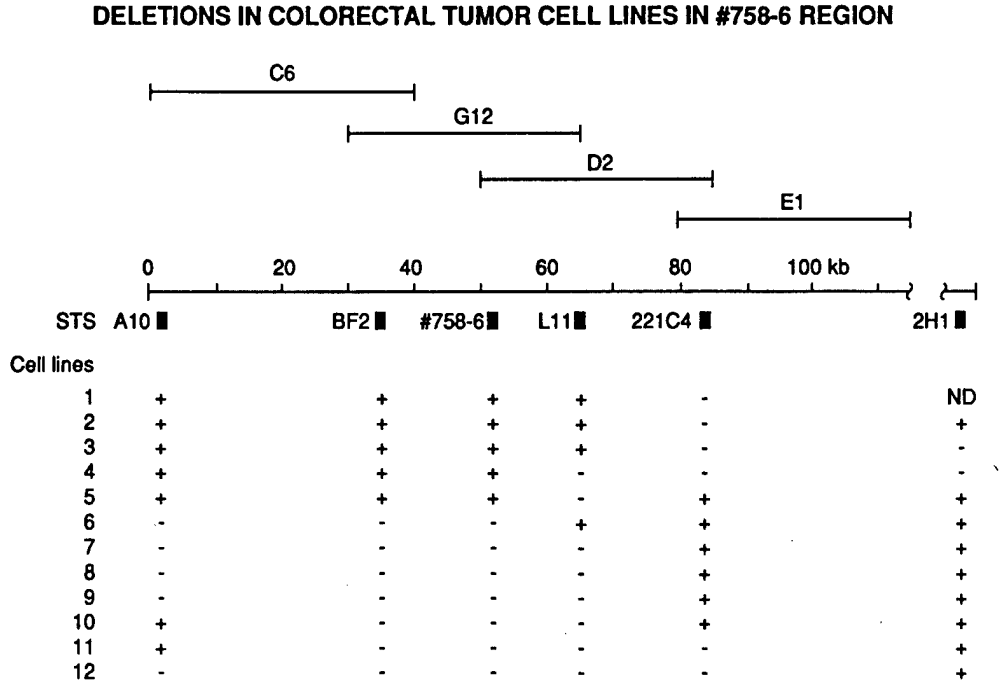


Figure 1. Physical map of a chromosome 3p region. On top are shown cosmids from a region isolated by chromosomal walking. Patterns of the homozygous loss of STSs (thick bars) detected by PCR analysis are depicted on the bottom. Pluses and minuses indicate presence or absence of the probe in DNAs from colorectal cancer cells.

ACKNOWLEDGMENTS

We thank J. Garnes for providing the chromosome-3 library constructed at the Human Genome Center, Lawrence Livermore National Laboratory, Livermore, California, under the auspices of the National Laboratory Gene Library Project sponsored by the U.S. Department of Energy; J. Gnarra and W.M. Linehan for providing DNA samples and cell lines from renal cell carcinomas; J. Willson and S. Markowitz for preparation of DNA from colorectal cancers; C. Sanchez and B. Reid for providing DNAs from FACS sorted nuclei; S. Bader and J. Minna for PCR screening of a collection of DNAs from lung cancers; Natalia Lisitsina for performing RDA; Linda Rodgers and Michael Riggs for DNA sequencing; and Patricia Bird for preparation of this manuscript.

REFERENCES

Lasko, D., W. Cavenee, and M. Nordenskjold. 1991. Loss of constitutional heterozygosity in human cancer. *Annu. Rev. Genet.* **25:** 281.

Lisitsyn, N.A., N.M. Lisitsina, and M. Wigler. 1993. Cloning the differences between two complex genomes. *Science* **259:** 946.

Lisitsyn, N.A., N.M. Lisitsina, G. Dalbagni, P. Barker, C.A. Sanches, J. Gnarra, W.M. Linehan, B.J. Reid, and M. Wigler. 1995. Comparative genomic analysis of tumors: Detection of DNA losses and amplification. *Proc. Natl. Acad. Sci.* (in press).

Presti, J.C., Jr., P.H. Rao, Q. Chen, V.E. Reuter, F.P. Li, W.R. Fair, and S.C. Jhanwar. 1991. Histopathological, cytogenetic, and molecular characterization of renal cortical tumors. *Cancer Res.* **51:** 1544.

Salomon, E., J. Bozzow, and A.D. Goddard. 1991. Chromosome aberrations and cancer. *Science* **254:** 1153.

Reprint Series
28 March 1997, Volume 275, pp. 1943–1947

SCIENCE

PTEN, a Putative Protein Tyrosine Phosphatase Gene Mutated in Human Brain, Breast, and Prostate Cancer

Jing Li,* Clifford Yen,* Danny Liaw,* Katrina Podsypanina,* Shikha Bose, Steven I. Wang, Janusz Puc, Christa Miliaresis, Linda Rodgers, Richard McCombie, Sandra H. Bigner, Beppino C. Giovanella, Michael Ittmann, Ben Tycko, Hanina Hibshoosh, Michael H. Wigler, and Ramon Parsons†

DAMD17-94-4247
APPENDIX Final Report

PTEN, a Putative Protein Tyrosine Phosphatase Gene Mutated in Human Brain, Breast, and Prostate Cancer

Jing Li,* Clifford Yen,* Danny Liaw,* Katrina Podsypanina,* Shikha Bose, Steven I. Wang, Janusz Puc, Christa Miliaresis, Linda Rodgers, Richard McCombie, Sandra H. Bigner, Beppino C. Giovanella, Michael Ittmann, Ben Tycko, Hanina Hibshoosh, Michael H. Wigler, Ramon Parsons†

Mapping of homozygous deletions on human chromosome 10q23 has led to the isolation of a candidate tumor suppressor gene, *PTEN*, that appears to be mutated at considerable frequency in human cancers. In preliminary screens, mutations of *PTEN* were detected in 31% (13/42) of glioblastoma cell lines and xenografts, 100% (4/4) of prostate cancer cell lines, 6% (4/65) of breast cancer cell lines and xenografts, and 17% (3/18) of primary glioblastomas. The predicted *PTEN* product has a protein tyrosine phosphatase domain and extensive homology to tensin, a protein that interacts with actin filaments at focal adhesions. These homologies suggest that *PTEN* may suppress tumor cell growth by antagonizing protein tyrosine kinases and may regulate tumor cell invasion and metastasis through interactions at focal adhesions.

As tumors progress to more advanced stages, they acquire an increasing number of genetic alterations. One alteration that occurs at high frequency in a variety of human tumors is loss of heterozygosity (LOH) at chromosome 10q23. This change appears to occur late in tumor development: although rarely seen in low-grade glial tumors and early-stage prostate cancers, LOH at 10q23 occurs in ~70% of glioblastomas (the most advanced form of glial tumor) and ~60% of advanced prostate cancers (*1*, *2*). This pattern of LOH, and the recent finding that wild-type chromosome 10 suppresses the tumorigenicity of glioblastoma cells in mice, suggest that 10q23 encodes a tumor suppressor gene (*3*).

To identify this putative tumor suppressor gene, we performed representational difference analysis (RDA) on 12 primary breast tumors (*4*). A probe, CY17, derived from one of the tumors was mapped to chromosome 10q23, near markers WI-9217 and WI-4264, on the Whitehead-MIT radiation hybrid map (*5*). To map the location of CY17 more precisely, we isolated three yeast artificial chromosomes (YACs) containing CY17 that are present on the

sequence tagged site (STS)–based map of the human genome (6, 7). These YACs placed CY17 slightly centromeric to the position determined by radiation hybridization and precisely identified its location (Fig. 1A). Analysis of 32 primary invasive breast cancers revealed LOH in this region in about 50% of the samples. No homozygous deletions of CY17 were detected in a panel of 65 breast tumor cell lines (25) and xenografts (40) (8), so eight additional markers were analyzed in the 10q23 region (D10S579, D10S215, AFMA086WG9, D10S541, AFM280WE1, WI-10275, WI-8733, WI-6971). We identified homozygous deletions of AFMA086WG9 in two xenografts, Bx11 and Bx38 (Figs. 1B and 2A) and then screened a bacterial artificial chromosome (BAC) library with this marker (9). Using new STSs from four independent BAC clones, we determined that the minimal region of deletion was within BAC C (Fig. 1B) (10). Homozygous deletions of AFMA086WG9 were also detected in two of eight glioblastoma cell lines, three of 34 glioblastoma xenografts, and two of four prostate cancer cell lines (11). One of the glioblastoma samples, cell line A172, had the same deletion pattern as the original breast xenografts; the deletions in the other samples were larger (Fig. 1B).

To confirm the presence of homozygous deletions, we hybridized a Southern (DNA) blot with a 3-kb probe derived from a genomic clone spanning the region of deletion (12). Xenografts anticipated to have a homozygous deletion did not hybridize to this probe; the control xenografts hybridized to the expected 3-kb band (Fig. 2B).

We identified genes within the 10q23 region by exon trap analysis of BACs C and D (Fig. 1B) (13). Two trapped exons, ET-1 and ET-2, had sequences that were perfect matches to an unmapped UNIGENE assembly of expressed sequence tags (ESTs) as well as several unassembled ESTs (6). Clones containing the ESTs were sequenced and used to assemble an open reading frame (ORF) of 403 amino acids (Fig. 3A). To verify the location of this cDNA, we obtained the intronic sequence around ET-1 by directly sequencing BAC C. An STS primer pair (ET-1) was generated that mapped back to BACs A, B, and C (Fig. 1B). In addition, we screened the Map Panel #2 monochromosome human-rodent hybrid panel to confirm the unique location of this exon on chromosome 10 (14).

Our entire panel of tumor xenografts and cell lines was screened with this primer pair, and we identified an additional glioblastoma cell line (DBTRG-05MG) with a deletion of 180 base pairs (bp) (Fig. 1B) (Fig. 2C). Sequence analysis revealed that the deletion had removed 180 bp of exonic sequence and the splice donor site from this 225-bp exon. This deletion was not present in 52 normal blood samples or in more than

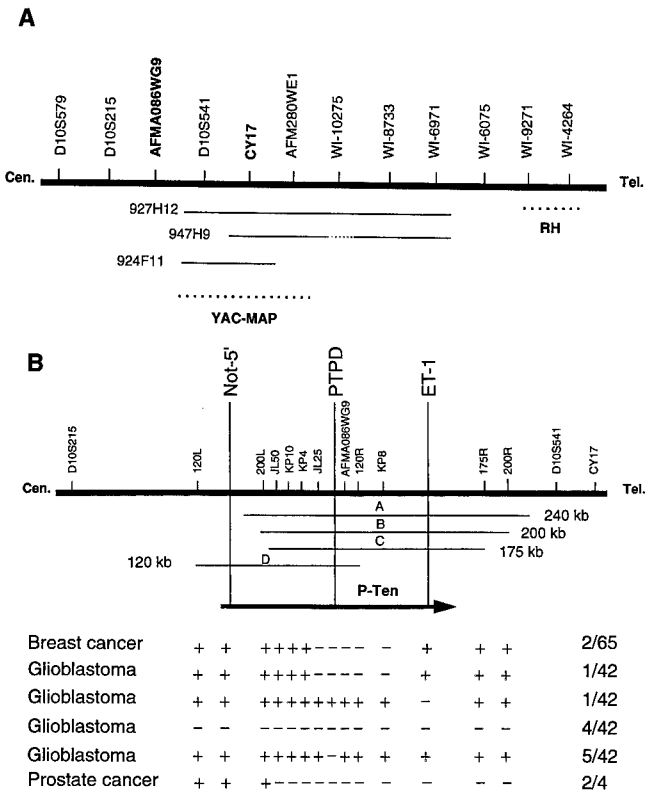


Fig. 1. Region of homozygous deletion on chromosome 10q23. (**A**) The STS-based YAC map of the region surrounding CY17. Marker locations are taken from the Whitehead STS-based map. RH indicates the radiation hybrid interval for CY17. CY17 positive YAC addresses are indicated. YAC map indicates the interval containing CY17 inferred from the YAC addresses. Cen., centromere; Tel., telomere. (**B**) Map of homozygous deletions on 10q23, showing the STS markers spanning the deleted region, the four BACs overlapping the region, and the location of *PTEN* with respect to the STS markers. STS markers Not-5′, PTPD, and ET-1 contain exonic sequences of the *PTEN* gene. Absence of homozygous deletion is indicated with a "+" and presence of homozygous deletion with a "−." Numbers to the right indicate the fraction of tumor cell lines and xenografts with the deletion. The two breast cancer samples with a deletion are xenografts Bx11 and Bx38. Glioblastoma line A172 has a deletion encompassing markers JL25 through KP8 and glioblastoma line DBTRG-05MG has a deletion affecting only ET-1. The glioblastoma samples with a deletion across the entire region are the cell line U105 and xenografts 2, 3, and 11, and the samples with deletion of only PTPD, which contains the phosphatase domain, are xenografts 22, 23, 24, 25, and 32. The prostate cancer cell lines with homozygous deletions are NCI H660 and PC-3. The 5′ end of the *PTEN* cDNA was determined to be coincidental with the Not I site 20 kb from the centromeric end of BAC D by sequence analysis. These maps are not drawn to scale.

Table 1. Summary of *PTEN* mutations in tumor cell lines and primary tumors.

Tumor sample	Tissue of origin	Codon	Mutation*	Predicted effect
LNCaP	Prostate	6	AAA to A	Frameshift
534T†	Glioblastoma	15	AGA to AGAGA	Frameshift
U87MG	Glioblastoma	54	49 bp deletion	Frameshift
MDA-MB-468	Breast	70	44 bp deletion	Frameshift
132T†	Glioblastoma	129	GGA to AGA	Gly to Arg
DU145	Prostate	134	ATG to TTG	Met to Leu
U373MG	Glioblastoma	241	TTT to TTT TT	Frameshift
BT549	Breast	274	GTA AAT to TAA AT	Stop
DBTRG-05MG‡	Glioblastoma	274–342	Delete 204 bp	In-frame deletion
134T†	Glioblastoma	337	4 bp deletion	Frameshift

*Mutations are indicated in the sense orientation. †Primary tumors. All other samples are tumor cell lines. The mutations in the primary tumors were not found in matched blood DNA. ‡DBTRG-05MG has a genomic deletion of 180 bp within exon ET-1, which includes the splice donor site. Because of this deletion, the transcript contains an in-frame deletion of codons 274 to 342.

J. Li, D. Liaw, K. Podsypanina, S. I. Wang, J. Puc, C. Miliaresis, R. Parsons, Department of Pathology and Department of Medicine, College of Physicians & Surgeons, Columbia University, 630 West 168 Street, New York, NY 10032, USA.
C. Yen, L. Rodgers, R. McCombie, M. Wigler, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA.
S. H. Bigner, Department of Pathology, Duke University Medical Center, Durham, NC 27710, USA.
B. Giovanella, Stehlin Foundation for Cancer Research, St. Joseph Hospital, Houston, TX 77003, USA.
M. Ittmann, New York VA Medical Center and Department of Pathology, New York University, 423 East 23 Street, New York, NY 10010, USA.
S. Bose, B. Tycko, H. Hibshoosh, Department of Pathology, College of Physicians & Surgeons, Columbia University, New York, NY 10032, USA.
*These authors contributed equally to this work.
†To whom correspondence should be addressed.

125 other primary tumors, xenografts, and cell lines tested.

Sequence analysis of the ORF revealed a protein tyrosine phosphatase domain (Fig. 3B) and a large region of homology (~175 amino acids) to chicken tensin and bovine auxilin (Fig. 3C). We therefore call the gene *PTEN* for Phosphatase and Tensin homolog deleted on chromosome Ten. The phosphatase domain of the P-TEN protein contained the critical (I/V)-H-C-X-A-G-X-X-R-(S/T)-G motif found in tyrosine and dual-specificity phosphatases (*15*). The phosphatase domain exon mapped within all four BACs and was deleted in all of the samples with homozygous deletions except for DBTRG-05MG. These results thus placed this exon within the region of homozygous deletion near JL25 and AFMA086WG9 (Fig. 1B). We then screened the remaining xenografts and cell lines for additional homozygous deletions and identified five more glioblastoma xenografts lacking this exon. These data indicate that the phosphatase domain encoded by *PTEN* was targeted for mutations in tumor xenografts and cell lines.

The phosphatase domain of P-TEN is most related in sequence to those of CDC14, PRL-1 (phosphatase of regenerating liver), and BVP (baculovirus phosphatase) (Fig. 3B). CDC14 and BVP are dual-specificity phosphatases that remove phosphate groups from tyrosine as well as serine and threonine (*16*). These phosphatases can be distinguished from the better characterized VH1-like enzymes by sequence differences outside of the core conserved domain. Both PRL-1 and CDC14 are involved in cell growth, and CDC14 appears to play a role in the initiation of DNA replication (*17*). In contrast to P-TEN, these phosphatases do not have extensive homology to tensin and auxilin. P-TEN is also homologous to the protein tyrosine phosphatase domains of three ORFs (Y50.2, PTP-IV1, CPTPH) whose protein products have not been characterized. Of these hypothetical proteins, only the putative yeast phosphatase Y50.2 has significant homology to tensin. Although tensin and auxilin are not expected to have phosphatase activity, they both contain elements of the protein tyrosine phosphatase signature sequence (*18*), which suggests that they may share a tertiary structure with these enzymes (*19*).

If *PTEN* is a tumor suppressor gene, the *PTEN* allele retained in tumor cells with LOH should contain inactivating mutations. To search for such mutations, we performed a protein truncation test on 20 breast, six glioblastoma, and two prostate tumor cell lines (*20*). Two truncating mutations in *PTEN* were identified in the breast samples (Table 1). BT549 cells had a 1-bp deletion of

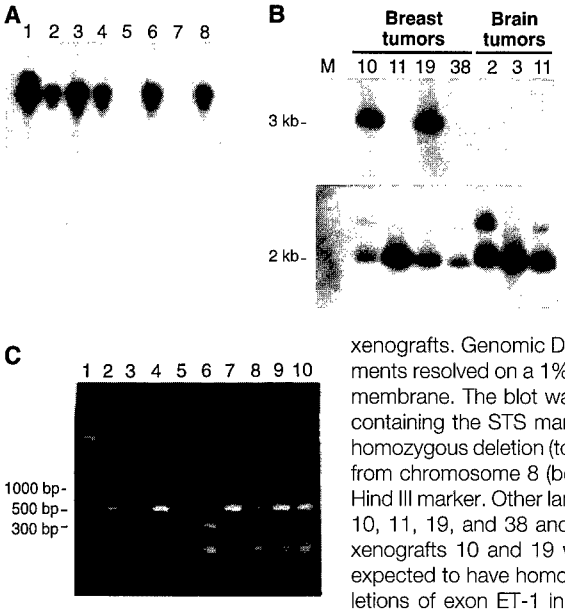


Fig. 2. Homozygous deletions in tumor cell lines and xenografts. (**A**) A 6% polyacrylamide sequencing gel showing the products of PCR amplification of AFMA086WG9 from breast cancer cell lines (lanes 1 to 4) and xenografts (lanes 5 to 8). Lane 1, MDA-MB-330; lane 2, MDA-MB-157; lane 3, MDA-MB-134-VI; and lane 4, MDA-MB-435S; lane 5, Bx11; lane 6, Bx15; lane 7, Bx38; and lane 8, Bx39. (**B**) Southern blot analysis of tumor xenografts. Genomic DNA was digested with Eco RI, the fragments resolved on a 1% agarose gel, and transferred to a nylon membrane. The blot was probed with a 3-kb Eco RI fragment containing the STS marker JL25, which is within the region of homozygous deletion (top), or to a second 2-kb Eco RI fragment from chromosome 8 (bottom). Lane M, bacteriophage lambda Hind III marker. Other lanes contain DNA from breast xenografts 10, 11, 19, and 38 and brain xenografts 2, 3, and 11. Breast xenografts 10 and 19 were loaded as controls and were not expected to have homozygous deletions. (**C**) Homozygous deletions of exon ET-1 in glioblastoma cell lines. Genomic DNA samples were PCR amplified using intronic primers that amplify exon ET-1. The products were resolved on a 1.2% agarose gel and then stained with ethidium bromide. Lane 1 contains a DNA marker. The remaining lanes contain PCR products from control templates and seven glioblastoma cell lines: lane 2, lymphocyte DNA; lane 3, water; lane 4, U118MG; lane 5, A172; lane 6, DBTRG-05MG; lane 7, U373; lane 8, T-98G; lane 9, U-87MG; and lane 10, U138MG. Full-length products are present for all templates except water, A172, and DBTRG-05MG.

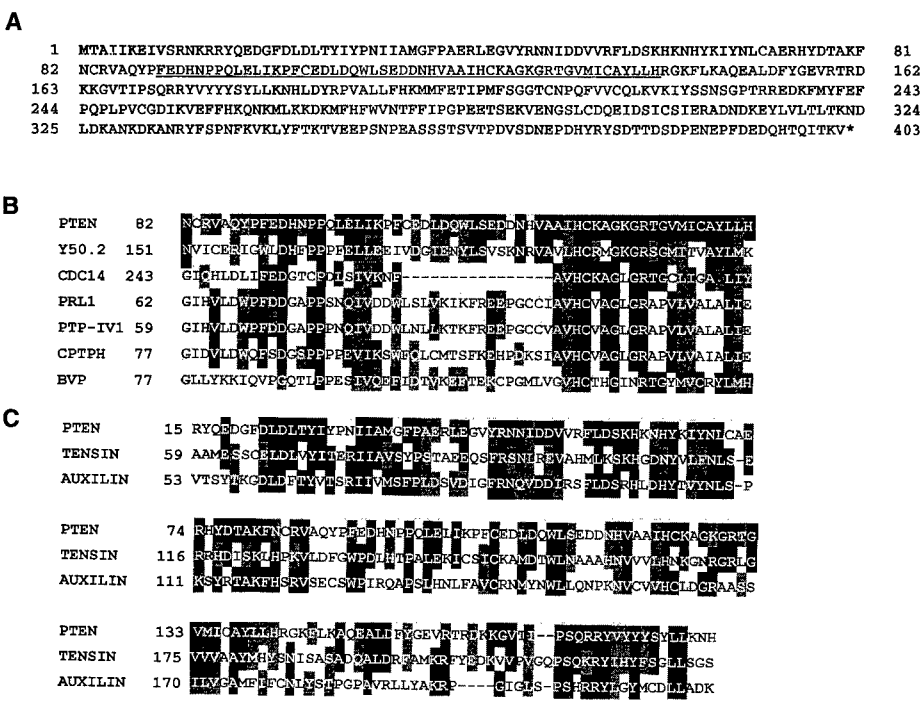


Fig. 3. (**A**) Predicted amino acid sequence of P-TEN. The putative phosphatase domain is underlined. The nucleotide sequence has been deposited in GenBank (accession number U93051). Abbreviations for amino acids are A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. (**B**) Homology of P-TEN to protein tyrosine phosphatases. The sequence alignment was performed by ClustalW (http://dot.imgen.bcm.tmc.edu:9331/multi-align/Options/clustalw.html). The National Center for Biotechnology Information (NCBI) ID numbers are P53916 (Y50.2), M61194 (CDC14), A56059 (PRL1), 1246236 (PTP-IV1), 1125812 (CPTPH), and P24656 (BVP). Black boxes indicate amino acid identities and gray boxes indicate similarities. (**C**) Homology of P-TEN to chicken tensin and bovine auxilin. Alignment was performed as in (B) over the region of highest homology. NCBI ID numbers are A54970 (tensin) and 485269 (auxilin).

a G, leading to the formation of a stop codon TAA (Fig. 4A), and MDA-MB-468 cells had a deletion of 44 bp at codon 70, which resulted in a frameshift on the amino terminal side of the tyrosine phosphatase domain. Mutations in *PTEN* were also identified in three of the six glioblastoma cell lines: DBTRG-05MG cells had an in-frame deletion of 204 bp caused by the genomically deleted exon ET-1 (Fig. 4B), U373MG had a 2-bp insertion at codon 242, and U87MG had a frameshift at codon 54. Both of the prostate tumor cell lines had *PTEN* mutations: LNCaP cells had a 2-bp deletion at codon 6, leading to a frameshift (Fig. 4C), and DU145 cells had a Met → Leu substitution at codon 134, within the phosphatase domain. The latter mutation was detected by a change in the pattern of in vitro translation initiation and was not found in >50 other alleles tested. However, Met-134 is not required for phosphatase activity (Fig. 3B), so this alteration could be a polymorphism. With one exception (DU145), all of the cell lines retained a mutant *PTEN* allele and lost the other allele, indicating that these cells are null for *PTEN*.

To determine whether *PTEN* mutations are present in primary tumors, we screened genomic DNA from 18 primary glioblastomas for mutations in three exons (*21*). Mutations in *PTEN* were found in three of these tumors: a 2-bp insertion at codon 15 (534 T), a point mutation resulting in a Gly → Arg change at codon 129 (132T), and a 4-bp frameshift mutation at codon 337 (134T) (Table 1 and Fig. 4D). The mutation at codon 129 is within the signature sequence for tyrosine phosphatases (Fig. 3B). All three tumors appeared to have LOH in the *PTEN* region since the wild-type allele was substantially reduced in intensity. In addition, the tumor mutations were not detected in paired blood DNA.

In summary, we detected homozygous deletions, frameshift, or nonsense mutations in *PTEN* in 63% (5/8) of glioblastoma cell lines, 100% (4/4) of prostate cancer cell lines, and 10% (2/20) of breast cancer cell lines. These frequencies are likely to be underestimates since the cell lines were not systematically screened for point mutations. We screened xenografts only for homozygous deletions in *PTEN* and detected them in 24% (8/34) of glioblastoma xenografts and 5% (2/40) of breast cancer xenografts. Finally, we detected *PTEN* mutations in 17% (3/18) of primary glioblastomas; this frequency is also likely to be an underestimate since the entire coding sequence was not analyzed. The results of these preliminary screens suggest that a large fraction of glioblastomas and advanced prostate cancers may harbor *PTEN* mutations, whereas the mutation frequency in breast cancer may be lower. Future systematic analysis of all tumor types will be of interest.

The likely function of the P-TEN tumor suppressor as an enzyme that removes phosphate from tyrosines is intriguing, given that many oncoproteins function in the reverse process—to phosphorylate tyrosines (*22*). P-TEN and tyrosine kinase oncoproteins may share substrates and the tight control of these substrates through phosphorylation is likely to regulate a critical pathway that is altered late in tumor development. The homology of P-TEN to tensin is also of interest. Tensin appears to bind actin filaments at focal adhesions—complexes that contain integrins, focal adhesion kinase (FAK), Src, and growth factor receptors (*23*). Integrins have been implicated in cell growth regulation (*24*) and in tumor cell invasion, angiogenesis, and metastasis (*25*), so it is conceivable that *PTEN* regulates one or more of these processes. Finally, the identification of P-TEN as a likely tumor suppressor raises the possibility that this protein and its substrates will be useful targets for the development of new therapeutics for cancer.

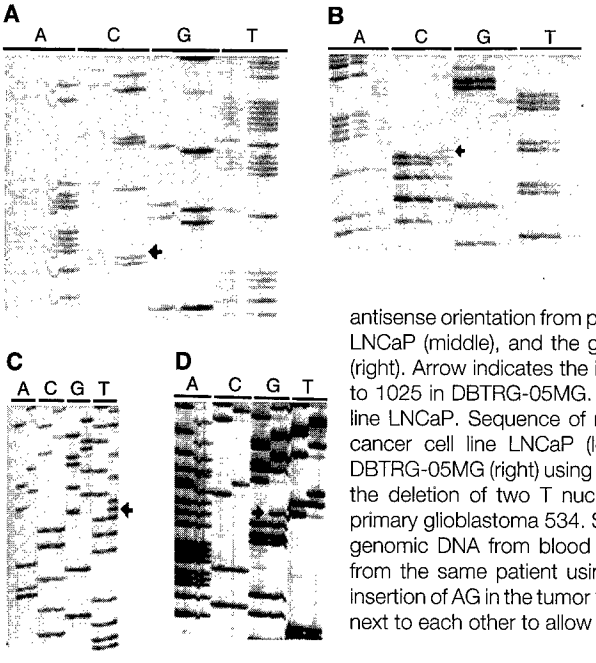


Fig. 4. Mutations of *PTEN* in cancer cell lines and primary tumors. (**A**) Mutation in breast cancer cell line BT549. Sequence of nucleotides 831 to 785 (bottom to top) using an antisense primer shows a deletion of a C (arrow) in sample on the right (BT549) but not in the sample on the left (breast cancer cell line ZR-75-30). (**B**) Mutation in glioblastoma cell line DBTRG-05MG. Sequence of nucleotides 1039 to 1010 in the antisense orientation from prostate cancer cell lines DU145 (left), LNCaP (middle), and the glioblastoma cell line DBTRG-05MG (right). Arrow indicates the in-frame deletion of nucleotides 822 to 1025 in DBTRG-05MG. (**C**) Mutation in prostate cancer cell line LNCaP. Sequence of nucleotides 34 to 2 of the prostate cancer cell line LNCaP (left) and the glioblastoma cell line DBTRG-05MG (right) using an antisense primer. Arrow indicates the deletion of two T nucleotides in LNCaP. (**D**) Mutation in primary glioblastoma 534. Sequence of nucleotides 26 to 63 of genomic DNA from blood (left) and primary tumor 534 (right) from the same patient using a sense primer. Arrow indicates insertion of AG in the tumor DNA. A, C, G, and T lanes are loaded next to each other to allow better detection of mutations.

REFERENCES AND NOTES

1. S. H. Bigner *et al.*, *Cancer Res.* **48**, 405 (1988); C. D. James *et al.*, *ibid.*, p. 5546; B. K. A. Rasheed *et al.*, *Oncogene* **10**, 2243 (1995).
2. I. C. Gray *et al.*, *Cancer Res.* **55**, 4800 (1995); M. Ittmann, *ibid.* **56**, 2143 (1996); T. Trybus, A. Burgess, K. Wojno, T. Glover, J. Macoska, *ibid.*, p. 2263.
3. S. Hsu *et al.*, *ibid.* **56**, 5684 (1996).
4. N. A. Lisitsyn and M. Wigler, *Science* **259**, 946 (1993); M. Schutte *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5950 (1995). RDA was performed as described in N. A. Lisitsyn *et al.* (*ibid.*, p. 151). Diploid and aneuploid nuclei from primary breast cancer cells were separated with a fluorescence-activated cell sorter. DNA (100 ng) from each fraction was digested with Bgl II and used to prepare amplicons for 12 separate RDA reactions. Probe CY17 was isolated from one of these reactions. CY17 was 236 bp long and was present in the diploid but not in the aneuploid amplicon from which it was derived. Hybridization of CY17 to normal genomic DNA samples digested with Bgl II revealed no evidence of restriction length polymorphism.
5. D. R. Cox, M. Burmeister, E. R. Price, S. Kim, R. M. Myers, *Science* **250**, 245 (1990). We generated primers to amplify CY17 and screened the GeneBridge4 radiation hybrid panel. The primers were 5′-ATCTAGTGAGTTGGGGGACAGAGG-3′ and 5′-CTGGGTTAGGGATTCTGCTCAG-3′. Amplification conditions were 95°C for 30 s, 56°C for 1 min, and 70°C for 1 min for 35 cycles.
6. T. J. Hudson *et al.*, *Science* **270**, 1945 (1995).
7. The CEPH (Centre d'Etude du Polymorphisme Humain) B library (Research Genetics, Huntsville, AL) was screened by PCR using a series of tiered pools to identify unique clones.
8. The forward PCR primer was labeled with [^{32}P]ATP and used to amplify 40 ng of genomic DNA. The samples were then subjected to electrophoresis and autoradiography. The samples included 25 human breast tumor cell lines available from American Type Culture Collection (ATCC) as well as 40 human primary breast tumors xenografted into nude mice. The cell lines were HS578T, SK-BR-3, UACC812, UACC893, MDA-MB-453, MDA-MB-175-VII, MDA-MB468, MDA-MB-361, MDA-MB-231, MDA-MB-436, MDA-MB-415, MDA-MB-330, MDA-MB-157, MDA-MB-134-VI, MDA-MB-435S, ZR 75-30, ZR 75-1, BT-549, BT-483, T-47D, BT-474, DU4475, CAMA1, MCF7, and BT-20.
9. U. Kim *et al.*, *Genomics* **34**, 213 (1996). Clones were isolated from a BAC library (Research Genetics) using AFMA086WG9 as an STS probe.
10. BAC DNA was prepared using the Nucleobond kit (Nest Group, Southboro, MA). BACs were digested with Not I and subjected to electrophoresis on a field inversion apparatus. BACs A, B, C, and D were 240, 200, 175, and 120 kb, respectively (see Fig. 1B). A Not I site was present 20 kb from one end of BAC D. Twelve new STS sites were generated by sequencing both ends of BACs B, C, and D, and shotgun cloning Eco RI fragments. Plasmid DNA was prepared from the cloned Eco RI fragments. DNA was cycle sequenced with appropriate primers using a [^{33}P]ddNTP cycle sequencing kit (Amersham). STS primers were designed and the relative location of the STSs determined by testing for their presence in

the BAC contig. Primer sequences are available upon request.
11. The glioblastoma lines included U105, U118MG, A172, DBTRG-05MG, U373MG, T-98G, U-87MG, and U138MG and 34 glioblastoma xenografts. The prostate cancer cell lines tested were DU145, LNCaP, NCI H660, and PC-3, and microsatellite analysis revealed that each was unique. With the exception of U105, all lines were obtained from ATCC.
12. DNA (10 μg) was digested with Eco RI, resolved on a 1% agarose gel, and transferred to nylon. The JL25 3-kb probe and the 2-kb control probe were randomly labeled and hybridized to the blot consecutively.
13. A. J. Buckler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4005 (1991); G. Lennon, C. Auffray, M. Polymeropoulos, M. B. Soares, *Genomics* **33**, 151 (1996). BACs C and D were digested with Bam HI, Bgl II, or both enzymes and then ligated into the trapping vector pSPL3. Libraries were transfected into COS1 cells with lipofectamine and polyadenylated RNA was extracted after 2 days. An exon trapping kit was purchased from GIBCO/BRL.
14. The map panel #2 monochromosome panel was purchased from the National Institute of General Medical Science (NIGMS) Human Mutant Genetic Cell Repository.
15. N. K. Tonks and B. G. Neel, *Cell* **87**, 365 (1996).
16. E. B. Fauman and M. A. Saper, *Trends Biochem. Sci.* **21**, 413 (1996); Z. Sheng and H. Charboneau, *J. Biol. Chem.* **268**, 4728 (1993).
17. R. H. Diamond, D. E. Cressman, T. M. Laz, C. S. Abrams, R. Taub, *Mol. Cell. Biol.* **14**, 3752 (1994); E. Hogan and D. Koshland, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 3098 (1992).
18. D. T. Haynie and C. P. Ponting, *Protein Sci.* **5**, 2643 (1996).
19. J. M. Denu, J. A. Stuckey, M. A. Saper, J. E. Dixon, *Cell* **87**, 361 (1996).
20. S. M. Powell *et al.*, *N. Engl. J. Med.* **329**, 1982 (1993); P. A. M. Roest, R. G. Roberts, S. Sugino, J. B. van Ommen, J. T. den Dunnen, *Hum. Mol. Genet.* **2**, 1719 (1993). Randomly primed cDNA was prepared from each of the cell lines studied. Reverse transcription (RT)–PCR reactions were performed with two overlapping primer pairs to screen the entire ORF. Primer pairs were as follows: 5′-GGATCCTAATACGACTCACTATAGGGAGACCACCATGGAGTCGCCTGTCACCATTTC-3′ and 5′-TTCCAGCTTTACAGTGAATTG-3′; 5′-GGATCCTAATACGACTCACTATAGGGAGACCACCATGGGATTTCCTGCAGAAAG-3′ and 5′-TTTTTTCATGGTGTTTTATCCCTC-3′. In vitro transcription and translation were performed with the T7 TNT kit (Promega, Madison, WI) and the translation products resolved by electrophoresis. The RT-PCR products that generated truncated proteins were directly cycle sequenced (20 ng each) to identify potential mutations. All mutations were verified by repeating the RT-PCR and mutation analysis. Complementary DNA primer sequences are available upon request.
21. STSs Not-5′, PTPD, and ET-1 were amplified from primary glioblastoma DNA and blood DNA and the exonic regions were sequenced.
22. T. Hunter, *Cell* **50**, 823 (1987).
23. J. A. Wilkins, M. A. Risinger, S. Lin, *J. Cell Biol.* **103**, 1483 (1986); J. Z. Chuang, D. C. Lin, S. Lin, *ibid.* **128**, 1095 (1995); S. Miyamoto *et al.*, *ibid.* **131**, 791 (1995); S. Miyamoto, S. K. Akiyama, K. M. Yamada, *Science* **267**, 883 (1995).
24. X. Zhu, M. Ohtsubo, R. M. Bohmer, J. M. Roberts, R. K. Assoian, *J. Cell Biol.* **133**, 391 (1996); K. K. Wary, F. Mainiero, S. J. Isakoff, E. E. Marcantonio, F. E. Giancotti, *Cell* **87**, 733 (1996).
25. S. K. Akiyama, K. Olden, K. M. Yamada, *Cancer Metastasis Rev.* **14**, 173 (1995).
26. We thank B. Vogelstein, N. Tonks, and E. Marcantonio for their comments and S. Kalachikov and R. Hauptschein for helpful suggestions. R.P. is a James S. McDonnell Scholar. M.H.W. is an American Cancer Society Research Professor and is supported by the Department of the Army (DAMD 17-94-I4247), NCI (5R35 CA39829), Amplicon Corporation, and the "1 in 9" breast cancer organization. This work is dedicated to the memory of Richard K. Parsons and Richard P. Sanchez.

3 February 1997; accepted 27 February 1997

Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 9052–9057, August 1997
Biochemistry

P-TEN, the tumor suppressor from human chromosome 10q23, is a dual-specificity phosphatase

(cancer/tyrosine phosphorylation/signal transduction/protein tyrosine phosphatase)

MICHAEL P. MYERS*, JAVOR P. STOLAROV*, CHARIS ENG†, JING LI‡, STEVEN I. WANG‡, MICHAEL H. WIGLER*, RAMON PARSONS‡, AND NICHOLAS K. TONKS*§

*Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724; †Dana–Farber Cancer Institute, Harvard Medical School, 44 Binney Street, Boston, MA 02115; and ‡Departments of Pathology and Medicine, College of Physicians and Surgeons, Columbia University, 630 West 168th Street, New York, NY 10032

Contributed by Michael H. Wigler, June 11, 1997

ABSTRACT **Protein tyrosine phosphatases (PTPs) have long been thought to play a role in tumor suppression due to their ability to antagonize the growth promoting protein tyrosine kinases. Recently, a candidate tumor suppressor from 10q23, termed P-TEN, was isolated, and sequence homology was demonstrated with members of the PTP family, as well as the cytoskeletal protein tensin. Here we show that recombinant P-TEN dephosphorylated protein and peptide substrates phosphorylated on serine, threonine, and tyrosine residues, indicating that P-TEN is a dual-specificity phosphatase. In addition, P-TEN exhibited a high degree of substrate specificity, showing selectivity for extremely acidic substrates *in vitro*. Furthermore, we demonstrate that mutations in P-TEN, identified from primary tumors, tumor cells lines, and a patient with Bannayan–Zonana syndrome, resulted in the ablation of phosphatase activity, demonstrating that enzymatic activity of P-TEN is necessary for its ability to function as a tumor suppressor.**

A variety of techniques have been used to identify genes involved in the etiology of cancer. From these studies, a surprisingly large number of protein tyrosine kinases (PTKs) have been implicated in carcinogenesis (1). The PTKs are activated by amplification, deletion, or mutation of important negative regulatory domains, or by genetic rearrangements that result in the production of activated fusion proteins (1, 2). Further support for the importance of tyrosine phosphorylation in oncogenesis comes from the finding that expression of v-crk, a small adaptor protein that does not contain intrinsic PTK activity, results in an increase in the levels of cellular phosphotyrosine and cellular transformation (3). The critical role that tyrosine phosphorylation plays in oncogenesis has led to the suggestion that many protein tyrosine phosphatases (PTPs) would act as tumor suppressors. Although PTPs have been linked to the inhibition of cell proliferation, there had been no clear cut examples of these enzymes functioning as tumor suppressors.

P-TEN, a candidate tumor-suppressor gene identified on chromosome 10, also known as MMAC1, shares homology with the PTP family, as well as with the cytoskeletal protein tensin (4, 5). P-TEN was isolated from a locus on chromosome 10, 10q22-23, which is deleted in a large number of tumors, especially glioblastomas (4, 5). Subsequently, it was shown that P-TEN is deleted or mutated in a significant fraction of glioblastomas and prostate tumors (4, 5). Importantly, germline mutations in P-TEN give rise to Cowden disease, which is typified by the formation of multiple, benign tumors and an increased susceptibility to malignant cancers. Mutations in P-TEN were also found in patients suffering from disorders similar to Cowden disease, such as Lhermitte–Duclos disease, which has additional pathologies, including ataxia, macrocephaly, and dysplastic cerebellar gangliocytomatosis (6, 7). The detection of germ-line mutations in four of five Cowden/Lhermitte–Duclos kindreds verified that P-TEN functions as a tumor suppressor and also suggests that P-TEN plays a role in the proper development and formation of certain tissues (8). Recent data also suggest that germ-line mutations of P-TEN are also responsible for Bannayan–Zonana syndrome (C.E., unpublished data), an autosomal dominant disorder, that in addition to the mental retardation, macrocephaly, and thyroid disease shared with Cowden disease, is characterized by lipomatosis, speckled penis, and an early onset of the neoplastic disease (9).

All PTPs contain the catalytic signature motif HCXXGXXRS/T (10). The cysteine residue in this motif is absolutely required for catalysis, as it acts as a nucleophile to attack the phosphorous atom in the phosphate moeity of its substrate, forming a thiol-phosphate intermediate (11). Mutation of this cysteine to serine or alanine results in the complete loss of phosphatase activity (12). The dual-specificity phosphatases, which catalyze the hydrolysis of phospho-seryl, -threonyl, and -tyrosyl residues, also contain the canonical PTP catalytic motif (13). The crystal structure of PTP1B, the prototypic PTP, reveals that the stringent amino acid selectivity of the PTPs is determined, in part, by the location of the catalytic cysteine at the base of a cleft (11, 14). The depth of the cleft (9 Å) matches the length of a phosphotyrosine residue and the shorter phospho-seryl and -threonyl residues are not able to reach the catalytic cysteine. In addition, the deep cleft is lined with hydrophobic residues that help stabilize the interaction with the hydrophobic tyrosine moiety (12). In contrast to PTP1B, the phosphate-binding loop in the dual-specificity phosphatase VHR is at the base of a much shallower cleft, which can accommodate all three phosphorylated hydroxyl amino acids (15). Regardless of the depth of the catalytic pocket, both classes of enzymes proceed through similar steps of catalysis. The initial nucleophilic attack of the cysteine residue results in the formation of an enzyme–substrate complex. This complex is disrupted by the catalytic acid (Asp-181 in PTP1B and Asp-92 in VHR) which protonates the phenolic oxygen of the tyrosyl group, releasing the dephosphorylated substrate from the complex. The phosphorous atom remains associated with the active site cysteine as a thiol–phosphate. In the case of the PTPs, the active enzyme is

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: PTP, protein tyrosine phosphatase; PTK, protein tyrosine kinase; GST, glutathione *S*-transferase; MBP, myelin basic protein; MAP, mitogen-activated protein; RCML, reduced carboxyamidomethylated and maleylated lysozyme.

§To whom reprint requests should be addressed. e-mail: tonks@cshl.org

regenerated by the hydrolysis of the thiol–phosphate bond by a water molecule, possibly activated by the same aspartate residue (11, 14, 15). The efficiency of this reaction is illustrated by the finding that PTP1B can undergo ≈2,000 catalytic cycles per minute (12).

Recent data indicate that the PTPs exhibit a great deal of substrate specificity *in vivo* and are not simply unregulated antagonists of the signals mediated by the PTKs (12, 16). In fact, PTPs can exert both positive and negative effects on signaling pathways, indicating that they do not simply function as "off switches" (17, 18). In spite of the large number of PTKs that have been shown to play a role in oncogenesis (1), there were no examples of PTPs that function as classical tumor suppressors. Although P-TEN has been identified as a tumor suppressor and displays some of the structural features of the PTP family, its phosphatase activity has not been characterized. We have expressed P-TEN in *Escherichia coli* as a glutathione *S*-transferase (GST) fusion protein and demonstrated that P-TEN is catalytically active. Purified GST–P-TEN showed a strong preference for only the most acidic substrates. Furthermore, P-TEN was able to dephosphorylate serine, threonine, and tyrosine residues, establishing it as a dual-specificity phosphatase. Finally, we demonstrated that the phosphatase activity of P-TEN is necessary for its ability to function as a tumor suppressor, because a variety of point mutations, derived from tumor samples or Bannayan–Zonana syndrome, ablated P-TEN activity.

MATERIALS AND METHODS

Expression and Purification of GST–P-TEN. A full-length P-TEN cDNA was generated by ligating the *Not*I–*Bgl*II fragment from EST264611 with the *Bgl*II–*Eco*RI fragment of EST365465 into pBluescript digested with *Not*I–*Eco*RI. The resulting full-length P-TEN cDNA was amplified by PCR using pfu polymerase (Stratagene) and primers that add a 5′ *Bam*HI site (5′-CGCGGATCCATGACAGCCATCATCAAAGAGATCGTTAGC) and a 3′ *Eco*RI site (5′-CGCGAATTCTCAGACTTTTGTAATTTGTGTATGC). The resulting fragment was subcloned into pGEX2T (Pharmacia), and the sequence was verified by automated sequencing. Expression of P-TEN was induced in 500 ml of mid-log phase bacteria (A_{600} = 0.600) by the addition of isopropyl β-D-thiogalactoside to 200 μM. The culture was shifted to room temperature and expression was allowed to proceed for 12 h. The bacteria were harvested by centrifugation, the supernatant was removed, and the bacterial pellet was frozen at −80°C. The frozen pellets were resuspended in 5 ml of ice cold 20 mM Tris, 150 mM NaCl, and 5 mM EDTA (pH 8.0) supplemented with lysozyme (1 mg/ml), aprotinin (5 μg/ml), leupeptin (5 μg/ml), and benzamidine (1 mM) and incubated on ice for 15 min. The bacteria were lysed by sonicating three times for 1 min each with a Branson model 450 sonifier, power setting 4, 70% duty cycle. The lysate was cleared by centrifugation at 30,000 × *g* for 10 min, diluted with an equal volume of HBS (50 mM Hepes/150 mM NaCl, pH 7.4). Glutathione-Sepharose 4B (300 μl) was added, and the resulting slurry was incubated at 4°C on a rocking platform for 1–2 h. The glutathione-Sepharose was washed five times, each with 10 ml of ice cold HBS, and the washed fusion proteins were eluted with a solution containing 20 mM glutathione, 50 mM Hepes, and 30% glycerol (pH 8.0). Protein concentrations were determined by the method of Bradford, using BSA as a standard, and the integrity of the fusion proteins were verified by SDS/PAGE. Mutations identified from tumor samples and cell lines [refs. 4 and 5; and L. Hedrick (Johns Hopkins University) personal communication] or identified from Cowden disease (8) and Bannayan–Zonana (C.E., unpublished data) were introduced into pGEX2T–P-TEN using the Quickchange mutagenesis kit as described by the manufacturer (Stratagene). For all mutations, the entire P-TEN ORF was sequenced to confirm that no other mutations had been introduced.

Substrate Preparation. All tyrosine-phosphorylated substrates were phosphorylated with the cytoplasmic fragment of the β subunit of the insulin receptor kinase (β-IRK) and purified as described (19). Serine-phosphorylated substrates were phosphorylated with recombinant protein kinase A (New England Biolabs) or with recombinant casein kinase II (a gift from D. Litchfield, University of Western Ontario) in a reaction mixture consisting of 50 mM Hepes (pH 7.2), 10 mM $MgCl_2$, 2 mM DTT, 2 mM ATP, and 1 mCi of [γ-^{32}P]ATP (1 Ci = 37 GBq) in a total volume of 1 ml. Casein was used at a concentration of 10 mg/ml, myelin basic protein (MBP) at 4 mg/ml, and the peptides RRRDDDSDDD (DSD) and RRREEETEEE (ETE) were used at a concentration of 0.5 mg/ml. Protein substrates were precipitated by the addition of ammonium sulfate to 80%, incubated on ice for 30 min, and harvested by centrifugation. The precipitated proteins were washed three times with 80% ammonium sulfate and then resuspended in 500 μl of 1 M Hepes (pH 7.5). The solubilized proteins were dialyzed against several changes of 50 mM imidizole (pH 7.2). Peptide substrates were purified using a Sep-pak C18 reverse phase cartridge (Waters) as described (20). The purified peptides were lyophilized to dryness and resuspended in 50 mM imidizole (pH 7.2). Random copolymers of glutamate and tyrosine, with a 4:1 ratio of glutamate to tyrosine ($polyGlu_4Tyr_1$) or with a 1:1 ratio ($polyGlu_1Tyr_1$), were purchased from Sigma and phosphorylated with β-IRK at a final polymer concentration of 1 mg/ml and purified, as described, using Sep-pak C18 reverse phase chromatography. Phosphorylated ERK2 was produced in *E. coli* by coexpression of activated MEK and was a gift from D. Barford (Oxford University).

Phosphatase Assays. The standard phosphatase assay contained 10 μM substrate, 50 mM Hepes (pH 7.0), 10 mM $MgCl_2$, and 10 mM DTT. The reaction was initiated by the addition of enzyme, typically 1–2 μg, to prewarmed (30°C) substrate mix, resulting in a final volume of 60 μl. The reactions were allowed to proceed at 30°C for the indicated times and stopped by the addition of a suspension of activated charcoal in 900 mM HCl, 90 mM NaPPi, and 2 mM NaPi (21). Dephosphorylation of ERK2 was performed essentially as described for the radioactive substrates; at the indicated times duplicate aliquots were removed and stopped by the addition of 5× Laemmli sample buffer and processed for immunoblot analysis as described (20). Immunoblots were probed with a 1:5,000 dilution of anti-mitogen-activated protein (MAP) kinase ascites or with a mixture of anti-phosphotyrosine antibodies [1:2,000 dilution of G-104 and G-2-98 ascites (22)] and were developed with ECL reagents (Amersham) in conjunction with an horseradish peroxidase-labeled rat anti-mouse κ-chain antibody (Zymed).

RESULTS

Although P-TEN displays structural features of members of the PTP family of enzymes an important, and essential, step toward understanding the function of P-TEN is the characterization of its enzymatic activity. To test whether P-TEN encodes an active phosphatase, we expressed it as a fusion protein in *E. coli*. As a control, we assayed all substrates with a mutant of P-TEN in which the essential cysteine from the signature motif had been mutated to serine (P-TENC124S). This mutant would be expected to be catalytically inactive, and ensures that any activity seen was the result of P-TEN and not of a copurifying bacterial phosphatase.

PTP Activity of P-TEN. Initially, we assayed P-TEN against a number of tyrosine-phosphorylated proteins and peptides, including reduced carboxyamidomethylated and maleylated lysozyme (RCML), MBP, $polyGlu_4Tyr_1$, and $polyGlu_1Tyr_1$, as

well as the peptide EDNDYINASL. The activity of P-TEN toward the classical substrates RCML and MBP was weak (<80 pmol/min per mg). However, P-TEN, exhibited robust phosphatase activity when polyGlu$_4$Tyr$_1$ was used as a substrate (4,840 pmol/min per mg) (Table 1 and Fig. 1*A*). Significantly, addition of unphosphorylated polyGlu$_4$Tyr$_1$ to reactions using RCML as a substrate did not result in an increase in activity toward RCML (data not shown), indicating that P-TEN was not activated by the polyanionic character of polyGlu$_4$Tyr$_1$. In fact, inclusion of unphosphorylated polyGlu$_4$Tyr$_1$ in reactions containing RCML resulted in the inhibition of the already limited dephosphorylation of RCML, suggesting that even unphosphorylated polyGlu$_4$Tyr$_1$ was capable of binding to P-TEN, displacing the more weakly interacting RCML. We also observed that P-TEN dephosporylated polyGlu$_1$Tyr$_1$. When the stoichiometries of phosphorylation were normalized, no significant differences were detected in the rates of dephosphorylation of polyGlu$_4$Tyr$_1$ and polyGlu$_1$Tyr$_1$, suggesting that the acidic character of these substrates is an important determinant of substrate recognition. P-TEN, however, exhibited reduced activity when assayed with the acidic peptide EDNDYINASL (Table 1), suggesting that the mere presence of acidic residues was not sufficient to ensure that a peptide would be a substrate. cdc14, a dual-specificity phosphatase that is closely related to P-TEN (8), does not discriminate between RCML and polyGlu$_4$Tyr$_1$ (Fig. 1*B*), indicating that polyGlu$_4$Tyr$_1$ is not a universal substrate for the dual-specificity phosphatases, and that the substrate specificity of P-TEN is likely to be unique feature determined by structural motifs separate from the catalytic motif. Caution should be exercised when comparing the activity of P-TEN to that of other dual-specificity phosphatases because of the inherent substrate specificity exhibited by these enzymes.

Dual-Specificity Phosphatase Activity of P-TEN. To test whether P-TEN falls into the class of dual-specificity phosphatases, we assayed activity using a number of proteins and peptides phosphorylated on serine and threonine residues. Similar to the findings with tyrosine-phosphorylated substrates, P-TEN dephosphorylated serine/threonine residues in substrates that had a preponderance of acidic residues. Specifically, P-TEN dephosphorylated two peptide substrates (DSD and ETE) with the highest efficiency (Table 1 and Fig. 2). Furthermore, P-TEN showed specificity even among acidic serine/threonine substrates, exhibiting a reduced activity when casein, phosphorylated by casein kinase II or protein kinase A, was used as substrate (Table 1 and Fig. 2). As might be anticipated in light of these properties, P-TEN exhibited almost undetectable activity when assayed with polybasic substrates, such as MBP or Kemptide (LRRASLG) (Fig. 2 and data not shown). The finding that P-TEN dephosphorylated

Table 1. Activity of P-TEN measured with tyrosine, serine, and threonine phosphorylated substrates

Substrate	Activity, pmol/min per mg
Phosphotyrosyl substrates	
polyGlu$_4$Tyr$_1$	4,840 ± 140
RCML	88 ± 6.6
EDNDYINASL peptide	51 ± 4.0
MBP	21 ± 1.2
Phospho-seryl and -threonyl substrates	
DSD peptide	210 ± 4.2
ETE peptide	161 ± 1.8
Casein (casein kinase II)	14.2 ± 1.2
Casein (protein kinase A)	27.5 ± 2.2
MBP	11.3 ± 0.6

P-TEN activity was measured with the indicated substrates in triplicate. The activity is expressed as the mean pmol phosphate released per min per mg enzyme ± SD.

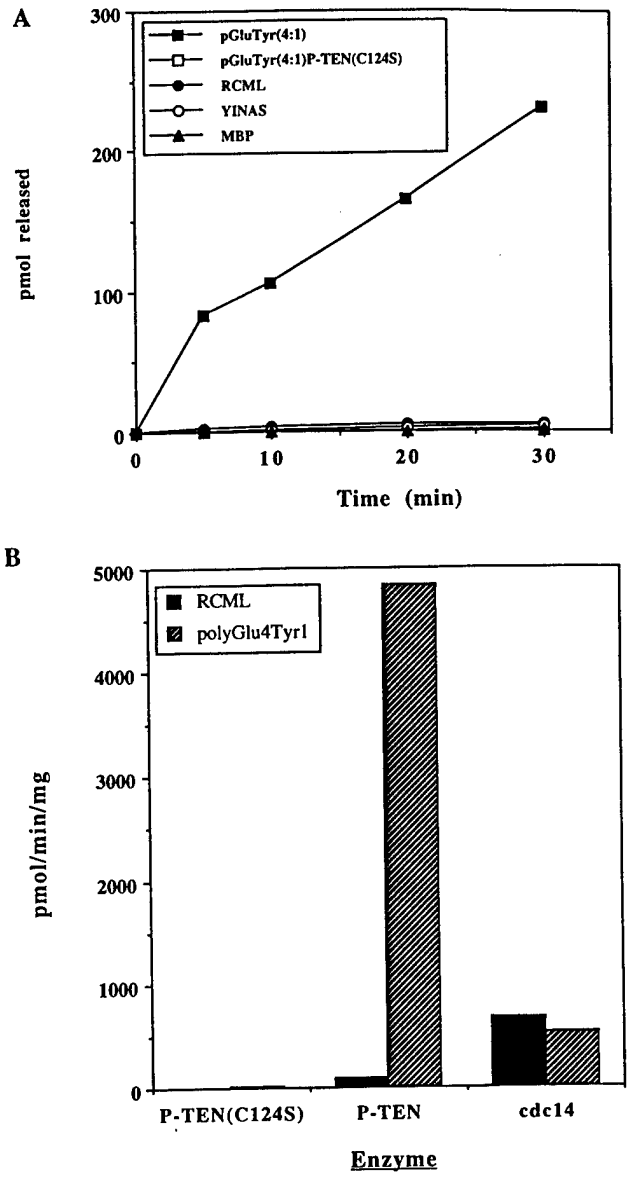


FIG. 1. Tyrosine phosphatase activity of purified P-TEN. (*A*) P-TEN was tested for protein phosphatase activity using the indicated tyrosine-phosphorylated substrates. Activity is expressed as pmol of phosphate released. A catalytically inactive mutant of P-TEN (P-TENC124S) was included as a control to rule out the possibility of contaminating bacterial phosphatases. (*B*) Comparison of P-TEN and cdc14 activities. P-TEN and cdc14 were assayed as above with RCML or polyGlu$_4$Tyr$_1$, and the activity is expressed as pmol of phosphate released per min per mg.

two peptide substrates of casein kinase II also indicates that the inability of P-TEN to dephosphorylate EDNDYINASL does not simply reflect its inability to dephosphorylate small peptide substrates. Although many dual-specificity phosphatases show a preference for tyrosine residues, it is possible that the reduction in the activity of P-TEN toward serine/threonine substrates may be the result of these residues being located in a suboptimal substrate backbone.

Dephosphorylation of ERK2. A number of dual-specificity phosphatases show a preference for members of the MAP kinase family (13, 23, 24). Therefore, we tested whether P-TEN could dephosphorylate ERK2. We could not detect dephosphorylation of ERK2 by P-TEN, either by changes in anti-phosphotyrosine antibody reactivity or by changes in the electrophoretic mobility of ERK2. In contrast, MAP kinase

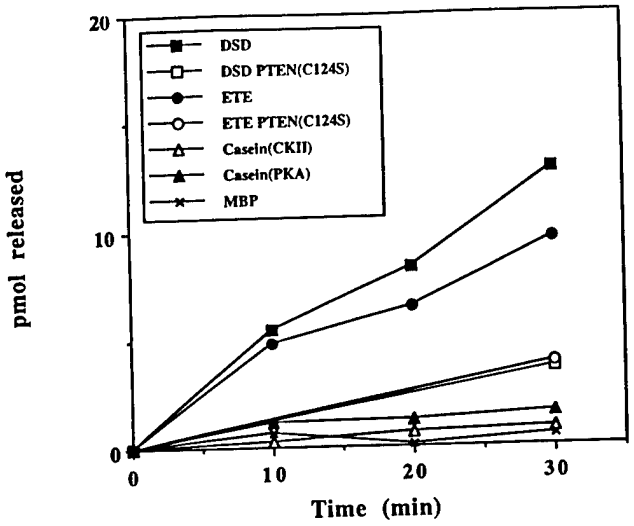


FIG. 2. Dual-specificity phosphatase activity of purified P-TEN. P-TEN was tested for protein phosphatase activity using the indicated serine/threonine-phosphorylated substrates. Activity is expressed as pmol of phosphate released. A catalytically inactive mutant of P-TEN (P-TENC124S) was included to control for contaminating bacterial phosphatases. Casein was phosphorylated, as indicated, with casein kinase II (CKII) or protein kinase A (PKA).

phosphatase 1 (MKP-1) quickly and completely dephosphorylated ERK2, as shown by the removal of phosphotyrosine from ERK2 and an increase in its electrophoretic mobility (Fig. 3 *A* and *B*, respectively).

Effects of Point Mutations on the Activity of P-TEN. To test whether the activity of P-TEN was altered during tumorigenesis, we assessed the effect on activity of a variety of point mutations that occur in P-TEN isolated from tumor specimens. Many of these mutations occur in or near the catalytic motif of P-TEN, including His-123 → Tyr (H123Y), Gly-129 → Arg (G129R) or Gly-129 → Glu (G129E), and Met-134 → Leu

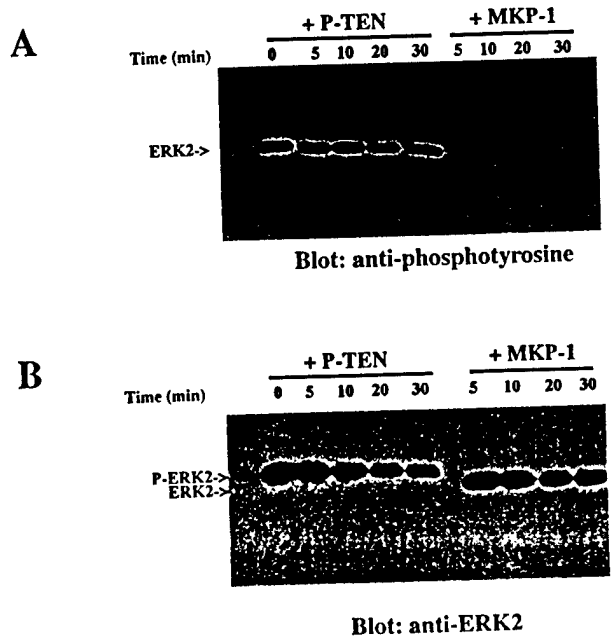


FIG. 3. P-TEN does not dephosphorylate ERK2. Phosphorylated ERK2 was incubated with P-TEN or MKP-1 for the indicated times and ERK2 assayed for residual phosphotyrosine by immunoblotting (*A*) with an anti-phosphotyrosine antibody or (*B*) by immunoblotting with anti-MAP kinase antibodies to visualize the change in electrophoretic mobility.

(M134L). In addition, several mutations were found that occurred outside the conserved catalytic motif, including Leu-57 → Trp (L57W) and a cluster of mutations C terminal to the catalytic loop [Gly-165 → Arg (G165R), Thr-167 → Pro (T167P), and Ser-170 → Arg (S170R)]. The exact positions of these mutations in P-TEN are indicated in Fig. 4. We produced GST fusion proteins in which P-TEN was mutated at each of these positions to mimic the mutant alleles, and the phosphatase activity of the resulting recombinant proteins was measured using $polyGlu_4Tyr_1$ as a substrate (Fig. 5). With the exceptions of the M134L and the G129E mutation, all the mutations tested resulted in a dramatic decrease in the activity of P-TEN (Fig. 5). These data indicate that the catalytic activity of P-TEN has been disrupted in the majority of these tumors.

The activity exhibited by the G129E mutation was surprising, especially in light of the finding that mutation of this same residue to arginine (G129R) resulted in a significant reduction of P-TEN activity. Other potential effects of this mutation on P-TEN, such as changes in protein stability in the complex environment of the mammalian cell or changes in substrate specificity, have not been addressed. On the other hand, the essentially wild-type activity exhibited by the M134L mutation was not totally unexpected because many other dual-specificity phosphatases contain a leucine residue at this position (25).

DISCUSSION

The most common chromosomal deletion in glioblastoma occurs around 10q22-23, suggesting the presence of a tumor suppressor at this locus (26). Mapping of the deletions at 10q22-23, as well as representational difference analysis (27), led to the identification of P-TEN as the tumor suppressor residing at this locus (4, 5). P-TEN shares sequence homology with the cytoskeletal protein tensin and with the family of PTPs (4, 5). A substantial fraction of all glioblastoma samples tested have either deleted or mutated P-TEN alleles. In addition, P-TEN has been shown to be disrupted in a large number of breast and prostate tumor samples (4, 5). Importantly, germ-line mutations in P-TEN give rise to a variety of genetic disorders, most notably Cowden disease, a disorder characterized by the formation of multiple benign tumors (hamartomas), as well as an increased susceptibility to malignant cancers of the breast and thyroid (28), confirming that P-TEN functions similarly to classical tumor-suppressor genes (8). Significantly, many of the mutations isolated from tumor samples, as well as from Cowden disease and Bannayan–Zonana syndrome, were predicted to disrupt the phosphatase domain of P-TEN. These studies, however, did not address whether P-TEN is a functional phosphatase or whether these mutations abrogate activity.

To address these issues, we measured the enzymatic activity of recombinant P-TEN. Initially, we were frustrated by the low activity of the fusion proteins when assayed with proteins that are commonly used as PTP substrates, such as RCML or MBP. Recently, another group has also reported weak phosphatase activity when RCML was used as a substrate (29). However, we demonstrate that P-TEN exhibits robust phosphatase activity when assayed with a random copolymer of glutamate and tyrosine, $polyGlu_4Tyr_1$, suggesting that it exhibits an unusual substrate specificity. P-TEN did not efficiently dephosphorylate the acidic peptide EDNDYINASL, suggesting that the presence of acidic residues may not be sufficient to create an optimal substrate and that higher order structures, such as those found in protein substrates, may be required.

On the basis of the presence of the "AYLL/M" motif found in many dual-specificity phosphatases, as well as the absence of some of the sequence motifs found in PTPs, particularly those that contribute to creating a deep, substrate-binding cleft (30), one would predict that P-TEN is a dual-specificity

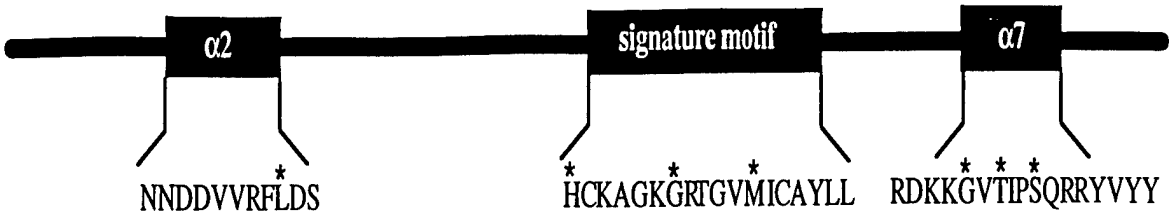


FIG. 4. Location of P-TEN mutations. A diagram of P-TEN showing the locations of the point mutations, indicated by an *, that were tested in this study. In addition, the predicted structural motifs (see text) in which these mutations lie is also indicated.

phosphatase. P-TEN was shown to exhibit activity against serine/threonine, as well as tyrosine-phosphorylated proteins (Table 1), verifying that it is a dual-specificity phosphatase. Like the tyrosine-phosphorylated substrates, P-TEN dephosphorylated only the most acidic serine/threonine-phosphorylated substrates tested. Even the best serine/threonine substrates, the ETE and DSD peptides, were not as efficiently dephosphorylated as $polyGlu_4Tyr_1$ (Table 1). Casein, phosphorylated by casein kinase II, was dephosphorylated poorly by P-TEN (Table 1 and Fig. 2), suggesting that P-TEN may require multiple acidic residues positioned both N and C terminally to the phosphorylated residue. It is currently unclear if the apparent amino acid selectivity of P-TEN is an intrinsic characteristic or if it is caused by differences in the affinity of P-TEN for $polyGlu_4Tyr_1$ and the ETE or DSD peptides, and may reflect a preference for protein over peptide substrates.

Mutations that occur in P-TEN during tumorigenesis fall into three large classes: (*i*) genomic deletions encompassing all or most of P-TEN, (*ii*) frameshift mutations resulting in the production of truncated P-TEN proteins, and (*iii*) point mutations resulting in the substitution of one amino acid for another (4, 5). We have chosen to introduce a variety of the point mutations found in tumor samples into P-TEN, because they are less likely than the frameshift deletions to cause unpredictable changes in the conserved secondary structure

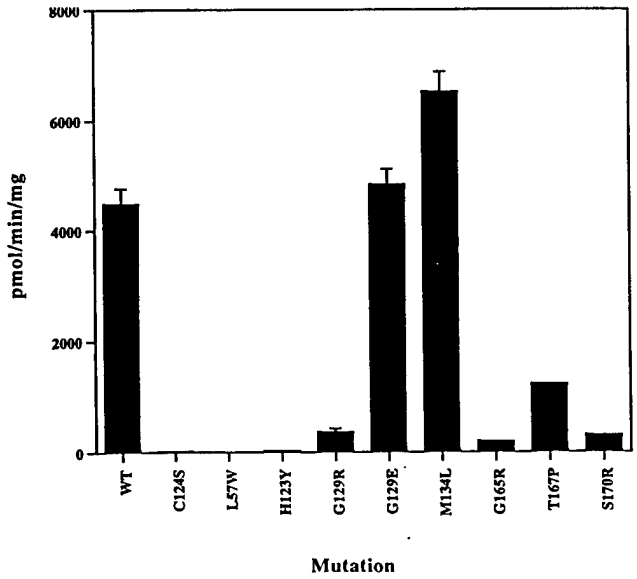


FIG. 5. Disruption of P-TEN activity by point mutations found in tumor samples. The indicated point mutations were introduced into recombinant P-TEN and their effects on phosphatase activity were determined. Assays were performed with $polyGlu_4Tyr_1$ for 15 min. Activity is expressed as pmol of phosphate liberated per min per mg of P-TEN. Assays were performed in triplicate and are expressed as the mean ± SD. The catalytically inactive mutant of P-TEN (P-TENC124S) was included as control to rule out the possibility of contaminating bacterial phosphatases.

found in all PTPs (30). Comparisons of P-TEN with other phosphatases, whose crystal structures have been solved, aids in predicting how these point mutations might disrupt P-TEN activity. For example, the mutation of His-123 → Tyr (H123Y) and Gly-129 → Arg, two residues located in the catalytic motif, resulted in the complete loss of phosphatase activity. One would have predicted that the H123Y mutation, detected in an endometrial cancer, would be catastrophic to activity because of the critical importance of this histidine in correctly orienting the catalytic cysteine so it can act as a nucleophile (11). Although, Gly-129 is not as highly conserved as other residues found in the catalytic motif, the substitution of the large, positively charged side chain of arginine, a mutation found in a glioblastoma cell line, for the much smaller glycine, is likely to have a deleterious effect on the overall structure of the phosphate-binding loop and is likely to impede the binding of phosphate. In contrast, a mutation that was identified in a prostate tumor cell line, in which Met-134 was changed to Leu (M134L), had no effect on P-TEN activity. However, among the dual-specificity phosphatases, leucine is commonly found at this position (25), suggesting that the M134L allele may represent a naturally occurring polymorphism rather than a mutation.

Several other mutations that reside outside of the catalytic motif were also tested for their effect on P-TEN activity. A point mutation discovered in a glioblastoma sample that changes Leu-57 → Trp (L57W) also eliminated P-TEN phosphatase activity. This amino acid is located in a conserved α-helix (α2), which is found in both PTP1B and VHR. Although not directly involved in catalysis, this helix helps form the overall secondary structure of the enzyme (14, 15). Similarly, mutation of residues in this helix of LAR, a receptor PTP, also resulted in a significant loss of phosphatase activity (31). A second cluster of point mutations was discovered in the last conserved structural motif found in most PTPs and dual-specificity phosphatases, which also is an α-helix (14, 15). In YopH, a PTP isolated from the causative agent of bubonic plague, a hydrophobic residue in this α-helix is important for coordinating the water molecule necessary for regenerating the active enzyme (15). All three mutations found in this region—Gly-165 → Arg (found in a glioblastoma), Thr-167 → Pro (found in a breast cancer), and Ser-70 → Arg (found in a patient with a Bannayan–Zonana syndrome)—resulted in the loss of P-TEN activity. The G165R and the S170R mutations would result in substitution of relatively small, uncharged amino acids for a much larger positively charged residue, potentially disrupting important interactions between this α-helix and surrounding structures. Moreover, the substitution of Thr-167 → Pro is also likely to disrupt these interactions by interrupting the proper folding of this α-helix. These data indicate that this conserved helix (α7 in YopH and VHR) is a required motif in the dual-specificity phosphatases.

All of the mutations tested were derived from tumor samples that, with two exceptions, resulted in the complete loss of, or greatly reduced, enzymatic activity of P-TEN, even when the mutations resided well outside of the conserved catalytic motif. These observations indicate that the inhibition of en-

zymatic activity was required for the progression of these cells to a cancerous state.

M134L was one of two mutations found in which activity was not lost. However, leucine is commonly found at this position in other dual-specificity phosphatases (25) and the wild-type allele was not lost in the tumor from which this mutation derived. Hence, M134L is likely to be a naturally occurring polymorphism.

We were surprised by the finding that the G129E mutation, a mutation found in the germ line of patients suffering from Cowden disease, did not disrupt P-TEN activity, whereas a mutation of this same residue to arginine (found in a glioblastoma) ablated P-TEN activity. Although we were unable to detect a change in the activity of P-TEN carrying the G129E mutation, it is likely that this mutation effects the ability of P-TEN to function *in vivo*, as this mutation has been isolated from two independent Cowden disease kindreds but not from unaffected patients (8). Moreover, deletions of the wild-type allele were found in the tumors that develop in this Cowden disease kindred (8). Although the G129E mutation does not effect activity when measured *in vitro* using artificial substrates, this mutation may interfere with the ability of P-TEN to dephosphorylate its physiological targets. Furthermore, this mutation may decrease the stability or half-life of P-TEN in the complex environment of the mammalian cell. A full understanding of the effects of this mutation on P-TEN will require the identification of its physiological substrates, regulators and perhaps even the elucidation of the structural changes elicited by this amino acid substitution.

Even though the sample size is small, there does appear to be a correlation between the severity in the disruption of P-TEN activity and the pathology of the disease. In contrast to the G129E mutation, the S170R mutation exhibited a significant reduction in activity and was isolated from a patient suffering from a severe disorder, Bannayan–Zonana, that shares many features with Cowden disease, but also manifests additional pathologies, including early onset of the disease at birth (9).

To understand how P-TEN functions as tumor suppressor, it will be necessary to identify its physiological substrates. Most dual-specificity phosphatases dephosphorylate and inactivate the MAP kinases (13, 23, 24), and moreover, the MAP kinases are found to be hyperphosphorylated in breast cancer (32), a cancer that frequently contains P-TEN mutations or deletions (8). Therefore, we tested if P-TEN could dephosphorylate the MAP kinase ERK2. However, our data indicate that P-TEN was incapable of dephosphorylating ERK2 *in vitro*, strongly suggesting that the MAP kinases are not regulated by P-TEN *in vivo*. Thus, the physiological substrates of P-TEN are still unknown. The unique substrate specificity of P-TEN, however, should aid in identifying its *in vivo* substrates. Importantly, the requirement for acidic residues may be satisfied by the presence of phosphorylated residues, suggesting that multiply phosphorylated proteins may serve as P-TEN substrates.

We thank Drs. D. Barford, K. Lerea, and H. Charbonneau for providing reagents and Dr. L. Hedrick for providing information on P-TEN mutations in endometrial cancer. We are grateful to M. Daddario for expert technical support and to T. Tiganis, S.H. Zhang and especially to H. Charbonneau for helpful discussions. We also thank Drs. E. Fischer and B. Vogelstein for critical reading of the manuscript. This work was supported by grants to N.K.T. from the National Institutes of Health (CA53840 and GM 55989). M.P.M. was supported by a National Cancer Institute training grant (5T32 CA09311-18). C.E. is the Lawrence and Susan Marx Investigator in human cancer genetics. M.H.W. is an American Cancer Society Research Professor and is supported by the U.S. Department of the Army (DAMD-17-94-14247), the National Cancer Institute (5R35 CA39829), Amplicon Corporation, and the "1 in 9" breast cancer organization.

1. Rosen, N. (1995) in *Molecular Basis of Cancer*, eds. Mendelsohn, V., Howley, P. M., Israel, M. A. & Liota, L. A. (Saunders, Philadelphia), pp. 105–140.
2. Gauwerky, C. E. & Croce, C. M. (1995) in *Molecular Basis of Cancer*, eds. Mendelsohn, V., Howley, P. M., Israel, M. A. & Liota, L. A. (Saunders, Philadelphia), pp. 18–37.
3. Barker, K. & Hanafusa, H. (1990) *Mol. Cell. Biol.* **10**, 3813–3817.
4. Steck, P. A., Perhouse, M. A., Jasser, S. A., Yung, W. K. A., Lin, H., Ligon, A. H., Lauren, A. L., Baumgard, M. L., Hattier, T., Davis, T., Frye, C., Hu, R., Swedlund, B., Teng, D. H. F. & Tavtigan, S. V. (1997) *Nat. Genet.* **15**, 356–362.
5. Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang, S., Puc, J., Milliaresis, C., Rodgers, L., McCombie, R., Bigner, S. H., Giovanella, B. C., Ittman, M., Tycko, B., Hibshoosh, H., Wigler, M. H. & Parsons, R. (1997) *Science* **275**, 1943–1946.
6. Eng, C., Murday, V., Seal, S., Mohammed, S., Hodgson, S. V., Chaudray, M. A., Fentiman, I. S., Ponder, B. A. J. & Eeles, R. A. (1994) *J. Med. Genet.* **31**, 458–461.
7. Eng, C. (1997) *J. Genet. Counsel.* **6**, 181–191.
8. Liaw, D., Marsh, D. J., Li, J., Dahia, P. L. M., Wang, S. I., Zheng, Z., Bose, S., Call, K. M., Tsou, H. C., Peacocke, M., Eng, C. & Parsons, R. (1997) *Nat. Genet.* **16**, 64–67.
9. Grolin, R. J., Cohen, M. M., Condon, L. M. & Burke, B. A. (1992) *Am. J. Med. Genet.* **44**, 307–314.
10. Charbonneau, H. & Tonks, N. K. (1992) *Annu. Rev. Cell Biol.* **8**, 463–493.
11. Barford, D., Flint, A. J. & Tonks, N. K. (1994) *Science* **263**, 1397–1404.
12. Flint, A. J., Tiganis, T., Barford, D. & Tonks, N. K. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 1680–1685.
13. Sun, H., Charles, C. H., Lau, L. F. & Tonks, N. K. (1993) *Cell* **75**, 487–493.
14. Stuckey, J. A., Schubert, H., Fauman, E. B., Zhang, Z.-Y., Dixon, J. E. & Saper, M. A. (1994) *Nature (London)* **370**, 571–575.
15. Yuvaniyama, J., Denu, J. M., Dixon, J. E. & Saper, M. A. (1996) *Science* **272**, 1328–1331.
16. Garton, A. J., Flint, A. J. & Tonks, N. K. (1996) *Mol. Cell. Biol.* **16**, 6408–6418.
17. Hertog, J., Tracy, S. & Hunter, T. (1994) *EMBO J.* **13**, 3020–3032.
18. Diamond, R. H., Cressman, D. E., Laz, T. M., Abrams, C. S. & Taub, R. (1994) *Mol. Cell. Biol.* **14**, 3752–3762.
19. Flint, A. J., Gebbink, M. F. G. B., Franza, B. R., Hill, D. E. & Tonks, N. K. (1993) *EMBO J.* **12**, 1937–1946.
20. Myers, M. P., Murphy, M. B. & Landreth, G. E. (1994) *Mol. Cell. Biol.* **14**, 6954–6961.
21. Zhang, S.-H., Eckberg, W. R., Yang, Q., Samatar, A. A. & Tonks, N. K. (1995) *J. Biol. Chem.* **270**, 20067–20072.
22. Garton, A. J., Flint, A. J. & Tonks, N. K. (1996) *Mol. Cell. Biol.* **16**, 6408–6418.
23. Kwak, S. P., Hakes, D. J., Martell, K. J. & Dixon, J. E. (1994) *J. Biol. Chem.* **269**, 3569–3604.
24. Mourey, R. J., Vega, Q. C., Campbell, J. S., Wenderoth, M. P., Hauschka, S. D., Krebs, E. G. & Dixon, J. E. (1996) *J. Cell Biol.* **271**, 3795–3802.
25. Barford, D. (1995) *Curr. Opin. Struct. Biol.* **5**, 728–34.
26. Fults, D., Pedone, C. A., Thomas, G. A. & White, R. (1990) *Cancer Res.* **50**, 5784–5789.
27. Lisitsyn N. A., L. N. M., Dalbagni G., Barker P., Sanchez C. A., Gnarra J., Linehan W. M., Reid B. J. & Wigler M. H. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 151–155.
28. Mallory, S. B. (1995) *Dermatol. Clin.* **13**, 27–31.
29. Li, D.-M. & Sun, H. (1997) *Cancer Res.* **57**, 2124–2129.
30. Barford, D., Jia, Z. & Tonks, N. K. (1995) *Nature Struct. Biol.* **2**, 1043–1053.
31. Streuli, M., Krueger, N. X., Thai, T., Tang, M. & Saito, H. (1990) *EMBO J.* **9**, 2399–407.
32. Sivaraman, V. S., Wang, H., Nuovo, C. J. & Malbon, C. C. (1997) *J. Clin. Invest.* **99**, 1478–1483.

Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 3764–3769, March 1998
Genetics

Rapid isolation of cDNA by hybridization

(gene/bacterial artificial chromosome/exon trapping/sequencing)

MASAAKI HAMAGUCHI, ELIZABETH A. O'CONNOR, TONG CHEN, LARRY PARNELL, RICHARD W. MCCOMBIE, AND MICHAEL H. WIGLER*

Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724

Contributed by Michael H. Wigler, January 21, 1998

ABSTRACT **The isolation of genes from a given genomic region can be a rate-limiting step in the discovery of disease genes. We describe an approach to the isolation of cDNAs that have sequences in common with large genomic clones such as bacterial artificial chromosomes. We applied this method to loci both amplified and deleted in cancer, illustrating its usage in the identification of both oncogenes and tumor suppressor genes, respectively. The method, called rapid isolation of cDNAs by hybridization (RICH), depends on solution hybridization, enzymatic modification, and amplification/selection of sequences present in both cDNA populations and the genomic clones. The method should facilitate the development of transcription maps for large genomic clones, possibly even yeast artificial chromosomes.**

Powerful methods have facilitated the localization of disease genes to regions of the genome. Typically, candidate regions are contained on large yeast or bacterial cloning vectors, and these vectors must be searched assiduously by various means for candidate genes. This step has often proved to be an obstacle in gene discovery. The problem has been attacked in roughly three ways: by hybridization (1–3), sequence analysis (4, 5), and exon trapping methods (6–8). Each method has its own particular advantages, but no current method is without serious problems.

We present herein a method we call rapid isolation of cDNAs by hybridization (RICH) based on the identification of sequences in common between a cDNA library and a large clone of genomic DNA. The method selects and amplifies those restriction endonuclease fragments of cDNAs that hybridize precisely at one end to the end of a similarly cleaved genomic DNA fragment. Before hybridization, the cDNA and genomic fragments are modified with different adaptors. Those cDNAs that form hybrids with genomic DNA at at least one end are ligated to a "selection adaptor" that is complementary to the genomic adaptor and contains an additional sequence complementary to an RNA polymerase site. Such cDNAs can be selectively amplified by successive treatments with RNA and DNA polymerases.

We illustrate the basic method with two series of experiments: (*i*) a search for transcripts from the c-*MYC* locus in cDNAs from a breast cancer cell line and (*ii*) a search for transcripts from the *PTEN* tumor suppressor locus in cDNAs from normal breast tissue. Although the method is complex, in that many different enzymes (restriction endonucleases, various DNA ligases, RNA polymerase, various DNA endo- and exonucleases, reverse transcriptase, and various DNA polymerases) are used, they are all robust enzymes that are readily available. Only 3 days are required to yield candidates for further study.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Materials. The oligonucleotides which were synthesized for this research are listed in Table 1 and obtained from Biosynthesis (Lewisville, TX). P1 clone 8001 (Genome Systems, St. Louis) is an 80-kb genomic clone that contains the exons 1–3 of the c-*MYC* gene. A bacterial artificial chromosome (BAC) clone, 60C5, containing exons 4–9 of *PTEN* (containing nucleotides 1,244–2,246 of the published cDNA, GenBank accession no. U92436) was obtained from Genome Systems. The plasmid pUC18, digested with *Bam*HI and treated with bacterial alkaline phosphatase was supplied by Amersham. pCR-Script SK(+) and Epicurian Coli XL2-Blue cells were obtained from Stratagene. SKBr3 is a breast cancer cell line from which the poly(A)$^+$ RNA was extracted with the Fast-Track kit (Invitrogen). cDNA was synthesized from the poly(A)$^+$ RNA of SKBr3 and commercially available poly(A)$^+$ RNA of mammary gland tissue (CLONTECH) by using the Copy kit (Invitrogen). The Megascript kit was from Ambion (Austin, TX).

The enzymes and their specific buffers used in RICH were obtained from the following suppliers: *Sau*3AI and T4 DNA ligase from New England Biolabs; Stoffel fragments, AmpliTaq, and AmpliTaq Gold from Perkin–Elmer; Ampligase from Epicentre Technologies (Madison, WI); *Pfu* DNA Polymerase and RNase-free Dnase from Stratagene; and λ exonuclease from Amersham.

Glycogen was obtained from Boeringer Mannheim. GeneQuant G-50, S-300HR, and S-400HR columns and Sephaglas BandPrep kit were obtained from Amersham. RNase-free water was supplied by Ambion. Maxi prep kit was supplied by Qiagen (Hilden, Germany). Phenol was prepared as described elsewhere (9).

RICH Standard Protocol. To facilitate description of the RICH protocol, we have broken up the procedure into discrete units, labeled those units with an alphabetic letter, and assigned that letter both to the procedure and its final product (Fig. 1).

A. Preparation of genomic DNA. For the preparation of genomic DNA, 1 μg of BAC DNA is digested with 20 units of *Sau*3AI. The digests are purified by phenol/chloroform extraction and ethanol precipitation. The whole digests are mixed with 100 pmol of pR-24 and 100 pmol of R-12 oligonucleotides (Table 1) in 20 μl of 1× T4 DNA ligase buffer. The double-strand cDNA and the oligonucleotides are heated at 65°C for 5 min, annealed by cooling down the mixture to 4°C gradually, and then ligated by overnight incubation with 2,000 units of T4 DNA ligase at 11°C. The excess oligonucleotides are removed from the ligated fragments with S-400HR columns.

B. Preparation of selection adaptor. The selection adaptor is synthesized, phosphorylated at its 5′end, and purified by PAGE. The working concentration of the adaptor is 50 nM.

Abbreviations: RICH, rapid isolation of cDNA by hybridization; BAC, bacterial artificial chromosome; RIGHT, rapid isolation of genes by hybridization to transcripts.

*To whom reprint requests should be addressed. e-mail: wigler@cshl.org.

Table 1. Oligonucleotide sequences for RICH

Oligo	Sequence	5′ end
N-12	GATCTTCCCTCG	Dephosphorylated
R-12	GATCTGCGGTGA	Dephosphorylated
pN-24	AGGCAACTGTGCTATCCGAGGGAA	Phosphorylated
pR-24	AGCACTCTCCAGCCTCTCACCGCA	Phosphorylated
pR(+)SP6	CTGCGGTGAGAGGCTGGAGAGTGCTCTATAGTGTCACCTAAAT	Phosphorylated
SP6	TATTTAGGTGACACTATAGAGCA	Dephosphorylated

C. Preparation of cDNA. Double-stranded cDNA is synthesized from 5 μg of poly(A)$^+$ RNA with the Copy kit. One microgram of cDNA is digested with 20 units of *Sau* 3AI and ligated to 100 pmol of pN-24 and 100 pmol of N-12 oligonucleotides in a 20-μl reaction volume as described above. A partial fill-in mixture is made by combining 21 μl of water, 5 μl of 10× Stoffel buffer, 1 μl of 10 mM dATP, 1 μl of 10 mM dGTP, and 1 μl of 10 mM dTTP and is added to the ligation mixture. After incubation at 72°C for 5 min, 2 units of Stoffel fragment are added to the mixture and incubated at 72°C for additional 5 min. The fragments are extracted twice with 50 μl of phenol/chloroform. The aqueous phase is transferred to a new tube.

D. Hybridization. Ten microliters of genomic fragments (components **A**), 5 μl of R(+)SP6 (components **B**), and 25 μl of cDNA fragments (components **C**) are mixed and then extracted twice with 40 μl of phenol/chloroform. Ten microliters of 10 M ammonium acetate, 1 μl of glycogen (20 mg/ml), and 92 μl of ethanol are added. After centrifugation and washing with 70% ethanol, the DNA is dissolved in 4 μl of 3× EE buffer (10). The solution is overlaid with 30 μl of mineral oil and denatured by incubation in boiling water for 5 min. One microliter of 5 M NaCl is added and the DNA is reannealed for 16 h at 67°C.

E. Ligation of selection adaptor. To ligate the selection adaptor to the cDNA fragments, 6 μl of 10× Ampligase buffer and 1 μl of Ampligase (100 units/μl) are added to 48 μl of water, and this ligation mixture is incubated at 67°C for 3 min before adding to the solution (component **D**). The total mixture is then incubated for 4 h at 67°C. After ligation, the DNA is extracted with 60 μl of phenol/chloroform twice, precipitated by ethanol as above, and dissolved in 50 μl of TE buffer (10 mM Tris·HCl, pH 7.5/1 mM EDTA). Excess R(+)SP6 is removed by S-400HR after denaturation.

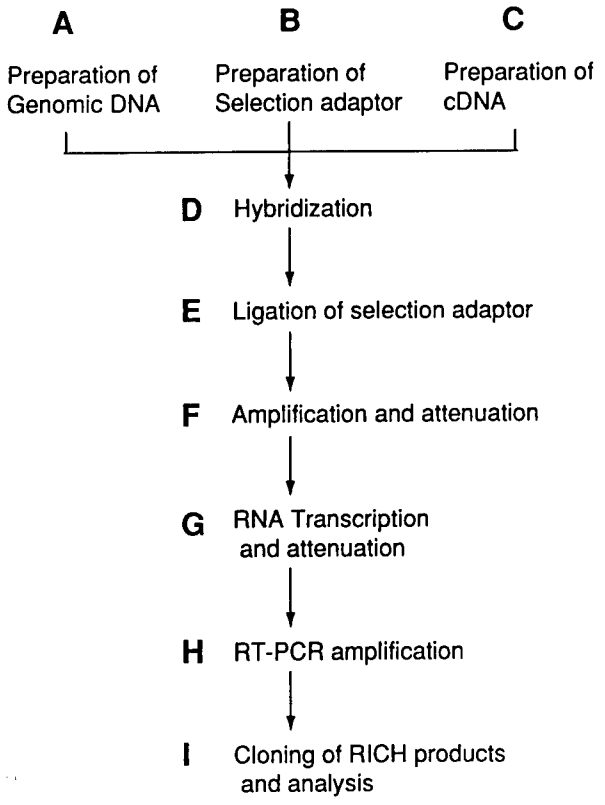


FIG. 1. Flowchart of RICH. Each boldface letter represents a unit of the RICH procedure described in the standard RICH protocol.

F. Amplification and attenuation. Five microliters of the DNA solution (component **E**) is mixed with 50 μl of reaction mixture [1× AmpliTaq buffer/all four dNTPs (each at 200 nM)/1 μM pN-24/1 μM SP6/1 unit of AmpliTaq Gold]. The DNA is amplified by a PCR program that specifies an incubation for 10 min at 94°C followed by 10 cycles of 1 min at 94°C, 1 min at 60°C, and 3 min at 72°C. After purification with S-300 HR columns, the DNA is digested at 37°C with 10 units of λ exonuclease. Eighty microliters of TE is added to the reaction mixture and incubated in boiling water for 5 min. The 100-μl PCR mixture contains 8 μl of the DNA, all four dNTPs (each at 200 μM), 1 μM pN-24, 1 μM SP6, and *Pfu* buffer at a 1× final concentration. After initial denaturation of 94°C for 5 min, 2.5 units of *Pfu* polymerase is added to the mixture and 20 cycles at 94°C for 1 min and 70°C for 4 min are performed, followed by a final extension at 72°C for 10 min. The PCR products are purified with phenol/chloroform extraction and ethanol precipitation. The DNA is dissolved in 25 μl of RNase-free water after washing with 70% ethanol.

G. RNA transcription and attenuation. RNA is transcribed from 8 μl of the DNA template (component **F**) with the Megascript kit. After a 6-h incubation at 37°C, 20 units of RNase-free Dnase I is added and the incubation is continued for an additional 10 min. The RNA is purified with phenol/chloroform extraction and dissolved in 15 μl of RNase-free water after ethanol precipitation.

H. Reverse transcription-coupled PCR amplification. To synthesize cDNA, 1.3 μl of 50 μM pN-24 is added to the mixture (component **G**), incubated at 65°C for 10 min, and then placed at room temperature for 5 min. Double-stranded cDNA is synthesized by the Copy kit following the supplier's protocol. After the products are denatured by heating to 94°C for 10 min, they are purified with S-300HR columns. The 200-μl PCR mixture contains 10 μl of the cDNA, all four dNTPs (each at 200 μM), 1 μM pN-24, 1 μM pR-24, 5 units of AmpliTaq Gold, and AmpliTaq buffer at a 1× final concentration. After initial denaturation of 94°C for 10 min, 30 cycles at 94°C for 1 min and 72°C for 3 min were performed, followed by a final extension at 72°C for 10 min.

I. Cloning and sequencing. The RICH products were purified with GeneQuant G-50 columns and digested with 40 units of *Sau*3AI in a 120-μl reaction volume at 37°C for 4 h. The digests were size-fractionated by electrophoresis through an agarose gel for Southern blot analysis and for cloning. The DNA bands were excised, purified with the Sephaglas BandPrep kit, and cloned into pUC18 digested with *Bam*HI and treated with bacterial alkaline phosphatase. Epicurian Coli XL2-Blue cells were transformed with the plasmid. The plasmid DNA was purified with the Maxi prep kit and sequenced by the Cycle Sequencing system (Perkin–Elmer).

J. Southern blot analysis. A part of RICH product was electrophoresed through a 4% agarose gel and transferred to

Hybond N^+ (Amersham). Hybridization with the PTEN cDNA probe was detected by the ECL system (Amersham).

K. Making representations of genomic clones. In one variation of the basic method, we used representations of genomic clones. To make the representation, 100 ng of BAC DNA is digested with 5 units of *Sau*3AI. The digests are purified by phenol/chloroform extraction and ethanol precipitation. The whole digests are mixed with 100 pmol of pR-24 and 100 pmol of R-12 oligonucleotides (Table 1) in 20 μl of 1× T4 DNA ligase buffer. The double-strand cDNA and the oligonucleotides are heated at 65°C for 5 min, annealed by cooling the mixture to 4°C gradually, and then ligated by overnight incubation with 2,000 units of T4 DNA ligase at 11°C.

RESULTS

The Basic Method. For convenience, we break the method into two parts. In part one, we add a selection adaptor to only those cDNA fragments that hybridize at one end to an end of a genomic fragment. In part two, we selectively amplify only those modified cDNA fragments. The complexity of the method arises from the need to suppress the amplification of self-annealed cDNA and genomic homoduplexes. Fig. 2 outlines the basic schema.

Part One. First, both the large genomic clone and cDNA are separately cleaved with *Sau*3AI and modified with T4 DNA ligase by the addition of oligonucleotide adaptors (pR-24 for genomic DNA and pN-24 for cDNA) at their 5′ ends only. These adaptors have phosphate groups at their 5′ ends, for reasons that will become apparent in part two.

Next, the 3′ ends of the adapted cDNA molecules are partially filled-in with the three nucleotides guanosine, adenosine, and thymidine by using the Stoffel fragment (a *Taq* DNA polymerase lacking both 3′ and 5′ exonuclease activity). This step enables us to distinguish later the homoduplexes of genomic DNA from the cDNA–genomic heteroduplexes. An alternative to partial filling is presented later.

The cDNA and genomic DNAs are mixed at equal mass ratios, and the selection adaptor pR(+)SP6 is added. pR(+)SP6 is complementary to pR-24 but has an extra cytidine at its 5′ end and a sequence complementary to the SP6 RNA polymerase promoter sequence at its 3′ end.

The mixture is heat-denatured and allowed to reanneal. During reannealing, two and three part structures will form, as indicated in Fig. 2*A*. The reannealed mixture is treated with a thermostable DNA ligase, Ampligase, at 68°C to ensure that perfect matches are preferentially ligated.

Ligation of pR(+)SP6 to cDNA will occur only when the latter is annealed to a matching genomic fragment, forming a perfect contiguous substrate for the ligase. The extra cytidine at the 5′ end of pR(+)SP6 is needed to fill the gap of the partially filled-in cDNA fragment. pR(+)SP6 will not be ligated to genomic homoduplexes because those three part structures will have a 3-nucleotide gap. Neither will pR(+)SP6 be ligated to cDNA homoduplex because the latter, lacking pR-24 and having only a 1-base overhang, will not base-pair to pR(+)SP6.

Thus only cDNAs with homology to genomic fragments at one end will have the pN-24 adaptors at their 5′ ends and the pR(+)SP6 selection adaptor at their 3′ ends.

Part Two. The selectively modified cDNA fragments are amplified, in the following manner. First, after ligation, we denature the mixture and use *Taq* DNA polymerase and an oligonucleotide primer containing the SP6 promoter sequence to make the selectively modified cDNAs double-stranded. These SP6 primers are not phosphorylated at their 5′ ends. This step is repeated 10 times to give an arithmetic increase in the number of complementary strands (see Fig. 2*B*).

Next, the entire mixture is treated with λ exonuclease, which degrades double-stranded fragments from their phosphory-

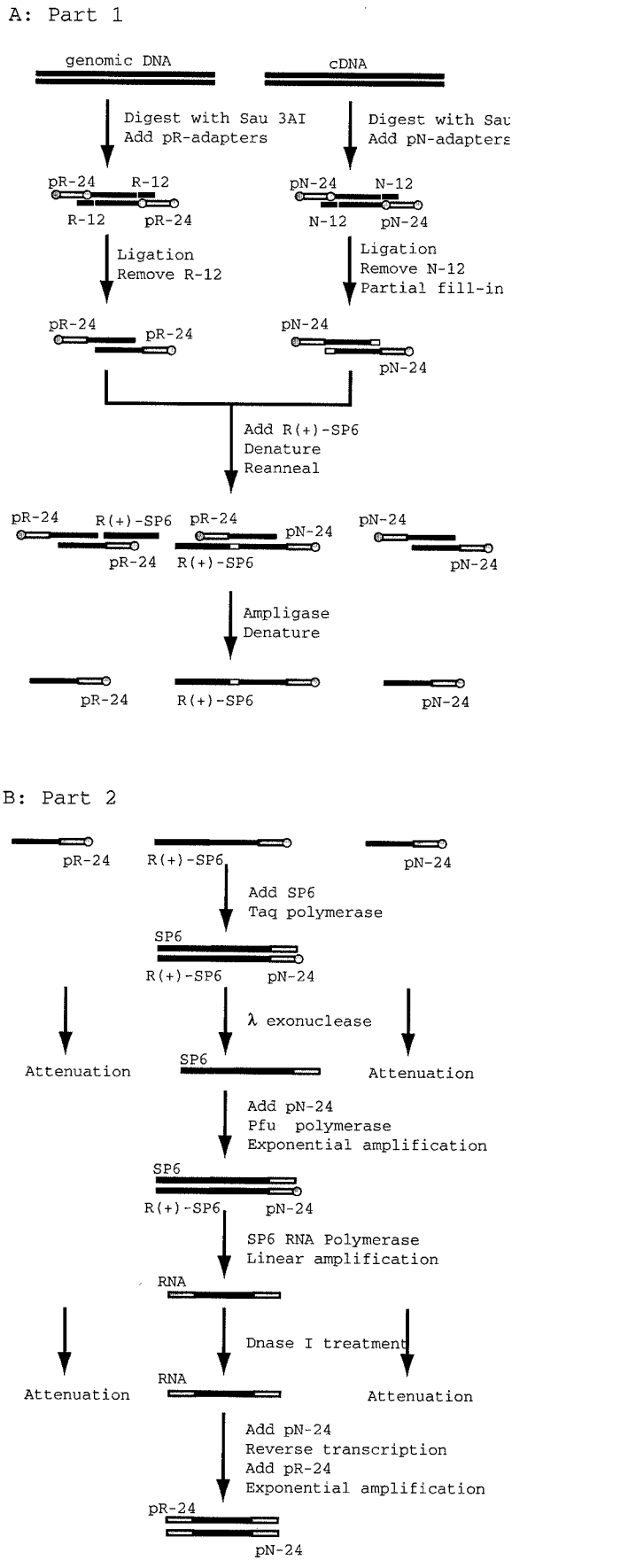


FIG. 2. RICH procedure. (*A*) Part 1. (*B*) Part 2. A circle represents a phosphorylated 5′ end. Open boxes represent filled-in bases. Oligonucleotides and their antisense sequences are drawn as dark and light shaded boxes.

lated 5′ ends. The strands synthesized from the SP6 primers are thus protected. A small number of duplexes will be formed during the reannealing of the previous step and, when filled-in by *Taq* polymerase, would contaminate the subsequent reactions. These contaminants are destroyed by λ exonuclease because the adaptors pR-24 and pN-24 have phosphorylated 5′ ends.

To amplify the SP6-primed strand, we perform a PCR with *Pfu* DNA polymerase and SP6 and pN-24 oligonucleotides as primers. We use *Pfu* polymerase because its products are blunt-ended, whereas *Taq* polymerase sometimes adds an extra nucleotide to the 3′ end of its product, and the SP6 RNA polymerase used in the next step does not work well on substrates with 3′ protruding ends.

In this amplification step, it is likely that cDNA homoduplexes that escaped λ exonuclease treatment will be amplified. Moreover, a significant amount of genomic homoduplexes would be amplified in later steps. Therefore, we use SP6 RNA polymerase (Ambion) to create RNA transcripts of those molecules that contain the SP6 promoter sequence and digest any surviving DNA molecules with DNase I.

Finally, reverse transcription with avian myeloblastosis virus reverse transcriptase (Invitrogen) using pN-24 as primer, followed by PCR with *Taq* DNA polymerase and pN-24 and pR-24 as primers, yields products for cloning and analysis.

Testing the c-*MYC* Locus. Our first mock experiments, not shown, used serial dilutions of the bacterial plasmid pUC18 into a cDNA library, and DNA from a 160-kbp BAC (60C5) to which was added an equimolar amount of the same plasmid. These experiments demonstrated to our satisfaction that the procedure would work well with cDNA species containing as little as 0.01% of a cDNA library. Success of the method was not evident when the pUC18 was present in cDNA at the 0.0001% level. We proceeded next with actual test cases.

In our first test case, we used an 80-kb P1 bacterial cloning vector containing the entire c-*MYC* gene and used randomly primed double-stranded cDNA that was prepared from the poly(A)$^+$ RNA of the breast cancer cell line SKBr3, in which the c-*MYC* locus is amplified. Northern blot analysis indicated that the level of c-*MYC* expression from SKBr3 was in the middle of the range for a panel of breast cancer cell lines, including BT20, Du4475, HS578T, MDA134, MDA231, MDA436, SKBr3, UACC812, UACC893, ZR75–1, and ZR75–30. We estimated that c-*MYC* expression was 3-fold higher in SKBR3 than in normal breast tissue and 3-fold lower than in the highest expressing tumor cell line tested, ZR75–30.

Clones were prepared from the RICH products by either blunt-ended ligation to pCR-Script SK(+) or cleavage with *Sau*3AI followed by ligation to pUC18. Vector inserts were sequenced. All blunt-ended inserts contained both the pN-24 and pR-24 sequences and the adjoining GATC sequences, indicating that they all derived from cDNA–genomic heteroduplexes.

cDNA insert sizes varied from 61 to 259 bp. Roughly one-quarter of clones either had no matches in the data bases or were repeat sequences, mainly of the *alu* family. The remaining three-quarters had identity to sequences found on c-MYC mRNA, as indicated in Fig. 3. One type of RICH product derived from exon 1 of c-*MYC* entirely, one spanned exons 2 and 3, and two derived entirely from exon 3. Shorter fragments, less than 100 bp, and a larger fragment, 880 bp spanning exon 1 and 2, were not obtained in the first 81 clones examined. No clones from either the 3′ or 5′ ends of c-*MYC* were obtained, and this is predicted by theory because a cDNA fragment would need to have *Sau*3AI sites at both ends to be found.

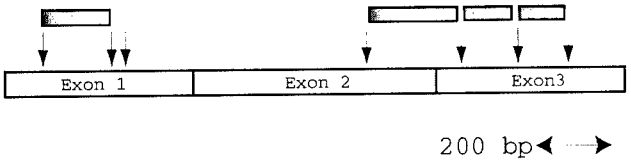


FIG. 3. RICH products and the c-*MYC* gene. Arrows indicate *Sau*3AI recognition sites. Shadowed boxes represent RICH products, the sizes of which are 150, 250, 166, and 135 bp from the left. The large restriction fragment that we failed to isolate lies between third and forth arrows from the left and is 880 bp.

Testing the *PTEN* Locus. In a second test of the method, we used a 160-kb BAC, 60C5, containing *PTEN*, a tumor suppressor gene that encodes a mixed-specificity protein phosphatase expressed ubiquitously but at a low level in normal tissues (11–13). We prepared double-stranded cDNA from normal human breast tissue by random priming of poly(A)$^+$ RNA. As a control we used another BAC from a different region of the genome. By following the same protocol as above, we obtained RICH products from each BAC and probed the products with PTEN cDNA by Southern blotting. Only RICH products from the PTEN BAC yielded fragments that hybridized with PTEN sequences (Fig. 4). In total at least three PTEN products were observed, about 200, 350, and 570 bp long. We expected fragment sizes of 153, 298, and 520 bp without adaptors and 201, 346, and 568 bp with adaptors, all of which span exons. We failed to observe a range of smaller fragments (Fig. 5 and see *Discussion*).

To estimate the ratio of the products that derived from PTEN cDNA to products deriving from repeat sequences, a collection of clones were analyzed by filter hybridization with both PTEN cDNA and total human DNA as probes. Six of 120 clones hybridized to PTEN cDNA and about 60 hybridized to total human genomic sequences.

To extend the utility of RICH, we tested a variation in the method: we used the PCR product from a large insert cloning vector as our source of genomic DNA. For this purpose, we gel-purified PTEN BAC and cleaved it with *Sau*3AI, ligated pR-24 to the 5′ ends, filled-in the 3′ ends, and used PCR to amplify fragments with pR-24 as primer. We used this amplified material in an otherwise identical protocol and obtained results that, by Southern blotting, appeared as satisfactory as when we began with restriction endonuclease-cleaved BAC DNA.

We analyzed in greater depth the RICH products of this experiment by blunt-ended cloning into pCR-Script SK(+). Ninety-six individual clones were analyzed. They were tested for the presence of inserts by PCR using primers derived from the

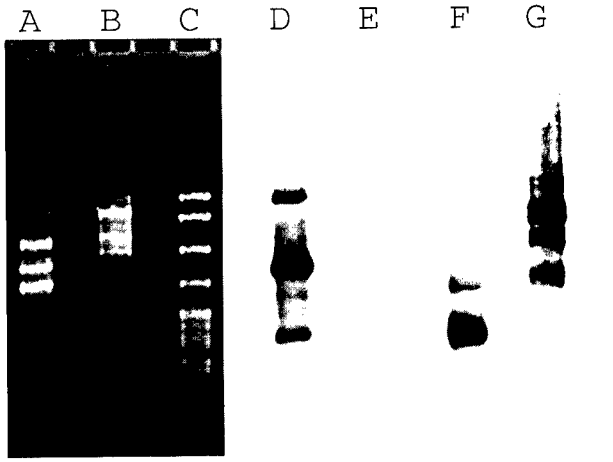


FIG. 4. RICH products from the *PTEN* gene. Lanes: A, RICH products obtained from the PTEN BAC, 60C5; B, RICH products from an unrelated BAC; C, pBR322 DNA digested with *Msp*I. Lanes A and B were transferred to a nylon membrane and probed with PTEN cDNA (lanes D and E) or total human DNA (lanes F and G). Lanes: D and F, RICH products from the PTEN BAC; E and G, RICH products from the unrelated BAC.

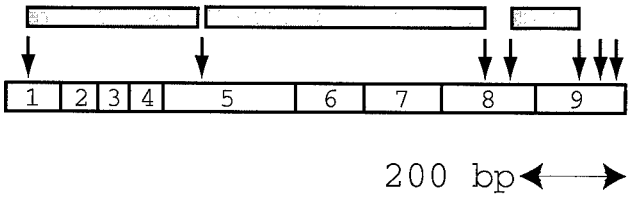


FIG. 5. RICH products and the *PTEN* gene. Arrows indicate *Sau*3AI recognition sites. The shadowed boxes represent RICH products, the sizes of which are 298, 520, and 153 bp from the left. The restriction fragments in exon 8 and exon 9 that we failed to isolate are 81, 33, 27, and 29 bp from the left.

vector cloning sites and for the presence of the *Sau*3AI sites by cleavage of the PCR product. By this simple test, only 17 had *Sau*3AI sites at the expected sites within the inserts. Sequence analysis indicated that all 17 of these contained both the pN-24 and pR-24 primers at the proper sites (this could also have been determined by PCR). Of the 17, five derived from PTEN cDNA, representing the 153-bp (three times) and the 298-bp (twice) *Sau*3AI fragments. The larger fragment, 520 bp long, was not obtained in this group. The remaining 12 inserts contain a *Sau*3AI site within human repeat sequences. Two of these contained repeats we have not yet found in the sequence of the PTEN BAC, the sequence of which is now 90% complete.

DISCUSSION

Finding genes in large chromosomal regions has been approached in three ways: exon trapping, DNA sequencing analysis, and direct hybridization selection. Exon trapping works when the gene in question contains splicing sites that are efficiently recognized by the host cell (6–8). But it fails when introns are absent or the intron–exon borders are not recognized; and exon trapping yields a background of false candidates derived from cryptic splice sites. DNA sequencing is effective when the gene in question has homology to a known expressed sequence. Even without homology, computational methods for predicting genes also have promise (4). But DNA sequencing on a massive scale is still costly. Direct hybridization selection (1) has also found use, but it diminishes in usefulness with rare messages and suffers from the vagaries of physical selection methods and background problems with repetitive sequences. We have described an additional approach to this problem: an effective protocol for selecting cDNA fragments that are homologous at one of their ends to one of the ends from a collection of genomic fragments. This method should work whenever a cDNA population is available that contains transcripts from the gene in question.

A protocol (end ligation coincident sequence cloning, EL-CSC) similar to ours, has been presented by Brookes *et al.* (14). Like ours, their protocol is based upon heteroduplex formation between DNA fragments made from two populations, and the use of ligation (with what they call "capture oligonucleotides") to distinguish heteroduplex from homoduplex. Unlike our procedure, their method requires heteroduplex formation at both ends between cDNA and genomic fragments, because both ends of the heteroduplex must be ligated. Thus, cloning of cDNA fragments that span introns is much less likely. In our procedure, we use a selection adaptor that allows us to generate an RNA intermediate, and so we can isolate heteroduplexes that have formed at only one end. Moreover, EL-CSC uses physical trapping through biotin–avidin complex formation to enrich for products. We have experienced difficulty with protocols incorporating such methods and have avoided them in RICH. The report of Brookes *et al.* (14) does not contain sufficient information to enable us to make quantitative comparisons of our methods nor have we found their procedure used in the published literature.

Although the yield of c-*MYC* fragments in the RICH products was very satisfactory, the yield of PTEN products was less so. An additional prescreening of RICH products from *PTEN* was needed: namely, the verification of the adaptors and the *Sau*3AI sites that should be present upon proper priming and ligation. We believe that this is due to the reduced level of expression of the *PTEN* gene compared with c-*MYC*. Moreover, an analysis of our products from the PTEN BAC revealed a higher proportion of repeat-containing sequences than were found for c-*MYC*, presumably for the same reason.

Not all fragments from the same cDNA will have the same yield. For example, the 880-bp fragment from c-*MYC* was not obtained as a RICH product. We speculate this may be due to the difficulty of obtaining long transcripts from the SP6 polymerase. Also, we did not clone the 520-bp fragment of PTEN that was seen upon Southern blotting to be present in reduced amount. Most strikingly, we do not observe very short fragments in the RICH products. Possibly, this deficit is partially due to the slower kinetics of hybridization of shorter fragments (15). Gel fractionation before cloning might overcome some of these problems of underrepresentation. Alternatively, these limitations can be overcome by using different restriction endonucleases during the protocol.

We have introduced one variation in our method: amplifying genomic DNA fragments before use. This step was incorporated for two reasons. (*i*) If genomic fragments are amplified and not cleaved, it should be unnecessary to partially fill-in the ends of restriction endonuclease cleaved cDNA, because the selection adaptor cannot be ligated to PCR-amplified genomic DNA. Thus, with this variation, any restriction endonuclease can be chosen, if used both for cDNA cleavage and amplification of genomic fragments, overcoming limitations in the discovery of cDNAs that might result from the use of *Sau*3AI discussed above. (*ii*) With genomic amplification the user can initiate the search for transcripts with very small amounts of genomic DNA. In fact, we have performed RICH starting from small amounts of gel-purified BAC DNA. It may be possible to extend this method to gel-purified yeast artificial chromosome DNAs.

We cannot obtain the 3′ or 5′ ends of cDNA transcripts with RICH because only cDNA fragments with restriction endonuclease cleavage sites at both ends can be selectively amplified. However, RICH can be used in reverse (rapid isolation of genes by hybridization to transcripts, RIGHT) to identify genomic clones with homology to cDNAs. The directional modifications of genomic fragments and cDNAs used in RICH can be essentially reversed, rendering genomic fragments that form heteroduplex with cDNA the only selectable RIGHT products. When unamplified cDNAs are used, the RIGHT protocol should yield the 3′ and 5′ ends of a transcription unit. In addition, RIGHT might facilitate the confirmation of cDNAs found by RICH and aid in the determination of intron-exon boundaries.

We thank Robert Lucito for useful discussions; Eric Green, Raju Kucherlapati, and Susan Naylor for critical reading of the manuscript; Scott Powers for providing RNA; Mike Riggs for DNA sequencing; Jim Duffy for artwork; and Patricia Bird for secretarial assistance. This work was supported by grants to M.H.W. from the National Institutes of Health (OIG-CA39829 and 5P50-CA-68425); the U.S. Army (DAMD17-94-J-4247); 1 in 9: The Long Island Breast Cancer Action Coalition; and Tularik. M.H.W. is an American Cancer Society Research Professor.

1. Lovett, M., Kere, J. & Hinton, L. M. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 9628–9632.
2. Mancini, M., Sala, C., Rivella, S. & Toniolo, D. (1996) *Genomics* **38,** 149–154.
3. Parimoo, S., Patanjali, S. R., Shukla, H., Chaplin, D. D. & Weissman, S. M. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 9623–9627.
4. Uberbacher, E. C. & Mural, R. J. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 11261–11265.

5. Zhang, M. Q. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 565–568.
6. Duyk, G. M., Kim, S. W., Myers, R. M. & Cox, D., R. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 8995–8999.
7. Buckler, A. J., Chang, D. D., Graw, S. L., Brook, J. D., Haber, D. A., Sharp, P. A. & Housman, D. E. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 4005–4009.
8. Hamaguchi, M., Sakamoto, H., Tsuruta, H., Sasaki, H., Muto, T., Sugimura, T. & Terada, M. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 9779–9783.
9. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed., p. B.4.
10. Straus, D. & Ausubel, F. M. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 1889–1893.
11. Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang, S. I., Puc, J., Miliaresis, C., Rodgers, L., McCombie, R. *et al.* (1997) *Science* **275,** 1943–1947.
12. Steck, P. A., Pershouse, M. A., Jasser, S. A., Yung, W. K., Lin, H., Ligon, A. H., Langford, L. A., Baumgard, M. L., Hattier, T., Davis, T. *et al.* (1997) *Nat. Genet.* **15,** 356–362.
13. Myers, M. P., Stolarov, J. P., Eng, C., Li, J., Wang, S. I., Wigler, M. H., Parsons, R. & Tonks, N. K. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 9052–9057.
14. Brookes, A. J., Slorach, E. M., Morrison, E., Qureshi, A. J., Blake, D., Davies, K. & Porteous, D. J. (1994) *Hum. Mol. Genet.* **11,** 2011–2017.
15. Wetmur, J. G. & Davidson, N. (1968) *J. Mol. Biol.* **31,** 349–370.

Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 4487–4492, April 1998
Genetics

Genetic analysis using genomic representations

ROBERT LUCITO*, MARIKO NAKIMURA*, JOSEPH A. WEST*, YING HAN*, KOEI CHIN†, KENDALL JENSEN*, RICHARD MCCOMBIE*, JOE W. GRAY†, AND MICHAEL WIGLER*‡

*Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724; and †Department of Laboratory Medicine, University of California, San Francisco, Box 0808 MCB 230, San Francisco, CA 94143-0808

Contributed by Michael Wigler, February 3, 1998

ABSTRACT **Analysis of the genetic changes in human tumors is often problematical because of the presence of normal stroma and the limited availability of pure tumor DNA. However, large amounts of highly reproducible "representations" of tumor and normal genomes can be made by PCR from nanogram amounts of restriction endonuclease cleaved DNA that has been ligated to oligonucleotide adaptors. We show here that representations are useful for many types of genetic analyses, including measuring relative gene copy number, loss of heterozygosity, and comparative genomic hybridization. Representations may be prepared even from sorted nuclei from fixed and archived tumor biopsies.**

Analysis of the genetic changes in human tumors is often problematical because of the presence of normal stroma. Although either microdissection or flow cytometry can produce small samples highly enriched for tumor cells or nuclei, the extracted DNA is of insufficient quantity for most uses. Nevertheless, we have successfully performed a complex protocol, representational difference analysis (RDA), on such small samples. RDA is a subtractive DNA hybridization technique that discovers the differences between paired normal and tumor genomes (1). The first step of RDA is the preparation of "representations," which are highly reproducible reformattings and amplifications of DNA populations. Typically, a representation is a set of restriction endonuclease fragments of a limited size range amplified by PCR. As much as 100 μg of DNA can be prepared from as little as 3 ng of DNA ($\approx 1 \times 10^3$ cells).

In RDA, a representation with much lower complexity than the starting population is needed to enable a subtractive hybridization step to proceed effectively. Such low complexity representations (LCRs) do not "capture" enough (typically, $\leq$7%) of the genome to be useful for many of the more common types of analyses. However, we demonstrate here that high complexity representations (HCRs) can provide ample amounts of DNA in a sufficiently reproducible manner suitable for most conventional studies. We demonstrate one type of HCR that captures $\approx$70% of the genome and illustrate its use for determining gene copy number, deletion mapping, loss of heterozygosity (LOH), and comparative genomic hybridization (CGH). HCRs may be a generally useful means of "immortalizing" and archiving DNA for later analysis from nonrenewable sources.

MATERIALS AND METHODS

Materials. Restriction endonucleases as well as T4 DNA ligase, T4 DNA polymerase, and T4 polynucleotide kinase were supplied by New England Biolabs. AmpliTaq was supplied by Perkin–Elmer. Oligonucleotide adaptors RBgl-12 and RBgl-24 [as used in Lisitsyn *et al.* (1)], oligonucleotides used for PCR of World-Wide Web probes, and tetranucleotide repeat D17S695 were synthesized by BioSynthesis (Lewisville, TX). Genescreen *Plus* was purchased from DuPont. Radioactive nucleotides, *redi*prime labeling kit, dNTPs, and hyperfilm were purchased from Amersham Pharmacia Biotech (Uppsala, Sweden). Cell lines used were obtained through American Type Culture Collection, grown in culture, and DNA-prepared. Normal human placenta DNA was obtained from CLONTECH. Oligonucleotides and probes for use with the Applied Biosystems 7700 Sequence Detector were synthesized by Applied Biosystems. Nusieve agarose gels used for analysis of PCR products were purchased from FMC. World-Wide Web probe sequences were downloaded from the Massachusetts Institute of Technology human genome sequencing database (http://www-genome.wi.mit.edu/).

Production of Representations. Genomic DNA (3–10 ng) was digested by the desired restriction endonucleases (*Dpn*II and *Bgl*II) under conditions suggested by the supplier. The digest was purified by phenol extraction and then precipitated in the presence of 10 μg of tRNA. The digested DNA was resuspended by the addition of 444 pmol of each adaptor (RBgl24 and RBgl12), T4 DNA ligase buffer (diluted to 1×, provided with enzyme), and water to bring the volume to 30 μl. The reaction was placed in a 55°C heat block, and the temperature was decreased slowly to 15°C by placing the heat block at 4°C (for $\approx$1 hr). On reaching 15°C, the RBgl24 adaptor was ligated by the addition of 400 units of T4 DNA ligase and by incubation at 15°C for 12–18 hr. The ligated material was divided into two tubes, and the following was added: 80 μl of 5× PCR buffer [335 mM Tris·HCl, pH 8.8/20 mM $MgCl_2$/80 mM $(NH_4)_2SO_4$/50 mM β-mercaptoethanol/0.5 mg/ml of BSA], 2′-deoxynucleoside 5′-triphosphates to a final concentration of 0.32 mM, RBgl24 adaptor to a final concentration of 0.6 μM, and H_2O to bring the volume to 400 μl. Each reaction was overlaid with 100 μl of mineral oil. The reaction was placed in a thermal cycler preheated at 72°C, and 15 units of AmpliTaq was added to the tubes. The thermal cycler was set to continue at 72°C for 5 min to allow for filling in the 3′ ends of the ligated molecules. This step was followed by 20 cycles lasting 1 min at 95°C and 3 min at 72°C, with an additional extension of 10 min at 72°C after the last cycle. The PCR was divided into two tubes (now a total of four tubes), and the following reagents were added: 40 μl 5× PCR buffer (as above), dNTP to a final concentration of 0.32 mM, RBgl24 adaptor to a final concentration of 0.6 μM, water to bring the volume to 400 μl, and 100 μl of mineral oil. The tubes then were amplified for an additional five cycles with extension according to the above conditions. The reactions were purified by phenol-chloroform and then precipitated by the addition of

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HCR, high complexity representation; LCR, low complexity representation; RDA, representational difference analysis; CGH, comparative genomic hybridization; LOH, loss of heterozygosity.

‡To whom reprint requests should be addressed at: Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724. e-mail: wigler@cshl.org.

1/10th the reaction volume of sodium acetate (3M pH 5.2) and by the addition of one reaction volume of isopropanol.

PTEN Tumor Suppressor Lori Genomic Sequencing. Bacterial artificial chromosome DNA was purified by detergent lysis and polyethylene glycol precipitation. DNA (5 μg) was sheared and then repaired by using T4 DNA polymerase. Fragments in the range of 1.5–2 kb were isolated by gel fractionation and inserted into M13 by ligation. Sequencing reactions were done by using dye primer chemistry that uses energy transfer primers (2) and thermosequenase polymerase (3). Reactions were run on an Applied Biosystems 377 sequencer for 6- to 8-hr sequence runs for analysis. Base calling was carried out by using the program PHRED (P. Green, University of Washington). The collection of initial data was assembled into a contig by using the automated assembly program PHRAP (P. Green). The database generated was converted to XGAP format and edited by using XGAP (4, 5). After completion of ≈6-fold coverage, the finishing process was carried out by using the program FINISH (G. Marth, Genome Sequencing Center at Washington University) to fill gaps in the contig.

Southern Blotting. Normal DNA from a tumor cell line or representations derived from sorted nuclei were digested with either *Dpn*II or *Bgl*II. Digested genomic DNA (5.0 μg), digested HCR (5.0 μg), or digested LCR (0.5 μg) was loaded on a 1.5% agarose gel. The gel was transferred to Genescreen *Plus* after electrophoresis was completed. After transfer, the blot was hybridized in GIBCO prehybridization solution at 67°C for 2 hr. The probe was produced by random priming with the *redi*prime kit in the presence of ^{32}PαdCTP. After 12–18 hr of hybridization, the blots were washed with standard saline phosphate/EDTA buffer as described (6). The blots were used for producing either auto-radiograms and/or phosphorimages.

LOH Analysis. LOH was assessed by using the tetranucleotide repeat marker D17S695 (listed as UT269 in the National Center for Biotechnology Information Web Site http://www.ncbi.nlm.nih.gov/Entrez/nucleotide.html), which maps to the 17p13 region of the human genome and has been used to determine the state of LOH at the p53 locus (7, 8). The reverse primer was labeled by phosphorylation in the presence of ^{32}PγATP as described (7, 8). The labeled reverse primer in combination with both forward and reverse primers was used for PCR by using 50 ng of HCR as DNA template. PCRs were performed as described in the National Center for Biotechnology Information Web Site, and gel electrophoresis was performed as described (7, 8).

CGH. CGH was performed according to standard procedure (9) by using HCR DNAs and genomic DNAs prepared from BT474 and MCF7 as target and normal female human lymphocyte as the reference. The genomic and HCR DNA samples were labeled by nick-translation with fluorescein isothiocyanate–12-dUTP and Texas Red 5-dUTP to produce DNA fragments ranging in size between 500 and 2,000 bp.

Quantitative PCR. Primer and probe sequences were produced according to the Applied Biosystems software PRIMER EXPRESS. These primers and probes were used in a PCR that was placed in the Applied Biosystems 7700 Sequence Detector and amplified as described (10, 11). Samples used were HCRs produced from DNA derived from nuclei of sorted primary tumor biopsies. After the PCR, the data of fluorescence corresponding to cycle number were downloaded from the Applied Biosystems 7700 Sequence Detector to Microsoft EXCEL and analyzed to give the graphs shown later.

RESULTS

The Basic Method. DNA from samples, usually paired tumor cells or nuclei and normal cells or nuclei, were processed in parallel to prepare representations. DNA was cleaved with a restriction endonuclease, such as *Dpn*II or *Bgl*II, that is not blocked by 5-methylcytosine, and double-stranded, cohesive, adaptor oligonucleotides were ligated to the fragment ends. The adaptors were not phosphorylated and therefore could not be self-ligated to form interfering dimers. Because adaptors can be ligated only to the 5′ ends of the cleavage fragments, the ligated product then was treated with DNA polymerase to fill in the 3′ ends, forming the primer binding site for the subsequent PCR. Amplification was performed in stages, each stage using the adaptor oligonucleotides as primers and proceeding for no more than 20 rounds per stage. PCR does not amplify all fragments equally well, and high molecular weight fragments in particular are very poorly amplified. Thus, restriction endonucleases that cleave infrequently were used to prepare LCRs, and enzymes that cleave frequently were used to prepare HCRs (see Fig. 1). From 3 ng of starting material, we could obtain 40 μg of total HCR from a total of 25 rounds. We have not tested the usefulness of an HCR beyond a total of 35 rounds.

Sampling the Complexity and Reproducibility of *Dpn*II Representations. We first sampled the reproducibility and complexity of *Dpn*II HCRs. We analyzed 14 different HCRs, each made from 5 ng of DNA prepared from diploid nuclei

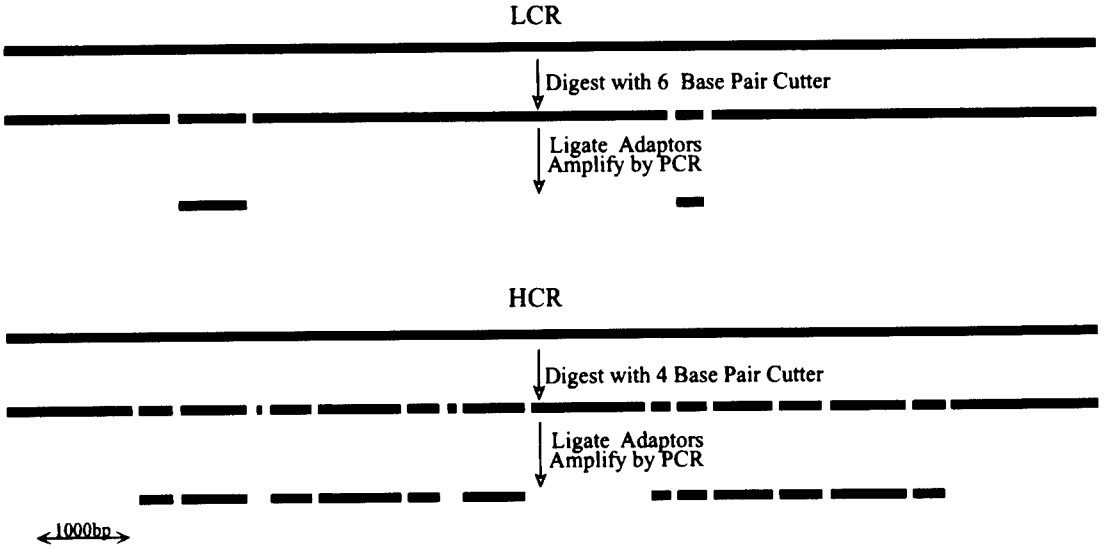


FIG. 1. Schematic comparison of LCR and HCR, illustrating the complexity reduction that occurs after cleavage with rare and frequent cutters, respectively.

separated from tumor biopsies by flow cytometry and each amplified by PCR for 25 rounds. In our first sampling, we designed pairs of PCR primers to detect Web Site sequence-tagged sites. We picked sequence-tagged sites that were not cleaved by *Dpn*II and used primer pairs that amplified a single band from total genomic DNA controls. Of these, 18 of 25 pairs (72%) were able to amplify the same molecular weight fragment from each HCR, and 7 generally failed to amplify from any HCR. Our results suggest that *Dpn*II HCRs reproducibly contain the same elements and ≈70% of the genome.

We performed a similar sampling with primer pairs derived from the locus encoding the PTEN tumor suppressor gene, for which we had the complete nucleotide sequence (unpublished data). In this way, we were able to use primers derived from *Dpn*II fragments of known size. *Dpn*II fragments were chosen at random, and PCR primer pairs were designed for each. Primer pairs (22 pairs) amplified single fragments by PCR from control genomic DNAs. These pairs were used with the same panel of 14 HCRs used above. Primer pairs (20 pairs) amplified the expected fragment from all HCRs, and 2 pairs failed to amplify from any. The fragments that were not in the HCRs were the largest, 3,916 bp, and one of the smallest, 97 bp. Totaling all fragment lengths, 16,039 bp were included in the HCRs, and 4,013 bp were excluded. Thus, assuming that our initial selection of *Dpn*II fragments was random, the HCRs contained ≈75% of the PTEN region.

If *Dpn*II cleavage is nearly complete during the preparation of an HCR, we expect that no PCR primer pairs should readily amplify from an HCR when the amplified sequence has an internal *Dpn*II site. To test this, we chose four primer pairs from the PTEN locus that amplified a single fragment containing a single internal *Dpn*II site. All four pairs amplified fragments from genomic DNA controls, and none amplified detectable fragments from the 14 HCRs. We conclude that HCRs prepared in parallel from samples processed in parallel are reasonably reproducible and represent ≈70% of the human genome.

Measuring Gene Copy Number in HCRs. Tumor genomes often contain either extra copies of sequences caused by gene amplification or missing sequences caused by gene deletion. To explore the usefulness of representations for measuring gene copy number, we first compared Southern blots of genomic DNA to blots of HCRs and LCRs. For this purpose, we used human placental DNA as normal and prepared genomic DNA from tumor cell lines amplified at c-erbB2 (BT-474) (12) or c-myc (SK-BR-3) (13). HCRs and LCRs were made from cell line or placental DNAs by using *Dpn*II or *Bgl*II, respectively. As probes we used small *Bgl*II fragments that we cloned from P1s containing inserts from the designated loci. The blots, shown in Fig. 2*A*, were quantitated by phosphorimaging. To normalize for loading differences, the blots were stripped and rehybridized with a single copy sequence probe. The normalized ratios of signal from tumor and normal are tabulated in Fig. 2*B*. The same relative copy number (tumor to normal) was determined from blots of representations as was determined from blots of genomic DNAs, indicating that there is a quantitatively reproducible amplification of these test sequences during the preparation of either HCRs or LCRs prepared in parallel from similar starting materials. Similar results were obtained for the cyclin D locus (data not shown).

To explore the usefulness of HCRs for deletion mapping, we blotted both genomic and HCR DNAs from tumor cell lines for deletion at the 20p11 locus. This locus was discovered initially by using RDA and subsequently was found to be deleted frequently in gastrointestinal cancers (R.L., unpublished data). Fig. 2*C* illustrates that the probe hybridized to sequences in the HCRs when and only when it hybridized to sequences in the respective genomic DNA. An unrelated probe detects sequences present in all genomic samples and their representations.

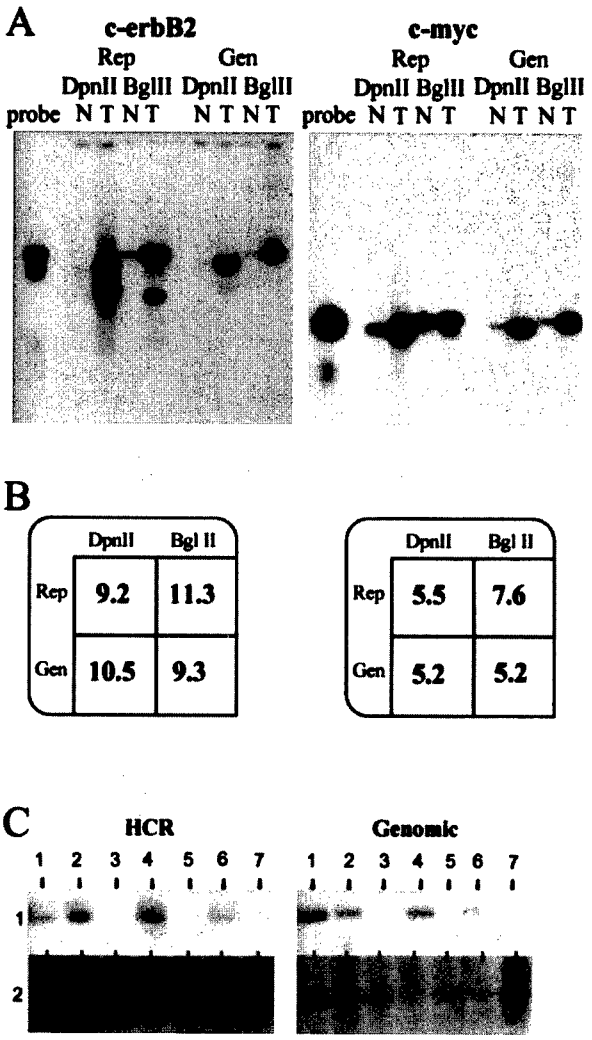


FIG. 2. Analysis of copy number using representations (Rep) and genomic DNA (Gen) for c-erbB2 and c-myc. (*A*) Southern blot comparing tumor cell lines (T) to normal (N) cell lines. *Dpn*II was used to prepare HCRs (*Dpn*II), and *Bgl*II was used to prepare LCRs (*Bgl*II). As a reference, free probe also was run alongside the samples (probe). The hybridization probes were derived from small *Bgl*II fragments isolated from P1 clones specific for each locus respectively. (*B*) Quantitation of the Southern blot. Rep, representation; Gen, genomic DNA; *Dpn*II, either *Dpn*II HCR or *Dpn*II digest of genomic DNA; *Bgl*II, either *Bgl*II HCR or *Bgl*II digest of genomic DNA. Blots in *A* were stripped and reprobed with a single copy probe to analyze loading differences (data not shown). Phosphorimage analysis of this probe was used to normalize the amplification differences. After this analysis, the resulting intensity of tumor was divided by the intensity of normal to give the fold amplification displayed. (*C*) Deletion mapping using HCRs and the deletion mapping of seven tumor cell lines comparing blots of HCRs (HCR) with blots of the corresponding genomic DNA (Genomic), each designated 1–7. The probe used for hybridization is from the human genomic region 20p11 (HCR). The Genomic panel shows the same DNAs blotted with an unrelated probe.

We tested the value of the HCRs made from limited amounts of DNA for quantitation of copy number. HCRs prepared from aneuploid and diploid nuclei sorted from several breast cancer biopsies were blotted for c-erbB2. Fig. 3*A* illustrates that c-erbB2 is amplified in the HCRs made from the aneuploid nuclei of some biopsy samples. Fig. 3*B* shows the blots with an unrelated probe. We obtained confirmation of the validity of the c-erbB2 amplifications by demonstrating that probes adjacent to but distinct from the c-erbB2 probe also were amplified in the same samples (M.N., unpublished results).

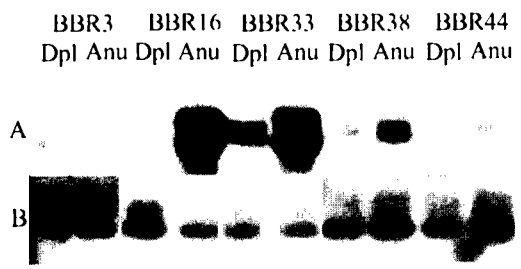


FIG. 3. Comparison of HCRs from primary tumor biopsies by Southern blotting. (*A*) HCRs from the designated tumor biopsies (denoted by an identification number preceded by BBR) were prepared from diploid (Dpl) and aneuploid (Anu) nuclei and compared by Southern blot analysis. The c-erbB2 probe was the same as that used in Fig. 2. (*B*) The same DNAs blotted with an unrelated probe.

Finally, we tested some of the same samples by quantitative PCR. For this purpose, we used the dual-labeled fluorogenic hybridization probes and the Applied Biosystems 7700 Sequence Detector (10, 11) to compare HCR DNAs prepared from aneuploid and diploid nuclei pairs. The results, shown in Fig. 4, indicate that differences in copy number were detected by probes for the c-erbB2 oncogene and the p16 tumor suppressor. On the other hand, no differences in copy number were detected by probes from an uninvolved region on chromosome 3. The curve of amplification of the c-erbB2 fragment arises four cycles sooner in one aneuploid HCR than it does in the paired diploid HCR, indicating a higher copy number for c-erbB2 (≈16-fold higher in the aneuploid HCR sample). The curve for the aneuploid HCR deleted for p16 arises three cycles later than the paired diploid HCR (approximately one-eighth as much in the aneuploid HCR), probably reflecting ≈10% contamination of the aneuploid nuclei with diploid nuclei after sorting. These results were confirmed by Southern blotting (data not shown). One tumor/normal pair showed a shift of a single cycle for primer pairs detecting the p16 gene, which might reflect loss of a single allele in the tumor or experimental error.

Detection of LOH in HCRs. LOH is a common lesion found in cancer cells and may be indicative of genomic instability and/or the loss of function of a specific tumor suppressor gene (14–17). The detection of LOH often is obscured by the presence of normal stroma; hence, we tested whether HCRs prepared from minute amounts of samples highly enriched for tumor nuclei could be used for LOH analysis. PCR primers that amplify microsatellites and detect fragment length polymorphisms are used frequently for LOH mapping, and we chose to examine a primer pair that amplifies a highly polymorphic tetranucleotide repeat near the p53 locus (7, 8).

In preliminary experiments, we established that these PCR primers detected the same allele pattern in both genomic and HCR DNAs prepared from cell lines. Next, we examined 12 pairs of HCRs prepared from aneuploid and diploid nuclei. LOH at this locus was detected clearly in 9 of 10 informative pairs (see Fig. 5 for representative cases), which is greater than the reported proportion of LOH at this locus in breast cancer (50%) (14, 18, 19) but may reflect the selection of the highly aneuploid tumors that we sorted and/or the purity of our samples.

Comparative Genome Hybridization with HCRs. CGH is a powerful tool for analyzing the global genomic changes of tumors (20–25). It has been reported that chromosome paint-

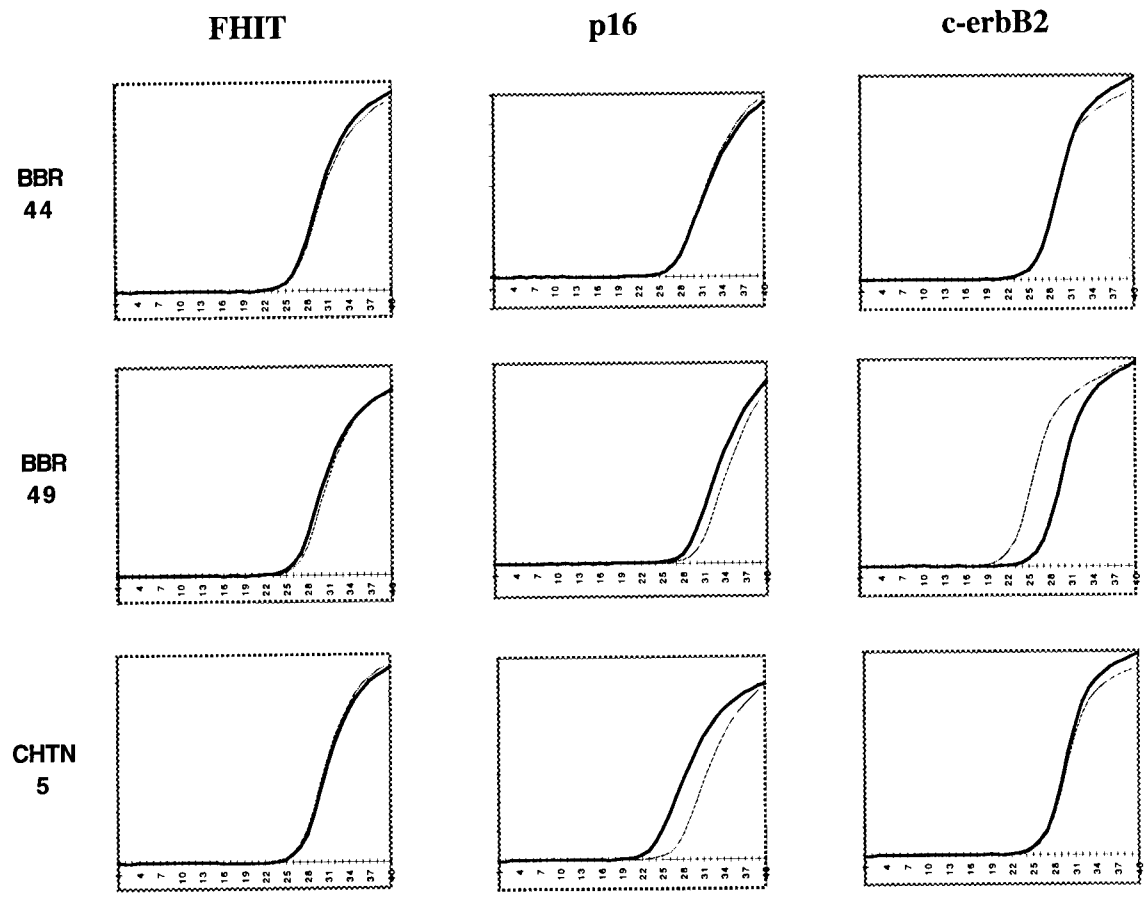


FIG. 4. Quanitative PCR analysis of HCRs. Diploid (black) and aneuploid (gray) HCRs derived from sorted primary tumor biopsies were used as template for quanitative PCR analysis. Probes from several genomic loci (FHIT, p16, and c-erbB2) were used to determine copy number in three primary tumor pair HCRs (BBR44, BBR49, and CHTN5). The data from the Applied Biosystems 7700 Sequence Detector was analyzed with Microsoft EXCEL to produce the graphs shown. (*x* axis, the cycle number during the reaction; *y* axis, the fluorescence detected.)

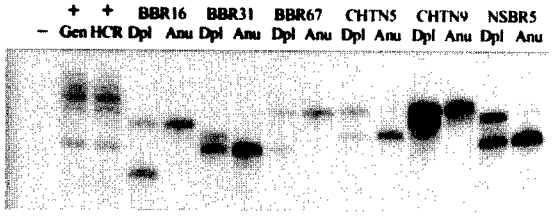


FIG. 5. Using HCRs for LOH analysis. LOH analysis was carried out on HCRs derived from sorted primary tumor biopsies. Dpl, diploid HCR; Anu, aneuploid HCR. The tumors are labeled as either BBR, CHTN, or NSBR followed by an identification number. The primers used in the reaction amplify a tetranucleotide repeat within the p53 locus. A mixed population of normal DNAs was used as a positive control. (+Gen, normal genomic DNA used as template; +HCR, HCR produced from this normal DNA used as template. –, reaction in which no template was added.)

ing could be performed with representations produced from each chromosome (26). We therefore tested whether CGH could be performed with HCRs. For this experiment, we chose to examine tumor cell lines so that we could directly compare CGH performed with genomic DNA with CGH performed with HCR. Little difference between HCR and genomic DNA could be discerned with either of the two cell lines examined, BT-474 and MCF7. Fig. 6 shows a sample of the chromosomal scanning profiles obtained with each DNA source.

DISCUSSION

The idea of capturing essential features of the genome as a PCR product is not new. In 1989, Kinzler and Vogelstein (27) described "whole genome" PCRs to select for DNA sequences that were binding sites for DNA binding proteins. Their method was used in 1995 (28) for the same purpose. In the experiments from 1989, PCR adaptors were blunt-end ligated to the cleavage fragments of total genomic DNA. The usefulness of the resulting PCR products for genomic analysis was not explored; nor was its efficiency explored when starting with tiny amounts of material. An alternate approach to whole genome PCR is to use random priming with degenerate oligonucleotides (29–33). This method can produce large amounts of DNA starting from very minute amounts of sample, but the complexity of the DNA produced, the reproducibility of the representation, and the loss of natural restriction sites at the ends of the amplified material make the usefulness of this method somewhat limited. Restriction endonuclease-based representations have major advantages: Their complexity can be regulated by the choice of restriction enzyme; they can be readily reamplified; they can be analyzed by Southern blotting; and they are highly reproducible.

In this report, we have explored the uses of HCRs, including quantitative assessment of copy number, LOH, and CGH. All of these analytical methods require a high level of reproducibility in the representation. To achieve this level of reproducibility, we have prepared paired samples of HCRs from the same amount of starting material, used genomic DNAs extracted in the same manner, and performed PCR at the same time, under the same conditions, and in the same thermal cycler.

In principle, HCRs can be prepared from normal and tumor tissue stored as fixed, paraffin-embedded, archived biopsies, which would extend greatly the usefulness of such samples. We have made HCRs from DNAs extracted from pairs of aneuploid and diploid nuclei sorted from such sources (R.L., unpublished data). More rounds of PCR are required to obtain workable amounts of DNA, and HCRs from DNA extracted from fixed specimens have a markedly lower size distribution than HCRs prepared from fresh sources. Moreover, there is enormous variability between the HCRs from different specimens, reflected as varying size distributions. This diversity is probably caused by the variation in the quality of the DNA that can be extracted from specimens fixed and stored under different conditions (19, 34–36). Despite the variability between specimens, we found that the HCRs prepared from aneuploid and diploid nuclei from the same fixed specimen are similar to each other. Indeed, we have found that HCRs prepared from sorted nuclei of fixed specimens are useful for CGH (unpublished observations). These results suggest that normal and tumor HCRs will be useful even when prepared for analysis from microdissected specimens.

Both HCR and LCR render paired tumor and normal DNAs from minute specimens in a stable format that can be analyzed or further amplified at a later date. We have emphasized the usefulness of HCRs for gene copy, LOH, and global genomic analysis of tumor specimens. Most probes, chosen at random, will be present in an HCR but will not be present in the LCRs. However, precisely because LCRs are of lower complexity,

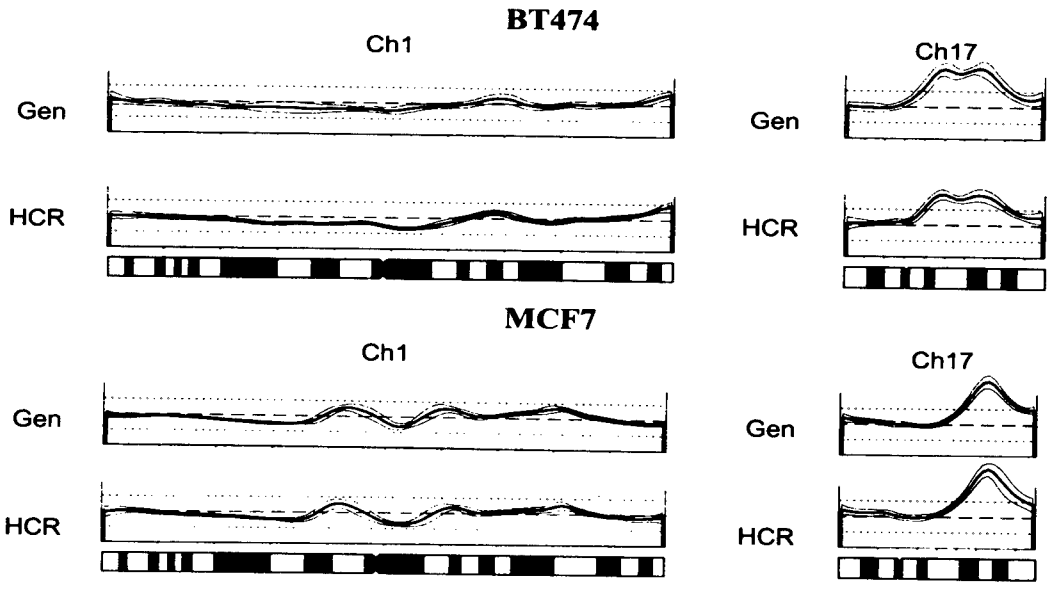


FIG. 6. CGH analysis using HCRs. Shown are two representative chromosome spreads (Ch 1, and Ch17) comparing the genomic (Gen) with the HCR for two different cell lines, BT474, and MCF7. Deviation from the dashed line represents genomic change, a peak represents gain in copy number, and a trough represents loss of copy number. Underneath the profiles is an idiogram of each chromosome as a reference.

hybridization-based assays that depend on completeness of hybridization are easier to perform. This ease is apparent in blotting analysis of LCRs. For the same reason, some global genomic analyses, such as microchip array analysis (37–40) or CGH (20–25), that depend on hybridization kinetics should be facilitated by the use of LCRs because reannealing times should be reduced and signal-to-noise ratios enhanced.

We thank Masaaki Hamaguchi, Eli Hatchwell, and Clifford Yen for useful discussions and probe sequences; Bert Vogelstein, Manuel Perucho, and Nikolai Lisitsyn for critical reading of the manuscript; Linda Rodgers and Mike Riggs for technical assistance; Jim Duffy for artwork, and Patricia Bird for secretarial assistance. This work was supported by grants to M.W. from the National Institutes of Health (OIG-CA39829 and 5P50-CA-68425); the United States Army (DAMD17-94-J-4247); 1 in 9: The Long Island Breast Cancer Action Coalition; and Tularik. M.W. is an American Cancer Society Research Professor.

1. Lisitsyn, N., Lisitsyn, N. & Wigler, M. (1993) *Science* **259,** 946–951.
2. Ju, J., Ruan, C., Fuller, C. W., Galazer, A. N. & Mathies, R. A. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 4347–4351.
3. Tabor, S. & Richardson, C. C. (1995) *Proc. Natl. Acad. Sci. USA* **93,** 6339–6343.
4. Dear, S. & Staden, R. (1991) *Nucleic Acids Res.* **19,** 3907–3911.
5. Gleeson, T. J. & Staden, R. (1991) *Comput. Appl. Biosci.* **7,** 398.
6. Sambrook, J., Fritsch, E. F. & Maniatis, T (1989) in *Molecular Cloning: A Laboratory Manual*, eds. Ford, N., Nolan, C. & Ferguson, M. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 9.38–9.56.
7. Gerken, S. C., Albertsen, H., Elsner, T., Ballard, L., Holik, P., Lawrence, E., Moore, M., Zhao, X. & White, R. (1995) *Am. J. Hum. Genet.* **56,** 484–499.
8. Stack, M., Jones, D., White, G., Liscia, D. S., Venesio, T., Casey, G., Crichton, D., Varley, J., Mitchell, E., Heighway, J., *et al.* (1995) *Hum. Mol. Genet.* **4,** 2047–2055.
9. DeVries, S., Gray, J. W., Pinkel, D., Waldman, F. M. & Sudar, D. (1995) in *Current Protocols in Human Genetics,* ed. Boyle, A. L. (Wiley, New York), Suppl. 6, Unit 4.6, pp. 1–18.
10. Gibson, U. E., Heid, C. A. & Williams, P. M. (1996) *Genome Res.* **6,** 995–1001.
11. Isono, K. (1997) *Rinsho Byori* **45,** 218–223.
12. Szollosi, J., Balazs, M., Feuerstein, B. G., Benz, C. C. & Waldman, F. M. (1995) *Cancer Res.* **55,** 5400–5407.
13. Feo, S., Di, L. C., Jones, T., Read, M. & Fried, M. (1994) *Oncogene* **9,** 955–961.
14. Deng, G., Chen, L. C., Schott, D. R., Thor, A., Bhargava, V., Ljung, B. M., Chew, K. & Smith, H. S. (1994) *Cancer Res.* **54,** 499–505.
15. Patel, U., Chen, H. C. & Banerjee, S. (1994) *Cell. Mol. Biol. Res.* **40,** 683–691.
16. Wada, C., Shionoya, S., Fujino, Y., Tokuhiro, H., Akahoshi, T., Uchida, T. & Ohtani, H. (1994) *Blood* **83,** 3449–3456.
17. Wieland, I., Ammermuller, T., Bohm, M., Totzeck, B. & Rajewsky, M. F. (1996) *Oncol. Res.* **8,** 1–5.
18. Chen, L. C., Neubauer, A., Kurisu, W., Waldman, F. M., Ljung, B. M., Goodson, W., III, Goldman, E. S., Moore, D., II, Balazs, M., Liu, E., *et al.* (1991) *Proc. Natl. Acad. Sci. USA* **88,** 3847–3851.
19. Chen, Y. H., Li, C. D., Yap, E. P. & McGee, J. O. (1995) *J. Pathol.* **177,** 129–134.
20. Bryndorf, T., Kirchhoff, M., Rose, H., Maahr, J., Gerdes, T., Karhu, R., Kallioniemi, A., Christensen, B., Lundsteen, C. & Philip, J. (1995) *Am. J. Hum. Genet.* **57,** 1211–1220.
21. Carter, N. P., Ferguson, S. M., Perryman, M. T., Telenius, H., Pelmear, A. H., Leversha, M. A., Glancy, M. T., Wood, S. L., Cook, K., Dyson, H. M., *et al.* (1992) *J. Med. Genet.* **29,** 299–307.
22. Cher, M. L., MacGrogan, D., Bookstein, R., Brown, J. A., Jenkins, R. B. & Jensen, R. H. (1994) *Genes Chromosomes Cancer* **11,** 153–162.
23. Hermsen, M. A., Meijer, G. A., Baak, J. P., Joenje, H. & Walboomers, J. J. (1996) *Hum. Pathol.* **27,** 342–349.
24. Kallioniemi, O. P., Kallioniemi, A., Piper, J., Isola, J., Waldman, F. M., Gray, J. W. & Pinkel, D. (1994) *Genes Chromosomes Cancer* **10,** 231–243.
25. Storkel, S., Simon, R., Brinkschmidt, C., Gronwald, J. & Bocker, W. (1996) *Pathologe* **17,** 189–194.
26. Vooijs, M., Yu, Y.-C., Johnson, D., Tkachuk, D., Pinkel, D. & Gray, J. (1993) *Am. J. Hum. Genet.* **52,** 586–597.
27. Kinzler, K. W. & Vogelstein, B. (1989) *Nucleic Acids Res.* **17,** 3645–3653.
28. Joulin, V. & Richard, F. H. (1995) *Eur. J. Biochem.* **232,** 620–626.
29. Telenius, H., Carter, N. P., Bebb, C. E., Nordenskjold, M., Ponder, B. A. & Tunnacliffe, A. (1992) *Genomics* **13,** 718–725.
30. Xu, K., Tang, Y., Grifo, J. A., Rosenwaks, Z. & Cohen, J. (1993) *Hum. Reprod.* **8,** 2206–2210.
31. Kristjansson, K., Chong, S. S., Van den Veyver, I. B., Subramanian, S., Snabes, M. C. & Hughes, M. R. (1994) *Nat. Genet.* **6,** 19–23.
32. Sun, F., Arnheim, N. & Waterman, M. S. (1995) *Nucleic Acids Res.* **23,** 3034–3040.
33. Xiao, Y., Slijepcevic, P., Arkesteijn, G., Darroudi, F. & Natarajan, A. T. (1996) *Cytogenet. Cell Genet.* **75,** 57–62.
34. Ben, E. J., Johnson, D. A., Rossi, J., Cook, N. & Wu, A. (1991) *J. Histochem. Cytochem.* **39,** 351–354.
35. Greer, C. E., Lund, J. K. & Manos, M. M. (1991) *PCR Methods Appl.* **1,** 46–50.
36. Greer, C. E., Peterson, S. L., Kiviat, N. B. & Manos, M. M. (1991) *Am. J. Clin. Pathol.* **95,** 117–124.
37. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. (1995) *Science* **270,** 467–470.
38. Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O. & Davis, R. W. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 10614–10619.
39. Schena, M. (1996) *Bioessays* **18,** 427–431.
40. Shalon, D., Smith, S. J. & Brown, P. O. (1996) *Genome Res.* **6,** 639–645.

Revised: September 11, 1998

The lipid phosphatase activity of PTEN is critical for its tumor supressor function

Michael P. Myers, Ian Pass*, Ian H. Batty*, Jeroen Van der Kaay*, Javor P. Stolarov, Brian A. Hemmings†, Michael H. Wigler, C. Peter Downes* and Nicholas K. Tonks‡.

Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. *Department of Biochemistry, Medical Sciences Institute, University of Dundee, Scotland. †Friedrich Meischer Institute, Basel, Switzerland.

‡Corresponding author: Nicholas K. Tonks
Cold Spring Harbor Laboratory
1 Bungtown Road
Cold Spring Harbor, NY 11724-2208
Tel: (516)367-8846
FAX: (516)367-6812
e-mail: tonks@cshl.org

Abstract

Since their discovery, protein tyrosine phosphatases have been speculated to play a role in tumor suppression due to their ability to antagonize the growth promoting protein tyrosine kinases. Recently, a tumor suppressor from human chromosome 10q23, called PTEN or MMAC1, has been identified that shares homology with the protein tyrosine phosphatase family. Germline mutations in PTEN give rise to several related neoplastic disorders, including Cowden disease. A key step in understanding the function of PTEN as a tumor suppressor is to identify its physiological substrates. Here we report that a missense mutation in PTEN, PTEN-G129E, which is observed in two Cowden disease kindreds, specifically ablates the ability of PTEN to recognize inositol phospholipids as a substrate, suggesting that loss of the lipid phosphatase activity is responsible for the etiology of the disease. Furthermore, expression of wild type or substrate trapping forms of PTEN in HEK293 cells altered the levels of the phospholipid products of phosphatidylinositol 3-kinase and ectopic expression of the phosphatase in PTEN-deficient tumor cell lines resulted in the inhibition of PKB/Akt and regulation of cell survival.

Introduction

Glioblastoma is one of the most common and malignant forms of cancer. It is often characterized by the constitutive activation of EGF-dependent signalling pathways due to the amplification of members of the EGF receptor family of protein tyrosine kinases (PTKs). The products of tumor suppressor genes may attenuate these signalling pathways and, therefore, their loss through deletion or mutation may contribute to tumor progression. The freqeunt loss of heterozygosity at tumor suppressor loci is often used to identify tumor suppressor genes. The 10q23 region of human chromosome 10 is frequently deleted or mutated in a wide variety of tumor types, most frequently in glioblastoma, endometrial cancer and prostate cancer, indicating the presence of a tumor suppressor gene at this locus. PTEN was subsequently identified as this tumor suppressor and was found to contain the catalytic signature motif detected in all members of the protein tyrosine phosphatase (PTP) family (1, 2). Importantly, PTEN appears to be lost frequently in advanced cancers suggesting that its deletion may not be the transforming event but that PTEN may inhibit other cellular functions necessary for tumor progression (3, 4).

Importantly, germline transmission of mutations in PTEN was shown to give rise to a related set of disorders, including Cowden disease, that are characterized by numerous small benign tumors and an increased incidence of other malignant growths (5-7). The detection of germline mutations in these neoplastic disorders verifies that PTEN is the tumor suppressor residing on

human chromosome 10q23. This conclusion is supported by the recent findings that PTEN knockout mice develop tumors similar to those found in Cowden disease (8).

A key step in establishing the cellular function of PTEN is the identification of its physiological substrates. Recently, two groups have identified different candidate substrates for PTEN. Tamura et al report that overexpression of PTEN results in changes in cell adhesion and spreading and suggest that these effects result from the the dephosphorylation of the protein tyrosine kinase FAK (9). In contrast, Maehama and Dixon have shown that PTEN recognizes phosphatidylinositol phosphate as a substrate (10). Our initial studies demonstrating catalytic activity in PTEN indicated that it prefered highly acidic or multiply phosphorylated substrates (11). This may explain why, in certain contexts PTEN will dephosphorylate poly-phosphorylated molecules such as FAK or phosphatidylinositol phosphate (11, 12). Nevertheless, the important question remains which multiply phosphorylated molecules, protein or otherwise, are the physiologically relevant targets of PTEN.

Significantly, we have demonstrated previously that the majority of missense mutations isolated from tumor and Cowden disease samples ablate the protein phosphatase activity of PTEN (11). One exception is a missense mutation that changes a glycine residue in the catalytic signature motif to a glutamate (PTEN-G129E) (11). Although there was no adverse effect on the protein phosphatase activity of PTEN, this G129E mutation was isolated from

two, independent Cowden disease kindreds, indicating that it abolishes the tumor suppressor activity of PTEN (5, 12). Therefore, this mutation can be used as an important indicator to determine whether a proposed function of PTEN is specific for its role as a tumor suppressor.

We analyzed the PTEN-G129E Cowden disease mutation and found that this mutation specifically inhibits the recognition of phosphatidylinositol phosphates (PtdInsP) by PTEN. Additionally, we show that expression of PTEN in mammalian cells results in changes in PtdIns(3,4,5)P_3 levels and expression of PTEN in two independent glioblastoma cell lines results in the disruption of signalling downstream of PI 3-kinase to PKB/Akt and Bad. Significantly, we also show that expression of PTEN in LnCaP cells, a prostate tumor cell line, abrogates cell survival and that this effect is inhibited by the expression of a constituitively activated form of PKB. Therefore, we believe that the physiological function of PTEN is to antagonize signalling downstream of PI 3-kinase by dephosphorylating phosphatidylinositol phosphates.

Methods

PTEN Phosphatase Assays. Recombinant wild type and mutant forms of PTEN were expressed in *E. coli* and purified by glutathione-affinity chromatography (11). The purified proteins were assayed with polyGluTyr as described (11). Release of $^{32}P_i$ from radiolabeled phosphatidylinositol phosphates was determined by performing a modified Bligh and Dyer extraction (13). The upper phase (containing inorganic phosphate) was removed, dried down and resuspended in a 1M TCA, 1% ammonium molybdate solution. Following extraction with 2 volumes of toluene:isobutylalchohol (1:1) the upper phase was removed and counted. Site selectivity was determined by incubating recombinant PTEN or SHIP (a gift from C. Erneux, Free University, Brussels) with radiolabeled PtdIns(3,4,5)P_3 and the lipid products of these reactions were analyzed by thin layer chromatography or HPLC (13). The phospholipids were extracted by performing the modified Bligh and Dyer extraction described and the resulting lower phases were dried down, resuspended in 20 µl of chloroform:methanol (2:1) and applied to an oxalate-activated silica 60 TLC plate. Plates were developed in methanol/chloroform/water/ammonia (100:75:25:15).

Determination of PtdIns(3,4,5)P_3 levels. HEK293 cells were transfected via calcium phosphate co-precipitation with 20 µg of DNA per 10 cm dish. The calcium phosphate:DNA co-precipitate was removed by washing with PBS 16 hours after addition and the cells were returned to growth medium for 36

hours before harvesting. Co-transfection of PTEN and p110 constructs were performed using 9.5 μg of each plasmid DNA. Transfection efficiency was determined by including a GFP expression plasmid (1 μg) in all transfections and was shown to be 80%. Phospholipids were extracted from the cells exactly as described (14). PtdIns(3,4,5)P_3 levels were assayed using a ligand displacement assay (14) and were normalized to total protein.

Expression of PTEN in Glioblastoma Cell lines. PTEN retroviral expression vectors were constructed in pBabePuro (15). Following transfection into packaging lines, the viral supernatants were harvested, diluted with growth media and incubated with U87MG or U373MG cells for 8 hours at 32^0 C in the presence of polybrene. Infected cells were selected with puromycin (2 μg/ml) and drug resistant colonies were expanded to generate clonal cell lines. Cells lysates and immunoblots were performed essentially as described (16). Protein levels were determined by the method of Bradford using BSA as a standard and equal protein was loaded in each lane. Antibodies to PTEN were generated in rabbits using a C-terminal peptide of PTEN (ENEPFDEDQHTQITKV) conjugated to KLH. Antibodies to PKB/AKT, phospho-PKB/Akt (specific to phosphorylated Ser473) and phospho-BAD were purchased from New England Biolabs.

Expression of PTEN in LnCaP Cells. Cells were transfected using cationic lipids (Transfast, Promega) at a DNA:lipid ratio of 1. Cells were incubated with the

DNA:lipid complexes for 12 hours and then transfected cells were identified based on the expression of a co-transfected GFP vector or by performing immunoflourescence using antibodies to the HA-epitope tag located at the N-terminus of PTEN, essentially as described (17).

Results

Activity of PTEN towards inositol phospholipids. Recombinant PTEN, produced in *E. coli* (11), was assayed for its ability to release $^{32}P_i$ from radiolabeled PtdIns(3,4,5)P_3 or polyGluTyr (Figure 1A). As expected wild type PTEN catalyzed the dephosphorylation of both substrates (Figure 1A). Remarkably, the activity of PTEN-G129E toward PtdIns(3,4,5)P_3 was reduced by ~90% relative to the wild type enzyme while still retaining activity towards polyGluTyr (Figure 1A). A catalytically-inactive mutant of PTEN (PTENC124S) was unable to dephosphorylate polyGluTyr or PtdIns(3,4,5)P_3 indicating that the phosphatase activity was not due to a bacterial contaminant. In addition, cdc14, a dual specificity phosphatase closely related to PTEN, was also unable to dephosphorylate PtdIns(3,4,5)P_3, demonstrating that recognition of this phospholipid substrate is not a general property of other, even closely related, dual specificity phosphatases. These data, demonstrating that the PTEN-G129E mutation specifically ablates the activity of PTEN towards phosphorylated inositol lipids, demonstrates that the lipid phosphatase, rather than the protein phosphatase, activity of PTEN is required for it to function as tumor suppressor and that loss of the lipid phosphatase activity results in Cowden disease.

Site Specificity of PTEN within PtdIns(3,4,5)P_3. In order to determine whether PTEN recognizes specific sites in the inositol ring of PtdIns(3,4,5)P_3, lipid substrate, labeled exclusively in the 3-position with ^{32}P, was incubated with

PTEN or SHIP, a well characterized 5-phosphatase (18). The products of these reactions were analyzed by thin layer chromatography. Incubation with SHIP yielded a radiolabelled product with the expected mobility of PtdIns(3,4)P_2 (Figure 1B). However, no labeled lipid products were generated following treatment with PTEN, under conditions where >50% of the ^{32}P was lost from the substrate (Figure 1B). The site selectivity of PTEN was confirmed by HPLC analysis of similar reactions using [^{3}H]- and [3-^{32}P]-labeled Ins(1,3,4,5)P_4 in which the only products detected were $^{32}P_i$ and [^{3}H]-Ins(1,4,5)P_3 (data not shown). In addition to PtdIns(3,4,5)P_3, PTEN also hydrolyzed the other potential products of PI 3-kinase, PtdIns(3)P and PtdIns(3,4)P_2, with the following rank order based on relative percent hydrolysis: PtdIns(3,4,5)P_3 = PtdIns(3,4)P_2 > PtdIns3P > Ins(1,3,4,5)P_4. The specificity of PTEN for 3-phosphorylated inositol lipids indicates that it may function as a negative regulator of PI 3-kinase-mediated signalling and distinguishes PTEN from other known lipid phosphatases.

PTEN Inhibits PI 3-kinase-dependent Signaling. The potential for PTEN to antagonize PI 3-kinase signalling was further investigated in mammalian cells (Figure 2). Transient expression of PTEN in HEK293 cells lowered the levels of PtdIns(3,4,5)P_3 (Figure 2) (from 1 pmol/mg to 0.7 pmol/mg). Importantly, PTEN-C124S, in which the catalytic cysteine has been mutated to serine, resulted in an increase in the levels of PtdIns(3,4,5)P_3 (Figure 2) (from 1 pmol/mg to 2 pmol/mg). The accumulation of PtdIns(3,4,5)P_3 in response to expression of PTEN-C124S is likely due to the ability of this form of PTEN to

behave as a "substrate trapping" mutant, resulting in a stable complex with the lipid substrate which protects it from dephosphorylation by endogenous phosphatases (19). The accumulation of PtdIns(3,4,5)P_3 in the presence of a substrate trapping mutant confirms that PtdIns(3,4,5)P_3 is a physiological target of PTEN (19, 20). Expression of a constitutively-activated PI 3-kinase (p110-CAAX) resulted in elevated levels of PtdIns(3,4,5)P_3 that were reduced ~60% following co-expression with PTEN and were increased ~4.5 fold when co-expressed with PTEN-C124S (Figure 2). The ability of PTEN to decrease PtdIns(3,4,5)P_3 levels produced by p110-CAAX further indicates that this phosphatase exerts its effects on the products of PI 3-kinase rather than by dephosphorylating the tyrosine residues responsible for recruiting the p85-p110 PI 3-kinase complex to the membrane.

Importantly, we detected defects in PI 3-kinase signalling in tumor cell lines that have lost their endogenous PTEN genes due to deletion or mutation. Signalling downstream of PI 3-kinase was assessed in two of these PTEN-deficient cell lines, U87 MG and U373 MG, by determining the phosphorylation status of PKB/Akt, a kinase whose activation is dependent on PtdIns(3,4,5)P_3 or PtdIns(3,4,)P_2, two products of PI 3-kinase (21). Clonal cell lines were generated in U87 MG and U373 MG in which PTEN expression was restored by infection with recombinant retroviruses (Figure 3). The parental cell lines exhibited high basal levels of phosphorylated PKB/Akt, whereas in every wild type PTEN-expressing clonal cell line the basal levels of activated PKB/Akt were significantly lowered (Figure 3). Expression of a catalytically-inactive mutant of

PTEN, PTEN-R130M, had no effect on basal PKB/Akt activation (Figure 3). Similarly, no effect on PKB/Akt activation was observed in a U373 MG (PTEN-SRVE #B9) clone infected with an unstable mutant of PTEN (Figure 3). Following reconstitution of PTEN expression in the tumor lines, activation of PKB/Akt by insulin or PDGF was unaffected (Figure 3), suggesting that there are regulatory mechanisms which allow for the generation of PtdIns(3,4,5)P_3 following growth factor stimulation. The activation of PKB/Akt was completely inhibited by wortmannin in the tumor cell lines (Figure 3) indicating that the increase in PKB/Akt phosphorylation was PI 3-kinase-dependent.

Regulation of Cell Survival. One of the downstream substrates of PKB/Akt is the death effector protein BAD (22), phosphorylation of which has been linked to the promotion of cell survival (23). Although expression of PTEN in both U87 MG and U373 MG resulted in the reduction of BAD phosphorylation (Figure 3C), we did not detect evidence of apoptosis in the PTEN-expressing clones. However, expression of wild type PTEN in LnCaP cells, a PTEN-deficient prostate cancer cell line, resulted in a decrease in the number of PTEN positive cells recovered relative to controls expressing the catalytically-inactive mutant PTEN-C124S (Figure 4). In this assay, the Cowden disease mutation PTEN-G129E behaved similarly to PTEN-C124S, consistent with the ablation of the tumor suppressor activity of PTEN by this mutation (Figure 4). Significantly, coexpression of a constitutively-active, membrane-targeted

PKB/Akt (24) completely reverted the PTEN phenotype, in that PTEN positive cells could now be recovered despite the presence of wild type PTEN (Figure 4).

Discussion

PTEN is a newly defined tumor suppressor gene that has been implicated in a wide variety of cancers and in a series of related disorders that are characterized by a predispostion to cancer(1, 2, 25-28). Although these studies have offered many insights into the biology of PTEN, they do not address the molecular mechanism of PTEN function which requires the identification of its physiological substrates. However, thishas become an area of controversy. Tamura et al have suggested that PTEN disrupts cell spreading and migration by dephosphorylating FAK (9). In contrast, Maehema and Dixon have shown that, in certain contexts, PTEN can dephosphorylate the lipid second messenger phosphatidylinositol phosphate (10). However, both of these studies fall short of demonstrating that the ability of PTEN to dephosphorylate FAK or phosphatidylinositol phosphate is related to its tumor suppressive function. In light of the facts that previous studes have shown that PTEN prefers highly acidic or multiply phosphorylated substrates (11) and that FAK and phosphatidylinositol phosphate share these properties, one must view these results with caution as they may result from non-specific interactions between PTEN and highly charged molecules (12).

Significantly, we have been able to show that a single point mutation in PTEN, PTEN-G129E, specifically ablates the ability of PTEN to recognize phospholipids as substrates. In contrast, Tamura et al found that the PTEN-G129E point mutation behaves essentially like wild type PTEN in their assays of cell spreading (9). Even though this mutation does not alter the ability of PTEN

to recognize proteinaceous substrates, the fact that it has been identified in two independent Cowden disease kindreds indicates that this mutation would be expected to impair the tumor suppressor activity of PTEN (5, 11). Therefore, these findings demonstrate that the lipid phosphatase activity rather than the protein phosphatase activity of PTEN is important for its tumor suppressive activity and that FAK, at least in this respect, is not a relavent target of the phosphatase.

Characterization of the lipid phosphatase activity of PTEN demonstrated that it shows specificity for phosphatidylinositols phosphorylated at the 3-position. This represents the first phosphatidylinositol phosphatase with specificity for the lipid products of PI 3-kinase and suggests that PTEN may function to antagonize the growth promoting signals generated by PI 3-kinase. Indeed, overexpression of PTEN in mammalian cells disrupted the PI 3-kinase-dependent production of PtdIns(3,4,5)P_3. Furthermore, expression of a catalytically-inactive mutant of PTEN, PTEN-C124S, which may function as a "substrate trap", results in the accumulation of PtdIns(3,4,5)P_3, indicating that PTEN may function *in vivo* to antagonize PI 3-kinase-dependent signalling.

Importantly, signalling downstream of PI 3-kinase is altered in two different glioblastoma cell lines, U87 MG and U373 MG, which have lost expression of PTEN due to mutation or deletion of both alleles. Both cell lines exhibited elevated levels of active, phosphorylated PKB/Akt, a serine/threonine kinase whose activation requires the production of PtdIns(3,4,5)P_3 or PtdIns(3,4)P_2 (21, 29). Restoration of PTEN expression by

retroviral transduction resulted in the reduction of activated PKB/Akt as is evident from the observed reduction in the phosphorylation of PKB/Akt and in reduced levels of phosphorylated BAD, an apoptosis regulator which is a known *in vivo* substrate of PKB/Akt (22). Significantly, insulin or PDGF stimulation of these cell lines resulted in the stimulation of PKB/Akt regardless of whether PTEN expression had been reconstituted, suggesting that there are mechanisms in place which, in response to growth factor stimulation, serve to regulate the activity of PTEN and allow the production of PtdIns(3,4,5)P_3. This is especially important given that PTEN is widely expressed, including cell lines which produce PtdIns(3,4,5)P_3 in response to insulin.

Although the reconstitution of PTEN decreased the levels of phosphorylated BAD, we were unable to detect evidence of apoptosis in these cell lines. PKB/AKT is thought to affect cell survival in many cell types by phosphorylating BAD or inactivating GSK3β (23, 33). Reconstitution of PTEN in LnCaP cells, a PTEN-deficient prostate cancer cell line, resulted in the inhibition of cell survival. Importantly, expression of the PTEN-G129E Cowden disease point mutation does not result in the abrogation of cell survival. This indicates that the lipid phosphatase activity of PTEN is required for these effects on cell survival. The properties of the PTEN-G129E mutant emphasize the requirement of the lipid phosphatase activity for the tumor suppressor function of PTEN. Furthermore, the PTEN-dependent induction of apoptosis was completely inhibited by co-expressing a constitutively-active,

membrane-targeted form of PKB/Akt indicating that PTEN, by dephosphorylating phosphatidylinositides, is an upstream regulator of PKB/Akt and, at least in LNCAP cells, serves to regulate survival signals. Thus, in some cell types, PTEN functions as a tumor suppressor to inhibit the PKB/AKT-dependent survival signals that are activated in response to 3-phosphorylated phosphatidylinositols.

The failure of PTEN to regulate cell survival in the glioblastoma cell lines suggests that the generation of phosphatidylinositol 3-phosphates plays other roles during tumor progression. In fact, PI 3-kinase has been implicated in the regulation of a wide variety of cellular processes (30). Therefore, by dephosphorylating the products of PI 3-kinase, PTEN is likely to be involved in the regulation of many of these processes. Significantly, both PI 3-kinase and PKB/Akt have been isolated as transforming oncogenes in retroviruses (31, 32). These observations indicate that PI 3-kinase and PKB/Akt function in a signalling pathway promoting growth and survival. Our data demonstrate that PTEN functions to suppress these growth promoting and survival signals by dephosphorylating the phospholipid products of PI 3-kinase. Our data provide a molecular basis for the development of Cowden disease, as loss of the lipid phosphatase activity of PTEN leads to the accumulation of the products of PI 3-kinase and results in the neoplastic pathologies characteristic of the disease.

Acknowledgments: This work was supported by grants CA53840 and GM55989 to NKT and by an MRC Program Grant to CPD. MHW is an American Cancer

Society Research Professor and is supported by the U.S. Department of the Army (DAMD-17-94-14247), the National Cancer Institute (5R35 CA39829), Amplicon Corporation and the "1 in 9" breast cancer organization. MPM was supported by a National Cancer Institute training grant (5T32 CA09311-18). IP was supported by an BBSRC CASE studentship. The authors wish to thank B. Neel and L. Faleiro for invaluable comments.

References

1. Steck, P. A., Perhouse, M. A., Jasser, S. A., Yung, W. K. A., Lin, H., Ligon, A. H., Lauren, A. L., Baumgard, M. L., Hattier, T., Davis, T., Frye, C., Hu, R., Swedlund, B., Teng, D. H. F. & Tavtigan, S. V. (1997) *Nature Genetics* **15,** 356-362.

2. Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang, S., Puc, J., Milliaresis, C., Rodgers, L., McCombie, R., Bigner, S. H., Giovanella, B. C., Ittman, M., Tycko, B., Hibshoosh, H., Wigler, M. H. & Parsons, R. (1997) *Science* **275,** 1943-1946.

3. Whang, Y. E., Wu, X., Suzuki, H., Reiter, R. E., Tran, C., Vessella, R. L., Said, J. W., Isaacs, W. B. & Sawyers, C. L. (1998) *Proc. Natl. Acad. Sci. U S A* **95,** 5246-50.

4. Rasheed, B. K., Stenzel, T. T., McLendon, R. E., Parsons, R., Friedman, A. H., Friedman, H. S., Bigner, D. D. & Bigner, S. H. (1997) *Cancer Res.* **57,** 4187-90.

5. Liaw, D., Marsh, D. J., Li, J., Dahia, P. L. M., Wang, S. I., Zheng, Z., Bose, S., Call, K. M., Tsou, H. C., Peacocke, M., Eng, C. & Parsons, R. (1997) *Nature Genetics* **16,** 64-67.

6. Marsh, D. J., Dahia, P. L. M., Zheng, Z., Liaw, D., Parsons, D., Gorlin, R. J. & Eng, C. (1997) *Nature Genetics* **16,** 334-334.

7. Eng, C., Murday, V., Seal, S., Mohammed, S., Hodgson, S. V., Chaudray, M. A., Fentiman, I. S., Ponder, B. A. J. & Eeles, R. A. (1994) *J. Med. Genet.* **31,** 458-461.

8. Di Cristofano, A., Pesce, B., Cordon-Cardo, C. & Pandolfi, P. P. (1998) *Nature Genetics* **19,** 348-55.

9. Tamura, M., Gu, J., Matsumoto, K., Aota, S., Parsons, R. & Yamada, K. M. (1998) *Science* **280,** 1614-7.

10. Maehama, T. & Dixon, J. E. (1998) *J. Biol. Chem.* **273,** 13375-8.

11. Myers, M. P., Stolarov, J. P., Eng, C., Li, J., Wang, S. I., Wigler, M. H., Parsons, R. & Tonks, N. K. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 9052-9057.

12. Myers, M. P. & Tonks, N. K. (1997) *Am. J. Hum. Genet.* **61,** 1234-8.

13. Batty, I. H., Nigel Carter, A., John Challis, R. A. & Hawthorne, J. (1997) *Receptor-Linked Phosphoinositide Metabolism.*

14. van der Kaay, J., Batty, I. H., Cross, D. A., Watt, P. W. & Downes, C. P. (1997) *J. Biol. Chem.* **272,** 5477-81.

15. Morgenstern, J. P. & Land, H. (1990) *Nucleic Acids Res* **18,** 3587-96.

16. Zhang, S. H., Kobayashi, R., Graves, P. R., Piwnica-Worms, H. & Tonks, N. K. (1997) *J. Biol. Chem.* **272,** 27281-7.

17. Tiganis, T., Bennett, A. M., Ravichandran, K. S. & Tonks, N. K. (1998) *Mol. Cell. Biol.***18,** 1622-34.

18. Lioubin, M. N., Algate, P. A., Tsai, S., Carlberg, K., Aebersold, A. & Rohrschneider, L. R. (1996) *Genes Dev.* **10,** 1084-95.

19. Sun, H., Charles, C. H., Lau, L. F. & Tonks, N. K. (1993) *Cell* **75,** 487-493.

20. Garton, A. J., Flint, A. J. & Tonks, N. K. (1996) *Mol. Cell. Biol.* **16,** 6408-6418.

21. Bos, J. L. (1995) *Trends Biochem. Sci.* **20,** 441-2.

22. Marte, B. M. & Downward, J. (1997) *Trends Biochem. Sci.* **22,** 355-8.

23. Datta, S. R., Dudek, H., Tao, X., Masters, S., Fu, H., Gotoh, Y. & Greenberg, M. E. (1997) *Cell* **91,** 231-41.

24. Andjelkovic, M., Alessi, D. R., Meier, R., Fernandez, A., Lamb, N. J., Frech, M., Cron, P., Cohen, P., Lucocq, J. M. & Hemmings, B. A. (1997) *J. Biol. Chem.* **272,** 31515-24.

25. Risinger, J. I., Hayes, A. K., Berchuck, A. & Barrett, J. C. (1997) *Cancer Res.* **57,** 4736-8.

26. Cairns, P., Okami, K., Halachmi, S., Halachmi, N., Esteller, M., Herman, J. G., Jen, J., Isaacs, W. B., Bova, G. S. & Sidransky, D. (1997) *Cancer Res.* **57,** 4997-5000.

27. Wang, S. I., Puc, J., Li, J., Bruce, J. N., Cairns, P., Sidransky, D. & Parsons, R. (1997) *Cancer Res.* **57,** 4183-6.

28. Tashiro, H., Blazes, M. S., Wu, R., Cho, K. R., Bose, S., Wang, S. I., Li, J., Parsons, R. & Ellenson, L. H. (1997) *Cancer Res.* **57,** 3935-40.

29. Alessi, D. R., James, S. R., Downes, C. P., Holmes, A. B., Gaffney, P. R., Reese, C. B. & Cohen, P. (1997) *Cur.r Biol.* **7,** 261-9.

30. Toker, A. & Cantley, L. C. (1997) *Nature* **387,** 673-6.

31. Bellacosa, A., Testa, J. R., Staal, S. P. & Tsichlis, P. N. (1991) *Science* **254,** 274-7.

32. Chang, H. W., Aoki, M., Fruman, D., Auger, K. R., Bellacosa, A., Tsichlis, P. N., Cantley, L. C., Roberts, T. M. & Vogt, P. K. (1997) *Science* **276,** 1848-50.

33. Pap, M. & Cooper, G.M. (1997) *J. Biol. Chem.* **273,** 19929-32.

Figure Legends

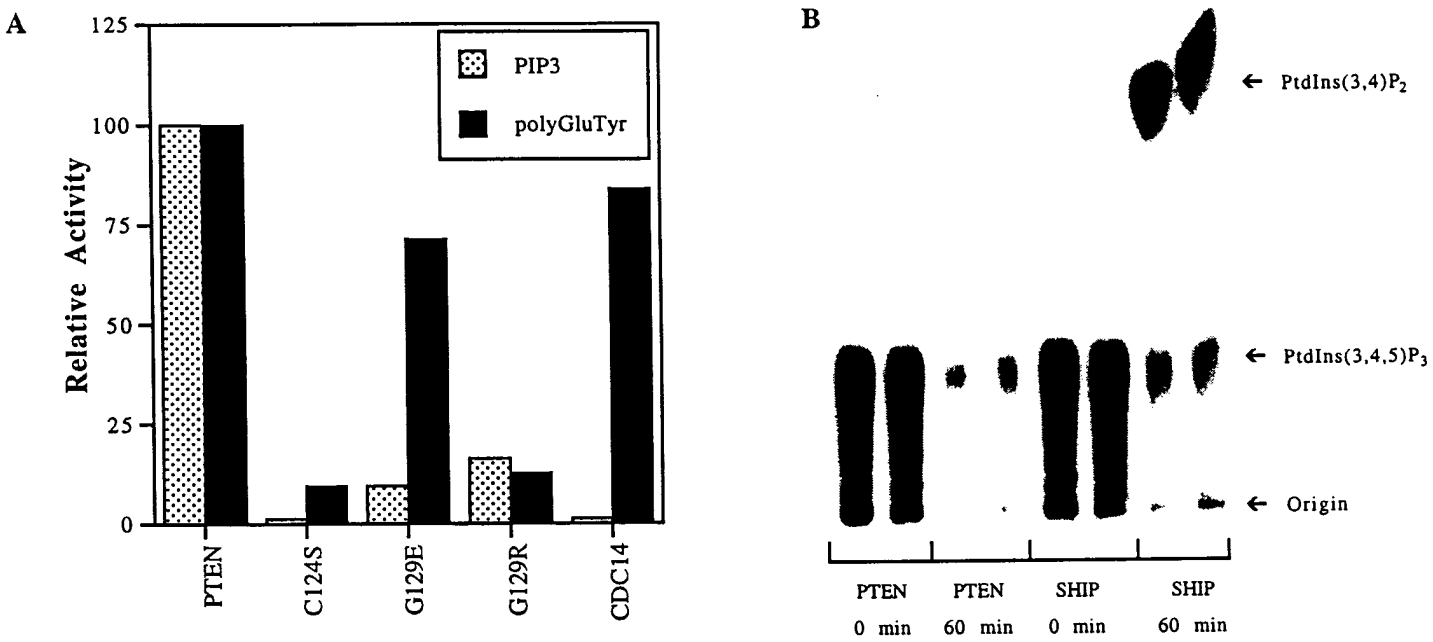
Figure 1: PTEN is a phosphatidylinositol 3-phosphatase **A)** Recombinant PTEN was incubated with radiolabeled polyGluTyr and PtdIns(3,4,5)P_3 and the release of $^{32}P_i$ was measured as described. **B)** PtdIns(3,4,5)P_3, labeled at the 3-position, was incubated with PTEN or SHIP for 0 or 60 minutes and the reaction products were resolved on TLC plates and visualized by autoradiography.

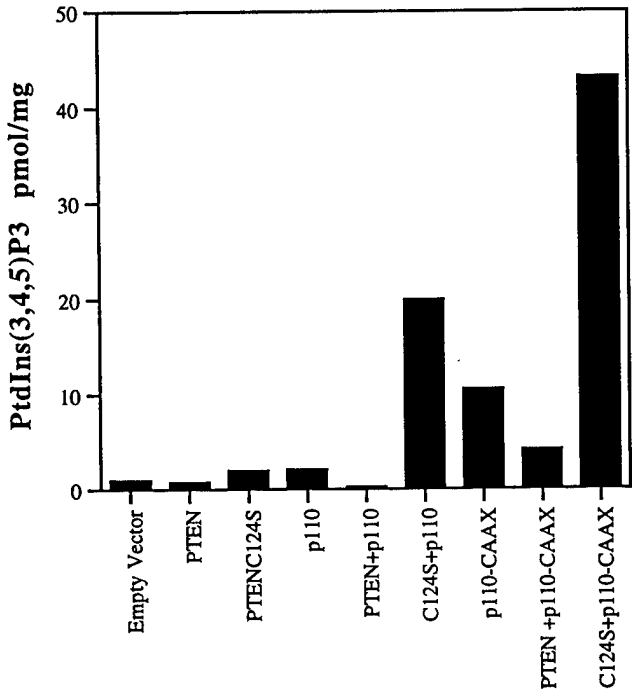
Figure 2: Expression of PTEN antagonized PI 3-kinase. HEK293 cells were transfected with PTEN or PTENC124S in combination with the p110 subunit of PI 3-kinase or an activated, membrane-bound PI-3 kinase (p110-CAAX). The resulting levels of PtdIns(3,4,5)P_3 were determined as described and were normalized to total protein. Data are expressed as pmol PtdIns(3,4,5)P_3 per mg protein. These data, from a single experiment, are representative of three experiments which yielded similar results.

Figure 3: Expression of PTEN in glioblastoma cell lines decreases the amount of activated PKB/AKT. PTEN expression was reconstituted in **A)** U87MG or **B)** U373MB glioblastoma cell lines by infecting with recombinant retrovirus. Confluent dishes were either left untreated, stimulated with insulin (10 μg/ml for 10 minutes), or PDGF (50 ng/ml for 10 minutes), or pretreated with Wortmannin (150 nM for 30 minutes) and then stimulated with both insulin (10 μg/ml) and PDGF (50 ng/ml) for 10 minutes, and then lysed. Equal

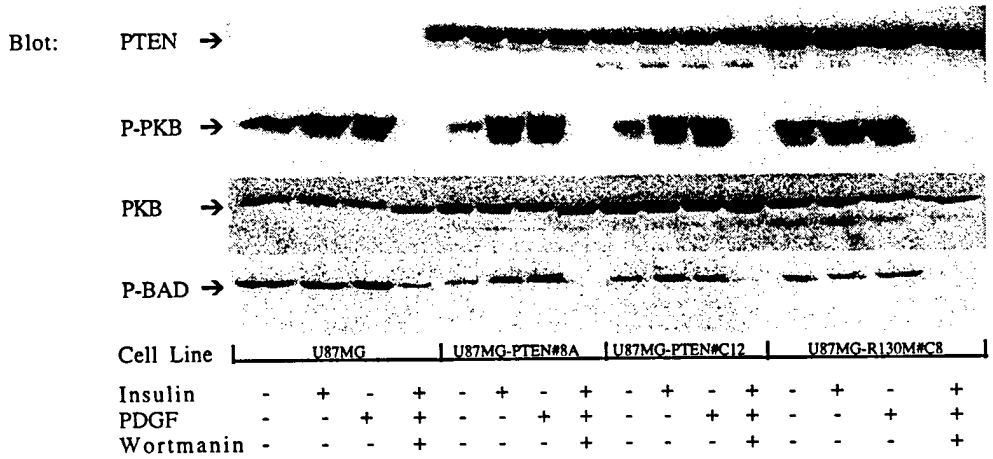
amounts of protein were loaded for each cell line and immunoblots were probed with antibodies to PTEN, PKB/Akt (PKB), phospho-PKB/Akt (P-PKB), or phosphorylated BAD (P-BAD) and visualized using ECL reagents.

Figure 4: Expression of PTEN inhibits cell survival in LnCaP cells. PTEN, PTENC124S or PTENG129E, were co-transfected with a green fluorescent protein (GFP) expression vector into LnCaP cells. The co-transfections also included either an expression vector for consitutively active PKB/Akt (black bars) or a control empty vector (striped bars) . Transfected cells were identified as GFP positive cells or by immunofluorescence microscopy following staining with antibodies to the transfected PTEN. Transfection efficiency was assessed by determining the percentage of GFP/PTEN positive cells (total number of cells was determined by counting nuclei (stained with DAPI)). Data are expressed as the mean transfection efficiency (+/- s.d. n=3).

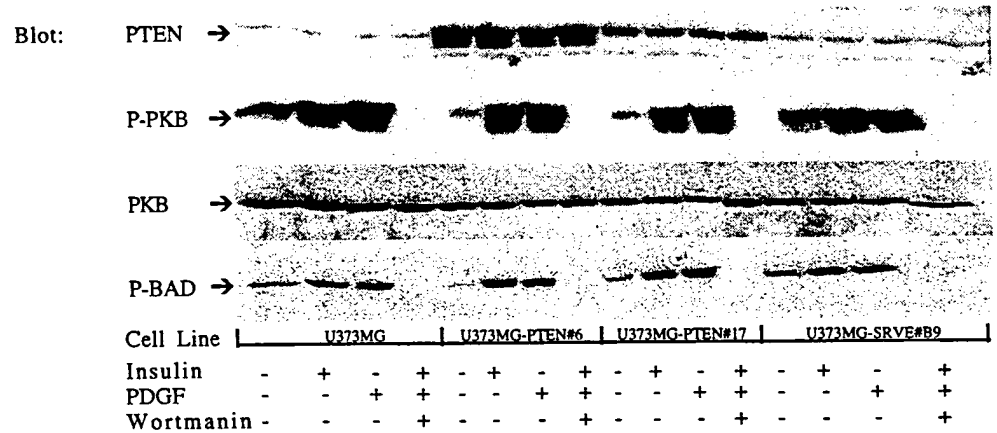

A
125
100
75
50
25
0
Relative Activity
PIP3
polyGluTyr
PTEN
C124S
G129E
G129R
CDC14
B
PtdIns(3,4)P2
PtdIns(3,4,5)P3
Origin
PTEN
0 min
PTEN
60 min
SHIP
0 min
SHIP
60 min


PtdIns(3,4,5)P3 pmol/mg
50
40
30
20
10
0
Empty Vector
PTEN
PTENC124S
p110
PTEN+p110
C124S+p110
p110-CAAX
PTEN +p110-CAAX
C124S+p110-CAAX

A


Blot:
PTEN
P-PKB
PKB
P-BAD
Cell Line
U87MG
U87MG-PTEN#8A
U87MG-PTEN#C12
U87MG-R130M#C8
Insulin - + - + - + - + - + - + - + - +
PDGF - - + + - - + + - - + + - - + +
Wortmanin - - - + - - - + - - - + - - - +

B


Blot:
PTEN
P-PKB
PKB
P-BAD
Cell Line
U373MG
U373MG-PTEN#6
U373MG-PTEN#17
U373MG-SRVE#B9
Insulin - + - + - + - + - + - + - + - +
PDGF - - + + - - + + - - + + - - + +
Wortmanin - - - + - - - + - - - + - - - +