DETECTION REAGENTS FOR TPC2 AND TPC3, TWO PROTEINS THAT ARE COEXPRESSED WITH TELOMERASE

Inventors: Bryant Villeponteau, San Carlos, CA (US); Junli Feng, San Carlos, CA (US); William H. Andrews, Richmond, CA (US); Robert R. Adams, Redwood City, CA (US)

Correspondence Address:
GERON CORPORATION
230 CONSTITUTION DRIVE
MENLO PARK, CA 94025

Related U.S. Application Data

Continuation of application No. 09/220,157, filed on Dec. 23, 1998, now Pat. No. 6,300,110, which is a continuation of application No. 08/710,249, filed on Sep. 13, 1996, now Pat. No. 5,858,777, which is a continuation-in-part of application No. 08/583,808, filed on Jan. 5, 1996, now abandoned.

Provisional application No. 60/003,492, filed on Sep. 9, 1995.

Publication Classification

Int. Cl. 7 .................. G01N 33/53; G01N 33/537; G01N 33/543; C12P 21/04; C07K 16/40

U.S. Cl. .................. 435/70.21; 435/331; 530/388.26; 435/7.92

ABSTRACT

Purified and recombinant proteins TPC2 and TPC3 and recombinant or synthetic oligonucleotides corresponding to those proteins or fragments thereof can be used to detect regulators of telomere length and telomerase activity in mammalian cells and for a variety of related diagnostic and therapeutic purposes.
FIG. 1A.
FIG. 1B.
**FIG. 1C.**

The diagram shows a comparison of TP53/GAPDH ratios across various cell lines, with the x-axis representing different cell lines and the y-axis representing the TP53/GAPDH ratio. The cell lines are categorized into two groups: mortal and immortal. The mortal group includes PFL, PFS, APP, PSF, FSP, IMR-90, BJ, and HUVEC, while the immortal group includes LEU, U2OS, MGH22, CT SW620, BREAST, HELA, 143TA, IDH4, and TESTES.
FIG. 2A.
FIG. 2C.
FIG. 3.
FIG. 4.

ProArgSerAlaAsnMetAlaAlaAlaThrValGlyArgAspThrLeuPro
1  CCGCGTCCGGCGGGAACATGGCAGCCGGGCGGACGTCGGGGCGGACCTTTA
GluHisTrpSerTyrGlyValCysArgAspGlyArgValPhePheIleAsn
52  GAGCATTGGCTCTACGGGTTGTGCGGGGATGGCAGCCGGGCTCTTTCATCA
AspGlnLeuArgCysThrThrTrpLeuHisProArgThrGlyGluProVal
103  GACCAGGCTCCGTCACGGACCTGGTGCACCCCCGCGAAGGGGAGGTC
AsnSerGlyHisMetIleArgSerAspLeuProArgGlyTrpGluGly
154  AACTCGGGCCACATGATCCGTCAGACCTGGCCCGGGCTGGAGAAGGC
PheThrGluGlyGlyAlaSerTyrPheIleAspHisAsnGlnGlnThrThr
205  TTCACGGAGGAGGCGCGAGCTACTTCTACGGACCATAACCAGAACCACA
AlaPheArgHisProValThrGlyGlnPheSerProGluAsnSerGluPhe
256  GCATTCCAGCATACTGTCAGGGACAGTTTCTCCAGAAAATAGTGAA
IleLeuGlnGluProAsnProHisMetSerLysGlnAspArgAsnGln
307  ATTTCTCAAGAGAGCCGCAATCCACATGTGCAAGCAAGCAAGCAAGCA
ArgProSerSerMetValSerGluThrSerThrAlaGlyThrAlaSerThr
358  AGACCGTCCACATGGTGCTAAGACATCCACAGGCTGGAGCCAGCTCCACC
LeuGluAlaLysProGlyProLysIleIleLysSerSerSerSerLysValHis
409  CTGGAGGCAAGCCTGAGAACAGATCATAAGTCCAGCAGTAAAGTGCCAC
SerPheGlyLysArgAspGlnAlaIleArgArgAsnProAsnValProVal
460  AGCTTGGAAAGAGAGCCAGCCCATAGGGAGAAACCAATGTTCCCTG
ValAlaArgGlyTrpLeuHisLysGlnAspSer GlyMetArgLeuTrp
511  STGGTGGAGGCGCTGCCACTAAGCAGGACAGCTTTGGGATGGAGCTGG
LysValArgTrpValLeuAlaAspTyrCysLeuPheTyrTyrLysAla
562  AAAAGGAGGCTTGTTGTGCTGTGTTACTGCTTTTTTTACTATAAGCC
GluLysLysArgSerSer SerIleProLeuPro TyrVal Ser
613  GAGAAGAAAGCGGTCTCGGAGGACATCCCTGTCCTGACCAG3TACGTGAT3TCT
ProValAlaProGluAspArgIleSerArgLysTyrSerPheLysAlaVal
664  CCTGTTGCCCCTGGAGGACATCATAAGCAGCAGGCAAATATTCTTTTTAAGGCTGG
HisThrGlyMetArgAlaLeuIleTyrAsnSerSerThrAlaGlySerGln
715  CACACGGGGATGGCAGGGCTCATCCTATAACAGCTCCAGCAGGGGCTCTCAG
FIG. 4.
PATENT APPLICATION

Figure 4.
AsnAspValGluGlnLeuLysGlnThrLeuGlnGluGlnHisArgArgAla
2296 AATGATGTGGAACAGCTGAGCAGACCCCTGCCAGGAGCAACAGAAGAGCC
PhePhePheGlnGluLysSerGlnIleGlnLysAspLeuTrpArgIleGlu
2347 TTTTTTCCCCAGGAAATCCGAGATACAGAAGATCTATGGAGAAAA
AspValThrAlaGlyLeuSerAlaAsnLysGluAsnPheArgIleLeuVal
2398 GATGTCACCTGCGGCTGAGTGAAATAAAGAGAACATGAAATCTAGT
GluSerValLysAsnProGluArgLysThrValProLeuPheProHisPro
2449 GAGTCAGTAAAAATCCGGAGAAGAAAAACGGTGCTTTGTTCCTCACCAG
ProValProSerLeuSerThrSerGluSerLysProProGlnProSer
2500 CCTGTCCTTCACCTCAAACTCTCTCGAGAGCAAGCGCCGCCACAGCAGCCAGT
ProProThrSerProValArgThrProLeuValGluArgPheProGln
2551 CCTCCACCAGCCCTGTCGGGACCCCTCTGGAGGGTTCGACTCTTCCACCAG
LeuGlnThrTyrValProTyrArgProHisProProGlnLeuArgLysVal
2602 CTGCAAACCTGCTGCGGATACCCGACCTCAACCCACCCAGCTGAGGAAATG
ThrSerProLeuGlnSerProThrLysAlaLysProLysValGlnGluAsp
2653 ACATCCCCCTTTCAGTCACCAACTAAGGGCAAGCCCAAGTTACAGGAAGAT
GluAlaProArgProProLeuProGluLeuTyrSerProGluAspGln
2704 GAAGCACCCTCCAGGCCCCGACTCCCCGGACCTACAGCCAGGACCCAG
ProProAlaValProProLeuArgGluAlaThrIleIleArgHisThr
2755 CCCCAGGTGCGCCGGCTCCTGCAGAAAGAGCCACCCACATCAAGCGGCACACA
SerValArgGlyLeuLysArgGlnSerAspGluArgLysArgAspArgGlu
2806 TCTGTCGGGGCCATCAGCCGAGTGGCAGGAGGGAAGCCGAGACCAGGAG
LeuGlyGlnCysValAsnGlyAspSerArgValGluLeuArgSerTyrVal
2857 CGGGGCAGTGCTGGAATGCGATCCAGGGTGGAGCTCGGCTGCGTATGTC
SerGluProGluLeuAlaThrLeuSerGlyAspMetAlaGlnProSerLeu
2908 AGTGAGCCGCTAGCTGCGCCGACCTCAGCGGGGACATGGCCACAGCCTCTCTA
GlyLeuValGlySerArgTyrGlnThrLeuProGlyArgGlyLeu
2959 GAGATTTGCGGCTGAGGAGAGCAGCCTACCAGACGGCTGGGCAGAGGCCCT
SerGlySerThrSerArgLeuGlnGlnSerThrIleAlaProTyrVal
3010 TCAGGGTCACCTGCAAGCTCCACGAGTGGCTCCACCATTGTCTCAGCTAGTC

FIG. 4.

ThrLeuArgArgGlyLeuAsnAlaGluSerSerLysAlaThrPheProArg
3061 AACAATCGGAGGGTGCTCTCAATGCGCAAGCAAGCAACACACCTCAGA

ProLysSerAlaLeuGluArgLeuTyrSerGlyAspHisGlnArgGlyLys
3112 CCTAAGAGCTGCTCGGAGCGCTGTGACTGAGGGATACCCAGCGAGGGCAAG

MetSerAlaGluGlnLeuGlueArgMetLysArgHisGlnLysAlaLeu
3163 ATGAGGTCCAGAGACGCTGCTGAAGCGACATGAAGCGACAGACAGGACAGCCTG

ValArgGluArgLysArgThrLeuGlyGlnGlyArgThrGlyLeuPro
3214 GTCCGAGAGCGCAAGAGACACTGCGCAAGCAGGAGAGCGAGGCGGCTGCCC

SerSerArgTyrLeuSerArgProLeuGlyAspLeuGlySerValCys
3265 TCAATCGCTACCTCAGGCCGCGCCTCGTGGAGATCTGCTGATGT

---

3316 TAGGGAGGGCCAGCCAGCGGCGAGGGGACAGGAGCCGAGTGGCCTCAG
3367 GTCCCCCAACACAAACAGCACATCACAATCCCGACTGAGGAGACGTTCAAT
3418 ACCTACATGGTTCAGAGAACACACACACGCGCTGGTTGTCGAGCAGCCAG
3469 CTTGAGGAAATGCGTCTCAGAGGTTGCTGAGCTGACAAGACACTCTGAGAG
3520 TCAACAGCCCCCCAGGCTGAGAGAGGAAATACCCAGTCTGATCCTCCAGAC
3571 GATTGGAGGGGCTTTTCCGCCGTTGGAATCTGCTGATATTCAGACC
3622 TGGGAAATACGAGGTTCTCCAGAAGAGGAGGAGAATGGTACATGATTGG
3673 TTAGACAGCACAAGCAAACTGATACAGCCACGCAGACCGCATGGCATGCTG
3724 TGTTGAGGTGATTGAGGACAGGCGGTCACACCCCTGCCCCAAAGGCGACTGG
3775 CCTCCTGGCTGCTGAGTCAAGGCGCTGCTGAGTGGGGTTGCTGGGAC
3826 GGGCTGTGCTCCTGGCCCTCGGTGCGGCTTCCGCTGCGGCGGTCCAAGC
3877 TCACCTAGAGGACCCTTTTGAGAAACATTCTCATTGACTTTTTCCTGTGGT
3928 AAAATCCCATGTCTCCCTAAACACCTGTGATGTTGTTTCTCTTCTAAATTC
3979 TTGCAACAGCTCTCTTGTGTTATAGTACGCCCTATTGTCCTCTATTTCC
4030 TGCTCTAGTTTGGTTACAGAAGCTCTCTGATATGGCATGATCGCTCAAAC
4081 TGCTCCTGCTGACTTTATACAGCCGGCGGTGAATACGCGTCTGGCTTACA
4132 GCCCTTGGAGTGGCTCCAGATGTGACTACACTTTTCTGCTGCGCCTCTTCATG
4183 CAGGCGCTACTGACTCATATAATCTACTCTTGTCCGTACCGCGCCGCAATTC

CODE MEANING
----- --------
ANY CODES NOT LISTED BELOW MATCHES A,C,G,OR T (ANY BASE)
1 A, C, G, OR T (ANY BASE)
2 A, C, G, OR T (ANY BASE)
3 PROBABLY CYTOSINE
4 PROBABLY THYMINE
5 PROBABLY ADENINE
6 PROBABLY GUANINE

FIG. 4.
FIG. 4.
FIG. 5.
ProAspProValGlnThrGlnLeu
ProProSerAlaProPheLeuSerGlyGluArgPheCysThrAsnPhePro
ValGluGlyGlySerAlaLeuSerGlnProLeuProSerLysThrArgPro
TrpSerArgAsnLeuGlnAlaAspAlaAlaMetGlnHisTyrGlyValAsn
GlyTyrSerLeuHisAlaMetAsnSerLeuSerAlaMetTyrAsnLeuHis
GlnGlnAlaAlaGlnAlaGlnAlaGlnHisAlaProAspTyrArgProSerVal
HisAlaLeuThrLeuAlaGluArgLeuAlaGlyCysThrPheGlnAspIle
IleLeuGluAlaArgTyrGlySerGlnHisArgLysGlnArgArgSerArg
ThrAlaPheThrAlaGlnGlnLeuAlaLeuGluGluLysThrPheGlnLys
ThrHisTyrProAspValValMetArgGluArgLeuAlaMetCysThrAsn
LeuProGluAlaArgValGlnValTrpPheAsnArgArgAlaLysPhe
ArgLysLysGlnArgSerLeuGlnGlnLeuGlnLysGlnLysGlu
AlaGluGlySerHisGlyGluGlyLysAlaGluAlaProThrProAspThr
GlnLeuAspThrGluGlnProProArgLeuProGlySerAspProProAla
FIG. 6.
FIG. 6.

GluLeuHisLeuSerLeuSerGluGlnSerAlaSerGluSerAlaProGlu
766 GAGCTTCACCTGAGTCTCTGAGCAGTCAGGCCAGTGAAGCTGGAGCCC
AspGlnProAspArgGluGluAspProArgAlaGlyAlaGluAspProLys
817 GATCAGCCGACCCTGAGGAGAGCCACCCAGGGGCAAGGCTGGAGAA
AlaGluLysSerProGlyAlaAspSerLysGlyLeuGlyCysLysArgGly
868 GCTGAGAAAGGACCTGGGCTAGACAGCAAGGGCTGGGCTGGGATAGAC
SerProLysAlaSerGlySerLeuThrIleThrProValAlaPro
919 AGCCCAAGGAGATCTCCCCAGCCTGGATCACCATACTCTGTGGCC
GlyGlyGlyLeuGluGlyProSerHisSerTyrSerSerSerProLeuSer
970 GGGGGTGTCCTCTGAGCAGGAGCTCCAGCCACCTATGGGGGAGCA
LeuPheArgLeuGlnGluPheArgGlnHisMetAlaAlaThrAsnAsn
1021 CTCTTGCCCTTGAAGCAGCTATGGGGGCTCCAGCAGGAGCA
LeuValHisTyrSerSerPheGluValGlyGlyProAlaProAlaAlaAla
1072 CTGGTGCACTACTCTGCTCTGAGAGGTTGACCCGAGTTGGGCTGG
AlaAlaAlaAlaAlaValProTyrLeuGlyValAsnMetAlaProLeuGly
1123 GCGCGGGCGTCTGAGCGGGCGTACCTGAGGCTGGCaACATGGGGGCGTTGGG
SerLeuHisCysGlnSerTyrTyrGlnSerLeuSerAlaAlaAlaAlaAla
1174 TCACTGCAGTGTTACGGCATTTCCATGGCCCTACCTGTGCTGCTGGCC
HisGlnGlyValTrpGlySerProLeuLeuProAlaProAlaProAlaGlyLeu
1225 GCCGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCC
AlaProAlaSerAlaThrLeuAsnSerLysThrThrIleGluAsnAsn
1276 GTCCTGCATCAGCTACCCCTGACACGAAGTCAGAGACGACGACGAGG
ArgLeuArgAlaLysGlnHisAlaAlaSerLeuGlyLeuAspThrLeuPro
1327 CGGCTCCGGGCGACAGCAGAAGCGGGCCTGGGACTCGATACGCTGCC
Asn---
1378 AACTGACTGTCCTGGCTCTGACACCCAGGCTACGGGCTTACGCTTGCCT
1429 AGCCCTGTGTATCCCTAGGTCGGCTCTGAGGATTAACCCCATGAGGCC
1480 AGGGATCTAGGCTGGCTTTACGGCTGCTGCTGCCCTGGCCCCATACCC
1531 AGCCCAAGGCTGAGGACCCACACTACACACCTCTGCTCATTGGGCCCTGCT
1582 CCCCCTGGGAAGGGAGTTGGAGAGGCTGGCTGGGCTGGGCCCAAGCT
1633 TCCCTGAGGAAGGCTGGGCTCTCCCTGAGAAGTCTCTATCTCATAGAGC

FIG. 6.
FIG. 6.
3877 GGGCCTGGGAATCCAGGCTAAAGACCACACCTACATGTGGCAAGCACCAAG  
3928 ACAGGCATTTTGGGTTTCCAAATCCTCAGGTCTTTGGCTGGGTCTGGAA  
3979 TTTGGAAGGGGAATCCACCAGCCATTGGGGCATAGAGGAGGACTTAGGC  
4030 AGCGCTGTGGGAGGCTAGGAGTGACAGAGGAGGTTTTGTTTTGTTGTTGTTTGT  

**FIG. 6.**
CTGCAGAGGATAGAAAAAG0CCCTCTGATACCTCAAGTTAGTTTACCTTTA

-PST1-

AAGAAGGTCGGAAGTAAAGACGCAAAGCCTTTCCGAGGTGCAAGGCTAAG

CCGCCGCAGACTCAAGGGAAGTGAGCCGGGGTTGCCTGGAG

FIG. 9.

CTAGAGCAAAACAAAAATGTCAGCTGCTGGCCGTTGGCCCTCCGCCGAG

hTR

TGCGGCGGGTCGGCTGGCCAGCCCCGAACCCCGCTGGAGGGCGCGGTCGGC

CGGGGCTTCTCCGGAGGGCAACGGCAGTAGTGGGCTTCTGTA

GCCGCGGTTCTCGGGGGCCGAGGCGAGTTTAGGCCTCCAGGGGCAGGA

AGAGGAAAGCAGAGA6TCGCCGCGGCGGCGGATCCCTGAGCTGGGAGC

TGACCCCAAGAATGCTGCTGGGCAAGATGTTGCTTTTCTTGTTGGGAG

ACGCGGATCAGTGCGCATCCCTACCCCTTCGCAGGGCGAGTTGGGCTTGTGAA

CCAAACCTGACTGACTGGGCCAGTGTGCTGCAAATTGGCAGGAGACGTGAAG

GCACCTCAGGACTCGGGCAACAAAAATGAATGGGCAATGAGCAGGCGGGTTGCTGGAG
902 CCGTTCTGCGGTTCTCCCGCTTCCCGCTTTTTGTTGCTTTTATGGTTG

955 TATTACAACCTTAGTCCCTGCTCTGCA

-PST1-

FIG. 9.
DETECTION REAGENTS FOR TPC2 AND TPC3, TWO PROTEINS THAT ARE COEXPRESSED WITH TELOMERASE

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation of U.S. Ser. No. 09/220,157, filed Dec. 23, 1998 (pending); which is a continuation of 08/710,249, filed Sep. 13, 1996 (issued as U.S. Pat. No. 5,858,777 on Jan. 12, 1999); which is a continuation-in-part of U.S. Ser. No. 08/583,808, filed Jan. 5, 1996 (now abandoned), and claims the benefit of Provisional Application No. 60/003,492, filed Sep. 8, 1995. All the priority documents are hereby incorporated herein by reference in their entirety.

FIELD OF THE INVENTION

[0002] The present invention provides methods and reagents for regulating telomere length and modulating telomerase activity in mammalian cells as well as for detecting, diagnosing, and treating related diseases and conditions in humans and other mammals. In an important embodiment, the invention provides oligonucleotide probes and primers, polynucleotide plasmids, peptides, proteins, antibodies, and enzymes relating to genes and gene products that regulate telomere length and telomerase activity in mammalian cells. The invention has diverse applications and provides important advances in the fields of molecular biology, chemistry, pharmacology, and medical therapeutic and diagnostic technology.

BACKGROUND OF THE INVENTION

[0003] The DNA at the ends of the telomeres of chromosomes in mammalian cells consists of double- and single-stranded nucleic acid composed of many tandem repeats of a simple nucleotide sequence referred to as the telomeric repeat sequence. Telomeres help maintain chromosome structure and function; the loss of telomeric DNA can activate the cellular processes that detect and control DNA damage and monitor and control cell proliferation and senescence. The maintenance of telomeres and the regulation of telomere length are vital cellular functions involved in transmitting genetic information from generation to generation, aging, the control of cell growth, and cancer. See Harley, 1991, Mutation Research 256:271-282; and Blackburn, 1992, Annu. Rev. Biochem. 61:113-129, each of which is incorporated herein by reference (note: references cited herein are provided for convenience; such citations are not to be construed as an admission of prior invention).

[0004] The multi-component telomerase ribonucleoprotein enzyme catalyzes the synthesis of the first strand of telomeric DNA synthesized during telomerase elongation, using the RNA component of the enzyme as a template. Although the RNA component of human telomerase (hTR) and other mammalian telomerase enzymes has been identified, isolated, characterized, and described in the scientific literature, the protein components of the telomerase enzyme as well as most other cellular macromolecules involved in telomere maintenance and the regulation of telomere length and telomerase activity in mammalian cells have not. See Feng et al., 1995, Science 269:1236-1241; PCT patent publication No. 96/01835; and pending U.S. patent application Ser. Nos. 08/521,634, filed Aug. 31, 1995, and 08/330,123, filed Oct. 27, 1994, each of which is incorporated herein by reference.

[0005] Many useful methods and reagents relating to telomere and telomerase biology have been described. See, e.g., U.S. Pat. No. 5,489,508; PCT patent publication Nos. 95/23572, 95/13381, 95/13382, and 95/13383; and U.S. patent application Ser. No. 08/632,662, filed Apr. 15, 1996, each of which is incorporated herein by reference. Significant improvements to and new opportunities for telomere- and telomerase-mediated therapies as well as related assays, screens, diagnostic methods, and reagents could be realized and obtained, however, if additional cellular macromolecules involved in mammalian telomere maintenance and the regulation of telomere length and telomerase activity could be identified, characterized, and made available in pure or isolatable form. In particular, the characterization of the nucleotide and corresponding amino acid sequences of such macromolecules could provide new and useful recombinant expression vectors and plasmids, as well as related reagents useful in medical therapeutic and diagnostic technology.

SUMMARY OF THE INVENTION

[0006] The present invention provides methods and reagents for regulating telomere length and modulating telomerase activity in mammalian cells as well as for detecting, diagnosing, and treating related diseases and conditions in humans and other mammals.

[0007] In one embodiment, the invention provides recombinant mammalian host cells containing:

[0008] (i) a recombinant or synthetic nucleic acid comprising at least about 10 to 15 to 25 to 100 or more contiguous nucleotides corresponding to an open reading frame sequence of a human gene TPC2 contained in a human DNA insert of an ~3.5 kb NotI-BseEII restriction fragment of plasmid pGRN109 (on deposit with the American Type Culture Collection under the accession number ATCC 97708); or

[0009] a synthetic or recombinant peptide or protein comprising at least about 6 to 10 to 15 to 25 to 100 or more contiguous amino acids corresponding to an amino acid sequence encoded by said open reading frame sequence; and

[0010] (ii) a recombinant or synthetic nucleic acid comprising at least about 10 to 15 to 25 to 100 or more contiguous nucleotides corresponding to an open reading frame sequence of a human gene TPC3 contained in a human DNA insert of an ~1.4 kb EcoRI-BamHI restriction fragment of plasmid pGRN92 (ATCC 97707); or

[0011] a synthetic or recombinant peptide or protein comprising at least about 6 to 10 to 15 to 25 to 100 or more contiguous amino acids corresponding to an amino acid sequence encoded by said open reading frame sequence of gene TPC3;
said TPC2 and TPC3 genes characterized in coding for proteins that regulate telomere length or modulate telomerase activity and are present in human or other mammalian cells that express telomerase activity.

Other mammalian host cells provided by the invention include those that comprise either or both TPC2- and TPC3-derived recombinant or synthetic nucleic acids, peptides, or proteins. Furthermore, the invention also provides such cells further modified to contain a synthetic or recombinant nucleic acid comprising at least about 10 to 15 to 25 or more contiguous nucleotides corresponding to a contiguous nucleotide sequence of human hTR located in an −2.5 kb HindIII-SacI restriction fragment of pGRN33 (ATCC 75926).

The recombinant host cells of the invention have application in many useful methods also provided by the invention. For example, the invention provides recombinant host cells comprising novel expression vectors with expression control sequences operatively linked to nucleotide sequences encoding amino acids in a sequence substantially identical to the amino acid sequences encoded by the human TPC2 or TPC3 genes and, optionally, a recombinant hTR gene. These recombinant host cells are useful for producing recombinant human telomerase, for use in screens to identify agents that modulate telomerase activity or regulate telomere length, as well as for a variety of other purposes described more fully below. The recombinant host cells of the invention can also be incorporated into the germ line and/or somatic tissues of non-human transgenic mammals, as well as be administered to mammals for therapeutic purposes.

In another embodiment, the invention provides synthetic and recombinant oligonucleotides and nucleic acids in a variety of forms, i.e., isolateable, isolated, purified, or substantially pure, and for a variety of purposes, i.e., as probes or primers, as polynucleotide plasmids and vectors for introducing recombinant gene products that regulate telomere length or modulate telomerase activity in mammalian host cells, as restriction fragments for creating useful nucleic acids, and as reagents for therapeutic, diagnostic, and other applications. In particular, the invention provides recombinant or synthetic nucleic acids comprising at least about 10 to 15 to 25 to 100 or more contiguous nucleotides substantially identical or complementary in sequence to a contiguous nucleotide sequence located in either:

(i) an open reading frame sequence of a human gene TPC2 contained in a human DNA insert of an −3.5 kb NotI-BstEII restriction fragment of plasmid pGRN109; or

(ii) an open reading frame sequence of a human gene TPC3 contained in a human DNA insert of an −1.4 kb EcoRI-BamHI restriction fragment of plasmid pGRN92.

The novel oligonucleotide probes and primers of the invention typically comprise nucleotides in a sequence substantially identical or complementary to a sequence of nucleotides in a TPC2 or TPC3 gene or gene product to allow specific hybridization thereon in a complex mixture of nucleic acids. Such probes and primers therefore have useful application in a variety of diagnostic, therapeutic, and other applications.

The expression vectors of the invention typically comprise expression control sequences operatively linked to a nucleotide sequence encoding amino acids in a sequence identical to a sequence of amino acids in a TPC2 or TPC3 protein gene product. Such expression vectors have many useful applications, including in therapeutic methods of the invention as gene therapy vectors for modulating telomerase activity, to either activate or inhibit that activity, or for regulating telomere length, either to increase or decrease the length, in a target cell or tissue.

Gene therapy expression vectors of the invention also include those that encode variants or “muteins” of the TPC2 and/or TPC3 proteins, i.e., express proteins that differ from TPC2 and/or TPC3 by deletion, substitution, and/or addition of one or more amino acids. The gene therapy vectors of the invention may also, however, encode useful nucleic acids, such as hTR, or antisense nucleic acids or ribozymes that target the TPC2, TPC3, and/or hTR gene products, i.e., mRNA and telomerase RNA. Such vectors are useful in the therapeutic methods of the invention for treating or preventing diseases or conditions in which modulation of telomerase activity or telomere length can be of benefit. For example, in telomerase positive cancer cells, inhibition of telomerase activity can prevent telomere maintenance in those cells, inducing upon continued proliferation telomere loss, cell crisis, and death. For such purposes, the gene therapy vectors of the invention that express a non-functional TPC2 or TPC3 mutein or variant protein or other nucleic acid that can inhibit telomerase formation or telomere elongation by telomerase activity in the cell, such as by competing for RNA component or protein components, inhibition of endogenous gene expression, or other means, are preferred.

In another embodiment, the present invention provides peptides, proteins, antibodies, and enzymes, relating to genes and gene products that regulate telomere length and telomerase activity in mammalian cells. In particular, the invention provides synthetic or recombinant peptides or proteins comprising at least about 6 to 10 to 15 to 25 to 100 or more contiguous amino acids identical in sequence to an amino acid sequence encoded by an open reading frame sequence of a human gene located in either:

(i) an −3.5 kb NotI-BstEII restriction fragment of plasmid pGRN109; or

(ii) an −1.4 kb EcoRI-BamHI restriction fragment of plasmid pGRN92.

The present invention provides the proteins encoded by the TPC2 and TPC3 genes in isolatable form from host cells expressing recombinant TPC2 and/or TPC3 protein, as well as in purified and substantially pure form from synthesis in vitro or by purification from recombinant host cells or by purification of the naturally occurring proteins using antibodies or other reagents of the invention. Such proteins have application in methods for reconstituting in vitro telomerase or other enzymatic activities that maintain telomeres and regulate telomere length. These methods in turn have application in screens for therapeutic agents, for diagnostic tests, and for other applications. In addition, peptides corresponding to the amino acid sequences of TPC2 or TPC3 proteins can also be used to regulate telomere length and telomerase activity in mammalian cells.
The proteins and peptides of the invention can also be used to generate antibodies specific for TPC2 or TPC3 proteins or for particular epitopes on these proteins. Thus the invention provides polyclonal and monoclonal antibodies that specifically bind to TPC2 or TPC3 proteins. These antibodies can in turn be used to isolate TPC2 or TPC3 proteins from normal or recombinant cells and so can be used to purify the proteins as well as other proteins associated therewith. These antibodies also have important application in the detection of cells comprising TPC2 or TPC3 proteins in complex mixtures of cells. Such detection methods have application in screening, diagnosing, and monitoring diseases and other conditions, such as cancer, pregnancy, or fertility, because the TPC2 and TPC3 proteins are present in most cells capable of elongating telomeric DNA and expressing telomerase activity.

The immunogenic peptides and proteins of the invention can also be used in therapeutic immunization and vaccination procedures. See U.S. provisional patent application Ser. No. 60/008,949, filed Oct. 20, 1995, incorporated herein by reference. The invention provides a method of immunizing a subject, as well as vaccines useful in the method, against cells that maintain telomeres and express telomerase activity that comprises administering an immunostimulating amount of such peptides or proteins of the invention.

In another embodiment, the invention provides a subtraction hybridization differential display method to identify, isolate, and clone expressed sequence tags (ESTs) of mRNA species encoding rare proteins, such as those involved in telomere elongation and the regulation of telomere length and telomerase activity. This method comprises the steps of:

(i) obtaining mRNA from a first population of mammalian cells which contain said rare protein, i.e., a protein component of telomerase, and from a second population of mammalian cells which do not contain said rare protein;

(ii) subjecting such mRNA to reverse-transcription and second-strand synthesis to form first and second cDNA preparations, said first and second cDNA preparations differing from one another with respect to presence or absence of cDNA molecules encoding said rare protein and a label incorporated into one of said first and second cDNA preparations;

(iii) combining said cDNA preparations under conditions such that complementary strands of cDNA from said first and second cDNA preparations anneal to form a mixture of double-stranded and single-stranded cDNA; and

(iv) separating cDNA comprising said label from cDNA that does not, thereby forming an isolated preparation of cDNA from said first population that has been depleted from complementary cDNA in said second population and enriched for said cDNA encoding said rare protein. Steps (iii) and (iv) of the above method can be repeated as often as desired, and the cDNA isolated after completion of step (iv) can be amplified by PCR, to provide cDNA preparations greatly enriched for the desired cDNA.

These and other embodiments of the invention will be described in detail below.

FIG. 1, in parts A, B, and C, is a bar graph showing the results of RT-PCR analysis using primers specific for TPC2 (FIG. 1A) or TPC3 (FIG. 1B) cDNA. In this and the other bar graphs, the number over each bar is the numerical result obtained; for RT-PCR results, this number was generated by scanning autoradiograms or PhosphorImager™ screens (Molecular Dynamics) of the RT-PCR products after gel electrophoresis. Under these test conditions, TPC2 and TPC3 mRNA is absent or detectable only at very low levels in the telomerase negative cell lines tested (labeled “Mortal” in the Figure) and detectable in all (most at clearly detectable levels) telomerase positive cell lines tested (labeled “Immortal” in the Figure). FIG. 1C shows TPC3 mRNA levels normalized to GAPDH levels and illustrates the difference in TPC3 mRNA levels between mortal and immortal cells (the spaces marked “0.0” are provided merely as breaks in the graphed data). GAPDH mRNA was used as a control; due to its greater abundance, the RT-PCR of the GAPDH samples was allowed to complete fewer cycles of PCR than used for the TPC2 or TPC3 samples.

FIG. 2, in parts A, B, and C, is a bar graph showing the results of an RT-PCR analysis of hTR RNA and TPC2 and TPC3 mRNA levels as well as telomerase activity in a variety of cell lines. FIG. 2A shows TPC2 and TPC3 mRNA levels normalized to GAPDH mRNA levels in various cell lines, all of which are telomerase positive except IMR-90, and demonstrates a correlation in the levels of these two telomere length and telomerase activity regulatory proteins. FIG. 2B shows how TPC3 mRNA levels correlate with telomerase activity (as measured using the TRAP assay) in a variety of cell lines. The IMR90, HTB-153, W1-38 VA13, KMSF, and T0 (unactivated T cells; note that T7 represents activated T cells) express no or only very low levels of telomerase activity. FIG. 2C shows how hTR RNA levels correlate with telomerase activity levels in a variety of cell lines. Taken together, these results show that TPC2 and TPC3 mRNA levels correlate with hTR levels and with telomerase activity levels in a variety of mortal and immortal cell lines.

FIG. 3 shows a restriction site and function map of the ~7.2 kb plasmid pGJR109, which contains an ~3.5 kb NotI-BstEII restriction fragment that contains an ~3.3 kb open reading frame encoding the TPC2 protein (labeled “ORF” and “TPC2”).

FIG. 4 lists portions of the nucleotide (SEQ ID NO: 1) sequence and deduced amino acid sequence (SEQ ID NO:2) of the TPC2 open reading frame corresponding to the human TPC2 gene, mRNA, and protein products. In the Figure, as well as throughout the specification and Figures, nucleotides and amino acids are represented using standard abbreviations and designations; however, ambiguous nucleotides are represented as shown in the key at the bottom of FIG. 4. The initiating methionine codon is believed to be at nucleotides 16-18 of the sequence; the termination codon is marked with “—”. 
FIG. 5 shows a restriction site and function map of the ~8 kb plasmid pGRN92, which contains an ~1.4 kb EcoRI-BamHI restriction fragment that contains an ~1.1 kb open reading frame encoding the TPC3 protein (labeled “ORF” and “TPC3”).

FIG. 6 lists the nucleotide sequence (SEQ ID NO: 4) and deduced amino acid sequence (SEQ ID NO: 4) of the TPC3 open reading frame corresponding to the human TPC3 gene, mRNA, and protein products. The initiating methionine codon is marked with “****” and the stop codon with “...”.

FIG. 7 shows the results of an analysis of telomerase activity levels in stable recombinant HeTe7 clones expressing the sense or antisense mRNA of gene TPC3 or a control vector. The recombinant sense TPC3 mRNA reduced telomerase activity markedly in these cells.

FIG. 8 shows the results of an analysis of telomere length in stable recombinant HeTe7 clones expressing the sense or antisense mRNA of gene TPC3 or a control vector. The recombinant TPC3 sense mRNA decreased telomere length (mean TRF) in the cells.

FIG. 9 lists the nucleotide sequence (SEQ ID NO: 5) of the hTR gene and corresponding RNA transcript; the sequence shown is that of one strand of an ~1 kb PsI restriction fragment that can be isolated from plasmid pGRN33. The sequence of the mature hTR transcript, which serves as the template in the telomerase ribonucleoprotein, is masked with ***—the 3’ end of the transcript is marked with an “>”.

These Figures are discussed in more detail below, where a variety of preferred embodiments of the invention are described.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention provides methods and reagents for regulating telomere length and modulating telomerase activity in mammalian cells as well as for detecting, diagnosing, and treating related diseases and conditions in humans and other mammals. To facilitate understanding and practice of the invention in its many and diverse applications, this description is organized as shown below.

i. DEFINITIONS

i. CLONING AND CHARACTERIZATION OF THE TPC2 AND TPC3 GENES

III. RECOMBINANT HOST CELLS

IV. Oligonucleotides and Nucleic Acids

V. Peptides and Proteins

VI. Antibodies

VII. Methods

VIII. Examples

I. Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by those of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, preferred methods and materials are described. For purposes of the present invention, the following terms are defined below.

“Antibody” refers to naturally occurring and recombinant polypeptides and proteins encoded by immunoglobulin genes, or fragments thereof, that specifically bind to or “recognize” an analyte or “antigen”. Immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant region genes, as well as the variable immunoglobulin variable region genes. An antibody can exist as an intact immunoglobulin or as any one of a number of well characterized fragments, e.g., Fab’ and F(ab)’, fragments, produced by various means, including recombinant methodology and digestion with various peptidases.

“cDNA” refers to deoxyribonucleic acids produced by reverse-transcription and typically second-strand synthesis of mRNA or other RNA produced by a gene; if double-stranded, a cDNA molecule has both a coding or sense and a non-coding or antisense strand. “Complementary to” refers to a polynucleotide sequence that can hybridize specifically to another polynucleotide sequence; for example, a nucleic acid comprising nucleotides in the sequence “5'-TATAAC" is complementary to a nucleic acid comprising nucleotides in the sequence "5'-GTATA".

“Corresponds to” or “corresponding to” refers to (i) a polynucleotide having a nucleotide sequence that is substantially identical or complementary to all or a portion of a reference polynucleotide sequence or encoding an amino acid sequence identical to a amino acid sequence in a peptide or protein; or (ii) a peptide or polypeptide having an amino acid sequence that is substantially identical to a sequence of amino acids in a reference peptide or protein.

“Encoding” refers to the inherent property of specific sequences of nucleotides in a nucleic acid, such as a gene in a chromosome or an mRNA, to serve as templates for synthesis of other polymers and macromolecules in biological processes having a defined sequence of nucleotides (i.e., tRNA, tRNA, other RNA molecules) or amino acids and the biological properties resulting therefrom. Thus a gene encodes a protein, if transcription and translation of mRNA produced by that gene produces the protein in a cell or other biological system. Both the coding strand, the nucleotide sequence of which is identical to the mRNA sequence and is usually provided in sequence listings, and non-coding strand, used as the template for transcription, of a gene or cDNA can be referred to as encoding the protein or other product of that gene or cDNA. A nucleic acid that encodes a protein includes any nucleic acids that have different nucleotide sequences but encode the same amino acid sequence of the protein due to the degeneracy of the genetic code. Nucleic acids and nucleotide sequences that encode proteins may include introns.

“Expression control sequence” refers to nucleotide sequences in nucleic acids that regulate the expression (transcription and/or translation) of a nucleotide sequence
operatively linked thereto. Expression control sequences can include, for example and without limitation, sequences of promoters, enhancers, transcription terminators, a start codon (i.e., ATG), splicing signals for introns, and stop codons.

[0059] “Immunoassay” refers to an assay that utilizes an antibody to bind an analyte specifically. An immunoassay is characterized by the use of specific binding properties of a particular antibody to isolate, target, and/or quantify the amount of an analyte.

[0060] “Label” or “labeled” refers to a detectable marker and to the incorporation of such a marker into a nucleic acid, protein, or other molecule. The label may be detectable directly, i.e., the label can be a radioisotope (e.g., $^{35}$I, $^{125}$I, $^{38}$S, $^{131}$I) or a fluorescent or phosphorescent molecule (e.g., FITC, rhodamine, lanthanide phosphors), or indirectly, i.e., by enzymatic activity (e.g., horseradish peroxidase, beta-galactosidase, luciferase, alkaline phosphatase) or ability to bind to another molecule (e.g., streptavidin, biotin, an epitope). Incorporation of a label can be achieved by a variety of means, i.e., by use of radiolabeled or biotinylated nucleotides in polymerase-mediated primer extension reactions, epitope-tagging, or binding to an antibody. Labels can be attached directly or via spacer arms of various lengths to reduce steric hindrance.

[0061] “Naturally occurring” refers to a substance, typically an amino acid, nucleotide, nucleic acid, or protein, that exists in nature without human intervention. For example, deoxyribonucleic acid or DNA is naturally occurring.

[0062] “Oligonucleotide” refers to a polymer composed of a multiplicity of nucleotide units (ribonucleotides or deoxyribonucleotides) or related structural variants or synthetic analogs thereof linked via phosphodiester (or related structural variants or synthetic analogs thereof. Thus, while the term “oligonucleotide” typically refers to a nucleotide polymer in which the nucleotides and the linkages between them are naturally occurring; the term also refers to various analogs, such as, for example and without limitation, peptide-nucleic acids (PNAs), phosphoramidates, phosphorothioates, methyl phosphonates, 2-O-methyl ribonucleic acids, and the like. An oligonucleotide typically rather short in length, generally from about 10 to 30 nucleotides, but the term can refer to molecules of any length, although the term “polynucleotide” or “nucleic acid” is typically used for large oligonucleotides.

[0063] “Open reading frame” refers to a nucleotide sequence that encodes a polypeptide or protein and is bordered on the 5’-end by an initiation codon (ATG) or another codon that does not encode a stop codon and on the 3’-end by a stop codon but otherwise does not contain any in-frame stop codons between the codons at the 5’-border and the 3’-border.

[0064] “Pharmaceutical composition” refers to a composition suitable for pharmaceutical use in a mammal. A pharmaceutical composition comprises a pharmaceutically effective amount of an active agent and a pharmaceutically acceptable carrier. “Pharmacologically effective amount” refers to that amount of an agent effective to produce the intended pharmacological result. “Pharmacologically acceptable carrier” refers to any of the standard pharmaceutical carriers, buffers, and excipients, such as a phosphate buffered saline solution, 5% aqueous solution of dextrose, and emulsions, such as an oil/water or water/oil emulsion, and various types of wetting agents and/or adjuvants. Suitable pharmaceutical carriers and formulations are described in Remington’s Pharmaceutical Sciences, 19th Ed. (Mack Publishing Co., Easton, 1995). Preferred pharmaceutical carriers depend upon the intended mode of administration of the active agent. Typical modes of administration include enteral (i.e., oral) or parenteral (i.e., subcutaneous, intramuscular, or intravenous intraperitoneal injection; or topical, transdermal, or transmucosal administration).

[0065] “Physiological conditions” refer to temperature, pH, ionic strength, viscosity, and like biochemical parameters that are compatible with a viable organism and/or that typically exist intracellularly in a viable mammalian cell. For example, the intracellular conditions in a mammalian cell grown under typical laboratory culture conditions are physiological conditions. Suitable in vitro reaction conditions for PCR and many polynucleotide enzymatic reactions and manipulations are generally physiological conditions. In general, in vitro physiological conditions comprise 50-200 mM NaCl or KCl, pH 6.5-8.5, 20-45 degrees C., and 0.001-10 mM divalent cation (e.g., Mg++, Ca++); preferably about 150 mM NaCl or KCl, pH 7.2-7.6, 5 mM divalent cation, and, often, including 0.01-1.0 percent nonspecific protein (e.g., BSA). A non-ionic detergent (Twee, NP-40 Triton X-100) can also be present, usually at about 0.001 to 2%, typically 0.05-0.2% (v/v). Particular aqueous conditions may be selected by the practitioner according to conventional methods. For general guidance, the following buffered aqueous conditions may be applicable: 10-250 mM NaCl, 5-50 mM Tris HCl, pH 5-8, with optional addition of divalent cation(s) and/or metal chelators and/or nonionic detergents and/or membrane fractions and/or antifoam agents and/or scintillants.

[0066] “Polynucleotide” or “nucleic acid” refers to an oligonucleotide and is typically used to refer to oligonucleotides greater than 30 nucleotides in length. Conventional notation is used herein to portray polynucleotide sequences: the left-hand end of single-stranded polynucleotide sequences is the 5’-end; the left-hand direction of double-stranded polynucleotide sequences is referred to as the 5’-direction. The direction of 5’ to 3’ addition of nucleotides to nascent RNA transcripts is referred to as the transcription direction; the DNA strand having the same sequence as the mRNA is referred to as the “coding strand”; sequences on the DNA strand having the same sequence as an mRNA transcribed from that DNA and which are located 5’ to the 5’-end of the RNA transcript are referred to as “upstream sequences”; sequences on the DNA strand having the same sequence as the RNA and which are 3’ to the 3’ end of the coding RNA transcript are referred to as “downstream sequences”. Polynucleotides and recombinantly produced protein, and fragments or analogs thereof, may be prepared according to methods known in the art and described in Maniatis et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., (1989), Cold Spring Harbor, N.Y., and Berger and Kimmel, Methods in Enzymology, Volume 152, Guide to Molecular Cloning Techniques (1987), Academic Press, Inc., San Diego, Calif., which are incorporated herein by reference.
“Polypeptide,” “peptide” and “protein” are used interchangeably herein to refer to a polymer of amino acid residues and to variants and synthetic analogs of the same. Thus, these terms apply to amino acid polymers in which one or more amino acid residues is a synthetic non-naturally occurring amino acid, such as a chemical analog of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. Conventional notation is used herein to portray polypeptide sequences: the left-hand end of polypeptide sequences is the amino-terminus, the right-hand end of polypeptide sequences is the carboxy-terminus. The term “recombinant protein” refers to a protein that is produced by expression of a recombinant DNA molecule that encodes the amino acid sequence of the protein. Terms used to describe sequence relationships between two or more polynucleotides or polypeptides include “reference sequence”, “comparison window”, “sequence identity”, “percentage of sequence identity”, and “substantial identity”. A “reference sequence” is a defined sequence used as a basis for a sequence comparison and may be a subset of a larger sequence, i.e., a complete cDNA, protein, or gene sequence. Generally, a reference sequence is at least 12 but frequently 15 to 18 and often at least 25 nucleotides (or other monomer unit) in length. Because two polynucleotides may each comprise (1) a sequence (i.e., only a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a “comparison window” to identify and compare local regions of sequence similarity. A “comparison window” refers to a conceptual segment of typically at least 12 contiguous residues that is compared to a reference sequence; the comparison window may comprise additions or deletions (i.e., gaps) of about 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by computerized implementations of algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, Wis.) or by inspection, and the best alignment (i.e., resulting in the highest percentage of homology over the comparison window) generated by any of the various methods is selected.

“Primer” refers to an oligonucleotide, i.e., a purified restriction fragment or a synthetic oligonucleotide, that is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product complementary to a nucleic acid strand (the “template”) is induced, i.e., in the presence of nucleotides and an agent for polymerization such as DNA polymerase and at a suitable temperature and pH. The primer is preferably single-stranded for maximum efficiency in amplification but may alternatively be double-stranded. If double stranded, the primer may need to be treated to separate its strands before being used to prepare extension products. Primers are typically oligodeoxynucleotides, but a wide variety of synthetic and non-naturally occurring oligonucleotide primers can be used for various applications. A primer must be sufficiently long to prime the synthesis of extension products in the presence of the agent for polymerization. The length of a primer depends on many factors, including application, temperature to be employed, template, reaction conditions, other reagents, and source of primers. For example, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15-25 or more nucleotides, although it may contain fewer nucleotides. Short primer molecules generally require cooler temperatures to form stable hybrid complexes with template. Primers can be large polynucleotides, such as from about 200 nucleotides to several kilobases or more. A primer must be substantially complementary to the sequence on the template to which it is designed to hybridize to serve as a site for the initiation of synthesis but need not reflect the exact sequence of the template. For example, non-complementary nucleotides may be attached to the 55-end of the primer, with the remainder of the primer sequence being complementary to the template. Alternatively, non-complementary nucleotides or longer sequences can be interspersed into a primer, provided that the primer sequence has sufficient complementarity with the sequence of the template to hybridize therewith and thereby form a template for synthesis of the extension product of the primer.

“Probe” refers to a molecule that binds to a specific sequence or subsequence or other moiety of another molecule. Unless otherwise indicated, the term “probe” typically refers to an oligonucleotide probe that binds to another nucleic acid, often called the “target nucleic acid”, through complementary base pairing. Probes may bind target nucleic acids lacking complete sequence complementarity with the probe, depending upon the stringency of the hybridization conditions. Probes can be directly or indirectly labeled.

“Recombinant” refers to methods and reagents in which nucleic acids synthesized or otherwise manipulated in vitro are used to produce gene products encoded by those nucleic acids in cells or other biological systems. For example, an amplified or assembled product polynucleotide may be inserted into a suitable DNA vector, such as a bacterial plasmid, and the plasmid can be used to transform a suitable host cell. The gene is then expressed in the host cell to produce the recombinant protein. The transformed host cell may be prokaryotic or eukaryotic, including bacterial, mammalian, yeast, Aspergillus, and insect cells. A recombinant polynucleotide may serve a non-coding function (e.g., promoter, origin of replication, ribosome-binding site, etc.) as well.

“Recombinant host cell” refers to a cell that comprises a recombinant nucleic acid molecule, typically a recombinant plasmid or other expression vector. Thus, for example, recombinant host cells can express genes that are not found within the native (non-recombinant) form of the cell.

“Selected from” refers, in connection with sequences, to one sequence sharing identity with another sequence.

“Sequence identity” refers to sequences that are identical (i.e., on a nucleotide-by-nucleotide or amino acid-by-amino acid basis) over the window of comparison. The term “percentage of sequence identity” is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched
positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity.

[0074] “Specifically binds to” refers to the ability of one molecule, typically a macromolecule such as an antibody or oligonucleotide, to contact and associate with another specific molecule even in the presence of many other diverse molecules. For example, a single-stranded nucleic acid can “specifically bind to” a single-stranded oligonucleotide that is complementary in sequence, and an antibody “specifically binds to” or “is specifically immunoreactive with” its corresponding antigen. Thus, under designated immunonassay conditions, an antibody binds preferentially to a particular protein and not in a significant amount to other proteins present in the sample. Specific binding to a protein under such conditions requires an antibody selected for its specificity for a particular protein. To select antibodies specifically immunoreactive with a particular protein, one can employ a variety of means, i.e., solid-phase ELISA immunonassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane (1988), Antibodies, A Laboratory Manual, Cold Spring Harbor Publications, N.Y.

[0075] “Specific hybridization” refers to the formation of hybrids between a probe nucleicotide (e.g., a polynucleotide of the invention which may include substitutions, deletions, and/or additions) and a specific target polynucleotide (e.g., a polynucleotide having the sequence of a TPC2 or TPC5 gene or gene product), wherein the probe preferentially hybridizes to the specific target and not to other polynucleotides in the mixture that do not share sequence identity with the target.

[0076] “Substantial identity” or “substantially identical” denotes a characteristic of a polynucleotide or polypeptide that comprises a sequence that is at least 80 percent identical, preferably at least 85 percent and often 90 to 95 percent identical, more usually at least 99 percent identical, to a reference sequence over a comparison window of at least 20 nucleotide positions, frequently over a window of at least 25 to 50 nucleotides, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence, which may include deletions or additions that total 20 percent or less of the reference sequence, over the window of comparison. The reference sequence may be a subset of a larger sequence.

[0077] “Stringent conditions” refer to temperature and ionic conditions used in nucleic acid hybridization. The stringency required is nucleotide sequence dependent and also depends upon the various components present during hybridization. Generally, stringent conditions are selected to be about 5 to 20 degrees C. lower than the thermal melting point Tm for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength and pH) at which 50% of a target sequence hybridizes to a complementary probe.

[0078] “Substantially pure” means an object species is the predominant species present (i.e., on a molar basis, more abundant than any other individual macromolecular species in the composition), and a substantially purified fraction is a composition wherein the object species comprises at least about 50 percent (on a molar basis) of all macromolecular species present. Generally, a substantially pure composition means that about 80 to 90 percent or more of the macromolecular species present in the composition is the purified species of interest. The object species is purified to essential homoegry (contaminant species cannot be detected in the composition by conventional detection methods) if the composition consists essentially of a single macromolecular species. Solvent species, small molecules (<500 Daltons), stabilizers (e.g., BSA), and elemental ion species are not considered macromolecular species for purposes of this definition.

[0079] “Suitable reaction conditions” are those conditions suitable for conducting a specified reaction using commercially available reagents. Such conditions are known or readily established by those of skill in the art for a variety of reactions. For example, suitable polymerase chain reaction (PCR) conditions include those conditions specified in U.S. Pat. Nos. 4,683,202; 4,683,195; 4,600,159; and 4,965,188, each of which is incorporated herein by reference. As one example and not to limit the invention, suitable reaction conditions can comprise: 0.2 mM each dNTP, 2.2 mM MgCl2, 50 mM KCl, 10 mM Tris-HCl, pH 9.0, and 0.1% Triton X-100.

[0080] “telomere length regulatory protein” and “telomerase regulatory protein” refers to polypeptides involved in telomere metabolism and telomerase activity. Such proteins include telomerase, the protein components of telomerase, proteins that selectively bind nucleic acids containing telomere repeat sequences or telomeric ends, proteins required for telomere repair, maintenance, and/or elongation, and proteins necessary for expression or formation of active telomerase enzyme. Although the present invention relates to such proteins generally, mammalian telomerase, and particularly human telomerase, and related proteins are provided as preferred embodiments.

[0081] “Telomerase activity” refers to the ability of telomerase protein components to associate with one another and the RNA component of telomerase either in vivo or in vitro into a multi-component enzyme that can elongate telomeric DNA. A preferred assay method for detecting telomerase activity is the TRAP assay. See PCT patent publication No. 95/13381, supra. This assay measures the amount of radioactive nucleotides incorporated into elongation products, polynucleotides, formed by nucleotide addition to a telomerase substrate or primer. The radioactivity incorporated can be measured as a function of the intensity of a band on a PhosphorImager™ screen exposed to a gel on which the radioactive products are separated. A test experiment and a control experiment can be compared by visually using the PhosphorImager™ screens. See also the commercially available TRAP-eze™, telomerase assay kit (Oncor); and Morin, 1989, Cell 59:521-529.

[0082] II. Cloning and Characterization of the TPC2 and TPC3 Genes

[0083] The present invention provides methods and reagents for regulating telomere length and modulating telomerase activity in mammalian cells as well as for detecting, diagnosing, and treating related diseases and conditions in humans and other mammals. The present invention arose in part out of an effort to clone the protein components of telomerase and other protein components of macromolecules that regulate telomere length and telom-
erase activity in human and other mammalian cells. These rare proteins and the mRNAs that encode these proteins are present in very low abundance in mammalian cells, necessi-
tating the use of a novel mRNA isolation and identification method called “subtraction hybridization differential display.”

In brief, this method involves obtaining mRNA from a first population of mammalian cells which contain the rare or low abundant protein of interest and from a second population of mammalian cells that contain 10- to 100-fold lower levels of the rare protein. The two mRNA populations are then individually used to generate cDNA preparations by reverse-transcription and second-strand syn-
thesis to form first and second cDNA preparations. A detect-
able label is incorporated as well into the second cDNA preparation. The two cDNA preparations are then denatured and combined under conditions such that complementary strands of cDNA from the two cDNA preparations anneal to form a mixture of double-stranded and single-stranded cDNA. The mixture of cDNAs is then separated into two different populations, one comprising the label and one that does not, thereby forming an isolated, unlabeled preparation of cDNA that has been enriched for cDNA encoding the rare protein of interest. The steps of hybridization and separation can be repeated as often as desired, and the cDNA isolated after the separation step can be amplified by PCR, to provide cDNA preparations greatly enriched for the desired cDNA. Typically after two cycles of subtraction, cDNAs corre-
sponding to abundant transcripts are depleted more than in 100-fold and low abundant transcripts are enriched in the subtracted cDNA libraries. The reproducibility of the method is excellent, and the method can be used to identify low abundant gene products such as those encoding telomere length and telomerase regulatory proteins.

To isolate cDNAs corresponding to telomere length and telomerase regulatory proteins, cDNA libraries were prepared from six different cells lines or tissues, three of which were “telomerase positive” (i.e., the cells express telomerase activity; the IDH and 293 cell lines, and testes tissue), and three of which were “telomerase negative” (i.e., the cells do not express telomerase activity; the HUVEC, BJ, and IMR-90 cell lines). These cDNA libraries were sub-
jected to subtraction hybridization against the telomerase negative HUVEC cDNA library. Then, differential display was performed by first replicating each of the six subtracted cDNA libraries with either a single 5’-arbitrary primer or in a PCR with a 5’-arbitrary primer and a 3’-poly(dT) primer, separating the replication products by gel electrophoresis, and identifying and isolating the differentially expressed products (identified visually as bands on a gel).

This process generated a number of differentially expressed cDNAs. Two of these cDNAs that were present in the cDNA libraries generated from the telomerase positive cell lines but not present (or present at much lower levels) in the telomerase negative cell lines, and that were later identified as originating from the 3’-ends of mRNA produced by the TPC2 and TPC3 genes, were isolated, cloned, and characterized by DNA sequence analysis. The DNA sequence analysis was used to design oligonucleotide primers that, in turn, were used to perform reverse-transcription and PCR (RT-PCR) on mRNA prepared from each of the same panel of six cell lines used to prepare the subtracted cDNA libraries. This RT-PCR experiment was designed to confirm that the mRNA corresponding to the putatively differentially expressed cDNAs is expressed at much higher levels in telomerase positive cell lines. The results were as predicted: the RT-PCR generated products of the predicted size; for the primers specific for the TPC2 mRNA, a substantial amount of product was generated using IDH1 mRNA, while lower amounts of product were generated using 293 and testes mRNA, and product was almost undetectable in mRNA prepared from HUVEC, BJ, and IMR-90 cells; for the primers specific for the TPC3 mRNA, product was generated only using mRNA from the telomerase positive cell lines.

To extend the analysis of the expression pattern of TPC2 and TPC3 in various cell lines and tissues, RT-PCR with primers specific for nucleotide sequences in the cDNAs corresponding to the differentially expressed TPC2 and TPC3 mRNAs was performed on a variety of cell lines. As a control, RT-PCR with primers specific for nucleotide sequences in GAPDH mRNA (GAPDH is a “house-keeping” enzyme present in both telomerase positive and telomerase negative cell lines) was performed as well. In brief, the primers used for TPC2 were:

the primers used for TPC2 were:

\[
tpo-p1: 5'-ATGCGGATTCCAGGTTGACCT-3', \quad (SEQ ID NO: 6)
\]
\[
tpo-p4: 5'-ACCTGCTTCAGGGGCGGACT-3', \quad (SEQ ID NO: 7)
\]
and the primers used for TPC3 were:

\[
tpo-p13: 5'-TAAGACAAAGAACAAGTCACACNA-3', \quad (SEQ ID NO: 8)
\]
\[
tpo-p14: 5'-ATGGTCTTTAGAGTGTTGCCAG-3'. \quad (SEQ ID NO: 9)
\]

The RT-PCR was performed by making first strand cDNA made from total RNA with random hexamer primers and then PCR-amplifying the single-stranded cDNA with one of the two primer sets above, following the protocol of 16 to 28 cycles of PCR amplification (typically, 16 cycles for GAPDH mRNA, 25 cycles for TPC2 mRNA, and 27 cycles for TPC3 mRNA), with each cycle consisting of a step at 94 degrees C. for 45 sec., 65 degrees C. for 45 sec., and 72 degrees C. for 90 sec. Other illustrative RT-PCR primers and conditions are shown in Parts C and D of the Examples below.

FIG. 1A, in parts A, B, and C, shows the results of RT-PCR analysis using primers specific for the TPC2 (FIG. 1A) or TPC3 (FIG. 1B) cDNA. Under these test conditions, TPC2 and TPC3 mRNA is absent or detectable only at very low levels in the telomerase negative cell lines tested (labeled “Mortal” in the Figure) and detectable in all (most at clearly detectable levels) telomerase positive cell lines tested (labeled “Immortal” in the Figure). These results, which show that TPC2 and TPC3 mRNA is present in testes tissue as well as most tumor cell lines but absent or present at lower abundance in normal cell lines, demonstrate how the methods of the invention for detecting and quantitating TPC2 and/or TPC3 gene products can be used to detect immortal cells, especially telomerase positive cancer cells, and to diagnose cancer and other diseases and conditions in humans and other mammals. FIG. 1C shows TPC3 mRNA levels normalized to GAPDH levels and illustrates
the clear difference in TPC3 mRNA levels between mortal and immortal cells. This RT-PCR analysis also indicated that, as expected, the TPC2 and TPC3 mRNA is present in very low abundance even in telomerase positive cells TPC2 or TPC3 mRNA amplification products detected after ~25 cycles; GAPDH or HPRT detected after ~15 or ~20 cycles, respectively). Confirmatory evidence for the low abundance of TPC2 mRNA in telomerase positive cells was obtained in the cloning of a cDNA corresponding to one-half of the full length TPC2 mRNA, where a primary screen of a lambda GT11 cDNA library from telomerase positive 293 cells showed that only one of ~1.4 million plaques was positive, indicating a very rare transcript.

[0090] FIG. 2, in parts A, B, and C, is a bar graph showing the results of an RT-PCR analysis of hTR RNA and TPC2 and TPC3 mRNA levels as well as telomerase activity in a variety of cell lines. FIG. 2A shows TPC2 and TPC3 mRNA levels normalized to GAPDH mRNA levels in various cell lines, all of which are telomerase positive except IMR-90, and demonstrates a correlation in the levels of these two telomere length and telomerase activity regulatory proteins. FIG. 2B shows how TPC2 mRNA levels correlate with telomerase activity levels in a variety of cell lines. The IMR90, HTB-153, WI-38 VA13, KMSF, and TO (unactivated T cells; note that T7 represents activated T cells) express no or only very low levels of telomerase activity. FIG. 2C shows how hTR RNA levels correlate with telomerase activity levels in a variety of cell lines. The RT-PCR protocol for hTR RNA is described in Part D of the Examples; the nucleotide sequence of the hTR gene and transcribed RNA is shown in FIG. 9.

[0091] Taken together, these FIGS. show that TPC2 and TPC3 mRNA levels as well as hTR levels correlate with telomerase activity levels in a variety of mortal and immortal cell lines. These results demonstrate how the methods of the invention for detecting TPC2 or TPC3 gene products can be used to detect immortal cells, especially telomerase positive cancer cells, and so to diagnose cancer and other diseases and conditions in humans and other mammals. These results also demonstrate the utility of the methods of the invention in which the detection or quantitation of TPC2 or TPC3 gene products, together with measurements of other factors, shTR levels, can length, telomerase activity, or hTR levels, can be used to identify immortal cells, such as cancer cells, or to evaluate the proliferative capacity of a cell.

[0092] The absence or very low abundance of the TPC2 and TPC3 gene products in telomerase negative mortal cells and the low but clearly detectable abundance of those gene products in telomerase positive immortal cells demonstrate the utility of the methods and reagents of the invention for detecting the presence gene products that encode proteins such as the protein components of telomerase and other proteins that regulate telomere length and telomerase activity in mammalian cells. A comparison of telomere length by mean terminal restriction fragment (mean TRF) analysis of immortal cell lines with TPC2 mRNA levels indicates that TPC2 mRNA levels are inversely related to telomere length. In one test, ten immortal cell lines with relatively high TPC2 mRNA levels had mean TRFs of ~2.5 to 5.0 kb, whereas two immortal cell lines with very low TPC2 mRNA levels had mean-TRFs of ~17.5 to 35 kb (probability of this difference arising by chance is less than 1%). In general, TPC2 mRNA levels also correlate well with telomerase activity levels in most cell lines tested.

[0093] Tests such as those described above can also be used to determine the mechanism of action by which the TPC2 and TPC3 gene products serve to regulate telomere length and telomerase activity. The tests on TPC2 provide some indication that the TPC2 gene product functions, at least in part, by acting as an indicator of telomere length, much like the yeast EST1 protein. TPC2 is up-regulated in most tumor cell lines and in testes cells and down-regulated in normal cell lines. However, some cell lines with apparently high levels of telomerase activity and very long telomeres have low levels of TPC2 mRNA. As noted above, however, telomerase positive cell lines that have relatively low TPC2 levels also have relatively high mean TRFs, i.e., skin melanoma LOX (~35.2 kb TRF), testes embryonic carcinoma Tera-1 (~27.0 kb), lung carcinoma NCI-H23 (~17.5 kb). In contrast, skin melanoma lines SK-MEL2 (~2.3 kb), SK-MEL28 (~15.7 kb), SK-MEL5 (~5.0 kb) and testis tissue (~15 kb) have relatively lower mean TRFs and relatively higher TPC2 mRNA levels. Because all of these cell lines have relatively high telomerase activity and high hTR levels, the tests indicate that cell lines with relatively long telomeres in general have low TPC2 mRNA levels, suggesting that the TPC2 protein may encode a protein with a telomere-sensing function. The analysis of TPC3 mRNA levels and telomerase activity in the same cell lines indicates that the TPC3 gene product may act as a core component of the telomerase enzyme.

[0094] Significant additional information regarding the mechanism of action of the TPC2 and TPC3 gene products in the regulation of telomere length and telomerase activity can be derived by analysis of the nucleotide sequence and corresponding amino acid sequence of the open reading frames of the corresponding genes. The subtraction hybridization differential display identification and cloning generated only cDNAs corresponding to the 3’-ends of the TPC2 and TPC3 mRNA gene products, but the nucleotide sequence information generated from those cDNAs provided a means to attempt to identify and isolate clones in cDNA libraries prepared from telomerase positive cell lines that comprise additional portions of the mRNA.

[0095] Full length cDNA for the TPC2 and TPC3 gene products was obtained by a variety of methods, including the screening of subtracted and other specialized libraries and the use of 5’-RACE. Initially, a lambda GT11 cDNA library containing human cDNA from 293 cells (a telomerase positive human-transformed kidney cell line available from ATCC) was screened to identify lambda clones that hybridized to the short TPC2 and TPC3 cDNAs obtained by subtraction hybridization differential display. Then, after screening additional cDNA libraries and combining fragments from various subclones, full length open reading frames and genes were assembled into the plasmids pGRN92 (comprises the open reading frame of the TPC3 gene) and pGRN109 (comprises the open reading frame of the TPC2 gene).

[0096] For example, for TPC2, cDNA inserts in lambda clones were identified by screening with TPC2-specific probes and subcloned into plasmid pGEX and derivative vectors (Pharmacia) to yield plasmids that contained TPC2...
cDNA in various reading frames to test expression products and obtain partial nucleotide sequence and deduced amino acid sequence information about the open reading frame of the TPC2 mRNA. In the case of TPC3, for example, cDNA fragments were cloned into pBluescript Iks vector (Stratagene) to generate vectors for sequencing and analysis.

[0097] FIG. 3 shows a restriction site and function map of the ~7.2 kb plasmid pGRN109, which contains an ~3.5 kb NotI-BstEII restriction fragment that contains an ~3.3 kb open reading frame encoding the TPC2 protein (labeled “ORF” and “TPC2”). FIG. 4 lists portions of the nucleotide sequence and deduced amino acid sequence of the TPC2 open reading frame corresponding to the human TPC2 gene, mRNA, and protein products. FIG. 5 shows a restriction site and function map of the ~8 kb plasmid pGRN92, which contains an ~1.4 kb EcoRI-BamHI restriction fragment that contains an ~1.1 kb open reading frame encoding the TPC3 protein (labeled “ORF” and “TPC3”). FIG. 6 lists the nucleotide sequence and deduced amino acid sequence of the TPC3 open reading frame corresponding to the human TPC3 gene, mRNA, and protein products. The initiating methionine codon is marked with “****” and the stop codon with “—”. Plasmid pGRN92 does not comprise nucleotides 1-82 shown in FIG. 6.

[0098] Neither the TPC2 nor the TPC3 open reading frame or other gene sequences show significant homology to sequences in public databases other than to ESTs; however, both have motif signatures. TPC2 contains two WW domains and one L22 signature domain; TPC3 contains a homeobox domain. The “homeobox” is a protein domain of 60 amino acids (see Gehring, 1992, Trends Biochem. Sci. 17:277-280) first identified in a number of Drosophila homeotic and segmentation proteins and since found to be extremely well conserved in many animals, including vertebrates. This domain binds DNA through a helix-turn-helix type of structure. Proteins that contain homeobox domains are likely to play a role in development; most are known to be sequence specific DNA-binding transcription factors. Recent publications suggest that homeobox domains can bind RNA as well. See Dubnau and Struhl, Feb. 22, 1996, Nature 379:694. The homeobox domain in TPC3 is: LAMCTNLPEARQVYWWFKRRKAFK (SEQ ID NO: 10).

[0099] TPC2 contains two WW domains and an L22 ribosomal RNA signature domain. The ribosomal protein L22 is a protein component of the large ribosomal subunit that, in E. coli, binds 23S rRNA; the protein belongs to a family of ribosomal proteins. See Gantt et al., 1991, EMBO J. 10:3073-3078. For TPC2, this domain is: SSSKVTSHSF-GKRDQAIRKPNPVV (SEQ ID NO: 11). The WW domain, also known as RSP5 or WWP, is a short conserved region in a number of unrelated proteins, among them dystrophin, responsible for Duchenne muscular dystrophy. The domain spans about 35 residues, can be repeated up to 4 times in some proteins, and has been shown to bind proteins with particular proline-motifs, >AP1-P-P-P-AP1- Y-AP1-P-P-P-AP1 (SEQ ID NO: 12), and so somewhat resembles SH3 domains. The WW domain is frequently associated with other proteins in signal transduction processes and appears to contain beta-strands grouped around four conserved aromatic positions, generally Trp; the name WWP derives from the presence of these conserved Trp and Pro residues. For TPC2, this domain is represented by three amino acid residue sequences:

WGVCCRQGDVPFINDQCLCTWLEIP; (SEQ ID NO: 13)
WFLADVCFYTTAAXEERKESXEXXIP (SEQ ID NO: 14)
and
WEGSPTGESAVYFDHQRQCTAFFHP. (SEQ ID NO: 15)

[0100] The availability of plasmids encoding the TPC2 and TPC3 open reading frames provides a wide variety of benefits, including the benefit of recombinant host cells that express recombinant gene products comprising TPC2 and/or TPC3 open reading frame sequences or sequences encoding products that react specifically with TPC2 and/or TPC3 gene products.

[0101] III. Recombinant Host Cells

[0102] In one embodiment, the invention provides recombinant mammalian host cells containing:

[0103] (i) a recombinant or synthetic nucleic acid comprising at least about 10 to 15 to 25 to 100 or more contiguous nucleotides corresponding to an open reading frame sequence of a human gene TPC2 contained in a human DNA insert of an ~5.3 kb NotI-BstEII restriction fragment of plasmid pGRN109; or

[0104] a synthetic or recombinant peptide or protein comprising at least about 6 to 10 to 15 to 25 to 100 or more contiguous amino acids corresponding to an amino acid sequence encoded by said open reading frame sequence; and

[0105] (ii) a recombinant or synthetic nucleic acid comprising at least about 10 to 15 to 25 to 100 or more contiguous nucleotides corresponding to an open reading frame sequence of a human gene TPC3 contained in a human DNA insert of an ~1.4 kb EcoRI-BamHI restriction fragment of plasmid pGRN92; or

[0106] a synthetic or recombinant peptide or protein comprising at least about 6 to 10 to 15 to 25 to 100 or more contiguous amino acids corresponding to an amino acid sequence encoded by said open reading frame sequence of gene TPC3;

[0107] said TPC2 and TPC3 genes characterized in coding for proteins that regulate telomere length or modulate telomerase activity and are present in human or other mammalian cells that express telomerase activity.

[0108] Other mammalian host cells provided by the invention include those that comprise either or both TPC2- and TPC3-derived recombinant or synthetic nucleic acids, peptides, or proteins. Furthermore, the invention also provides such cells further modified to contain a synthetic or recombinant nucleic acid comprising at least about 10 to 15 to 25 to 100 or more contiguous nucleotides corresponding to a contiguous nucleotide sequence of human hTR located in an ~2.5 kb HindIII-SacI restriction fragment of pGRN33 (ATCC 75926).

[0109] The recombinant host cells of the invention have application in many useful methods also provided by the invention. For example, the invention provides recombinant
host cells comprising novel expression vectors with expression control sequences operatively linked to nucleotide sequences encoding amino acids in a sequence substantially identical to the proteins encoded by the human TP2C or TP3C genes, optionally with a recombinant hTTR gene as well. These recombinant host cells are useful for producing recombinant human telomerase, for use in screens to identify agents that modulate telomerase activity or regulate telomere length, as well as for a variety of other purposes described below. The recombinant host cells of the invention can also be incorporated into the germ line and/or somatic tissues of non-human transgenic mammals, as well as be administered to mammals for therapeutic purposes.

[0110] Thus, genomic clones of a gene that regulates telomere length or telomerase activity, such as the human TP2C or TP3C gene, or recombinant versions thereof, including versions that encode mulein TP2C or TP3C gene products, may be used to construct homologous targeting constructs for generating cells and transgenic nonhuman animals having at least one functionally disrupted (or otherwise altered) allele. Guidance for construction of homologous targeting constructs may be found in the art, including: Rahentulla et al., 1991, Nature 353:180; Jasmin et al., 1990, Genes Devel. 4:157; Koh et al., 1992, Science 256:1210; Molina et al., 1992, Nature 357:161; Grusby et al., 1991, Science 253:1417; and Bradley et al., 1992, Bio/Technology 10:534. See also U.S. Pat. Nos. 5,464,764 and 5,487,992. Transgenic cells and/or transgenic non-human animals may be used to screen for antineoplastic agents and/or to screen for potential carcinogens, as inappropriate expression of a protein that regulates telomere length or telomerase activity may result in a pre-neoplastic or neoplastic state or other disease state or condition. Homologous targeting can be used to generate so-called “knockout” mice, which are heterozygous or homozygous for an inactivated allele. Such mice may be sold commercially as research animals for investigation of immune system development, neoplasia, spermatogenesis, or as pets, or for animal products (food-stuff), or other purposes.


[0112] Additionally, a TP2C or TP3C cDNA or genomic clone may be used to construct transgenes for expressing polypeptides at high levels and/or under the transcriptional control of transcription control sequences which do not naturally occur adjacent to the gene (or vice-versa, i.e., the promoter of the TP2C or TP3C gene is positioned in front of a reporter gene for use in screening or other use). For example but not limitation, a constitutive promoter (e.g., an HSV-tk or pgk (phosphoglycerate kinase) promoter) or a cell-lineage specific transcriptional regulatory sequence (e.g., an CD4 or CD8 gene promoter/ enhancer) may be operably linked to a protein encoding nucleotide sequence to form a transgene (typically in combination with a selectable marker such as a neo gene expression cassette). Such transgenes can be introduced into cells (e.g., ES cells, hematopoietic stem cells, cancer cells), and transgenic cells, cell lines, and transgenic nonhuman animals may be obtained according to conventional methods therewith.

[0113] The recombinant host cells of the invention are often prepared using, or serve as a source of, valuable oligonucleotide and nucleic acid reagents provided by the present invention, such as the expression control vectors noted above. These nucleic acid reagents are described in more detail in the following section.

[0114] IV. Oligonucleotides and Nucleic Acids

[0115] In another embodiment, the invention provides synthetic and recombinant oligonucleotides and nucleic acids in a variety of forms, i.e., isolated, isolated, purified, or substantially pure, and for a variety of purposes, i.e., as probes or primers, as polynucleotide plasmids and vectors for introducing recombinant gene products that regulate telomere length or modulate telomerase activity in mammalian host cells, as restriction fragments for creating useful nucleic acids, and as reagents for therapeutic, diagnostic, and other applications. Isolated or purified polynucleotides of the invention typically are less than ~10 kb in size. In particular, the invention provides recombinant or synthetic nucleic acids comprising at least about 10 to 15 to 25 to 100 or more contiguous nucleotides substantially identical or complementary in sequence to a contiguous nucleotide sequence located in either:

[0116] (i) an open reading frame sequence of a human gene TP2C contained in a human DNA insert of an ~3.5 kb NotI-BstEII restriction fragment of plasmid pGR1109; or

[0117] (ii) an open reading frame sequence of a human gene TP3C contained in a human DNA insert of an ~1.4 kb EcoRI-BamHI restriction fragment of plasmid pGR992.

[0118] The novel oligonucleotide probes and primers of the invention typically comprise nucleotides in a sequence substantially identical or complementary to a sequence of nucleotides in a TP2C or TP3C gene or gene product, that allow specific hybridization thereto in a complex mixture of nucleic acids. Nucleotide substitutions, deletions, and additions may be incorporated into the polynucleotides of the invention. Nucleotide sequence variation may result from sequence polymorphisms of various alleles, minor sequencing errors, and the like. The minimum length of a polynucleotide required for specific hybridization to a target sequence depends on several factors: G/C content, positioning of mismatched bases (if any), degree of uniqueness of the sequence as compared to the population of target polynucleotides, and chemical nature of the polynucleotide (e.g., methylphosphonate backbone, polyamide nucleic acid, phosphorothioate, etc.), among others.

[0119] The probes and primers of the invention have useful application in a variety of diagnostic, therapeutic, and other applications. Because they are expressed differentially between immortal human cells lines, TP2C and TP3C genes and gene products serve as telomerase activity and tumor cell markers.
[0120] Oligonucleotides corresponding to unique TPC2 or TPC3 gene sequences can be used as primers or probes, may be attached to other nucleic acids, proteins, labels, etc., and are useful for a variety of purposes, including, for example, as diagnostic probes for tumor cells in clinical specimens. The oligonucleotides of the invention can be used as hybridization probes or PCR primers to detect the presence of TPC2 or TPC3 gene products, to diagnose a neoplastic disease characterized by the presence of an elevated or reduced TPC2 or TPC3 mRNA level in cells, to perform tissue typing (i.e., identify tissues characterized by the expression of telomerase or TPC2 or TPC3 mRNA), and to the like. Probes can be used to detect TPC2 or TPC3-specific nucleotide sequences in a DNA sample, such as for forensic DNA analysis or for diagnosis of diseases characterized by amplification, alteration, and/or rearrangements of the TPC2 or TPC3 genes. Certain preferred oligonucleotides of the invention typically comprise at least 8 to 10 to 15 to 25 to 99 to 250 to 1000 or more contiguous nucleotides capable of hybridizing under stringent hybridization conditions to nucleic acids corresponding to a nucleotide sequence in the 3.5 kb NotI-BstEII insert of pGRN109 or the 1.4 kb EcoRI-BamHI insert of pGRN92 and are useful as probes, primers, antisense therapeutics, and ribozyme therapeutics, for example.

[0121] Where expression of a polypeptide is not desired, polynucleotides of this invention need not encode a functional protein. Polynucleotides of this invention may serve as hybridization probes and/or PCR primers and/or LCR oligomers for detecting RNA or DNA sequences. Alternatively, polynucleotides of this invention may serve as hybridization probes or primers for detecting RNA or DNA sequences of related genes, for example, genes that encode structurally or evolutionarily related proteins. For such hybridization and other applications, such as those involving PCR, the polynucleotides of the invention need not encode a functional polypeptide. Thus, certain polynucleotides of the invention may contain substantial deletions, additions, nucleotide substitutions, and/or transpositions, so long as the ability of specific hybridization to or specific amplification of a TPC2 or TPC3 gene or mRNA gene product is retained.

[0122] As one example, antisense polynucleotides can include nucleotide substitutions, additions, deletions, or transpositions, so long as specific hybridization to the relevant target sequence, typically an mRNA, is retained as a functional property of the polynucleotide. Complementary antisense polynucleotides include soluble antisense DNA or RNA oligonucleotides that can hybridize specifically to mRNA species and genes and so prevent either transcription of the gene to produce the mRNA and/or translation of the mRNA. Antisense polynucleotides of various lengths may be used, although such antisense polynucleotides typically comprise a sequence of at least about 25 consecutive nucleotides that are substantially identical to a naturally occurring TPC2 or TPC3 gene sequence. Antisense polynucleotides may be produced from a heterologous expression cassette in a transfectant cell or transgenic cell, such as a transgenic pluripotent hematopoietic stem cell used to reconstitute all or part of the hematopoietic stem cell population of an individual. Alternatively, the antisense polynucleotides may comprise soluble oligonucleotides that are administered to the external milieu, either in the culture medium in vitro or in the circulatory system or interstitial fluid in vivo. Soluble antisense polynucleotides present in the external milieu have been shown to gain access to the cytoplasm and inhibit translation of specific mRNA species. In some embodiments the antisense polynucleotides comprise methylphosphonate or other synthetic moieties. For general methods relating to antisense polynucleotides, see Antisense RNA and DNA (1988), D. A. Melton, Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

[0123] The inhibitory nucleic acid also can be a so-called "sense" or other nucleic acid, i.e., a triplex-forming nucleic acid. As one example, expression of recombinant TPC3 mRNA in a cancer cell line resulted in the inhibition of telomerase activity by over 90%. In this example, the entire 1.1 kb coding sequence of the TPC3 gene was isolated as an EcoRI fragment (−2.1 kb) from vector pTA6PC3.9 and inserted into the EcoRI site of mammalian expression vector pBR32218 to give rise to two vectors: pGRN111, in which the sense strand of the TPC3 gene is operatively linked to the myelo proliferative sarcoma virus (MPSV) promoter, and pGRN112, in which the antisense strand is operatively linked to the MPSV promoter. Vector pTA6PC3.9 was constructed by ligation of TPC3 5'-RACE product (−2.1 kb) into pCRRII vector (Invitrogen). The sense and antisense vectors, as well as control vector pBR32218, were used to transform HeLaT7 cells by electroporation. The medium was changed to selection medium containing hygromycin (300 g/ml) and puromycin (0.2 g/ml) for four weeks to obtain individual clones. The individual clones were then isolated, expanded, and assayed for the expression of sense or anti- sense TPC3 gene product and vector transcription by RTPCR. The positive clones were then assayed for telomerase activity using the TRAP assay, and mean TRF values were measured at different time points.

[0124] FIG. 7 shows the results of the analysis of telomerase activity levels in recombinant HeLaT7 cells expressing the sense or antisense mRNA of gene TPC3 or a control vector. As noted above, presence of the recombinant sense mRNA reduced telomerase activity markedly in these cells. FIG. 8 shows the results of the analysis of telomere length in recombinant HeLaT7 cells expressing the sense or anti- sense mRNA of gene TPC3 or a control vector. The recombinant TPC3 sense mRNA decreased the mean TRF in the cells. Thus, the recombinant TPC3 gene product can regulate not only telomerase activity but also telomere length in these cells. This experiment shows how the recombinant nucleic acids of the invention can be expressed by transfecting the cell with an expression vector comprising expression control sequences operatively linked thereto. Fragments or analogs of TPC2 or TPC3 can also be expressed and function to compete with other active components of enzymes that regulate telomere length or telomerase activity. Assembly of ribonucleoproteins or other macromolecules with non-functional components results in non-functional complexes and subsequent decrease in associated activity, i.e., telomerase activity, telomere maintenance, and telomere length.

[0125] The expression vectors of the invention typically comprise expression control sequences operatively linked to a nucleotide sequence encoding amino acids in a sequence identical to a sequence of amino acids in a TPC2 or TPC3 protein gene product. The operably linked nucleotide sequence typically encodes at least 5 to 9 amino acids, or encodes all of or at least an active portion of the TPC2 or TPC3 proteins, or encode from 15 to 20 to 25 to 100 or more contiguous amino acids in a sequence selected from the
amino acid sequences of TPC2 or TPC3, or variant but related sequences thereto. For example, useful TPC2 and TPC3 variant proteins include fusion proteins, in which all or a portion of the TPC2 or TPC3 protein is fused to peptide or polypeptide that imparts some useful feature, such as a binding site for use in affinity purification, i.e., a polyhistidine tag of about six histidine residues or the maltose binding protein. Preferably, these amino acid sequences occur in the given order of the naturally occurring proteins (in the amino-terminal to carboxy-terminal orientation) but may comprise other intervening and/or terminal sequences; generally such polypeptides are less than 1000 amino acids in length, more usually less than about 500 amino acids in length, and frequently about 200 amino acids in length. The degeneracy of the genetic code gives a finite set of polynucleotide sequences encoding these amino acid sequences; this set of degenerate sequences may be readily generated by hand or by computer using commercially available software (Wisconsin Genetics Software Package Release 7.0). These and other expression vectors of the invention have many useful applications, including in therapeutic methods of the invention as gene therapy vectors for modulating telomerase activity, either to activate or inhibit that activity, or for regulating telomere length, either to increase or decrease the length, in a target cell or tissue.

[0126] Thus, the gene therapy expression vectors of the invention include those that encode or express, in part or whole, the TPC2 and/or TPC3 proteins, i.e., express proteins that differ from TPC2 and/or TPC3 by deletion, substitution, and/or addition of one or more amino acids. The gene therapy vectors of the invention may also, however, encode other useful nucleic acids, such as hTR, or antisense nucleic acids or ribozymes that target the TPC2, TPC3, and/or hTR gene products, i.e., mRNA and telomerase RNA. The vectors of the invention can also code for the expression of a protein which, when presented as an immunogen, elicits the production of an antibody that specifically binds to TPC2 or TPC3 proteins or cells expressing those proteins. Such vectors can also code for a structurally-related protein, such as a TPC2 or TPC3 protein fragment or analog. These vectors are useful in the therapeutic methods of the invention for treating or preventing diseases or conditions in which modulation of telomerase activity or telomere length can be of benefit. For example, in telomerase positive cancer cells, inhibition of telomerase activity can prevent telomere maintenance in those cells, inducing upon continued proliferation telomere loss, cell crisis, and death. For such purposes, the gene therapy vectors of the invention that express a non-functional TPC2 or TPC3 mimic or variant protein or other nucleic acid (i.e., over expression of TPC3 mRNA) that can inhibit telomerase formation or telomere elongation by telomerase activity in the cell, such as by competing for RNA component or protein components, inhibition of endogenous gene expression, or other means, are preferred.

[0127] Expression vectors of the invention comprise expression and replication signals compatible with the host cell of interest, i.e., sequences that facilitate transcription and translation (expression sequences) of the coding sequences, such that the encoded polypeptide product is produced. Construction of such polynucleotides is well known in the art and is described further in Muniatis et al., supra. For example, but not for limitation, such polynucleotides can include a promoter, a transcription termination site (polyadenylation site in eukaryotic expression hosts), a ribosome binding site, and, optionally, an enhancer for use in eukaryotic expression hosts, and, optionally, sequences necessary for replication of a vector. A typical eukaryotic expression cassette will include a polynucleotide sequence encoding a polypeptide linked downstream (i.e., in translational reading frame orientation; polynucleotide linkage) of a promoter such as the HSV tk, pgk, metallothionein, or any of a wide variety of other promoters suitable for use in mammalian cells, optionally linked to an enhancer and a downstream polyadenylation site (e.g., an SV40 large T Ag poly A addition site). Expression vectors useful for expressing the recombinant TPC2, TPC3, and other proteins of this invention include viral vectors such as retroviruses, adenoviruses and adenovirus-associated viruses, i.e., for therapeutic methods, plasmid vectors such as pCDNA1 (Invitrogen, San Diego, Calif.), in which the expression control sequence comprises the CMV promoter, cosmids, liposomes, and the like. Viral and plasmid vectors are often preferred for transfecting mammalian cells.

[0128] The nucleic acid reagents of the invention also include reagents useful in identifying, isolating, and cloning nucleic acids that encode proteins that interact with TPC2 and TPC3 gene products as well as mammalian (i.e., mouse) homologs of human TPC2 and TPC3 genes. Homologous DNA can be readily identified by screening a genomic or cDNA clone library prepared from the mammalian cells of interest, such as a mouse, rat, rabbit, or other cells, i.e., in yeast artificial chromosomes, cosmid, or bacteriophage lambda (e.g., Charon 35), with a polynucleotide probe comprising a sequence of about at least 24 (or in the range of 15 to 30 or more) contiguous nucleotides (or their complement) of the cDNA sequences of TPC2 or TPC3 disclosed herein. Typically, hybridization and washing conditions are performed at varying degrees of stringency according to conventional hybridization procedures. Positive clones are isolated and sequenced. For illustration and not for limitation, a full length polynucleotide corresponding to the open reading frame sequence of the TPC2 and TPC3 genes can be labeled and used as a hybridization probe to isolate genomic clones from a murine or other mammalian genomic clone or cDNA library (i.e., those available from Promega Corporation, Madison, Wis.).

[0129] The nucleic acids of the invention can also be employed to isolate and identify gene products that interact with or bind to TPC2 and/or TPC3 gene products. The yeast "two-hybrid" system (see Chien et al., 1991, Proc. Natl. Acad. Sci. (U.S.A.) 88:9578) utilizes expression vectors that encode the predetermined polypeptide sequence as a fusion protein and is used to identify protein-protein interactions in vivo through reconstitution of a transcriptional activator (see Fields and Song, 1989, Nature 340:245). Usually the yeast Gal4 transcription protein, which consists of separable domains responsible for DNA-binding and transcriptional activation, serves as the transcriptional activator. Polynucleotides encoding two hybrid proteins, one consisting of the yeast Gal4 DNA-binding domain fused to a polypeptide sequence of a first protein and the other consisting of the Gal4 activation domain fused to a polypeptide sequence of a second protein (either the first or second protein typically is a number of different proteins to be screened for ability to interact specifically with the other protein), are constructed and introduced into a yeast host cell. Intermolecular binding, if any, between the two fusion proteins reconstitutes the Gal4 DNA-binding domain with the Gal4 activation
domain, which leads to the transcriptional activation of a reporter gene (e.g., lacZ, HIS3) operably linked to the Gal4 binding site. Typically, the two-hybrid method is used to identify novel polypeptide sequences which interact with a known protein.

[0130] The invention also provides two- and three-hybrid systems, typically in the form of polynucleotides encoding a first hybrid protein comprising either TPC2 or TPC3, a second hybrid protein, and a reporter gene, wherein said polynucleotide(s) are either stably replicated or introduced for transient expression. The host organism can be a yeast cell (e.g., Saccharomyces cerevisiae) in which the reporter gene transcriptional regulatory sequence comprises a Gal4-responsive promoter (binding site). Yeast cells comprising (1) an expression cassette encoding a Gal4 DNA binding domain (or Gal4 activator domain) fused to a binding fragment of TPC2 or TPC3 protein; (2) an expression cassette encoding a Gal4 DNA activator domain (or Gal4 binding domain, respectively) fused to a member of a cDNA library; and (3) a reporter gene (e.g., betagalactosidase) comprising a cis-linked Gal4 transcriptional response element, can be screened to identify those cDNAs that bind to TPC2 and/or TPC3 proteins specifically. Yeast two-hybrid systems may be used to screen a mammalian (typically human) cDNA expression library, such as, for example, a cDNA library produced from human mature B cell line (Namalwa) mRNA (see Ambruş et al., 1993, Proc. Natl. Acad. Sci. (U.S.A.)). Once cDNAs encoding such interacting polypeptides are identified, the resulting polypeptides can be cloned, characterized, and used to screen compounds to identify compounds that can inhibit the binding interaction.

[0131] Notwithstanding the many and diverse application of the oligonucleotide and nucleic acid reagents of the invention, one important application relates to the production of recombinant peptides and proteins of the invention, as discussed in the following section.

[0132] V. Peptides and Proteins

[0133] In another embodiment, the present invention provides peptides, proteins, antibodies, and enzymes relating to genes and gene products that regulate telomere length and telomerase activity in mammalian cells. In particular, the invention provides synthetic or recombinant peptides or proteins comprising at least about 6 to 10 or 15 to 25 or 25 to 100 or more contiguous amino acids identical in sequence to amino acid sequence encoded by an open reading frame sequence of a human gene located in either:

[0134] (i) an ≈3.5 kb NotI-BstEII restriction fragment of plasmid pGRN109; or

[0135] (ii) an ≈1.4 kb EcoRI-BamHI restriction fragment of plasmid pGRN92.

[0136] The present invention provides the peptides and proteins encoded by the TPC2 and TPC3 genes, as well as fragments and analogs thereof, in isolatable form from eukaryotic or prokaryotic host cells expressing recombinant TPC2 and/or TPC3 protein, or from an in vitro translation system, as well as in purified and substantially pure form from synthesis in vitro or by purification from recombinant host cells or by purification of the naturally occurring proteins using antibodies or other reagents of the invention. Methods for expression of heterologous proteins in recombinant hosts, chemical synthesis of polypeptides, and in vitro translation are well known in the art and are described further in Maniatis et al. and Berger and Kimmel, supra. Such proteins have application in methods for reconstituting in vitro telomerase or other enzymatic activities that maintain telomerases and regulate telomere length. These methods in turn have application in screens for therapeutic agents, for diagnostic tests, and for other applications.

[0137] Because they are expressed differentially between immortal human cells lines, TPC2 and TPC3 genes and gene products serve as telomerase activity and tumor cell markers. Polypeptides having the full or partial amino acid sequence of TPC2 or TPC3 proteins are useful, for example, in the production of antibodies against TPC2 or TPC3 proteins and that are useful in the detection of TPC2 or TPC3 proteins in tumor cells. The invention provides purified TPC2 and TPC3 proteins having an amino acid sequence substantially identical to the amino acid sequences encoded by the open reading frames of the TPC2 and TPC3 genes. Such genes include human allelic variants or mammalian cognate genes that can be obtained in accordance with and using the reagents provided by the present invention.

[0138] The invention also provides TPC2 and TPC3 protein analogs, non-naturally occurring polypeptides comprising at least 5 to 10 to 15 to 20 to 25 to 100 or more amino acids in a contiguous sequence selected from the amino acid sequences of the TPC2 and TPC3 proteins but include one or more deletions or additions of amino acids, either at the amino- or carboxy-terminus, or internally, or both; analogs may further include sequence transpositions. Analogs may also comprise amino acid substitutions, preferably conservative substitutions. Analogs include active fragments as well as various muteins. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally occurring sequence. Preferred amino acid substitutions include those that: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter post-translational modification of the analog, possibly including phosphorylation, and (4) confer or modify other physicochemical or functional properties of such analogs. TPC2 or TPC3 protein analogs can be immunogenic for TPC2 or TPC3 proteins, i.e., when presented as an immunogen, the analog elicits the production of an antibody that specifically binds to TPC2 or TPC3 proteins. Active fragments can be identified empirically by generating fragments of the full length protein by deletion from either the amino-terminus or the carboxy-terminus or both, and testing the resulting fragments for activity.

[0139] Conservative amino acid substitution is a substitution of an amino acid by a replacement amino acid which has similar characteristics (e.g., those with acidic properties: Asp and Glu). A conservative (or synonymous) amino acid substitution does not substantially change the structural characteristics of the parent protein (e.g., a replacement amino acid should not tend to break a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Proteins, Structures and Molecular Principles (1984), Creighton (ed.), W. H. Freeman and Company, New York; Introduction to Protein Structure (1991), C. Branden and J. Tooze, Garland Publishing, New York; and Thornton et al., 1991, Nature 354:105; which are
incorporated herein by reference. The following six groups each contain amino acids that are conservative substitutions for one another: (1) Alanine (A), Serine (S), Threonine (T); (2) Aspartic acid (D), Glutamic acid (E); (3) Asparagine (N), Glutamine (Q); (4) Arginine (R), Lysine (K), (5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and (6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

[0140] Analogues may include heterologous sequences generally linked at the amino- or carboxy-terminus, wherein the heterologous sequence(s) confer a functional property to the resultant analog not shared by the native protein. Such analogs are referred to as fusion proteins and for purposes of the present invention typically comprise a TPC2 or TPC3 protein or analog and an additional peptide or protein moiety. Fusion proteins usefully combine properties of two different polypeptides or proteins, and can be used, for example, to confer a label, such as a polyhistidine polypeptide or a maltose binding protein, useful in affinity isolation of the fusion protein or to protect the fusion protein from degradation inside a cell. The fusion protein may comprise a limited peptide with desired properties, for example, a peptidase site that renders the TPC2 or TPC3 protein or analog cleavable from the remainder of the fusion protein. The fusion protein can also confer an antigenic epitope to the TPC2 or TPC3 protein of interest; antibodies that bind the epitope could then be used to immunoprecipitate the fusion protein for purification or to identify associated proteins.

[0141] Thus, the invention provides recombinant fusion proteins in which all or a portion of the TPC2 or TPC3 protein is fused to another polypeptide or protein of interest. For example, plasmids pGRN103, pGRN104, pGRN106, and pGRN110 express plasmids of the invention that code for the expression of novel fusion proteins of the invention that comprise a portion of either TPC2 or TPC3 protein and maltose binding protein (MBP). These vectors were created using the commercially available pMALc2 expression vector and system (New England Biolabs). Plasmid pGRN103 encodes a fusion protein comprising the amino-terminal portion of TPC3 protein and MBP and was prepared by replacing the Xmlnl-PsiI restriction fragment of plasmid pMALc2 with the PsiI-PsiI restriction fragment of plasmid pGRN92. Plasmid pGRN104 encodes a fusion protein comprising the carboxy-terminal portion of TPC3 protein and MBP and was prepared by replacing the Ecl136III-BamHI restriction fragment of plasmid pMALc2 with the BspE1 (treated with Klenow in the presence of dCTP and dGTP only)-BamHI restriction fragment of plasmid pGRN92. Plasmid pGRN106 encodes a fusion protein comprising the amino-terminal portion of TPC2 protein and MBP and was prepared by replacing the Sall-PsiI restriction fragment of plasmid pMALc2 with a Sall-SseS8 restriction fragment that can be isolated from plasmid pGRN109. Plasmid pGRN110 encodes a fusion protein comprising the carboxy-terminal portion of TPC2 protein and MBP and was prepared by inserting a restriction fragment containing the carboxy-terminal portion of the open reading frame of TPC2 into plasmid pMALc2 such that the fusion protein shown below (SEQ ID NO: 16) results from expression of the plasmid in E. coli W3110 cells (only the ends of the MBP and TPC2 proteins at the junction region are shown): #STR1## These and other fusion proteins of the invention can be isolated in accordance with standard procedures and then used to immunize animals, i.e., mouse and rabbits, for the production of polyclonal antiserum and monoclonal antibodies, as described in the following section.

[0142] TPC2 or TPC3 proteins, analogs, peptides, and polypeptides can also be prepared by chemical synthesis using well known methods. For example, various peptides with amino acid sequences corresponding to sequences of the TPC2 and TPC3 proteins can be chemically synthesized in vitro and used to generate antibodies that specifically bind to TPC2 and/or TPC3 proteins. Illustrative peptides of the invention include RGLKRRQDSERKRDRE (SEQ ID NO: 17) and KVTSPLQSPTKAKPK (SEQ ID NO: 18), which have been chemically synthesized in vitro and used to immunize animals to generate antibodies specific for TPC3 protein. Such peptides may correspond to structural and functional domains identified by comparison of the nucleotide and/or amino acid sequence data of a gene or protein to public or other sequence databases. Computerized comparison methods can be used to identify sequence motifs or predicted protein conformation domains that occur in other proteins of known structure and/or function. See Proteins, Structures and Molecular Principles (1984), Creighton (ed.), W. H. Freeman and Company, New York, incorporated herein by reference. Methods to identify protein sequences that fold into a known three-dimensional structure are known. See Bowie et al., 1991, Science 253:164. Recognized sequence motifs and structural conformations may be used to define structural and functional domains. Computer programs GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package (Genetics Computer Group, 575 Science Dr., Madison, Wis.) can be used to identify sequences in databases, such as GenBank/EMBL, that have regions of homology. Neural network methods, whether implemented in hardware or software, may be used to: (1) identify related protein sequences and nucleotide sequences, and (2) define structural or functional domains in polypeptides. See Brunak et al., 1991, J. Mol. Biol. 220:49, incorporated herein by reference.

[0143] Thus, one class of preferred peptides and proteins of the invention are fragments of the TPC2 or TPC3 proteins having amino- and/or carboxy-termini corresponding to amino acid positions near functional domain borders. Alternative fragments may also be prepared. The choice of the amino- and carboxy-termini of such fragments rests with the discretion of the practitioner and is based on considerations such as case of construction, stability to proteolysis, thermal stability, immunological reactivity, amino- or carboxyl-terminal residue modification, or other considerations.

[0144] The immunogenic peptides and proteins of the invention can be used in therapeutic immunization and vaccination procedures. See U.S. provisional patent application Ser. No. 60/008,949, filed Oct. 20, 1995, incorporated herein by reference. The invention therefore provides a method of immunizing a subject, as well as vaccines useful in the method, against cells that maintain telomeres and express telomerase activity, such as cancer cells, that comprise administering an immunostimulating amount of such peptides or proteins of the invention.

[0145] Peptides and proteins of the invention are suitably obtained in substantially pure form if at least about 50 percent (w/w) or more pure and substantially free of interfering proteins and contaminants. Preferably, these polypeptides are isolated or synthesized in a purity of at least 80 percent (w/w) or, more preferably, in at least about 95 percent (w/w), and are substantially free of other proteins or contaminants.
One important application of the peptides and proteins of the invention is the generation of antibodies that specifically bind to TPC2 or TPC3 proteins, as discussed in the following section.

VI. Antibodies

The proteins and peptides of the invention can also be used to generate antibodies specific for TPC2 or TPC3 proteins, or for particular epitopes on those proteins. TPC2 or TPC3 proteins, fragments thereof, or analogs thereof, can be used to immunize an animal for the production of specific antibodies. For example, but not for limitation, a recombinantly produced fragment of a TPC2 or TPC3 protein or a fusion protein can be injected into a mouse along with an adjuvant following immunization protocols known to those of skill in the art so as to generate an immune response. Alternatively, or in combination with a recombinantly produced polypeptide, a chemically synthesized peptide having an amino acid sequence corresponding to a TPC2 or TPC3 protein may be used as an immunogen to raise antibodies which bind a TPC2, TPC3, or another telomere or telomerase-related protein. Immunoglobulins that bind the target protein with a binding affinity of at least about 1 x 10^4 M^-1 can be harvested from the immunized animal as an antiseraum, and may be further purified by immunoadfinity chromatography or other means.

Additionally, spleen cells can be harvested from the immunized animal (typically rat or mouse) and fused to myeloma cells to produce a bank of monoclonal antibody-secreting hybridoma cells. The bank of hybridomas can be screened for clones that secrete immunoglobulins that bind the protein of interest specifically, i.e., with an affinity of at least 1 x 10^4 M^-1. Animals other than mice and rats may be used to raise antibodies, for example, goats, rabbits, sheep, and chickens may also be employed to raise antibodies reactive with a TPC2 or TPC3 protein. Transgenic mice having the capacity to produce substantially human antibodies also may be immunized and used for a source of antiseraum and/or for making monoclonal antibody secreting hybridomas.

Thus, the invention provides polyclonal and monoclonal antibodies that specifically bind to TPC2 or TPC3 proteins. Bacteriophage antibody display libraries may also be screened for phage able to bind peptides and proteins of the invention specifically. Combinatorial libraries of antibodies have been generated in bacteriophage lambda expression systems and may be screened as bacteriophage plaques or as colonies of lysogens. For general methods to prepare antibodies, see Antibodies: A Laboratory Manual (1988), E. Harlow and D. Lane, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., incorporated herein by reference.

These antibodies can in turn be used to isolate TPC2 or TPC3 proteins from normal or recombinant cells and so can be used to purify the proteins as well as other proteins associated therewith. Such antibodies are useful in the detection of TPC2 or TPC3 proteins in samples and in the detection of cells comprising TPC2 or TPC3 proteins in complex mixtures of cells. Such detection methods have application in screening, diagnosing, and monitoring diseases and other conditions, such as cancer, pregnancy, or fertility, because the TPC2 and TPC3 proteins are present in most cells capable of elongating telomeric DNA and expressing telomerase activity and are present in those cells at levels significantly higher than the levels observed in telomerase negative cells.

For some applications of the antibodies of the invention, such as identifying immuno-crossreactive proteins, the desired antiserum or monoclonal antibody(ies) is/are not monospecific. In these or other instances, it may be preferable to use a synthetic or recombinant fragment of a TPC2 or TPC3 protein as an antigen rather than the entire protein. More specifically, where the object is to identify immuno-crossreactive polypeptides that comprise a particular structural moiety, such as a DNA-binding domain, it is preferable to use as an antigen a fragment corresponding to part or all of a commensurate structural domain in the TPC2 or TPC3 protein.

Cationized or lipiddized antibodies reactive with TPC2 or TPC3 can be used therapeutically to treat or prevent diseases of excessive or inappropriate expression (e.g., neoplasia) of these proteins and the processes regulated thereby. Other methods of the invention are discussed in the following section.

VII. Methods

The various reagents of the invention described above have a wide variety of applications. The provision of polynucleotides capable of hybridizing to TPC2 or TPC3 cDNA and antibodies that specifically bind to TPC2 or TPC3 proteins allows one to detect expression of TPC2 and TPC3 in cells. The detection of TPC2 or TPC3 gene expression in cells suspected of being cancerous is useful in the diagnosis of cancer. Accordingly, this invention provides methods of detecting TPC2 or TPC3 mRNA or protein in a cell by hybridization or immunocassay methods. Hybridization methods can involve any of the routine methods including Northern blotting, Southern hybridization, amplification of target or probe nucleic acids by PCR, b-DNA, antibodies labeled with enzymes, LCR, Q-beta replicase, or 3SR; and the like, may also be used.

The polynucleotide sequences of the present invention can be used for forensic identification of individual humans, such as for identification of decedents, determination of paternity, criminal identification, and the like. The invention also provides TPC2 or TPC3 polynucleotide probes for diagnosis of disease states (e.g., neoplasia or pre-neoplasia) by detection of a TPC2 or TPC3 mRNA or rearrangements or amplification of the TPC2 or TPC3 gene in cells explanted from a patient, or detection of a pathognomonic TPC2 or TPC3 allele. Cells which contain an altered amount of TPC2 or TPC3 mRNA as compared to non-neoplastic or non-diseased cells of the same cell type(s) can be identified as candidate diseased cells in accordance with the methods of the invention. Similarly, the detection of pathognomonic rearrangements or amplification of the TPC2 or TPC3 gene locus or closely linked loci in a cell sample will identify the presence of a pathological condition or a predisposition to developing a pathological condition (e.g., cancer, genetic disease).

The isolation of three telomerase-related and telomere length regulatory components, TPC2, TPC3, and hTR, allows the production of recombinant telomerase comprising one or more of these components. In one method, recombinant telomerase is produced by expressing a TPC2
or TPC3 protein or active TPC2 or TPC3 analog and/or recombinant hTR in a cell. In another, telomerase is reconstituted in vitro. The recombinant RNA component of telomerase can be, for example, an RNA molecule derived from the sequence encoded by the ~2.5 kb HindIII-SacI insert of pGRN33 (ATCC 75926). Recombinant telomerase is useful, for example for screening assays to determine whether a compound modulates telomerase activity.

[0158] Telomerase- and telomere length-modulating agents which reduce a cell’s capacity to repair telomere DNA damage (e.g., by inhibiting endogenous naturally occurring telomerase) are candidate antineoplastic agents. Candidate antineoplastic agents are then tested further for antineoplastic activity in assays which are routinely used to predict suitability for use as human antineoplastic drugs. Examples of these assays include, but are not limited to, assays to measure the ability of the candidate agent (1) to inhibit anchorage-independent transformed cell growth in soft agar, (2) to reduce tumorigenicity of transformed cells transplanted into nu/nu mice, (3) to reverse morphological transformation of transformed cells, (4) to reduce growth of transplanted tumors in nu/nu mice, (5) to inhibit formation of tumors or pre-neoplastic cells in animal models of spontaneous or chemically-induced carcinogenesis, and (6) to induce a more differentiated phenotype in transformed cells.

[0159] Administration of an efficacious dose of an agent capable of specifically inhibiting telomere-maintenance or telomerase activity to a patient can be used as a therapeutic or prophylactic method for treating pathological conditions (e.g., cancer, inflammation, lymphoproliferative diseases, autoimmune disease, neurodegenerative diseases, and the like), which are effectively treated by modulating telomerase activity and telomere length. Additional embodiments directed to modulation of neoplasia or cell death include methods that employ specific inhibitory nucleic acids, e.g., sense or antisense polynucleotides corresponding to nucleotide sequences encoding TPC2, TPC3, or a cognate mammalian TPC2 or TPC3 protein.

[0160] The foregoing description of the preferred embodiments of the present invention has been presented for purposes of illustration and description and is not intended to be exhaustive or to limit the invention to the precise form disclosed but instead to illustrate the many modifications and variations possible in light of the invention and description and to include such modifications and variations as may be apparent to a person skilled in the art in light of this description within the scope of this invention and the claims thereto. All publications and patent documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication or patent document were so individually denoted.

VII. EXAMPLES

[0161] The following examples are given to illustrate but not limit the invention. Generally, the nomenclature used herein and many of the laboratory procedures in cell culture, molecular genetics, and nucleic acid chemistry and hybridization described below are those well known and commonly employed in the art. All percentages given throughout the specification and examples are based upon weight unless otherwise indicated. All protein molecular weights are based on mean average molecular weights unless otherwise indicated.

[0162] A. Methods In Molecular Genetics

[0163] Standard techniques are used for recombinant nucleic acid methods, polynucleotide synthesis, in vitro polyepitope synthesis, microbial culture and transformation (e.g., electroporation), and the like. Generally enzymatic reactions and purification steps using commercially available starting materials are performed according to the manufacturer’s specifications. The techniques and procedures are generally performed according to conventional methods in the art and various general references (see, generally, Sambrook et al. Molecular Cloning: A Laboratory Manual, 2d ed. (1989); Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., incorporated herein by reference) referenced herein.

[0164] Oligonucleotides can be synthesized on an Applied Bio Systems or other commercially available oligonucleotide synthesizer according to specifications provided by the manufacturer. Polynucleotide primers may be prepared using any suitable method, such as, for example, the phosphotriester and phosphodiester methods, or automated embodiments thereof. In one such automated embodiment, diethylphosphorodiamidites are used as starting materials and may be synthesized as described by Beaucage et al., 1981, Tetrahedron Letters 22:1859, and U.S. Pat. No. 4,588,006.

[0165] Methods for PCR amplification are known in the art (PCR Technology: Principles and Applications for DNA Amplification, Ed. Erlich, Stockton Press, New York, N.Y. (1989); PCR Protocols: A Guide to Methods and Applications, eds. Innis, Gelfand, Sninsky, and White, Academic Press, San Diego, Calif. (1990); Mattila et al., 1991, Nucleic Acids Res. 19:4967; Eckert and Kunkel, 1991, PCR Methods and Applications 1:17; and the U.S. Patents noted above. Optimal PCR and hybridization conditions will vary depending upon the sequence composition and length(s) of the targeting polynucleotide(s) primers and target(s) employed, and the experimental method selected by the practitioner. Various guidelines may be used to select appropriate primer sequences and hybridization conditions (see, Sambrook et al., supra). Generally PCR is carried out in a buffered aqueous solution, preferably at a pH of 7-9, most preferably about 8. The deoxyribonucleoside triphosphates dATP, dCTP, dGTP, and dTTP are also added to the synthesis mixture in adequate amounts, and the resulting solution is heated to about 85-100 degrees C. for about 1 to 10 minutes, preferably from 1 to 4 minutes. After this heating period, the solution is allowed to cool to about 20-40 degrees C., for primer hybridization. To the cooled mixture is added an agent for polymerization, and the reaction is allowed to occur under conditions known in the art. This synthesis reaction may occur at from room temperature up to a temperature just over which the agent for polymerization no longer functions efficiently. Thus, for example, if a heat-labile DNA polymerase is used as the agent for polymerization, the synthesis temperature is generally no greater than about 45 degrees C. The agent for polymerization may be any compound or system that will function to accomplish the synthesis of primer extension products, including enzymes. Suitable enzymes for this purpose include, for example, E. coli DNA polymerase I or the Klenow fragment thereof, Taq DNA polymerase, and other available DNA polymerases.
The newly synthesized strand and its complementary nucleic acid strand form double-stranded molecules used in the succeeding steps of the process. In the next step, the strands of the double-stranded molecule are separated using any of the procedures described above to provide single-stranded molecules. The steps of strand separation and extension product synthesis can be repeated as often as needed to produce the desired quantity of the specific nucleic acid sequence. The amount of the specific nucleic acid sequence produced will accumulate in an exponential fashion.

Both the subtractive hybridization method and the differential display method have disadvantages for isolating rare mRNAs that are differentially expressed. Subtractive hybridization is useful for enriching and producing a pool of low-abundance cDNA sequences, but conventional screening of the resultant library (ies), even if PCR amplified, is biased in favor of identifying species that are still abundant within the selected non-abundant cDNA pool, making difficult the isolation of very rare cDNA species with a conventional subtractive hybridization enrichment protocol. Differential display of mRNA amplification by PCR is biased by the initial abundance of the various mRNA species and often under-represents or fails to detect rare mRNA species among the many mRNA species that are more abundant and not substantially differentially expressed. The present invention provides a subtractive hybridization differential display method that is particularly preferred for isolating rare mRNAs, such as those expressed by the TCP2 and TCP3 genes. In brief, this method comprises the steps of: (1) one or more cycles of subtractive hybridization of two cDNA populations to generate a population of subtracted cDNA that is selectively enriched for cDNA species of low abundance mRNAs that are present at higher levels in one of the two cDNA populations, and (2) differential display of the electronophoretic gel and recovery of individual differentially expressed cDNAs by recovery from the gel. PCR amplification, under suitable PCR conditions, of said subtracted cDNA population with a 5' primer of arbitrary nucleotide sequence and optionally with a 3' primer comprising poly(dT) and/or poly(dT) and two or more arbitrary nucleotides at the 3' end to generate PCR products is typically used to replicate or amplify a subtracted library.

To accomplish the initial step(s) of subtractive hybridization, RNA prepared by conventional methods from a first cell population and RNA from a second cell population are separately reverse-transcribed and second-strand synthesized to form two pools of double-stranded cDNA, a tester pool comprising sequences of the mRNA species(s) for which enrichment is desired, and a driver pool comprising the sequences to be subtracted from the tester pool. The two pools may be fragmented by endonuclease digestion (restriction endonuclease or non-specific endonuclease) if desired to degrade cDNA consisting of tandem repeated sequences and to enhance hybridization efficiency. The driver pool is labeled, such as by photolabeling or attachment of another suitable recoverable label. The driver pool and tester pool are denatured and mixed together in a reaction mixture under hybridization conditions and incubated for a suitable hybridization period. The reaction mixture is contacted with a ligand which binds the recoverable label on the driver cDNA and which can be readily recovered from the reaction mixture (e.g., using avidin attached to magnetic beads), such that a substantial fraction of the driver cDNA and any tester cDNA hybridized thereto is selectively removed from the reaction mixture.

The remaining reaction mixture is enriched for tester cDNA species that are preferentially expressed in the first cell population as compared to the second cell population. The enriched (subtracted) tester cDNA pool may be subjected to one or more additional rounds of subtractive hybridization with a pool of labeled driver cDNA, which may be substantially identical to the initial pool of driver cDNA or which may represent a different cell population having mRNA species which are desired to be subtracted from the subtracted tester cDNA pool. A variety of means for accomplishing the subtractive hybridization(s) and suitable methodological guidance are available to the artisan. See Lee et al., 1991, Proc. Natl. Acad. Sci. (U.S.A) 88:2825; Mitter et al., 1995, Nucleic Acids Res. 23: 76; Liu et al., 1994, Anal. Biochem. 222: 102; Zebrowski et al., 1994, Anal. Biochem. 35 222: 285; Robertson et al., 1994, Genomics 23: 42; Rosenberg et al., 1994, Proc. Natl. Acad. Sci. (U.S.A.) 91: 6113; Li et al., 1994, Biotechniques 16: 722; Halvwoert et al., 1994, Nucleic Acids Res. 22: 878; Satoh et al., 1994, Mutat. Res. 316: 25; Marchal et al., 1993, Anal. Biochem. 208: 330; El-Deiry et al., 1993, Cell 78: 817; Har et al., 1991, Nucleic Acids Res. 19: 7097; and Herfort and Garber, 1991, Biotechniques 11: 598, each of which is incorporated herein by reference.

After the subtractive hybridization is completed, the subtracted tester cDNA is subjected to differential display. The general strategy involves amplification of cDNAs from the subtracted tester cDNA pool by PCR using one or a set of arbitrary sequence primers. Arbitrary primers are selected according to various criteria at the discretion of the practitioner so that each will amplify only a fraction of the DNAs in the subtracted cDNA pool so that the amplification products can be resolved and individually recovered on a separation system, such as a polyacrylamide gel. In part because the number and complexity of cDNA species represented in any particular subtracted tester pool may vary considerably depending upon the nature and complexity of the driver and tester pools, the selection of arbitrary primers and their sequence(s) is determined by the practitioner with reference to the literature. See U.S. patent application Ser. No. 08/235,180, filed Apr. 29, 1994; Linskens et al., 1995, Nucleic Acids Res. 23 (16): 3244-3251; Liang et al., 1993, Nucleic Acids Res. 21: 3269; Utsans et al., 1994, Proc. Natl.


The subtracted tester cDNA pool and a separate cDNA pool prepared in the same way from a cell line or tissue that does not express (or expresses at lower levels) the rare protein is amplified with suitable arbitrary primers(s) (i.e., primers having a predetermined sequence that is selected without reference to a sequence of a desired differentially expressed mRNA) for a suitable number of
amplification cycles to generate sufficient amplification product for display and recovery of desired species, as can be determined empirically. The primer(s) may comprise 5'-terminal sequences which serve to anchor other PCR primers (distal primers) and/or which comprise a restriction site (half-site or other ligatable end). The amplified products are usually labeled and are typically resolved by electrophoresis on a polyacrylamide gel; the location(s) where label is present in the subtracted tester cDNA but not present (or present at much lower levels) in the control cDNA are excised, and the labeled product(s) is (are) recovered from the gel portion, typically by elution.

[0175] The resultant recovered product species (typically an expressed sequence tag or EST cDNA) can be subcloned into a replicable vector with or without attachment of linkers, amplified further, and/or sequenced directly. Once the EST(s) is recovered, it can be used to obtain a substantially full length cDNA from a cDNA library. The EST(s) can be sequenced and the sequence information used to generate a primer for primer extension (5'-RACE), or the EST can be labeled and used as a hybridization probe to identify larger cDNA clones from a cDNA library. Genomic or full length cDNA clones corresponding to ESTs can be isolated from clone libraries (e.g., available from Clontech, Palo Alto, Calif.) using the labeled EST (e.g., by nick translation or end-labeling) or other hybridization probes with nucleotide sequences corresponding to those identified in the EST in conventional hybridization screening methods.

[0176] Thus, double stranded cDNA is made from total RNA purified by CsCl gradient centrifugation. In general, mix 5 g of total RNA, 0.5 g oligo dT (12 to 18 bases), and water (deionized water is routinely used) in a total of 7.1 g, denature RNA at 95 degrees C for 5 to 10 minutes, and placed on ice. The denatured RNA and oligo dT is then added to a tube containing 4.1 of 5xfirst strand synthesis buffer (BRL), 2.1 of 0.1 M DTT (BRL), 1.1 of dNTP (10 mM each), and 1.1 of RNasin (Pharmacia), and warmed for 2 minutes at 42 degrees C. About 5.1 of Superscript II™ reverse transcriptase (BRL) is added to the reaction mixture, and first strand cDNA synthesis is performed for 42 degrees C for 60 minutes. Then, the reaction mixture is placed on ice and is ready for the synthesis of second strand. The first strand cDNA is added to a tube containing 11.1 of water, 1.1 of 10xE. coli DNA ligase buffer, 3.1 of dNTP (10 mM each), 1.5 of E. coli DNA ligase (15 units, BRL), 7.7 of E. coli DNA polymerase (40 units, Pharmacia), and 0.7 of E. coli RNase H (BRL). The reaction mixture is incubated for two hours at 16 degrees C, and then 1.1 of T4 DNA polymerase (10 units, Pharmacia) is added. The incubation continues for 5 more minutes at the same temperature, and the reaction is stopped by the addition of 2.1 of 0.5 M EDTA and phenol/chloroform extraction, usually performed twice. The double-stranded cDNA is precipitated with ethanol and resuspended in 12.1 of TE buffer.

[0177] The cDNA is then modified by the addition of linkers. Mix 10.1 of cDNA prepared as above with 4.1 of 10xbuffer for Rsal, 21.1 of water, and 5.1 of Rsal (25-50 units), and incubate the mixture for two hours at 37 degrees C. For the T7 reaction, this is followed by 10 minutes at 37 degrees C and then 10 minutes at 30 degrees C to restrict the T7 restriction enzyme. For the Sau3A reaction, this is followed by 10 minutes at 37 degrees C along with the uncut cDNA for completion of digestion. The restriction enzyme is then inactivated for 10 min. at 65 degrees C.

[0178] The linkers are prepared as double stranded oligonucleotides by mixing 1 g of each of:

<table>
<thead>
<tr>
<th>Linker</th>
<th>SEQ ID NO.</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>NotA</td>
<td>SEQ ID NO: 19</td>
<td>5'-pATAGCCGGCCGCAAGAATTCA-NH2-3'</td>
</tr>
<tr>
<td>NotB</td>
<td>SEQ ID NO: 20</td>
<td>5'-TGAAATCTTGCGGCCGCTAT-3'</td>
</tr>
<tr>
<td>Ancol</td>
<td>SEQ ID NO: 21</td>
<td>5'-pCAAGAAGCTTGTTGATCCACACAG-NH2-3'</td>
</tr>
<tr>
<td>PCR02</td>
<td>SEQ ID NO: 22</td>
<td>5'-CTTGCTGATGACCAACCAAGCT-3'</td>
</tr>
</tbody>
</table>

[0179] PCR amplification of the tester and driver cDNA libraries is carried out by taking about 1.1 of each gel slice isolated as above (melted at 65 degrees C. before use) and mixing with 10.1 of NotI (for tester—this oligonucleotide serves as both the 5' and 3' primers) or PCR02 (for the driver), 5.1 of 10xPCR buffer, 6.1 dNTP (2.5 mM each), 1 unit of Taq polymerase (Boehringer Mannheim or Perkin Elmer), 1 unit of Pfu polymerase (Stratagene), 0.2 g of gene 32 protein (Boehringer Mannheim), and water to 50.1. PCR is performed for 20 cycles at 94 degrees C. for 45 sec., 60 degrees C. for 45 sec., and 72 degrees C. for 2 min., with a 5 min. extension at 72 degrees C. after completion of the last cycle. The driver is PCR amplified in multiple reactions to make enough DNA for photobiotinylation.

[0180] Photobiotinylation of the driver cDNA is conveniently accomplished as follows. About 100 g of driver cDNA in 1 mM EDTA is mixed with 100.1 of photo biotin (Vector). This mixture is placed on ice with the lid open and irradiated for 15 min. with a light source located about 10 cm away from the tube. After the irradiation, 30.1 of 1 M Tris-C1 (pH 9.1) is added to the tube, and the biotinylated DNA is extracted with water-saturated butanol several times (4x) until the orange color disappears from the aqueous phase. The extraction process is repeated once, and the biotinylated DNA is precipitated with ethanol and resuspended in TE buffer to a final concentration of 1 g/30
Subtraction hybridization is conveniently accomplished as follows. Mix 8 g of biotinylated driver DNA with 0.4 g of tester DNA (concentrations estimated by OD measurement and ethidium bromide staining of the gel). The mixed DNA is precipitated with ethanol and resuspended in 10 ml of HE buffer (10 mM HEPES, pH 7.3,1 mM EDTA). The DNA is denatured at 100 degrees C. for 4 min. and transferred to ice. About 10 ml of double hybridization solution containing 1.5 M NaCl, 50 mM HEPES, pH 7.3,10 mM EDTA, and 0.2% SDS is then added to the tube. Two drops of mineral oil are added, and the DNA is denatured again at 100 degrees C. for 4 min. and transferred to a water bath at 68 degrees C. The hybridization is performed at this temperature for 22 hours. Biotinylated DNA is removed with streptavidin MagneSphere™ Paramagnetic Particles (Promega), and the tester DNA remaining is recovered.

A second subtraction is performed by mixing recovered tester DNA (about 80 ml) with 8 ml (8 g) of biotinylated driver DNA and then precipitation with ethanol. The precipitated DNA pellet is resuspended in 10 ml of HE buffer. The denaturation, hybridization, and recovery are performed as above; however, the second hybridization is performed for only 2 hours at 68 degrees C. PCR amplifies the recovered DNA (0.3 ml) for 18 cycles in a reaction mixture containing 2 ml of 10X Pfu polymerase buffer, 2.5 ml of 2.5 mM dNTP, 0.2 ml of Taq polymerase (1 unit), 0.4 ml of Pfu polymerase (1 unit), 0.04 ml of T4 gene 32 protein, and water to 20 ml. The products are checked on a 1% agarose gel to confirm relative concentrations. The subtraction hybridization can be repeated on these samples. The final subtracted samples are PCR amplified (18 cycles) and diluted (1 to 10 or 1 to 15) and used for enhanced differential display.

Enhanced differential display of subtracted cDNA involves PCR amplification with 5 arbitrary primers and a 3' oligo dT primer with two randomized bases at the 3' end, recovery of bands identified as containing cDNA corresponding to differentially expressed mRNAs, and PCR amplification, sequencing, and/or cloning of the bands identified. Add 1 ml of one 5' primer (20 M stock) or two 5' primers (half of each) and 1.2 ml of one 5' primer (1 ml) and one 3' primer (0.2 ml) to the tube. Add 1 ml of subtracted DNA to the same tube. To this mixture, add 8 ml of cocktail mix containing 1 ml of 10X PCR buffer for Pfu polymerase (commercially available), 0.1 ml of dNTP (2.5 mM each), 0.3 mM alpha-32P-dATP, 0.1 ml of Taq polymerase, 0.2 ml of Pfu polymerase (Stratagene), 0.02 ml of T4 gene 32 protein (Boehringer Mannheim), and 5.38 ml water. Overlay one drop of mineral oil, and PCR amplify for 4 cycles at 94 degrees C. for 45 sec., 39 degrees C. for 1 min., and 72 degrees C. for 1 min., and then 22 cycles at 94 degrees C. for 45 sec., 60 degrees C. for 1 min., and 72 degrees C. for 1 min., with a final extension for 5 min. at 72 degrees C. About 5 ml of formamide dye is added to the PCR product, and the products are denatured at 95 degrees C. for 2.5 min. and loaded onto a prewarmed 6% polyacrylamide sequencing gel, which is run at 1900 to 2000 constant voltage (do not allow current to reach 50 ma) until the xylene cyanol dye is one inch from the bottom of the gel. The gel is dried under vacuum at 80 degrees C. for 45 min. and exposed to PhosphorImager™ screen (for notebook record) and/or then to X-ray film at room temperature for one or two days (tape the gel to the film and punch three holes at the three corner of the gel and film for easy identification of bands).

Differentially expressed gene fragments appear as bands on the screen or film that are present in the lanes on the gel corresponding to the cDNA of the tester cells but present at lower levels or absent from the lanes corresponding to the cDNA of the control lanes. The bands can be recovered from the gel by first aligning the gel with the film or screen (based on the three holes and marks) and then excising the bands of interest with a razor blade and transferring the gel slice to an Eppendorf™ tube. Rinse the razor blade between each cutting operation to avoid cross contamination. To remove the area and paper backing used with sequencing gels without substantial loss of the desired DNA, add about 900 ml of TE buffer to the tube containing the gel slice, incubate the tube at room temperature for 10 min., and then remove and discard the paper and TE buffer. To prepare a solution of the desired DNA from the gel slice, the gel slice is suspended in 40 ml of TE buffer containing 100 mM NaCl and heated for 10 min. at 95-98 degrees C. The liquid is collected (a short centrifugation collects the liquid at the bottom of the tube) and serves as a source of the desired DNA.

This DNA can be PCR-amplified by placing 1.3 ml of recovered DNA in a 50 ml total reaction volume in a reaction mixture containing 6 ml of total primer(s), 5 ml of 10X PCR buffer for Pfu polymerase, 0.1 ml of dNTP (2.5 mM each), 0.25 ml of Taq polymerase, 0.5 ml of Pfu polymerase, 0.05 ml of T4 gene 32 protein, and water. The PCR is performed for 25 cycles at 94 degrees C. for 45 sec., 60 degrees C. for 1 min., and 72 degrees C. for 1 min., with a 5 min. extension at 72 degrees C. at the end of the last cycle. The PCR products can be stored or further processed, i.e., subcloned and sequenced.

The availability of plasmids comprising restriction fragments corresponding to the open reading frames of the TCP2 and TCP3 genes makes possible the efficient isolation of these gene and gene products from other mammalian cells as well as the chemical synthesis of these genes and gene products and related reagents, i.e., peptides, oligonucleotides, antibodies, and antibody fragments.

RT-PCR Protocol for TCP3

Cell extracts are prepared using CHAPS, as described for the TRAP assay (TRAP-eze™ kit, Oncor). About 2 ml of cell extract are used per assay; typically 30-35 cycles of PCR are performed. Total RNA is prepared using the TRIzol™ RNA extraction method (Life Technologies) on cell pellets or CHAPS extracts. Each PCR tube contains: 15 ml of water; 2.5 ml of 25 M Mn(OAc)2; 5.5 ml of 5X EZ buffer (Perkin Elmer); 0.3 ml of 25 M dNTPs; 1 ml of 50th DNA polymerase buffer (Perkin Elmer); 0.1 ml (500 M) of primer 5'-CTCAGTGGACACTGGCTAGTTTC-3' (SEQ ID NO: 23); and 0.1 ml (500 M) of primer 3'-3'-CATCGTTGAACTCAAGAGTC-3' (SEQ ID NO: 24) in a total volume of ~25 ml. RT-PCR conditions include one minute treatment at 94 degrees C. to denature protein-RNA complexes; a thirty minute treatment at 65 degrees C. for the reverse transcription reaction; a 1.5 minute treatment at 94 degrees C. to denature DNA-RNA complexes; thirty cycles of PCR amplification with each cycle comprising a 30 second treatment at 94 degrees C. and a 30 second treatment at 65 degrees C.; and a final extension reaction by treatment for seven minutes at 60 degrees C. After PCR, the samples can be analyzed by gel electrophore-
sis using 1xTBE polyacrylamide gels and staining with SYBR-Green I. Tests showed that this primer set amplifies a band of correct size in both mortal and immortal cell lines and demonstrate that the TPC3 mRNA is expressed more abundantly in immortal cell lines.

[0190] D. RT-PCR Protocol for hTR

[0191] First strand cDNA synthesis is performed by mixing total RNA (1 μg) with 40 to 80 ng random hexamer in 11 μL, heating to 95 degrees C. for 5 min. to denature the nucleic acids (the thermal cycler may be used for this step), and then cooling on ice. The reaction mixture (8 μL) containing 4 μL of 5xbuffer (BRL; provided with the RTase), 2 μL of 0.1 M DTT, 1 μL of 10 mM dNTP (each), and 1 μL of RNase inhibitor (Pharmacia) is added to the denatured RNA and hexamer mixture and placed in a water bath at 42 degrees C. After a 1-2 min. incubation, 1 μL of Superscript II™ RTase (BRL) is added to the mixture and the incubation continued for 60 min. at 42 degrees C. The reaction is stopped by heating the tube containing the reaction mixture for 10 min. at 95 degrees C. The first strand cDNA is collected by precipitation and brief centrifugation and aliquoted to new tubes, in which it can be quickly frozen on dry ice and stored at ~80 degrees C., if necessary, for later use.

[0192] PCR amplification of hTR cDNA with specific primer sets can be generally accomplished as follows. About 1 μL of cDNA is used for each primer set. For a 10 μL PCR with 32P-dATP nucleotide, 1 μL of cDNA is mixed with 1 μL of 10xTaq buffer, 20 pmol of each primer, 1 μL of 2.5 mM dNTP, 5 Ci alpha-32P-dATP, 1 unit of Taq polymerase (Boehringer Mannheim), 1 unit of Taq antibody (Clontech), 0.2 μg of T4 gene 32 protein (Boehringer Mannheim), and water to 10 μL. One drop of mineral oil is then added to the tube. The conditions for PCR amplification for hTR are about 20 cycles of amplification, with each cycle comprising a treatment at 94 degrees C. for 45 sec., 60 degrees C. for 45 sec., and 72 degrees C. for 1.5 min. The primers used for the RT-PCR of hTR are shown below.

[0193] Upstream primer: F3b, 5'-TCTAACCCCTAATTGAGAAGGGCGTAG-3'(SEQ ID NO: 25);

[0194] Downstream primer: R3c, 5'-GTTTGCTCTAGAATGAACGTGGGAAG-3'(SEQ ID NO: 26);

[0195] Amplification of hTR with the F3b and R3c primer pair produces a 126 bp product. PCR products labeled with 32p can be conveniently detected by adding 5 μL of a formamide/dye mixture to the products, heating the products to denature the nucleic acids, separating the products by 6% urea polyacrylamide gel electrophoresis, and then exposing a Phosphorimag3™ cassette or X-ray film to the gel.

[0196] The invention has been described in terms of preferred embodiments and illustrated by way of example and is claimed below.

---

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 26

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4232 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..3215
(D) OTHER INFORMATION: /product= ”TPC2”

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 651
(D) OTHER INFORMATION: /note= ”N = probably C”

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 660
(D) OTHER INFORMATION: /note= ”N = probably C”

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1142
(D) OTHER INFORMATION: /note= ”N = probably T”
(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1160
(D) OTHER INFORMATION: /note= "N = maybe G"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1188
(D) OTHER INFORMATION: /note= "N = probably G"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1211
(D) OTHER INFORMATION: /note= "N = probably C"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1225
(D) OTHER INFORMATION: /note= "N = probably G"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1230
(D) OTHER INFORMATION: /note= "N = probably G"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1238
(D) OTHER INFORMATION: /note= "N = probably C"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1247
(D) OTHER INFORMATION: /note= "N = probably A"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1301
(D) OTHER INFORMATION: /note= "N = probably G"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1375
(D) OTHER INFORMATION: /note= "N = probably C"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1379
(D) OTHER INFORMATION: /note= "N = probably C"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1391
(D) OTHER INFORMATION: /note= "N = probably C"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1407
(D) OTHER INFORMATION: /note= "N = probably C"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1530
(D) OTHER INFORMATION: /note= "N = probably G"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1543
(D) OTHER INFORMATION: /note= "N = probably G"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1545
(D) OTHER INFORMATION: /note= "N = probably G"
-continued

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1596
(D) OTHER INFORMATION: /note= "N = probably G"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1598
(D) OTHER INFORMATION: /note= "N = probably G"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1652
(D) OTHER INFORMATION: /note= "N = probably G"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1689
(D) OTHER INFORMATION: /note= "N = probably C"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1688
(D) OTHER INFORMATION: /note= "N = probably G"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1707
(D) OTHER INFORMATION: /note= "N = probably C"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1719
(D) OTHER INFORMATION: /note= "N = probably C"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1725
(D) OTHER INFORMATION: /note= "N = probably C"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1734
(D) OTHER INFORMATION: /note= "N = probably C"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1789
(D) OTHER INFORMATION: /note= "N = probably G"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1796
(D) OTHER INFORMATION: /note= "N = probably A"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1816
(D) OTHER INFORMATION: /note= "N = probably G"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1824
(D) OTHER INFORMATION: /note= "N = probably G"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1929
(D) OTHER INFORMATION: /note= "N = probably G"

(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1935
(D) OTHER INFORMATION: /note= "N = maybe C"
(ix) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 1941
(D) OTHER INFORMATION: /note= "N = probably G"

(xi) FEATURE:
(A) NAME/KEY: unsure
(B) LOCATION: 3528
(D) OTHER INFORMATION: /note= "N = maybe A"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CCGGCTCGG CCAAATGAGC GGGGGCGAGG GTGGGGGCGGG ACAATTCAGG TGACAATGCG 60
TCCTACCGGG TGGGGCCGGA TTGCGCGCTC TTCTCAATC ACGCCACAGG CCGTCGAC 120
ACCTGCGGCG ACCTGCGGCG ACCTGCGGCG GTCAACTGCGG CCGCACATGAT CCGTCGACA 190
CTGCCCCCGG GCTGAGGAGG GGGCTCAGG GAGGAGGCG CGGGTACATT CATCGACCA 240
AACCGAGAGA CCAACACATT CAGGCAACTT GTGAGGGGAC AGTTTTCTCC AGAAAAATTAG 300
GAATCTCATC TTCAGAGAGA GCCGAATCCA CATATGCTGA AGCAAGACAG AAACCAAAAG 360
CGCTGACAGA TGCTGAGCGA AACATCAGCG GCTGGGACCG CTTCCACCTG GGAGCCCAA 420
CTGGACCCA AGATCATAAA GTCCACGCGT AAAAGCCCAA GCTTTGGAGA GAGAGCAGCA 480
GGATTGCGA GAAACCCCGA TTTCTCGGTG GTGTTAGGAA GTGCTCGGCA CAGCCAGGA 540
AGTTTTGCGA TGAGGCGCGG TGGTTTGTGC TGGCTGTATTC TGGCTTATG 600
TACTATAAAG CGGGGAGAGG GGGGCTCAGC CTTGCCCCAG NTACGTTAT 660
TCTCTGTGCG CCCCAGGCGA GGGCTCTCGG NGGAGCCTCG CTTGCCCCAG NTACGTTAT 720
GGAGAGGGAG CCCTCAGTGA TACAGCTCC ACGCCGGGCT CTGCGGCGCA CAGCTCAGG 780
ATGAGACCT ACTACAATGTC TCCGACACC CAGAGGAGCA TCAGACGCTG TGACGAGCG 840
ATGATCCAGG CTGACACAGGT CTGCTCTCGA TGGCTCAGA AGAGGGGATG GAGAGAGGT 900
GACCCGAGG CGTGCCCCGA GGGCCACCC ACAGATTCTG GTCCGAGAG TGGCCGAGGT 960
GGACCCGAG ATACGAGAGA TTGCTCTCTG CTGCGGCAATG ATGACATTGG CACACCTCG 1020
AGCCGAGAGG AGGGAGGAGG GGGGCCCCATC TCCGCGGCCG ACCACTGCGA GGGCCAGG 1100
GACCCGAGG AGGGAGGAGG GGGGCCCCATC TCCGCGGCCG ACCACTGCGA GGGCCAGG 1140
TNCAGGATG GCCGGAGAGA CAGCGGAGCA CACCCCCAGG GGGGCAGAA AATGGGAAG 1200
CTGGCTCGTY MTATGCGGCA AGAAGAACAC AAGGGGATGG GTGCGGCAAGC GGGCGCTT 1260
CYTGGAGAGA CCAACCCGGG AAAAACAGT CAAAAGGGA AAAATCGGAC TCGAGCTG 1320
CAGTGGCCGA GGGGGGAGAGG GGGGATCGAC AGGGAGCGTG CTGCTGCCAC GAGCCTTC 1380
GCCAGGGTCT NTGGGCCCCC CCTGCTCCTCC CGGAGAAAT CTGAGATTTG TCCGAGAGA 1440
ACGCCACGAC CCTGGGGGTC GTTCGCCGCA CYYCCGGAA GGCCAGAGA TCGGGAAATG 1500
TATTGGCAAG ACCACGGAGG CCACTGAAN ATGCTGAGG AANANCCGGC NGGCGAC 1560
GTAGCAGACG CGCGCCGGTA CTGTTACTNGT GAGGACCGCC GAGGAGTCC CGGGCGCC 1620
CCCACAGCC CCACTCGCC TTGGCTCAGC ANAGTCACCC GACCCGGGCG GGAGCGAG 1680
CATTGACAGG TCCTGGCGGCT CCAATTTGTG GCCATCAATG CATGNGGAC TCCGAGCTC 1740
AGGACCCCG CGGTTTGGCG CCGCGGCCAG CCAAGACAGA CCGCGAGAC GACGNCNA 1800
GAGGCCAGCG GAGCAGACAG RGAGTGGCGA CATCCTCGCTT GGGGATTCGC CATGCGTGT 1860
ATGAMCCAGA CGCTGCGCCGG TCCGAGTTGA CATGGAATAT CGGCTGATAG TACACGAC 1920
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAGCTGAAGA AARNACTGGA NTACTGTGAT CTAAGATGGA CAGGCCTGGG CACATCTCA</td>
<td>1980</td>
</tr>
<tr>
<td>GATCGAACGTC TGGAACCTGT GAAGACTGCT GAGAGGCAAC CTGAGCTCAA ACTGACCA</td>
<td>2040</td>
</tr>
<tr>
<td>TCCTUGAAC AGAGACGGGT CCTCAGGGAC TGGAAGAGCA AGATAGACGG CTTAAAAG</td>
<td>2100</td>
</tr>
<tr>
<td>AACAAGAAC ACCGAGTGACT TCTGCTGAGA GCTTACAGCA AGCAGTGAC CCAGTACC</td>
<td>2160</td>
</tr>
<tr>
<td>GAGGCCCCAC AAGCATTGGA GAAGATTGCG TACAGCAGAA GCTTCTGCAA GAGGGACC</td>
<td>2220</td>
</tr>
<tr>
<td>GTCTACATCC GAGCGTAGCT CTCAGAGGAG TTCACTTGGA TGGAAGAATTT TGGAAAGC</td>
<td>2280</td>
</tr>
<tr>
<td>TACCTGAAGT TGAGAATGA TGGAGAACAG TGGAAGACGC CCTTUCAGGA GCCAGCAAC</td>
<td>2340</td>
</tr>
<tr>
<td>AGAGGCTTTT TTTTTCCAGGA GAATGCAGAG ATACGAGAAAT ATCTAAGGGA AATGGAG</td>
<td>2400</td>
</tr>
<tr>
<td>GTCAGCTGAC GGGTGTGCCC AATAAAGAG AACTCCAGAA TTTCACTGGA GCAGATTA</td>
<td>2460</td>
</tr>
<tr>
<td>AATCCGGGAA GAAAAGGTGT GCTTTTTATT TTCTCAGCGGC GTGTCTGCTTC ACTCTACA</td>
<td>2520</td>
</tr>
<tr>
<td>TCTGACAGAG AGCGGCTCAGC ACAATAGGAC CTTCCAGCCA GCGCTGCTCA GACCGCTC</td>
<td>2580</td>
</tr>
<tr>
<td>GAGATCTGAC TCTGCTGCCCA GCTGCAACCC TACCTGCTGT ACCGAGCTCA CACACCCG</td>
<td>2640</td>
</tr>
<tr>
<td>CTGAAAGAGG CGTACATCCCA CCTCTCACTCA CCAACTAAGG CUAAGCCTACCG ATCTCAGG</td>
<td>2700</td>
</tr>
<tr>
<td>GATGAAGGAC CTGACCAGGCC CGACTCCGCC GAACCTTACA GCAAGCTAGG CAGCCGCC</td>
<td>2760</td>
</tr>
<tr>
<td>GCTGCTGCGG GCTGACAGAC AGAGGCCGCA ATACATGCTG GCCATCTGAG GCCGAGGGC</td>
<td>2820</td>
</tr>
<tr>
<td>AACGGCAAGGT CAGAGAACAGA GAAAGAGGAC AGAGGCTGCG GCCAGCTCTT GAATGGGGG</td>
<td>2880</td>
</tr>
<tr>
<td>TCGAGCAGGC AGGAACGACT GTACTGGTAC GAGCCAGCCT GCCGAGGCTT CAGCGGGG</td>
<td>2940</td>
</tr>
<tr>
<td>ATGGGCGAGC CCTCCTTGAG AGCTCTTGGG CCCGACAGCA GCTACAGAC GTCTGCTG</td>
<td>3000</td>
</tr>
<tr>
<td>AGAGGCTCTT CATGCGCTAC CGCAAGGCTT CGAGATGCTT CCACTACTG CACACCTAC</td>
<td>3060</td>
</tr>
<tr>
<td>ACATTGGCCAG ACGAAGACCT CGGGCGGCAC TCTGCTCTGGT ACGCAGAGC CACGGTCC</td>
<td>3120</td>
</tr>
<tr>
<td>GCTGCTGCGG GCCTCTACTCT AGGGGATCA CAGCCAGGCC AGATGAATGC AGAAGAGC</td>
<td>3180</td>
</tr>
<tr>
<td>CTTGAGGCGCA TGGAACCCGC AGCAGAGCCG CTTGCTGCGG AGCGACGAAG CACACTAG</td>
<td>3240</td>
</tr>
<tr>
<td>CAGGGGGAGA GCCAGGGGCG GCCTCTACTC CGACTTCCAC GCAGGGGCGC CCCTGGAG</td>
<td>3300</td>
</tr>
<tr>
<td>CTGAGCGGTC TATGTTAGGA GGCCAGGCGC AGCGGGACAG GAACAGAGC CCAAGTCC</td>
<td>3360</td>
</tr>
<tr>
<td>CTGACGGTAGG CCCAAACAGG ACAACATACAC ACCTCCAGGT GAGAGACCTG TCCATTGA</td>
<td>3420</td>
</tr>
<tr>
<td>TACAGGTGTG ACAAAAGACCC ACGAGGCGCT GTCTGCTGACG GACCGGCCTG GAAGAAT</td>
<td>3480</td>
</tr>
<tr>
<td>CTGCTGACAG AGTGGAGCCA ACCAAGAAGC CTGACGTGAGG GAGAGATGG ATCACAAGG</td>
<td>3540</td>
</tr>
<tr>
<td>AAGAATACCC CAGACGATTCC CTTCCAGGCC GATGGAGAGG GCCCTTTCGA GCCCTTCC</td>
<td>3600</td>
</tr>
<tr>
<td>ATCTCTGCGT CATTCCAGAC CTGGATAGCA ATGGGCTACCA AGGGAGAACG GGAAGATG</td>
<td>3660</td>
</tr>
<tr>
<td>GACAGCATTT GTGTTGGAAG ACAAGACCAA TGACGCTGCT CCCAGACCTG CGACCAGA</td>
<td>3720</td>
</tr>
<tr>
<td>CTGTTCTGAG GCTGGCGGAC AGGGGCTACCC ACCGGGCGCC CAAGGGCCAC TGCTTCCC</td>
<td>3780</td>
</tr>
<tr>
<td>GGGCTGACA TGCGAGGACG TCGGCTCTGT GGCGGTTGCT GGAGGAGNCT GTCTGCTG</td>
<td>3840</td>
</tr>
<tr>
<td>CTGTTCTGCGG TCTCTAGGCT TGACCTGCTG CCCAGAGCACC TCTGAGCTCA CTGACGCCC</td>
<td>3900</td>
</tr>
<tr>
<td>TCCATATTGA CTTTTCCCTG TGYYTGGAAA TGGAGTTTTT CCGTAAACCT CTAAGCCGG</td>
<td>3960</td>
</tr>
<tr>
<td>TGCTCTCTCC CTAATGCTG ACTCAGACCT CTTGGCTTTT AGTGTTACCC CTGATTCC</td>
<td>4020</td>
</tr>
<tr>
<td>CTCTATACTT GCCTGATGTT GCTTACAGAA GCTTCTGATT TGGATGGAAC ATGCTCAG</td>
<td>4080</td>
</tr>
<tr>
<td>TGCCCCGCTG TACATTATTAC CGGCTTTGGA ATAGCGGTTC TGGTCTTCAT AGGCTTGT</td>
<td>4140</td>
</tr>
</tbody>
</table>
GTUCCGAG ATGTGACTCA CCTTTCTGCT GCCCTCTTCA TUCAGGCTA CTGACTCA
ATTCCTTG GCTTGAGCC GGGCUGAT TC

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1105 amino acids
(B) TYPE: amino acid
(C) STRANDNESS: <Unknown>
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:
(A) NAME/KEY: Protein
(B) LOCATION: 1...1105
(D) OTHER INFORMATION: /note= "deduced amino acid sequence of TPC2 open reading frame"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
Pro Arg Ser Ala Asn Met Ala Ala Ala Thr Val Gly Arg Asp Thr Le
1  5  10 15
Pro Glu His Trp Ser Tyr Gly Val Cys Arg Asp Gly Arg Val Phe Ph
20  25  30
Ile Asn Asp Gln Leu Arg Cys Thr Thr Trp Leu His Pro Arg Thr Gl
35  40  45
Glu Pro Val Asn Ser Gly His Met Ile Arg Ser Asp Leu Pro Arg Gl
50  55  60
Trp Glu Glu Gly Phe Thr Glu Gly Ala Ser Tyr Phe Ile Asp Hi
65  70  75  80
Asn Gln Gln Thr Thr Ala Phe Arg His Pro Val Thr Gly Gln Phe Se
85  90  95
Pro Glu Asn Ser Glu Phe Ile Leu Gin Glu Pro Asn Pro His Me
100 105 110
Ser Lys Glu Asp Arg Asn Gln Glu Pro Ser Ser Met Val Ser Glu Th
115 120 125
Ser Thr Ala Gly Thr Ala Thr Leu Glu Ala Lys Pro Gly Pro Ly
130 135 140
Ile Ile Lys Ser Ser Ser Ser Val His Ser Phe Gly Lys Arg Asp Gl
145 150 155 160
Ala Ile Arg Arg Asn Pro Asn Val Pro Val Val Arg Gly Trp Le
165 170 175
His Lys Glu Asp Ser Xaa Gly Met Arg Leu Trp Arg Arg Arg Thr Ph
180 185 190
Val Leu Ala Asp Tyr Cys Leu Phe Tyr Tyr Lys Ala Glu Lys Lys Ar
195 200 205
Ser Ser Xaa Ser Ile Pro Leu Pro Xaa Tyr Val Xaa Ser Pro Val Al
210 215 220
Pro Glu Asp Arg Ile Ser Arg Lys Ser Phe Lys Ala Val His Th
225 230 235 240
Gly Met Arg Ala Leu Ile Tyr Asn Ser Ser Thr Ala Gly Ser Gin Al
245 250 255
Glu Gln Ser Gly Met Arg Thr Tyr Phe Ser Ala Asp Thr Gin Gl
260 265 270
Asp Met Asn Ala Trp Val Arg Ala Met Asn Gln Ala Ala Gin Val Le
275 280 285
<table>
<thead>
<tr>
<th>Residues</th>
<th>Amino Acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>299-303</td>
<td>Ser Arg Ser Leu Lys Arg Asp Met Glu Lys Val Glu Arg Gin Al</td>
</tr>
<tr>
<td>305-320</td>
<td>Val Pro Gln Ala Asn His Thr Glu Ser Cys His Glu Cys Gly Arg Va</td>
</tr>
<tr>
<td>325-330</td>
<td>Gly Pro Gly His Thr Arg Asp Pro His Arg Gly His Asp Asp Iil</td>
</tr>
<tr>
<td>340-350</td>
<td>Val Asn Phe Glu Arg Gin Gin Glu Gln Gly Glu Gin Tyr Arg Ser Gl</td>
</tr>
<tr>
<td>355-360</td>
<td>Arg Asp Pro Leu Glu Gly Lys Arg Asp Arg Ser Lys Ala Arg Ser Pr</td>
</tr>
<tr>
<td>370-375</td>
<td>Tyr Ser Pro Ala Glu Asp Ala Leu Phe Met Asp Xaa Pro Xaa Gl</td>
</tr>
<tr>
<td>380-390</td>
<td>Pro Arg Xaa Gin Gin Ala Gin Pro Gin Arg Ala Asn Gin Gly Me</td>
</tr>
<tr>
<td>395-400</td>
<td>Leu Pro Xaa Xaa Tyr Gly Pro Gly Xaa Xaa Asn Gin Xaa Gly Xa</td>
</tr>
<tr>
<td>405-410</td>
<td>Gin Arg Xaa Phe Xaa Pro Arg Thr Asn Xaa Gin Glu His Gin Ar</td>
</tr>
<tr>
<td>415-420</td>
<td>Lys Xaa Asn Leu Ala Gin Val Glu His Trp Ala Arg Ala Gin Lys Gl</td>
</tr>
<tr>
<td>425-430</td>
<td>Asp Ser Arg Ser Leu Pro Leu Asp Gin Thr Xaa Xaa Arg Gin Gly Xa</td>
</tr>
<tr>
<td>435-440</td>
<td>Gly Gin Ser Leu Xaa Phe Pro Gin Ala Tyr Gin Xaa Xaa Pro Lys Se</td>
</tr>
<tr>
<td>445-450</td>
<td>Thr Arg His Pro Ser Gin Xaa Xaa Ser Pro Xaa Pro Arg Asn Leu Pr</td>
</tr>
<tr>
<td>455-460</td>
<td>Ser Asp Tyr Lys Tyr Ala Gin Asp Arg Ala Ser His Leu Xaa Met Se</td>
</tr>
<tr>
<td>465-470</td>
<td>Ser Glu Xaa Arg Xaa Gly Ala Pro Gly Trp His Arg Val Ala Xaa Le</td>
</tr>
<tr>
<td>475-480</td>
<td>Xaa Xaa Ala Ala Pro Ala Val Pro Ala Gin Pro His Ser Al</td>
</tr>
<tr>
<td>485-490</td>
<td>His Leu Pro Trp Leu Pro Xaa Val His Arg Pro Gly Pro Glu Gin Gl</td>
</tr>
<tr>
<td>495-500</td>
<td>His Xaa Xaa Gin Gin Ala Gin Pro Leu His Gin Gin Gin Gin</td>
</tr>
<tr>
<td>505-510</td>
<td>His Xaa Ser Pro Arg Thr Pro Lys Gin Xaa Pro Thr Pro Ala Ala Xa</td>
</tr>
<tr>
<td>515-520</td>
<td>His Thr Ser Arg Xaa Ser Xaa Ser Glu Ala Thr Gly Pro Xaa Xaa Gl</td>
</tr>
<tr>
<td>525-530</td>
<td>Cys Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>535-540</td>
<td>Val Ser Ala Pro Ser Leu His Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>545-550</td>
<td>Gin Leu Xaa Lys Xaa Leu Xaa Tyr Leu Asp Leu Lys Met Thr Gin Ar</td>
</tr>
<tr>
<td>555-560</td>
<td>Asp Leu Leu Lys Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>565-570</td>
<td>Thr Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>575-580</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>585-590</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>595-600</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>605-610</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>615-620</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>625-630</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>635-640</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>645-650</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>655-660</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>665-670</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>675-680</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>685-690</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>695-700</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
</tbody>
</table>

Note: The sequence continues beyond the visible portion of the image.
Gln Asp Leu Glu Asp Lys Ile Arg Ala Leu Lys Glu Asn Lys Asp Gl
695 695 700
Leu Glu Ser Val Leu Glu Val Leu His Arg Gln Met Glu Gln Tyr Ar
705 710 715 720
Asp Gln Pro Gln His Leu Glu Lys Ile Ala Tyr Gln Gln Lys Leu Le
725 730 735
Gln Glu Asp Leu Val His Ile Arg Ala Glu Leu Ser Arg Glu Ser Th
740 745 750
Glu Met Glu Asn Ala Trp Asn Glu Tyr Leu Lys Leu Glu Asn Asp Va
755 760 765
Glu Gln Leu Gln Thr Leu Gln Glu Gln His Arg Arg Ala Phe Ph
770 775 780
Phe Gln Glu Lys Ser Gln Ile Gln Lys Asp Leu Trp Arg Ile Glu As
785 790 795 800
Val Thr Ala Gly Leu Ser Ala Asn Lys Glu Asn Phe Arg Ile Leu Va
805 810 815
Glu Ser Val Lys Asn Pro Gln Arg Lys Thr Val Pro Leu Phe Pro Hi
820 825 830
Pro Pro Val Pro Ser Leu Thr Ser Glu Ser Lys Pro Pro Pro Gl
835 840 845
Pro Ser Pro Pro Thr Ser Pro Val Arg Thr Pro Leu Glu Val Arg Le
850 855 860
Phe Pro Gln Leu Gln Thr Tyr Val Pro Tyr Arg Pro His Pro Gln
865 870 875 880
Leu Arg Lys Val Thr Ser Pro Leu Gln Ser Pro Thr Lys Ala Lys Pr
885 890 895
Lys Val Gln Glu Asp Glu Ala Pro Pro Arg Pro Pro Leu Pro Glu Le
900 905 910
Tyr Ser Pro Glu Asp Gln Pro Pro Ala Val Pro Pro Leu Pro Arg Gl
915 920 925
Ala Thr Ile Ile Arg His Thr Ser Val Arg Gly Leu Lys Arg Gln Se
930 935 940
Asp Glu Arg Lys Arg Asp Arg Glu Leu Gly Gln Cys Val Asn Glu As
945 950 955 960
Ser Arg Val Glu Leu Arg Ser Tyr Val Ser Glu Pro Glu Leu Ala Th
965 970 975
Leu Ser Gly Asp Met Ala Gln Pro Ser Leu Gly Leu Val Gly Pro Gl
980 985 990
Ser Arg Tyr Gln Thr Leu Pro Gly Arg Gly Leu Ser Gly Ser Thr S
995 1000 1005
Arg Leu Gln Gln Ser Thr Ile Ala Pro Tyr Val Thr Leu Arg Ar
1010 1015 1020
Gly Leu Asn Ala Glu Ser Ser Lys Ala Thr Phe Pro Arg Pro Lys S
1025 1030 1035 1040
Ala Leu Glu Arg Leu Tyr Ser Gly Asp His Gln Arg Gly Lys Met Se
1045 1050 1055
Ala Glu Glu Leu Glu Arg Met Lys Arg His Gln Lys Ala Leu Va
1060 1065 1070
Arg Glu Arg Lys Arg Thr Leu Gly Glu Gly Glu Thr Gly Leu Pr
1075 1080 1085
Ser Ser Arg Tyr Leu Ser Arg Pro Leu Pro Gly Asp Leu Gly Ser Va
1090 1095 1100
Cys
1105

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4080 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 79..1380
(D) OTHER INFORMATION: /product= "TPC3"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GUAAACCGGC TTTAAACACT CACGGCAAGGC CACGCTCGCC GTAGATTGCA CCAGGAGGCGG 60
CGCCTGGGCG CTGCGTGGCA CAA GAT CCC GTT CAA ACT CAG CTG CCA CCA AGT
Pro Asp Pro Val Glu Thr Glu Leu Pro Pro Ser 1 5 10
GCC CCT TTT TCT TCT GAA TTT GCA TCC TGG AGC AAC ATT TTT CCA GTT GAG
Ala Pro Phe Leu Ser Gly Leu Arg Phe Cys Thr Asn Phe Pro Val Glu 15 20 25
GGT GTG CAG CAG CAG CCT CTG CAG CCG CTC TGG AAC ACG CGG CCT TGG
Gly Gly Ser Ala Leu Ser Glu Pro Leu Pro Ser Lys Thr Arg Pro Trp 30 35 40
TCT ACG CCT CAG GCG GAT CAG CAC TAC GGC ATG CAC TAC 255
CAG GCC GCC ATG CAG CAC TAC GCC 50 55
GGT TCA CTG CAC GCC ATG AAT TAA CTA TAT CAA CAG CTC CAG CAC GCA GCC 303
Gly Tyr Ser Leu His Ala Met Asn Ser Leu Ser Ala Met Tyr Asn Leu 60 65 70 75
CAC CAG CAG CAC CAG CAC CAG CAG CAC CCC CAC CAG CAG CAC 351
His Glu Glu Glu Glu Glu His Ala Pro Tyr Arg Pro 80 85 90
GTC TAC CAG GCA GAT CTT GTG CAG CAG CTC CAG CAG CTC 399
Ser Val His Ala Leu Thr Leu Ala Glu Arg Leu Ala Gly Cys Thr Phe 95 100 105
CAG AAT CTC CAG TGG GAG GCC ATT CGT CCT ATG TAT CAC CAG CAC CGC CAG 447
Gln Asp Ile Ile Leu Glu GCC ATT CGT CCT ATG TAT CAC CAG CAC CGC CAG 110 115 120
GCC ATT CGT CCT ATG TAT CAC CAG CAC CGC CAG 495
Gln Asp Ser Arg Thr Ala Phe Thr Ala Gln Leu Glu Ala Leu Glu
125 130 135
AAG ACC TTT CAG AAG ACT CAC TAC CCA GAT GGT GGT ATG CTG GAG 543
Lys Thr Phe Gln Lys Thr His Tyr Pro Asp Val Val Met Arg Glu Arg 140 145 150 155
CTG GCC ATG CGC TAC ACC CTG CCT GAG GCC CGG GTG CAG TGG TGC 591
Leu Ala Met Cys Thr Asn Leu Pro Glu Ala Arg Val Glu Val Trp Phe 160 165 170
AAG ACC CAG CGC AAG TAC CAG AAG AAG CAG CTT CAG CAG CAG AAG 639
Lys Asn Arg Ala Lys Phe Arg Lys Lys Gln Arg Ser Leu Gln Lys 175 180 185
-continued

GAA CAG CTC CAG AAC CAG AAG GAG GCT GAG GCC TCC CAT GCG GAA GGC
  190   195
Glu Gin Leu Gin Lys Gin Lys Gin Ala Glu Gly Ser His Gly Glu Gly
Glu 200

AAC GCC GAG GCC CCC ACT CCA GAT CAG CTG CAG ACT GAG CAG CCC
  687
Lys Ala Glu Ala Pro Thr Pro Asp Thr Gin Leu Asp Thr Gin Leu Pro

GCA GCT CTG CCT GCC ACC GAC CCC GCT GCT GAG CTG CAC CTC AGT CTC
  205   210   215
Pro Arg Leu Pro Gly Ser Asp Pro Pro Ala Glu Leu His Leu Ser Leu

TCT GAG CAG TCA GCC AGT GAG TCA GCC CCT GAG GAT CAG CCC GAC CTT
  735
Ser Glu Gin Ser Ala Ser Glu Ser Ala Pro Glu Asp Glu Pro Asp Arg

TGG GAG GCC ACC AGG GCA GGG CTT GAG GAC CCC AAA GCT GAG CCC AAG
  240   245   250
Glu Glu Asp Pro Arg Ala Gly Ala Lys Ser

CCT GGU GCT GAC AGC AAG GGG CTG GCC TGC AAG AGG GCC GAC ACC CCC
  879
Pro Gly Ala Asp Ser Lys Leu Gly Lys Cys Lys Arg Gly Ser Pro Lys

GCA GAT TCC CCA GCC ACC CTG ACC ATC ACT CCT GTC GCC CCA GGG GCT
  927
Ala Asp Ser Pro Gly Ser Leu Thr Thr Pro Val Ala Pro Gly Gly

GCC CTC CTG GCC CCC TCC CAC TCC TAT TCC TCG TGC CAG GTC AGC CTC
  240   245   250
Gly Leu Leu Gly Pro Ser His Ser Tyr Ser Ser Ser Pro Leu Ser Leu

CTC GGT CAG CAA TCC CAG CTG CAT ATG GCG GCC ACC AAC AAC
  1023
Glu Arg Leu Glu Glu Gin Phe Arg Arg Pro Ala Ala Thr Aan Aen

CTG GTG CAC TAC TCG TCC TTC TAC GAA GTA GGU GGT CGC GCC CCT GCT
  1119
Leu Val His Tyr Ser Ser Phe Glu Val Gly Gly Pro Ala Pro Ala Ala

GCA GCC GCG GGT GCT GCT GTG CCC TAC CTG GCC GTC ACC ATG GCC CCG
  1167
Ala Ala Ala Ala Ala Val Pro Tyr Leu Glu Val Asn Met Ala Pro

CTG GCC TCA CTG CAC TCG TCC TGC TAC TCG TCC TGG TGC GCA GCC
  1215
Leu gly Ser Leu His Gly Ser Tyr Gin Ser Leu Ser Ala Ala

GCT GCT GCC CAC GAG GGT GGT TGC GCT CCT CTC GCA CCC
  1263
Glu Ala Ala His Glu Val Trp Gly Ser Pro Leu Pro Leu Pro Leu Leu

CCA GCA GCC CTT GCT GTC CAC GCT ACC AGT AAA ACC ACA
  1311
Pro Ala Gly Leu Ala Ala Pro Ser Thr Thr Gin Cys Ala Ser Ala

AGC ATC GAG AAC CTG CGG CTC CGG GCC AAG CAG CAC GCG GCC TCC CTG
  1359
Ser Ile Glu Asn Leu Arg Leu Arg Ala Lys Gin His Ala Ala Ser Leu

GGA CTC GAT AGC CTG CCC AAC TGC ACT GTC TGC GCC CAA CAC GGG GCT
  1410
Gly Leu Asp Thr Pro Asn

GTCATTTAC CTCATCTCC AAG GAG TAT CAG GCC CAT ACC AAT GAG TTA AAG
  1470
CATGACCCCA GAGATCTAG GGCCTGGGCT CCTGCCATCT CTCATACAC

ACCGCAGC TAA CGCCCA CACATCCA ACCATCG CTTGCTGAC ACCCATG
  1530
GCCGAAATG GGGAGGATGG CGACTGCTG GCCGTCCCTG AAGTGT

GGCTGCCAG CAGGTCATC TCCATCCAA AAGAC AAT CACT CAGGTCACCAT CTTGTCCTC
  1650

TGCCGCTGCA TAAAGTCTCG TGCTGGAGT CATGATATATA TAAATTATA TAAAGCTA
  1770
TTTCAGGCT CTGGCCGCTC CCCAGCCCCC TCCCCACCTT CTATCCTTCT TCCTCCCTG 1830
CCCCTCCCGT CAGCGCTCCGC GCTTCACTCG AGCACTCCCT TCTATTTTCT CTTCTCTCC 1990
CTCTCTTCTT CCTGCTGACT CCCCCTCCTC GTGCTGTCTG TGCTCCCGTG CCCGCTCT 1950
GCCCAGCCCT TGTATTTCTG CCTGCCTGAC CCCGCTCGCC TTCTCCCTGC TTCCCTG 2010
TTCTCTCTTT AGTTGGCTCT GTGTGGATAT CACCACTTCT TGAAGCTATAT TTGGTTTG 2070
GTGTTGGCTG CATTGCTTTT TATTAAAATT GGACTTCCC CTGCTCTTGCT TCTATCC 2130
TTCTCTCTTC CCTGCTGCTC TTCCCTGACCC TGCGCTAAAAC CTAAGAGACT GTTCTCTTTC 2190
ATGCCAATAG GAACCACAGA GGCCTAGAGC GGAATCTCTG AGGAGTGTGA GCTGCAGG 2250
GAAGTGGATG CTGAGCTGCA GCCGCCACCC TGGCCGGGCT GGAACCTAGA AGGAGCTC 2310
AAATTGCTTTG TCTTTGCTCT CTTCAATTTT GAGCCCTCTCA TTCCCAAGAT TGAGAGG 2370
GTAGTTTATCC TGGCGAACGG TGGACAGGAG CAGGTATCTT GGCCTTAAGT CTCCCGCA 2430
CAGAGGTGCTG TGAGCACAGT TCAGCCAGAA TGACCCGCAAC GGCCTCTAGG CTCTCTTTT 2490
TGAGTTGTGG CTTGAGTGAGC AGGCAGCAGG GGAGGAAGAG GGCTCTCAGA GACGGCATG 2550
CTGAGTCCTG CTTCAGACAA AGCTAGCTCC AGCCCTGCAA GATACCTGCC AAGGCAGAG 2610
TACGGCTGAC AGGAAGTGGG GATTGAACCT TCCTCTTGGC CACGGTGCGG GCAAACAGC 2670
GAGCTGACAA TCCAGCTTCC AGTGGGTCAG TTCTCGTCTC CAGACGCGAT TATGGAGA 2730
AGGCGGCTCC TCACAGCTACT CATGGGTATAA GAAAAAGGAA AGCTTTGGCT GCCTTTCTAC 2790
AACCCAGGCT AGCAAACTCC TTCTCAAGGC TGCTGTGACC GGTGGGAGTG GTGCCCTCA 2850
CCCTAAATAC AAGCCAGCTT TGGACGCAAC GCGAGCAACT CTGCAAGGCTG AAGAGTTCC 2910
GACCCAGCTG CTGAGATGAT TGAATAATAT AATAAAATTG AATATTGCTG CGCTGCGTG 2970
ACATCCTCTG ATCTCAGGCT AATCTGGAAGT CTTAGTTGGG AAGATCTCTG CAATCTGG 3030
GAGGGAATTC GCATGAGGCG GAGATCTAGC CACTCCAGCC CAGCCTGGCG GACGGAGT 3090
GACCTCCTCT CAAAAAAAGAAAAA ATATATACAAA ATAAAAAGCTG TGTGACCTTG GCACACAGC 3150
TACGTGCTCTG TGGCGCTTTT CCCCCTCTGG GCCTGGTCAGC TGTAAAGGCT CTGTTAGGCT 3210
CTACATCCTG CATTTGCGAG CGACCTGTTA AAGACAGATG TTATGATTA AATTGCAAA 3270
GAAGGCCTTC AAAACGCTAA TATATGAGT AGAACAAAAT TATATGATTA CTTGTGAGG 3330
AGGCGGCTTC AGCTGCTGAA GGCCTCGCTT CAGCTTTGGC CTGTTAGGCT CACAGCCATG 3390
TGGCGCCTCTT CTAGTTCCAG CAATCGTGG TACATGTGGT AAGCCAGCTG GACCCACTC 3450
AGCAGCCTCA TTTCTTCAAC CTTTTACTAT GTGACATGAC GAGCCGGCTC CTGGCGAGAG 3510
GCTGGCGAGC ATACCTGGTG ATGTCCTCCCA AACTGTTGGT GCTGACGACT GACACACA 3570
AACCCCTGCT ATCAGAAAGAG ATTTATTTGG CACGGCAAGG AGGAGATGCT AAGCGCTG 3630
TCTCTCCCCA GTGGAGAGAG AGAGAATGCT TGGTTTGCCT GTTGTGAATG CCTGGAGG 3690
AAGCCTGATG CACAGCCTCC TGGTGGGGGT CAGGCAAGGC TGGCGGAGAT GAGGAAGG 3750
TTGGACAGGA AAAAATGAGA CAGAGAAGCT GTCTAAGCCA GGGAAGATGT TGGAACAA 3810
CATCAGAGTT GAGAATCACT AGGACAGGGT AGGAGGAGTG CTGCTTTCCAC GAGGAGTG 3870
GCTAAGGCTGG GGAGAATCCC AGGCTAAACA CACACCCTAT ACAGGCAAG AAGACAGA 3930
GGCAAGTGGG GGTGGTCCAA CTTCAGGCTC TCTCTGATGG GTGGAGAAAT TGGAAGGG 3990
ATCCACACGT CAGGGGCGCA TCAGAGAGAA GACCAAGAGT GCCTGGTGGG AGGGTGCG 4050
ATTCCAGGAC TGACAGGAG GOTTTTTGT

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 434 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Pro Asp Pro Val Gln Thr Gln Leu Pro Pro Ser Ala Pro Phe Leu Ser 1 5 10 15
Gly Leu Arg Phe Cys Thr Asn Phe Pro Val Glu Gly Gly Ser Ala Leu 20 25 30
Ser Gln Pro Leu Pro Ser Lys Thr Arg Pro Trp Ser Arg Asn Leu Gln 35 40 45
Ala Asp Ala Ala Met Gln His Tyr Gly Val Asn Gly Tyr Ser Leu His 50 55 60
Ala Met Asn Ser Leu Ser Ala Met Tyr Asn Leu His Gln Gln Ala Ala 65 70 75 80
Gln Gln Ala Gln His Ala Pro Asp Tyr Arg Pro Ser Val His Ala Leu 85 90 95
Thr Leu Ala Glu Arg Leu Ala Gly Cys Thr Phe Gln Asp Ile Ile Leu 100 105 110
Glu Ala Arg Tyr Gly Ser Gln His Arg Lys Gln Arg Arg Ser Arg Thr 115 120 125
Ala Phe Thr Ala Gin Gin Leu Glu Ala Leu Glu Thr Phe Gln Lys 130 135 140
Thr His Tyr Pro Asp Val Val Met Arg Glu Arg Leu Ala Met Cys Thr 145 150 155 160
Asn Leu Pro Glu Ala Arg Val Gln Val Trp Phe Lys Asn Arg Arg Ala 165 170 175
Lys Phe Arg Lys Lys Gin Arg Ser Leu Gin Lys Glu Gin Leu Gin Lys 180 185 190
Gln Lys Glu Ala Glu Gly Ser His Gly Glu Lys Ala Glu Ala Pro 195 200 205
Thr Pro Asp Thr Gin Leu Asp Thr Glu Gln Pro Arg Leu Pro Gly 210 215 220
Ser Asp Pro Pro Ala Glu Leu His Leu Ser Leu Ser Glu Gin Ser Ala 225 230 235 240
Ser Glu Ser Ala Pro Glu Asp Gin Pro Asp Arg Glu Gin Asp Pro Arg 245 250 255
Ala Gly Ala Glu Asp Pro Lys Ala Glu Lys Ser Pro Gly Ala Asp Ser 260 265 270
Lys Gly Leu Gly Cys Lys Arg Gly Ser Pro Lys Ala Asp Ser Pro Gly 275 280 285
Ser Leu Thr Ile Thr Pro Val Ala Pro Gly Gly Leu Leu Gly Pro 290 295 300
Ser His Ser Tyr Ser Ser Ser Pro Leu Ser Leu Phe Arg Leu Gln Glu 305 310 315 320
AATCTAGTTC CTGCTCTGCA G

(2) INFORMATION FOR SEQ ID NO: 6:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
ATGGGATTC CAGGGTGGAG CT

(2) INFORMATION FOR SEQ ID NO: 7:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:
ACCTGCTCTC AGGGCCCACA AGT

(2) INFORMATION FOR SEQ ID NO: 8:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:
TAAGACAAG AACAGTCAC AACA

(2) INFORMATION FOR SEQ ID NO: 9:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:
ATTTGTGCTT AGAGGTCGTG CCAG

(2) INFORMATION FOR SEQ ID NO: 10:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: <Unknown>
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: peptide
Leu Ala Met Cys Thr Asn Leu Pro Glu Ala Arg Val Gln Val Trp Ph 1 5 10 15
Lys Asn Arg Ala Lys Phe Arg 20

(2) INFORMATION FOR SEQ ID NO: 11:

  (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 25 amino acids
      (B) TYPE: amino acid
      (C) STRANDEDNESS: <Unknown>
      (D) TOPOLOGY: linear

  (ii) MOLECULE TYPE: peptide

  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Ser Ser Ser Lys Val His Ser Phe Gly Lys Arg Asp Gln Ala Ile Ar 1 5 10 15
Arg Asn Pro Asn Val Pro Val Val Val 20 25

(2) INFORMATION FOR SEQ ID NO: 12:

  (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 5 amino acids
      (B) TYPE: amino acid
      (C) STRANDEDNESS: <Unknown>
      (D) TOPOLOGY: linear

  (ii) MOLECULE TYPE: peptide

  (ix) FEATURE:
      (A) NAME/KEY: Modified-site
      (B) LOCATION: 1
      (D) OTHER INFORMATION: /product= "OTHER"
      /note= "Xaa = Ala or Pro"

  (ix) FEATURE:
      (A) NAME/KEY: Modified-site
      (B) LOCATION: 4
      (D) OTHER INFORMATION: /product= "OTHER"
      /note= "Xaa = Ala or Pro"

  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Xaa Pro Pro Xaa Tyr 1 5

(2) INFORMATION FOR SEQ ID NO: 13:

  (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 26 amino acids
      (B) TYPE: amino acid
      (C) STRANDEDNESS: <Unknown>
      (D) TOPOLOGY: linear

  (ii) MOLECULE TYPE: peptide

  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Trp Ser Tyr Gly Val Cys Arg Asp Gly Arg Val Phe Phe Ile Asn As 1 5 10 15
Gln Leu Arg Cys Thr Thr Trp Leu His Pro 20 25
-continued

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: <unknown>
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Trp Phe Val Leu Ala Asp Tyr Cys Leu Phe Tyr Tyr Lys Ala Glu Ly
1 5 10 15
Lys Arg Ser Ser Xaa Ser Ile Pro
20

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: <unknown>
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Trp Glu Glu Gly Phe Thr Glu Glu Gly Ala Ser Tyr Phe Ile Asp Hi
1 5 10 15
Asn Gln Gin Thr Thr Ala Phe Arg His Pro
20 25

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 100 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: <unknown>
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Pro Asn Ile Pro Gin Met Ser Ala Phe Trp Tyr Ala Val Arg Thr Al
1 5 10 15
Val Ile Asn Ala Ala Ser Gly Gin Thr Val Asp Glu Ala Leu Ly
20 25 30
Asp Ala Gin Thr Asn Ser Ser Asn Asn Asn Asn Asn Asn Asn Asn
35 40 45
Asn Asn Leu Gly Ile Gly Arg Gly Ile Ser Glu Phe Ala Ala Ala Se
50 55 60
Thr Leu Asp Leu Lys Met Thr Gly Arg Asp Leu Leu Lys Asp Arg Se
65 70 75 80
Leu Lys Pro Val Lys Ile Ala Glu Ser Thr Asp Thr Val Leu Leu Se
85 90 95
Ile Phe Cys Glu
100
(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 15 amino acids
   (B) TYPE: amino acid
   (C) STRANDEDNESS: <Unknown>
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

Arg Gly Leu Lys Arg Gln Ser Asp Glu Arg Lys Arg Asp Arg Glu
1  5  10  15

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 15 amino acids
   (B) TYPE: amino acid
   (C) STRANDEDNESS: <Unknown>
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Lys Val Thr Ser Pro Leu Gln Ser Pro Thr Lys Ala Lys Pro Lys
1  5  10  15

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 20 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:
   (A) NAME/KEY: modified_base
   (B) LOCATION: 1
   (D) OTHER INFORMATION: /mod_base= OTHER
   /note= "N = 5'-phosphorylated adenine (p-A)"

(ix) FEATURE:
   (A) NAME/KEY: modified_base
   (B) LOCATION: 20
   (D) OTHER INFORMATION: /mod_base= OTHER
   /note= "N = adenine substituted at the 3' position of deoxyribose with an amino group (A-NH-2)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

NTAGCGGCCGCAAGAATTCN

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 20 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

TGAATTCTGCGGCCTAT
(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 1
(D) OTHER INFORMATION: /mod_base= OTHER /note= "N = 5'-phosphorylated cytosine (p-C)"

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 25
(D) OTHER INFORMATION: /mod_base= OTHER /note= "N = guanine substituted at the 1' position of deoxyribose with an amino group (G-NH-2)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

NAGAAGCTTG GTTGGATCCA GCAAN 25

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

CTTGCTGGAT CCAACCAAGC TTCTG 25

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

CTCACTGTAG ACACTGCCTC AGTTTC 26

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

CAGAAGCTTG CACTGGACT CAAGATC 27
The invention claimed is:

1. A monoclonal or isolated polyclonal antibody that specifically binds human TPC2 or TPC3 protein, as shown in SEQ. ID NO:2 or 4, respectively.

2. The antibody of claim 1, which specifically binds human TPC2 protein (SEQ. ID NO:2).

3. The antibody of claim 1, which specifically binds human TPC3 protein (SEQ. ID NO:4).

4. A method for obtaining an antibody according to claim 1, comprising collecting antiserum from a subject immunized with a peptide comprising at least 10 contiguous amino acids of SEQ. ID NO:2 or SEQ. ID NO:4.

5. A method for obtaining an antibody according to claim 1, comprising collecting spleen cells from an animal immunized with a peptide comprising at least 10 contiguous amino acids of SEQ. ID NO:2 or SEQ. ID NO:4.

6. A method for obtaining an antibody according to claim 1, comprising screening a phage antibody display library with a peptide comprising at least 10 contiguous amino acids of SEQ. ID NO:2 or SEQ. ID NO:4.

7. An antibody obtained according to the method of claim 4.

8. A host cell secreting a monoclonal antibody according to claim 1.

9. A composition effective for obtaining a TPC2-specific antibody according to the method of claim 4, comprising at least 10 contiguous amino acids of SEQ. ID NO:2.

10. A composition effective for obtaining a TPC3-specific antibody according to the method of claim 4, comprising at least 10 contiguous amino acids of SEQ. ID NO:4.

11. A method of determining a condition in a subject associated with a high level of TPC2 (SEQ. ID NO:2), comprising combining the antibody of claim 2 with a sample obtained from the subject, and correlating protein detected by the antibody with the condition of the subject.

12. A method of determining a condition in a subject associated with a high level of TPC3 (SEQ. ID NO:4), comprising combining the antibody of claim 3 with a sample obtained from the subject, and correlating protein detected by the antibody with the condition of the subject.

13. An isolated, recombinant or synthetic nucleic acid encoding a peptide immunogenic for a TPC2-specific antibody according to claim 2, comprising at least 25 contiguous amino acids of SEQ. ID NO:1.

14. The nucleic acid of claim 13, comprising at least 100 consecutive nucleotides of SEQ. ID NO:1.

15. A host cell containing a nucleic acid according to claim 13.

16. A method of obtaining an antibody according to claim 2, comprising expressing the nucleic acid of claim 13 in a host cell.

17. An isolated, recombinant or synthetic nucleic acid encoding a peptide immunogenic for a TPC3-specific antibody according to claim 3, comprising at least 25 contiguous amino acids of SEQ. ID NO:3.

18. The nucleic acid of claim 17, comprising at least 100 consecutive nucleotides of SEQ. ID NO:3.

19. A host cell containing a nucleic acid according to claim 17.

20. A method of obtaining an antibody according to claim 3, comprising expressing the nucleic acid of claim 17 in a host cell.