

(11) Número de Publicação: **PT 1370684 E**

(51) Classificação Internacional:

C12Q 1/68 (2007.10) **C12N 15/11** (2007.10)
A61K 31/00 (2007.10) **G01N 33/574** (2007.10)
G01N 33/68 (2007.10)

(12) **FASCÍCULO DE PATENTE DE INVENÇÃO**

(22) Data de pedido: **2001.06.15**

(30) Prioridade(s): **2000.06.15 US 211835 P**

(43) Data de publicação do pedido: **2003.12.17**

(45) Data e BPI da concessão: **2008.05.28**
175/2008

(73) Titular(es):

NOVARTIS VACCINES AND DIAGNOSTICS, INC.
4560 HORTON STREET EMERYVILLE, CA 94608
US

(72) Inventor(es):

CHRISTOPH REINHARD **US**
GIULIA C. KENNEDY **US**
SANMAO KANG **US**
ANNE BENNET JEFFERSON **US**

(74) Mandatário:

FRANCISCO NOVAES CUNHA BRITO SOTO MAIOR DE
ATAYDE
AV. DUQUE D'AVILA, N.º 32, 1º ANDAR 1000-141 LISBOA PT

(54) Epígrafe: **POLINUCLEÓTIDOS LIGADOS AO CANCRO DO COLON**

(57) Resumo:

DESCRIÇÃO

EPÍGRAFE:	<u>"POLINUCLEÓTIDOS LIGADOS AO CANCRO DO COLON"</u>
-----------	--

Campo da Invenção

A presente invenção está relacionada com genes expressos diferencialmente no cancro do cólon e na displasia. Mais especificamente, está relacionada com polinucleotídeos que são regulados diferencialmente no cancro do cólon e nos produtos genéticos codificados.

Antecedentes da Invenção

O cancro do cólon é a segunda causa de morte relacionada com o cancro nos Estados Unidos. A Sociedade Americana do Cancro estima que surgirão aproximadamente 94.700 novos casos de cancro do cólon nos Estados Unidos em 1999, e que o cancro do cólon será responsável por cerca de 47.900 mortes. O cólon tem quatro secções: o cólon ascendente, o cólon transversal, o cólon descendente e o cólon sigmóide, e termina no recto. Pólipos adenomatosos ou adenomas, lesões benignas comuns que evoluem para carcinomas podem desenvolver-se em qualquer das quatro partes do cólon ou no recto. Mais de 95% dos cancros do cólon são adenocarcinomas, ou cancros das células que revestem o interior do cólon. O cancro do cólon metastiza frequentemente para o fígado e para os pulmões.

Ao contrário do cancro do pulmão, no qual o fumo foi identificado como o principal factor etiológico responsável pela

doença, os mecanismos associados ao aparecimento do cancro do cólon são complexos e ainda não completamente compreendidos. Acredita-se que a alimentação promove a carcinogénese, especialmente um grande consumo de gorduras. Ao nível molecular, suspeita-se do envolvimento de um processo em vários passos envolvendo uma série de mutações na transformação de adenomas em tumores do cólon (Vogelstein et al. (1988) N. Engl. J. Med. 319:_525-532). O desenvolvimento e progressão do cancro do cólon é devido a uma sequência de mutações em três tipos de genes: oncogenes, genes supressores de tumores e gene reparadores de emparelhamentos inadequados, que controloam a velocidade de mutação de outros genes, incluindo oncogenes e genes supressores de tumores. Essas mutações ocorrem como resultado de uma predisposição genética (mutações germinais) ou em resposta a factores ambientais (mutações somáticas).

Foram identificadas várias mutações associadas ao cancro do cólon. Mutações germinais ligadas ao cancro do cólon hereditário ou familiar incluem o gene supressor de tumores APC (adenomatous polyposis coli) (Lengauer et al. (1991) Science 253:_665-669) e os genes reparadores de emparelhamentos inadequados MutL e MutS (Modrich (1995) Phil. Trans. R. Soc. Lond. B 347:_89-95; Kolodner (1996) Genes Dev. 10:_1433-1442). Genes APC defeituosos foram implicados na polipose adenomatosa familiar (FAP) e MutL e MutS no cancro colorectal não poliposo hereditário (HNPCC). Mutações somáticas identificadas associadas ao cancro do cólon esporádico incluem os oncogenes K-ras, o-myc, e os genes supressores de tumores p53, APC, proteína de activação da GTPase de neurofibromatoses do tipo 1 (NF 1 INTERVALO), "*deleted in colon cancer*" (DCC) e "*mutated in cancro do cólon*" (MCC) (Midgley et al. (1999) Lancet 353:391-399). WO 99/33963 e WO 00/22130

descrevem genes marcadores expressos diferencialmente que são crescentemente regulados (up-regulated) no cancro do cólon. O número de adesão AAZ 52884 descreve o isolamento de uma sequência muito semelhante de uma biblioteca de cDNA do tumor da próstata humano.

O cancro do cólon é bastante fácil de tratar e frequentemente curável quando detectado e tratado nos estágios iniciais. Procedimentos de diagnóstico convencionais incluem procedimentos invasivos, tais como toque rectal, sigmoidoscopia, colonoscopia e enema de bário, e procedimentos não invasivos, tais como teste de sangue oculto fecal e o rastreio genético. A utilização de marcadores tumorais é particularmente indicada para a identificação da doença hereditária assim como para o diagnóstico de reincidências. Por exemplo, a detecção do antígeno carcino-embriónico (CEA) é utilizada no diagnóstico da reincidência assintomática. Os novos métodos de diagnóstico incluem a imagiologia de fluorescência induzida por laser que consegue detectar células cancerosas na superfície epitelial ou no interior da parede do cólon (ver, p. ex., von Rueden et al. (1993) J. Surg. Oncol. 53:43-46).

Os tratamentos convencionais para o cancro do cólon incluem a extracção cirúrgica, radio e quimioterapia, incluindo terapia adjuvante. A terapêutica genética inclui transferência de citoquina ou genes dos antígenos imunitários, transferência de sistemas enzima-profármaco (ver, p. ex., Huber et al (1993) Cancer Res. 53:4619-4626) e substituição de genes supressores de tumores (ver, p. ex., Venook et al. (1998) Proc. ASCO 17:431a) utilizando vectores virais (Zwacka et al. (1998) Hematol. Oncol. Clin. North Am. 12:595-615).

Enquanto que vários genes associados ao cancro do cólon já foram identificados, a identificação de genes ligados ao desenvolvimento (ou inibição do desenvolvimento) do cancro do cólon podem fornecer novas ferramentas de diagnóstico bem como novos alvos terapêuticos. A identificação de genes expressos diferencialmente no cancro do cólon é particularmente importante para a evolução do desenvolvimento de fármacos e das técnicas de diagnóstico, bem como para a compreensão da evolução e natureza do cancro do cólon. A invenção permite a identificação de tais genes expressos diferencialmente.

Sumário da Invenção

Esta invenção está relacionada com polinucleotídeos que representam genes expressos diferencialmente no cancro do cólon, p. ex., pólipos adenomatosos, carcinoma colorectal, elevado potencial metastático tumor do cólon e cancro metastático do cólon. A invenção está também relacionada com meios de diagnóstico e terapêuticas compreendendo tais polinucleotídeos, os genes ou produtos genéticos correspondentes, incluindo sondas, nucleotídeos *antisense* e anticorpos.

De certo modo, a invenção proporciona um método de identificação de células **cancerosas** do cólon, envolvendo a detecção de pelo menos um produto genético expresso diferencialmente, sendo o produto genético codificado por um gene contendo a sequência SEQ ID NO:1 numa amostra, sendo essa amostra proveniente de uma célula suspeita de ser uma célula **cancerosa** do cólon, e a comparação do nível de expressão do produto genético expresso diferencialmente com o nível de expressão do produto genético expresso diferencialmente numa amostra de controlo, a qual provém de uma célula do cólon **cancerosa**. A detecção de um

nível de expressão do produto genético expresso diferencialmente na amostra semelhante ao nível de expressão do produto genético na amostra de controlo indica que a célula da amostra é uma célula do cólon **cancerosa**. Numa modalidade, a detecção é conseguida por hibridação da amostra com uma matriz (array) de referência, a qual compreende uma sequência identificativa SEQ ID NO:1.

A invenção proporciona ainda um método de identificação células **cancerosas** do cólon, envolvendo a detecção de pelo menos um produto genético expresso diferencialmente, feita pela detecção da hibridação de um polinucleotídeo compreendendo a sequência SEQ ID NOS:1 numa amostra proveniente de uma célula suspeita de ser uma célula cancerosa do cólon, e comparação do nível de hibridação do produto genético expresso diferencialmente detectado com o nível de hibridação do produto genético expresso diferencialmente numa amostra de controlo, a qual é proveniente de uma célula cancerosa do cólon. A detecção de um nível de hibridação do produto genético expresso diferencialmente na amostra semelhante ao nível de hibridação do produto genético na amostra de controlo indica que a célula de amostra é uma célula cancerosa do cólon. Numa modalidade, a detecção é conseguida através da hibridação da amostra com uma matriz de referência, compreendendo a matriz de referência uma sequência identificativa SEQ ID NO:1.

A invenção proporciona ainda um polinucleotídeo contendo a sequência SEQ ID NO:1 ou variantes degeneradas da mesma. Em alguns aspectos, a invenção proporciona matrizes e células hospedeiras recombinantes contendo um polinucleotídeo da invenção. Numa modalidade o polinucleotídeo inclui a sequência

nucleotídica de uma inserção contida num clone aqui descrito e depositado com o ATCC.

Num outro aspecto a invenção proporciona um polipeptídeo isolado codificado por um gene expresso diferencialmente da invenção, assim como anticorpos que se ligam eespecificaamente a tais polipeptídeos.

Num outro aspecto, a invenção proporciona composições terapêuticas contendo um princípio activo para modulação e expressão de um gene expresso diferencialmente em células cancerosas do cólon. Por exemplo, o princípio activo da composição terapêutica pode provocar uma diminuição da actividade biológica de um produto genético codificado por um gene sobre-expresso numa célula cancerosa relativamente a uma célula normal, ou pode provocar um aumento da actividade biológica de um produto genético codificado por um gene sob-expresso numa célula cancerosa relativamente a uma célula normal.

A invenção proporciona também uma biblioteca de genes expresso diferencialmente, a qual inclui a sequência informação do polinucleotídeo SEQ ID. NO:1.

A biblioteca pode ser disponibilizada como uma matriz de ácidos nucleicos ou em formato digital, e pode incluir quantidades relativas do polinucleotídeo SEQ ID NO:1, as quais são representativas das quantidades relativas do polinucleotídeo encontradas numa célula do cólon doente.

Um objectivo principal da invenção é proporcionar polinucleotídeos que correspondam a genes expressos diferencialmente e aos seus fragmentos, que são úteis no diagnóstico do cancro do cólon, assim como no desenvolvimento racional de fármacos e terapêuticas.

Vários aspectos e modalidades da invenção tornar-se-ão rapidamente evidentes a qualquer especialista que leia a descrição aqui fornecida.

Breve Descrição das Figuras

A FIG. 1 é um gráfico que ilustra os níveis de mensagem do gene correspondente a SK2. (c9083, SEQ ID NO:3) nas linhagem celulares indicadas.

A FIG. 2 é um gráfico que ilustra o efeito dos oligonucleotídeos *antisense* SK2 (9083) sobre os níveis de mensagem para o gene correspondente a SK2 (SEQ ID NO:3).

Os gráficos das FIGS. 3 e 4 ilustram o efeito dos oligonucleotídeos *antisense* SK2 (9083) sobre a proliferação de células SW620 (Fig. 3) e uma linhagem celular não pertencente ao cólon, HT1080 (Fig. 4).

A FIG. 5 é um gráfico que ilustra o efeito dos oligonucleotídeos *antisense* no gene correspondente ao cluster 378805 por crescimento de células SW620 (31-4as: *antisense*; 31-4rc: controlo inverso; WT: controlo selvagem (sem oligonucleotídeo)).

As FIGS. 6-8 são gráficos que ilustram o resultado do ensaio de proliferação com SW620 para verificar o efeito da expressão do K-Ras (controlo, Fig. 6), do gene correspondente a c3376 (CHIR11-4), e do gene correspondente a 402380 (CHIR33-4).

A Fig. 9 é um gráfico que ilustra os efeitos da expressão dos genes correspondentes a K-Ras (controlo) e a 402380 (CHIR33-4) sobre a formação de células SW620 do cólon em agar mole (valores normalizados a WST 1).

Descrição Detalhada da Invenção

Antes de uma descrição mais detalhada do objecto da invenção, deve ser entendido que a invenção não se limita às modalidades específicas da invenção descritas abaixo, pois podem ser feitas variações das modalidades particulares que ainda se enquadram no âmbito das reivindicações anexas. Deve também ser entendido que a terminologia empregue se destina à descrição das modalidades específicas, e não deve ser entendida como limitativa. Pelo contrário, o âmbito da presente invenção será estabelecido pelas reivindicações anexas.

Nesta especificação e nas reivindicações anexas, as formas singulares "um", "uma", "o", "a" incluem referências plurais, excepto se o contexto indicar claramente o contrário. Excepto se definido, em contrário, todos os termos técnicos e científicos aqui utilizados têm o significado comumente atribuído pelos especialistas da área desta invenção.

A invenção relaciona-se com polinucleotídeos contendo as sequências de nucleotídeos referidas, com cDNA completo, com sequências genómicas de mRNA, com genes correspondentes a essas sequências e variantes degeneradas dos mesmos, e com polipeptídeos codificados pelos polinucleotídeos da invenção e variantes de polipeptídeos. A descrição detalhada que se segue descreve as constituições de polinucleotídeos abrangidos pela invenção, métodos de obtenção de cDNA ou DNA genómico codificando um produto genético completo, a expressão desses polinucleotídeos e genes, a identificação de motivos estruturais dos polinucleotídeos e genes, a identificação da função de um produto

genético codificado por um gene correspondente a um polinucleotídeo da invenção, a utilização dos polinucleotídeos fornecidos como sondas e em mapeamento e caracterização de tecidos, a utilização dos polipeptídeos correspondentes e de outros produtos genéticos na criação de anticorpos, e a utilização dos polinucleotídeos e dos produtos genéticos por eles codificados para fins terapêuticos e de diagnóstico.

Definições

Os termos "polinucleotídeo" e "ácido nucleico", aqui utilizados indistintamente, referem-se a formas poliméricas de nucleotídeos de qualquer comprimento, sejam ribonucleotídeos ou desoxinucleotídeos. Assim, esses termos incluem, embora não se limitem a DNA ou RNA de cadeia simples, dupla ou múltipla, híbridos genômicos de DNA, cDNA, DNA e RNA, ou a um polímero contendo bases purínicas e pirimidínicas ou outras bases nucleotídicas naturais, modificadas química ou bioquimicamente, não naturais ou derivatizadas. Esses termos incluem ainda, mas não se limitam a mRNA ou cDNA contendo sequências intrônicas (ver, p. ex., Niwa et al. (1999) Cell 99(7):691-702). O esqueleto do polinucleotídeo pode conter açúcares e grupos fosfato (tal como se encontra normalmente no RNA ou no DNA), ou açúcares ou grupos fosfato substituídos ou modificados. Alternativamente, o esqueleto do polinucleotídeo pode conter um polímero de subunidades sintéticas tais como fosforamiditos e portanto ser um fosforamidato de oligodesoxinucleosídeo ou um oligômero misto fosforamidato-fosfodiéster. Peyrottes et al. (1996) Nucl. Acids Res. 24:_1841-1848; Chaturvedi et al. (1996) Nucl. Acids Res. 24:_2318-2323. Um polinucleotídeo pode conter nucleotídeos modificados, tais como nucleotídeos metilados e análogos de

nucleotídeos, uracilo, outros açúcares, e grupos de ligação tais como fluororibose e tioato, e ramificações de nucleotídeos. A sequência de nucleotídeos pode ser interrompida por componentes não nucleotídicos. Um polinucleotídeo pode ainda ser modificado após polimerização, por exemplo por conjugação com um componente de marcação. Outros tipos de modificação incluídos nesta definição são "caps", a substituição de um ou mais dos nucleotídeos naturais por um análogo, e a introdução de meios para ligar o polinucleotídeo a proteínas, íons metálicos, componentes marcação, outros polinucleotídeos ou suportes sólidos.

Os termos "polipeptídeo" e "proteína", aqui utilizados indistintamente, referem-se a uma forma polimérica de aminoácidos de qualquer comprimento, que podem incluir aminoácidos codificados ou não, aminoácidos modificados química ou bioquimicamente ou derivatizados e polipeptídeos contendo cadeias peptídicas modificadas. O termo inclui proteínas de fusão, incluindo mas não limitado a proteínas de fusão com uma sequência de aminoácidos heteróloga, fusão com sequências líder homólogas e heterólogas, com ou sem resíduos de metionina N-terminais, proteínas imunologicamente marcadas e similares.

"Heterólogos" significa que os materiais são provenientes de diferentes fontes (p. ex., de diferentes genes, de diferentes espécies, ect.).

O termo "gene expresso diferencialmente" tenciona abranger um polinucleotídeo que representa ou corresponde a um gene que é expresso diferencialmente numa célula cancerosa do cólon quando comparado com uma célula não cancerosa do mesmo tipo. Esse gene expresso diferencialmente pode incluir uma sequência potencialmente codificante (*open reading frame* - janela aberta de

leitura) codificando um produto genético (p. ex., um polipeptídeo), assim como intrões desses genes e sequências nucleotídicas não codificadas 5' e 3' adjacentes envolvidas na regulação da expressão, até cerca de 20 kb para além da região de codificação, mas possivelmente mais em cada direcção. O gene pode ser introduzido num vector apropriado para manutenção extracromossomal ou para integração num genoma hospedeiro. Geralmente, uma diferença no nível de expressão associada com uma diminuição no nível de expressão de, pelo menos cerca de 25%, usualmente de pelo menos 50% a 75%, comumente de pelo menos 90% ou mais, é indicativa de um gene de interesse expresso diferencialmente., i.e., um gene que é sob-expresso, ou regulado decrescentemente (*down-regulated*), na amostra, relativamente à amostra de controlo. Para além disso, uma diferença no nível de expressão associada a um aumento na expressão de, pelo menos, cerca de 25%, usualmente de pelo menos cerca de 50% a 70%, comumente de pelo menos cerca de 90%, podendo ser pelo menos 1 vez, normalmente pelo menos cerca de 2 a 10 vezes, ou ainda cerca de 100 a 1.000 vezes maior, relativamente a uma amostra de controlo, é indicadora de um gene de interesse expresso diferencialmente, i.e., um gene sobre-expresso ou regulado crescentemente (*up-regulated*).

"Polinucleotídeo expresso diferencialmente", tal como é aqui utilizado significa uma molécula de ácido nucleico (RNA ou DNA) contendo uma sequência que representa um gene expresso diferencialmente, p. ex., o polinucleotídeo expresso diferencialmente, uma sequência (p. ex., uma janela aberta de leitura (open reading frame), codificando um produto genético) que identifica unicamente um gene expresso diferencialmente de modo que a detecção do polinucleotídeo expresso diferencialmente

numa amostra está relacionada com a presença de um gene expresso diferencialmente ou um produto genético de um gene expresso diferencialmente numa amostra. Por exemplo, a detecção de um polinucleotídeo numa amostra que hibrida (p. ex., em condições severas), para um polinucleotídeo expresso diferencialmente, é indicadora da presença do gene expresso diferencialmente correspondente na amostra. O termo "polinucleotídeos expressos diferencialmente" também é suposto abranger fragmentos dos polinucleotídeos apresentados, p. ex., fragmentos que mantenham a sua actividade biológica, assim como ácidos nucleicos homólogos, substancialmente parecidos, ou substancialmente idênticos (p. ex., com cerca de 90% da sequência idêntica) aos polinucleotídeos apresentados.

Os termos "corresponde a" ou "representa", quando usado no contexto de, por exemplo, um polinucleotídeo ou uma sequência que "corresponde a" ou "representa" um gene, significa que a sequência do polinucleotídeo está presente no gene ou no ácido nucleico do produto genético (p. ex., mRNA). O polinucleotídeo estar totalmente presente no seio de um exão de uma sequência genómica do gene, ou diferentes porções da sequência do polinucleotídeo podem estar presentes em diferentes exões (p. ex., de modo a que esteja presente a sequência polinucleotídica contígua num mRNA, seja pré- ou pós-*splicing*, que seja um produto da expressão do gene). Em algumas modalidades, o polinucleotídeo pode representar ou corresponder a um gene que se encontra modificado numa célula cancerosa relativamente a uma célula normal. Por exemplo, o gene na célula cancerosa pode ser modificado pela inserção de um retrovírus endógeno, um elemento transponível, ou outro ácido nucleico, natural ou não. Em tais casos, o polinucleotídeo pode incluir sequências do gene nativo

(p. ex., o gene sem a sequência heteróloga) e a sequência não nativa inserida.

"Gene" é normalmente utilizado para abranger um polinucleotídeo que codifica um produto genético, p. ex., uma sequência de ácidos nucleicos que definem uma janela aberta de leitura (*open reading frame*).

O termo "produto genético", tal como é aqui utilizado, pretende abranger o produto ou parte da expressão de um gene correspondente a um polinucleotídeo aqui descrito, incluindo, mas não necessariamente limitado a uma molécula de RNA ou a um polipeptídeo.

"Diagnóstico", tal como é aqui utilizado, inclui geralmente a determinação da susceptibilidade de um sujeito a uma doença ou disfunção, a determinação de se um sujeito se encontra afectado por uma doença ou disfunção, bem como o prognóstico de um sujeito afectado por uma doença ou disfunção. A presente invenção abrange o diagnóstico de sujeitos no contexto de cancro do cólon (p. ex., pólipos adenomatosos, carcinoma colorectal), bem como tais cancros em qualquer estágio (p. ex., estágios I a IV de gravidade).

"Cancro do cólon" tal como é aqui utilizado pretende abranger formas benignas ou malignas de cancros do cólon e rectal; formas não-metastáticas, pré-metastáticas e metastizadas do cancro do cólon; e qualquer tipo de cancro proveniente de células do cólon e do recto (p. ex., pólipos adenomatosos, carcinoma colorectal e similares).

Os termos "indivíduo," "sujeito," "hospedeiro," e "paciente," são aqui usados indistintamente e referem-se a qualquer mamífero para o qual o diagnóstico, tratamento ou terapia é desejável, em particular humanos. Outros sujeitos podem

incluir gado, cães, gatos, cobaias, coelhos, ratazanas, ratos, cavalos, etc.

O termo "amostra" ou "amostra biológica" abrange uma variedade de tipos de amostra obtidas de um organismo e que podem ser utilizadas num diagnóstico ou ensaio de monitorização. O termo abrange amostras de sangue e outros líquidos de origem biológica, amostras de tecidos sólidos, tais como amostras de biópsias ou culturas de tecidos ou células daí derivadas e sua progenia. O termo abrange amostras que tenham sido manipuladas de qualquer modo após a sua aquisição, seja por tratamento com reagentes, solubilização, ou enriquecimento com certos componentes. O termo abrange amostras clínicas, e inclui também células de culturas celulares, células sobrenadantes, lisados de células, soro, plasma, fluidos biológicos e amostras de tecidos.

Uma "célula hospedeira", tal como é usado aqui, refere-se a um microorganismo ou a uma célula eucariótica ou uma linhagem celular cultivada como uma entidade unicelular que pode ser ou foi utilizada como receptor para um vector recombinante ou outros polinucleotídeos de transferência, e inclui a progenia da célula original que foi transfectada. Entende-se que a progenia de uma célula não é necessariamente completamente idêntica em termos morfológicos ou genómicos ou no complemento total do DNA como a progenitora original, devido a causas naturais, acidentais ou mutações deliberadas.

Os termos "cancro", "neoplasia", "tumor", e "carcinoma", são aqui utilizados indistintamente para referir células que exibam um crescimento relativamente autónomo, de modo que exibam um fenótipo de crescimento anormal caracterizado por uma perda significativa do controlo do crescimento celular. Geralmente, as células de interesse para a detecção ou tratamento na presente

aplicação incluem células pré-cancerosas (p. ex., benignas), malignas, pré-metastáticas, metastáticas, e não-metastáticas. A detecção de células cancerosas reveste-se de particular interesse.

O termo "fenótipo canceroso" refere-se geralmente a qualquer tipo de fenómeno biológico característico de uma célula cancerosa, podendo tal fenómeno variar com o tipo de cancro. O fenótipo canceroso é normalmente identificado por anomalias, por exemplo, no crescimento e proliferação celulares (p. ex., crescimento ou proliferação descontrolados), na regulação dos ciclos celulares, mobilidade celular ou interacção intercelular.

"Alvo terapêutico" refere-se geralmente a um gene ou produto genético que, por modulação da sua actividade (p. ex., modulação da expressão, actividade biológica e afins), permite a modulação do fenótipo canceroso.

Tal como utilizado ao longo deste texto "modulação" refere-se a um aumento ou uma diminuição do fenómeno indicado (p. ex., modulação da actividade biológica refere-se a um aumento numa actividade biológica ou numa diminuição numa actividade biológica).

Panorâmica da Invenção

Em termos gerais, a invenção baseia-se na descoberta de polinucleotídeos que representam genes que são expressos diferencialmente em células cancerosas do cólon. A expressão diferencial de genes em células do cólon afectadas com cancro é determinada, por exemplo, pela detecção de genes expressos numa célula cancerosa do cólon, e comparação do nível de expressão desses genes (p. ex., qualitativa ou quantitativamente) com a

expressão desses mesmos genes numa célula do cólon normal (i.e., uma célula do cólon não afectada por cancro do cólon).

Os polinucleotídeos correspondentes a genes expressos, diferencialmente aqui descritos, foram identificados utilizando amostragens diferenciadas de células de cólon normais, células de tumor primário do cólon, células metastáticas de tumor do cólon e células de pólipos adenomatosos. As sequências de polinucleotídeos específicos que representam genes expressos diferencialmente estão representadas em SEQ ID NOS: 1, 3, 5, 7, 9, 11-13, 15, 16, 18, 20, 22, 24, 26, 27 e 29.

Composição dos Polinucleotídeos

O âmbito da invenção relativamente às composições dos polinucleotídeos inclui, mas não se limita necessariamente a polinucleotídeos contendo um conjunto de sequências enunciadas na SEQ ID NO: 1, polinucleotídeos obtidos dos materiais biológicos aqui descritos ou de outras fontes biológicas (em particular fontes humanas) por hibridação debaixo de condições severas (especificamente condições de elevada severidade); genes correspondentes aos polinucleotídeos fornecidos; variações dos polinucleotídeos fornecidos e os genes correspondentes, em particular aquelas variantes que mantêm a actividade biológica do produto genético codificado (p. ex., a actividade biológica atribuída a um produto genético correspondente aos polinucleotídeos fornecidos como resultado da atribuição do produto genético a uma(s) família(s) de proteínas e/ou da identificação de um domínio funcional presente no produto genético). Outras composições de ácidos nucleicos contempladas pela e no âmbito da presente invenção serão evidentes para os especialistas na área quando fornecidos na presente descrição.

Os termos "polinucleotídeo" e "ácido nucleico", tal como aqui usados indistintamente, referindo-se a ácidos nucleicos da composição não pretendem ser limitantes da estrutura ou tamanho do ácido nucleico excepto se indicado. Para além disso, os polinucleotídeos aqui descritos podem consistir essencialmente em sequências exões, p. ex., sequências que definem janela aberta de leitura (*open reading frame*) e codificam um produto genético ou uma fracção dele. Por "consistir essencialmentef" no contexto de um polinucleotídeo aqui descrito entende-se que o polinucleotídeo é composto por uma sequência que codifica uma janela aberta de leitura (*open reading frame*), cuja sequência pode ser flanqueada por qualquer sequência que não afecte fisicamente a(s) característica(s) básicas do produto genético codificado. Sequências de acompanhamento apropriadas incluem, mas não se restringem necessariamente a sequências iniciadoras, sequências potenciadoras, locais de iniciação (start) e/ou término (stop) da transcrição, constructo ou sequências de vector (p. ex., sequências que permitem a manipulação do polinucleotídeo no interior duma molécula linear ou circular, incluindo, mas não se restringindo necessariamente a sequências de replicação e manutenção do constructo ou vector, sequências que codificam produtos genéticos que proporcionam a selecção (p. ex., resistência ou sensibilidade a antibióticos, factores que afectam o crescimento em meios com ou sem suplementose afins), sequências que permitem a produção de uma proteína de fusão com o polinucleotídeo e um polipeptídeo heterólogo (i.e., um polipeptídeo codificado por um polinucleotídeo proveniente de uma fonte diferente do polinucleotídeo ao qual está operacinalmente ligado) e afins.

A invenção proporciona polinucleotídeos que são expressos em tecidos humanos, especificamente em tecidos do cólon humano. As composições de ácidos nucleicos da invenção de particular interesse incluem a sequência enunciada na SEQ ID NO:1 ou uma sequência identificativa da mesma. Uma " sequência identificativa "é uma sequência consecutiva de resíduos de pelo menos cerca de 10 a cerca de 20 nt de comprimento, normalmente pelo menos cerca de 50 nt a cerca de 100 nt de comprimento, que identifica unicamente uma sequência polinucleotídica, p. ex., exibe menos de 90%, normalmente menos de cerca de 80% a cerca de 85% de identidade de sequência com qualquer sequência nucleotídica consecutiva de mais de cerca de 20 nt. Assim, as composições de ácidos nucleicos incluem cDNAs ou mRNAs completos que abrangem uma sequência identificativa nucleotídeos consecutivos de SEQ ID NO:1.

Os polinucleotídeos da invenção incluem também polinucleotídeos com semelhança de sequência ou identidade de sequência. Ácidos nucleicos com semelhança de sequência são detectados por hibridação em condições de baixa severidade, por exemplo, a 50°C e 10XSSC (0.9 M salinos/ 0.09 M citrato de sódio) e continuam ligados quando sujeitos a lavagem a 55°C in 1XSSC. A identidade de sequência pode ser determinada por hibridação em condições severas, por exemplo, at 50°C ou mais e 0.1XSSC (9 mM salinos/ 0.9 mM citrato de sódio). Os métodos e condições de hibridação são bem conhecidos na especialidade, ver, p. ex., USPN 5,707,829. Os ácidos nucleicos que são substancialmente idênticos às sequências polinucleotídicas fornecidas, p. ex. variantes alélicas, versões geneticamente alteradas do gene, etc., ligam-se a sequências polinucleotídicas fornecidas (SEQ ID NO:1) em condições de hibridação severas. Por recurso a sondas, em

particular sondas marcadas de sequências de DNA, é possível isolar genes homólogos ou relacionados. A fonte de genes homólogos pode ser qualquer espécie, p. ex. espécies de primatas, em particular a humana; roedores, tais como ratas e ratos; caninos, felinos, bovinos, ovinos, equinos, leveduras, nemátodos, etc.

Em geral, a hibridação é feita utilizando pelo menos 15 nucleotídeos consecutivos (nt) da SEQ ID NO:1. Isto é, quando pelo menos 15 nt consecutivos de uma das SEQ ID NOS apresentadas é utilizada como sonda, a sonda vai hibridar preferencialmente com um ácido nucleico contendo a sequência complementar, permitindo a identificação e recuperação dos ácidos nucleicos que somente hibridam com a sonda escolhida. Sondas de mais de uma SEQ ID NO podem hibridar com o mesmo ácido nucleico se o cDNA do qual foram obtidas corresponder ao mesmo mRNA completo. Podem ser utilizadas sondas de mais de 15 nt, p. ex., sondas de desde cerca de 18 nt a cerca de 100 nt, mas 15 nt representa a sequência suficiente para uma identificação exclusiva.

Os polinucleotídeos da invenção incluem também variantes naturais das sequências nucleotídicas (p. ex., variantes degeneradas, variantes alélicas, etc.). Variantes dos polinucleotídeos da invenção são identificadas por hibridação de variantes putativas de sequências nucleotídicas aqui descritas, preferencialmente por hibridação em condições severas. Por exemplo, utilizando condições de lavagem apropriadas, variantes dos polinucleotídeos da invenção podem ser identificadas quando a variante alélica exibe no máximo cerca de 25-30% pares de bases (bp) inadequados relativamente ao polinucleotídeo escolhido como sonda. Em geral, as variantes alélicas contêm 15-25% bp

inadequados, e podem conter apenas 5-15%, ou 2-5%, ou 1-2% bp inadequados, assim como um só bp inadequado.

A invenção também abrange homólogos correspondentes aos polinucleotídeos de SEQ ID NO: 1, nos quais a fonte de genes homólogos pode ser qualquer espécie de mamífero, p. ex., espécies de primatas, em particular a humana; roedores, tais como ratazanas, caninos, felinos, bovinos, ovinos, equinos, leveduras, nemátodos, etc. Entre espécies de mamíferos, p. ex., humana e do rato, os homólogos têm geralmente uma sequência semelhança substancial, p. ex., pelo menos 75% de identidade de sequência, normalmente pelo menos 90%, mais comumente pelo menos 95%, entre sequências nucleotídicas. A semelhança de sequência é calculada com base numa sequência de referência, a qual pode ser uma parte de uma sequência maior, tal como um motivo conservado, uma região de codificação, uma região de acompanhamento, etc. Uma sequência de referência terá normalmente pelo menos cerca de 18 nt consecutivos, mais vulgarmente pelo menos cerca de 30 nt, podendo estender-se até toda a sequência com a qual está a ser comparada. São conhecidos na especialidade algoritmos para análise sequencial, tais como Intervaloped BLAST, descrito por Altschul, et al. Nucleic Acids Res. (1997) 25:_3389-3402.

Geralmente, variantes da invenção têm uma identidade de sequência maior do que pelo menos cerca de 65%, de preferência pelo menos cerca de 75%, desejavelmente pelo menos cerca de 85%, e podem ser maiores do que pelo menos cerca de 90% ou mais, tal como determinado pelo algoritmo Smith-Waterman de busca de homologia implementado no programa MPSRCH (Oxford Molecular). Para os fins desta invenção, um método melhor de cálculo da percentagem de identidade é o algoritmo Smith-Waterman, utilizando o seguinte. A identidade de sequência total do DNA tem

que ser maior do que 65% tal como determinado pelo algoritmo Smith-Waterman de procura de homologia implementado no programa MPSRCH (Oxford Molecular) utilizando uma busca de sequências afins com os seguintes parâmetros de busca: intervalo open penalty, 12; e intervalo extension penalty, 1.

Os ácidos nucleicos pretendidos podem ser cDNAs ou DNAs genómicos, assim como fragmentos dos mesmos, em particular fragmentos que codificam um produto genético biologicamente activo e/ou são úteis para os métodos aqui descritos (p. ex.; em diagnóstico, como identificadores exclusivos de um gene de interesse expresso diferencialmente, etc.). O termo "cDNA" tal como é aqui utilizado pretende incluir todos os ácidos nucleicos que partilham o arranjo de elementos de sequência encontrado em espécies maduras de mRNA nativo, nas quais os elementos de sequência são exões e regiões non-codificadas 3' e 5'. Normalmente, espécies de mRNA apresentam exões consecutivos, com os intrões intervenientes, quando presentes, sendo removidos por *splicing* de RNA, para criar uma janela aberta de leitura (*open reading frame*) contínua codificando um polipeptídeo da invenção.

Uma sequência genómica de interesse contém o ácido nucleico presente entre o codão de iniciação e o codão de término (stop), tal como definido nas sequências listadas, incluindo todos os intrões que estão normalmente presentes num cromossoma nativo. Pode também incluir as regiões 3' e 5' não traduzidas presentes no mRNA maduro. Pode ainda incluir sequências de transcrição e tradução regulatória específicas, tais como promotores, potenciadores, etc., incluindo cerca de 1 kb, possivelmente mais, de DNA genómico flanqueador de qualquer das extremidades 5' e 3' da região transcrita. O DNA genómico pode ser isolado como um fragmento de 100 kbp ou menor e substancialmente livre de

sequência cromossômica flanqueadora. O DNA genômico flanqueador da região de codificação, seja a 3' ou a 5', ou sequências regulatórias internas, tal como por vezes se encontra em intrões, contém sequências requeridas para a adequada expressão de tecido, de estado específico, ou de estado específico da doença.

As composições de ácidos nucleicos da invenção podem codificar total ou parcialmente os polipeptídeos pretendidos. Fragmentos de cadeia simples ou dupla podem ser obtidos da sequência de DNA por síntese química de oligonucleotídeos através de métodos convencionais, por digestão por enzimas de restrição; por **amplificação** por PCR, etc. Os polinucleotídeos e fragmentos de polinucleotídeos obtidos da invenção contêm pelo menos cerca de 10, cerca de 15, cerca de 20, cerca de 35, cerca de 50, cerca de 100, cerca de 150 a cerca de 200, cerca de 250 a cerca de 300, ou cerca de 350 nt consecutivos seleccionados do sequenciador de polinucleotídeos tal como ilustrado em SEQ ID NO:1. Para a maioria, os fragmentos serão de pelo menos 15 nt, normalmente pelo menos 18 nt ou 25 nt, e até pelo menos cerca de 50 nt consecutivos ou mais. Numa modalidade preferida, as moléculas de polinucleotídeo contêm uma sequência contínua de pelo menos 12 nt seleccionados do grupo consistindo no polinucleotídeo mostrado em SEQ ID NO:1.

Sondas específicas para os polinucleotídeos da invenção podem ser geradas utilizando as sequências polinucleotídicas apresentadas em SEQ ID NO:1. As sondas são de preferência fragmentos de pelo menos cerca de 12,15,16,18,20,22,24, or 25 nt de uma sequência contínua correspondente de SEQ ID NO:1 e podem ter menos de 2,1,0.5,0.1, ou 0.05 kb de tamanho. As sondas podem ser sintetizadas quimicamente ou podem ser geradas a partir de polinucleotídeos maiores por recurso a enzimas de restrição. As

sondas podem ser marcadas, por exemplo, com um marcador radioactivo, biotinilado, ou fluorescente. Preferencialmente, as sondas são desenhadas com base numa sequência identificativa de um polinucleotídeo de SEQ ID NO:1.

Mais desejavelmente, as sondas são desenhadas com base numa sequência contínua de um dos polinucleotídeos requeridos que permaneça sem máscara após aplicação de um programa de "*masking*" para mascarar a baixa complexidade (p. ex., XBLAST) da sequência., i.e., selecciona-se uma região sem máscara, tal como indicado pelos polinucleotídeos fora das poli-n cadeias da sequência produzida pelo programa de "*masking*".

Os polinucleotídeos da invenção são isolados e obtidos com uma pureza substancial, normalmente como outros que não os de um cromossoma intacto. Normalmente, os polinucleotídeos, sejam DNA ou RNA, são obtidos substancialmente livres de outras sequências naturais de ácidos nucleicos, apresentando normalmente um grau de pureza de pelo menos cerca de 50%, comumente pelo menos cerca de 90% e são tipicamente "recombinantes", p. ex., acompanhados por um ou mais nucleotídeos aos quais não está normalmente associado num cromossoma natural.

Os polinucleotídeos da invenção podem ser fornecidos como uma molécula linear ou no interior de uma molécula circular, e podem ser fornecidos no interior de moléculas que se replicam autonomamente (vectores) ou no interior de moléculas sem sequências de replicação. A expressão dos polinucleotídeos pode ser regulada pelas suas próprias ou por outras sequência regulatórias conhecidas na especialidade. Os polinucleotídeos da invenção podem ser introduzidos em células hospedeiras apropriadas utilizando uma variedade de técnicas disponíveis na especialidade, tais como transferência de DNA mediada pela

aplicação múltipla de transferrina, transfecção com ácidos nucleicos livres ou encapsulados, transferência de DNA mediada por lipossomas, transporte intracelular de esferas de latex cobertas com DNA, fusão de protoplastos, infecção viral, electroporação, pistolas de genes, transfecção mediada por fosfato de cálcio, ou outras similares.

As composições de ácido nucleico pretendidas podem ser utilizadas para, por exemplo, produzir polipeptídeos, como sondas para a detecção de mRNA da invenção em amostras biológicas (p. ex., extractos de células humanas) para produzir cópias adicionais dos polinucleotídeos, para produzir ribozimas ou oligonucleotídeos *antisense*, e como sondas de DNA de cadeia simples ou como oligonudeotídeos formadores de cadeia tripla. As sondas aqui descritas podem ser utilizadas para, por exemplo, determinar a presença ou ausência das sequências polinucleotídicas como ilustrado em SEQ ID NO:1 ou variantes das mesmas numa amostra. Essa e outras utilizações encontram-se descritas em maior detalhe adiante.

Utilização de Polinucleotídeos para Obter cDNA completo, Genes, e Região Promotora

[0057] Moléculas de cDNA completo contendo os polinucleotídeos apresentados são obtidas do seguinte modo. Um polinucleotídeo contendo uma sequência de SEQ ID NO:1, ou uma porção do mesmo com pelo menos 12, 15, 18, ou 20 nt, é usado como sonda de hibridação para detectar membros hybridizantes de uma biblioteca de cDNA utilizando métodos de desenho de sondas, métodos de clonagem, e técnicas de selecção de clones tal como as descritas na USP 5,654,173. As bibliotecas de cDNA são construídas a partir de tecidos seleccionados, tais como tecidos normais ou tumorais, ou

de tecidos de um mamífero tratados com, por exemplo, um fármaco. De preferência, o tecido é o mesmo tecido a partir do qual os polinucleotídeos da invenção foram isolados, uma vez que tanto os polinucleotídeos aqui descritos e o cDNA representam genes expressos. Desejavelmente, a biblioteca de cDNA é construída a partir do material biológico aqui descrito nos Exemplos. A escolha do tipo de células para a construção da biblioteca pode ser feita de acordo com a identidade da proteína codificada pelo gene correspondente ao polinucleotídeo da invenção. Isto indicará que tecido e tipos de célula deverão exprimir o gene relacionado, e portanto constituem uma fonte adequada de mRNA para produzir o cDNA. Sempre que os polinucleotídeos fornecidos são isolados de bibliotecas de cDNA, as bibliotecas são preparadas a partir de mRNA de células do cólon humanas, desejavelmente, células de cancro do cólon humanas, as quais podem ser obtidas de tecidos de pacientes ou podem ser linhagens de células do cólon, p. ex., Kml2L4-A.

As técnicas de produção e sondagem de bibliotecas de sequências de ácidos nucleico encontram-se descritas; Por exemplo, em Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY. O cDNA pode ser preparado por recurso a iniciadores baseados na sequência de SEQ ID NO :1. Numa modalidade, a biblioteca de cDNA pode ser construída apenas a partir de mRNA poli-adenilado. Por conseguinte, pode-se recorrer a iniciadores poli-T para preparar cDNA a partir do mRNA.

Os membros da biblioteca maiores do que os polinucleotídeos fornecidos, e que de preferência abranjam a sequência codificadora completa da mensagem original, são obtidos: Por forma a confirmar que se obteve a totalidade do cDNA, fazem-se as

seguintes experiências de protecção do RNA. Hibridação de um cDNA completo a mRNA irá proteger o RNA de degradação pela RNase. Se o cDNA não é completo, então as porções do mRNA não hibridadas ficarão sujeitas a degradação pela RNase. Isto é testado, tal como é conhecido na especialidade, através de variações de mobilidade na electroforese em gel de poliacrilamida, ou por detecção de monoribonucleotídeos libertados. Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Ed., (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY. Para se obter sequências adicionais 5' na extremidade de um cDNA parcial, pode efectuar-se 5' RACE (PCR Protocols: A Guide to Methods and Applications, (1990) Academic Press, Inc.).

O DNA genómico é isolado utilizando os polinucleotídeos fornecidos de modo similar ao isolamento de cDNAs completos. Em suma, os polinucleotídeos fornecidos, ou partes deles, são utilizados como sondas para bibliotecas de genómico DNA. De preferência, a biblioteca é obtida a partir do tipo de células utilizado na produção dos polinucleotídeos da invenção, embora não necessariamente. Desejavelmente, o genómico DNA é obtido do material biológico aqui descrito nos Exemplos. Tais bibliotecas podem encontrar-se em vectores adequados para o transporte de grandes segmentos de um genoma, tais como P 1 ou YAC, como descrito detalhadamente em Sambrook et al., 9.4-9.30. Adicionalmente, as sequências genómicas podem ser isoladas a partir de bibliotecas BAC humanas, comercialmente disponíveis em Research Genetics, Inc., Huntsville, Alabama, USA, por exemplo. De modo a obter sequências 5' ou 3' adicionais, foi feito *chromosome walking*, como descrito em Sambrook et al., de forma a isolar fragmentos de DNA genómico adjacentes e sobrepostos. As mesmas são mapeadas e fragmentadas em conjunto, tal como é do

conhecimento da especialidade, por recurso a enzimas de restrição e digestão e DNA ligase.

Recorrendo às sequências polinucleotídicas da invenção, podem isolar-se os genes completos correspondentes utilizando métodos clássicos ou PCR para construir e sondar bibliotecas de cDNA. Em qualquer dos métodos, são feitos, de preferência, *Northern blots*, numa série de tipos de células para determinar qual das linhagens celulares apresenta o nível de expressão mais elevado do gene de interesse. Os métodos clássicos de construção de bibliotecas de cDNA encontram-se em Sambrook et al., supra. Com esses métodos, pode produzir-se cDNA a partir de mRNA e inseri-lo em vectores virais ou de expressão. Tipicamente, as bibliotecas de mRNA contendo caudas poli(A) podem ser produzidas com *primers* poli(T). De igual modo, as bibliotecas de cDNA podem ser produzidas recorrendo às sequências instantâneas como *primers*.

Utilizam-se métodos de PCR para amplificar os membros de uma biblioteca de cDNA que contém a inserção pretendida. Neste caso, a inserção pretendida irá conter uma sequência do cDNA completo correspondente aos polinucleotídeos instantâneos. Tais métodos de PCR incluem *gene trapping* e métodos RACE.. O *gene trapping* implica a inserção de um membro de uma biblioteca de cDNA num vector. O vector é então desnaturado para produzir moléculas de cadeia simples. De seguida, é utilizada uma sonda ligada ao substrato, tal como um oligo biotinilado, para armadilhar as inserções de cDNA de interesse. As sondas biotiniladas podem ser ligadas a um substrato sólido ligado a avidina. Podem-se utilizar métodos de PCR para amplificar o cDNA armadilhado. Para armadilhar sequências correspondentes aos genes completos, a sequência da sonda etiquetada baseia-se nas sequências

polinucleotídicas da invenção. Podem utilizar-se *primers* aleatórios ou *primers* específicos do vector da biblioteca para amplificar o cDNA armadilhado. Tais técnicas de armadilha de genes encontram-se descritas em Gruber et al., WO 95/04745 e Gruber et al., USPN 5,500,356. Existem à venda kits para realização de experiências de armadilha de genes em, por exemplo, Life Technologies, Gaithersburg, Maryland, USA.

"Amplificação rápida de extremidades de cDNA," ou RACE, é um método de PCR para amplificação de cDNAs de uma variedade de RNAs. Os cDNAs são ligados a um oligonucleotídeo *linker*, e amplificadas por PCR utilizando dois *primers*. Um *primer* baseia-se na sequência dos polinucleotídeos sequências instantâneos, para o qual se deseja a sequência completa, e o segundo *primer* contém a sequência que hibrida com o oligonucleotídeo *linker* para amplificar o cDNA. Uma descrição destes métodos encontra-se em WO 97/19110. Nas modalidades preferidas de RACE, é concebido um *primer* comum para hibridar a uma sequência adaptora arbitrária ligada às extremidades do cDNA (Apte and Siebert, Biotechniques (1993) 15:890-893; Edwards et al., Nuc. Acids Res. (1991) 19:5227-5232). Quando se emparelha um *primer* de cadeia simples de RACE específico para um gene com o *primer* comum, ocorre amplificação preferencial de sequências entre o *primer* de cadeia simples específico do gene e o *primer* comum. Existem bancos comerciais de cDNA modificadas para utilização em RACE.

Um outro método baseado no PCR produz uma biblioteca de cDNA completo com extremidades ancoradas sem necessidade de conhecimento específico da sequência do cDNA. O método utiliza *primers* de ancoragem (I-VI), sendo que um *primer*, poli TV (I-III) fixa a cauda poli A de mRNA eucariótico provocando a síntese da primeira cadeia e um segundo *primer*, polyGH (IV-VI) fixa na cauda

polyC adicionada pela deoxinucleotidil transferase (TdT) terminal (ver, p. ex, WO 96/40998).

A região promotora de um gene situa-se geralmente 5' em relação ao local de início da RNA polimerase II. Centenas de regiões promotoras contêm a "TATA" box, uma sequência como TATTA ou TATAA, que é sensível a mutações. A região promotora pode ser obtida através de 5' RACE utilizando um *primer* da região codificadora do gene. Em alternativa, o cDNA pode ser usado como sonda para a sequência genómica, e a região 5' em relação à região codificadora é identificada por "walking up". Se o gene tem uma expressão elevada ou é expresso diferencialmente, o promotor do gene pode ser útil num constructo regulador para um gene heterólogo.

Uma vez obtido o completo cDNA ou gene, podem ser preparada variantes de codificação de DNA por mutagénese selectiva, descrita em detalhe em Sambrook et al., 15.3-15.63. A escolha do codão ou nucleotídeo a ser substituído pode ser baseada como aqui referido em alterações opcionais nos amino ácidos para produzir proteínas alteradas estrutural e/ou funcionalmente.

Como método alternativo de obtenção de DNA ou RNA de um material biológico, podem ser sintetizados ácidos nucleicos contendo nucleotídeos com a sequência de um ou mais polinucleotídeos da invenção. Por conseguinte, a invenção abranje moléculas de ácido nucleico desde 15 nt (correspondendo a pelo menos 15 nt consecutivos da SEQ ID NO:1) até um tamanho máximo adequado para uma ou mais manipulações biológicas, incluindo replicação e expressão, da molécula de ácido nucleico. A invenção inclui mas não se limita a (a) ácido nucleico com o tamanho de um gene completo, e contendo SEQ ID NO:1, (b) o ácido nucleico de (a) contendo ainda pelo menos um gene adicional, ligados

operacionalmente para permitir a expressão de uma proteína de fusão; (c) um vector de expressão contendo (a) ou (b); (d) um plasmídeo contendo (a) ou (b) ; e (e) uma partícula viral recombinante contendo (a) ou (b). Uma vez fornecidos os polinucleotídeos aqui referidos, a construção ou preparação de (a) - (e) é bem conhecida no seio da especialidade.

A sequência de um ácido nucleico contendo pelo menos 15 nt consecutivos de SEQ ID NO:1 de preferência toda a sequência de SEQ ID NO:1 não se encontra limitada e pode ser qualquer sequência de A, T, G, e/ou C (para o DNA) e A, U, G, e/ou C (para o RNA) ou bases aí modificadas, incluindo inosina e pseudouridina. A escolha da sequência dependerá da função desejada e pode ser ditada por regiões codificadoras desejadas, por regiões do tipo intrão desejadas, e por regiões reguladoras desejadas. Quando a totalidade da sequência de SEQ ID NO: 1 se encontra no ácido nucleico, o ácido nucleico obtido é aqui referido como um polinucleotídeo contendo a sequência de SEQ ID NO: 1.

Expressão de Polipeptídeo Codificado por cDNA Completo ou Gene Completo

Os polinucleotídeos fornecidos (p. ex., um polinucleotídeo contendo a sequência de SEQ ID NO:1), o cDNA correspondente, ou o gene completo são usados na expressão total ou parcial de um produto genético. Os constructos dos polinucleotídeos contendo o sequenciador SEQ ID NO:1 podem ainda ser gerados por via sintética. Em alternativa, a junção num só passo de um gene e um plasmídeo completo a partir de um número elevado de oligodeoxirribonucleotídeos é descrita por, p. ex., Stemmer et al., Gene (Amsterdam) (1995) 164(1):49-53. Neste método, é

descrita a PCR de junção (a síntese de longas sequências de DNA a partir de um elevado número de oligodeoxirribonucleotídeos (oligos)). O método baseia-se na recombinação do DNA (Stemmer, Nature (1994) 370:389-391), e não depende da DNA.ligase, dependendo antes da DNA polimerase para construir fragmentos de DNA sucessivamente maiores durante o processo de junção.

Constructos de polinucleotídeos adequados são purificados através de técnicas padronizadas de DNA recombinante tal como descrito em, por exemplo, Sambrook et al., Molecular Cloning: A Laboratory Manual 2nd Ed, (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY, e em regulamentos correntes descritos em United States Dept. of HHS; National Institute of Health (NIH) Guidelines for Recombinant DNA Research. O produto genético codificado por um polinucleotídeo da invenção é expresso em qualquer sistema de expressão, incluindo, por exemplo, sistemas bacterianos, leveduras, insectos, anfíbios e mamíferos. Vectores, células hospedeiras e métodos de obtenção de expressão dos mesmos são bem conhecidos da especialidade. Vectores e células hospedeiras apropriados são descritos em USPN 5,654,173.

Moléculas de polinucleotídeo contendo uma sequência polinucleotídica aqui fornecida são normalmente propagadas colocando a molécula num vector. São utilizados vectores virais e não virais, incluindo plasmídeos. A escolha do plasmídeo dependerá do tipo de célula na qual se deseja fazer a propagação e do objectivo da propagação. Certos vectores são úteis na amplificação e produção de grandes quantidades da sequência de DNA desejada. Outros vectores são apropriados para a expressão em células em meios de cultura. Ainda outros vectores são adequados para transferência e expressão em células de um animal ou pessoa. A escolha do vector adequado é bem conhecida no seio da

especialidade. Muitos desses vectores encontram-se disponíveis comercialmente. Métodos de preparação de vectores contendo a sequência desejada são bem conhecidos na especialidade.

Os polinucleotídeos enunciados em SEQ ID NO:1 ou os polinucleotídeos completos correspondentes são ligados a sequências reguladoras se adequado por forma a obter as propriedades de expressão desejadas. Estas podem incluir promotores (ligados à extremidade 5' da cadeia *sense* ou à extremidade 3' da cadeia *antisense*), potenciadores, terminadores, operadores, repressores, e indutores. Os promotores podem ser regulados ou constitutivos. Em algumas situações pode ser desejável o recurso a promotores condicionalmente activos, tais como promotores específicos para tecidos ou específicos para estadios de desenvolvimento. Estes são ligados à sequência nucleotídica desejada recorrendo às técnicas acima descritas para ligação a vectores. Podem ser utilizadas quaisquer técnicas conhecidas na especialidade.

Quando se utilizam quaisquer das células hospedeiras acima, ou outras células hospedeiras ou organismos adequados, para replicar e/ou expressar os polinucleotídeos ou ácido nucleicos da invenção, o resultante ácido nucleico replicado, RNA, proteína ou polipeptídeo expresso, encontra-se no âmbito da invenção como produto da célula hospedeira ou organismo. O produto é recuperado por qualquer das técnicas adequadas conhecidas na especialidade.

Uma vez identificado o gene correspondente a um polinucleotídeo escolhido, a sua expressão pode ser regulada na célula à qual o gene pertence. Por exemplo, um gene endógeno de uma célula pode ser regulado por uma sequência reguladora exógena tal como referido na USPN 5,641,670.

Identificação de Motivos Functionais e Estruturais de Rastreo de Genes em Bases de Dados Públicas

As traduções da sequência nucleotídica dos polinucleotídeos, cDNAs ou genes completos fornecidos podem ser alinhados com sequências individuais conhecidas. A semelhança com sequências individuais pode ser usada para determinar a actividade dos polipeptídeos codificados pelos polinucleotídeos da invenção. Além disso, sequências exibindo uma semelhança com mais de uma sequência individual podem exibir actividades características de cada uma ou de ambas as sequências individuais.

As sequências completas e fragmentos das sequências polinucleotídicas vizinhas podem ser usadas como sondas e *primers* para identificar e isolar a sequência completa correspondente aos polinucleotídeos fornecidos. Os vizinhos podem indicar um tecido ou tipo de célula a ser usada na construção de uma biblioteca de sequências completas correspondentes aos polinucleotídeos fornecidos.

Tipicamente, um polinucleotídeo seleccionado é traduzido nas seis janelas de leitura para determinar o melhor alinhamento com as sequências individuais. As sequências aqui referidas na Listagem de Sequências têm uma orientação de 5' para 3' e a tradução em três janelas de leitura pode ser suficiente (à excepção de alguns casos específicos descritos nos Exemplos). Essas sequências de amino ácidos são referidas geralmente, por sequências problema (*query*), as quais serão alinhadas com as sequências individuais. Bases de dados com sequências individuais encontram-se descritas em "Computer Methods for Macromolecular Sequence Analysis" Methods in Enzymology (1996) 266, Doolittle, Academic Press, Inc., uma divisão da Harcourt Brace & Co., San

Diego, California, USA. As bases de dados incluem GenBank, EMBL, e Base de Dados de DNA do Japão (DDBJ).

As sequências problema e individuais podem ser alinhados recorrendo aos métodos e programas de computador acima descritos, e incluir BLAST 2.0, disponível na world wide web no endereço <http://www.ncbi.nlm.nih.gov/BLAST/>. Ver também Altschul, et al. *Nucleic Acids Res.* (1997) 25:3389-3402. Outro algoritmo de alinhamento é o Fasta, disponível no pacote Genetics Computing Group (GCG), Madison, Wisconsin, USA, uma subsidiária do Oxford Molecular Group, Inc. Outras técnicas para alinhamento encontram-se descritas em Doolittle, supra. De preferência, utiliza-se um programa de alinhamento que permita intervalos na sequência para alinhar as sequências. O algoritmo Smith-Waterman é um tipo de algoritmo que permite intervalos em alinhamentos de sequência. Ver *Meth. Mol. Biol.* (1997) 70: 173-187. Além disso, o program INTERVALO que utiliza o método de alinhamento de Needleman Wunsch pode ser utilizado para alinhar sequências.

Uma estratégia de busca alternativa utiliza software MPSRCH, que corre num computador MASP. O MPSRCH utiliza um algoritmo Smith-Waterman para serializar sequências num computador paralelo em grande escala. Esta abordagem aumenta a capacidade de identificação de sequências que são emparelhamentos remotamente relacionados, e é particularmente tolerante a pequenos intervalos e erros na sequência nucleotídica. Sequências de aminoácidos codificadas pelos polinucleotídeos fornecidos podem ser utilizadas para procurar em bases de dados de proteínas e de DNA.

Os resultados dos alinhamentos de sequências individuais e problema podem ser divididos em três categorias: semelhança elevada, baixa semelhança, e sem semelhança. Os resultados de alinhamento individual variando de semelhança elevada a baixa

semelhança fornecem uma base de determinação de actividade e/ou estrutura de polipeptídeos. Os parâmetros para classificar resultados individuais incluem: percentagem da região de alinhamento onde se encontra o alinhamento mais forte, percentagem de identidade de sequências, e valor de p. A percentagem da região de alinhamento calcula-se por contagem do número de resíduos da sequência individual encontrada na região de alinhamento mais forte, p. ex., região contígua à sequência individual que contém o maior número de resíduos idênticos aos resíduos da região correspondente da sequência problema alinhada. Este número é dividido pelo total de resíduos da sequência problema para calcular uma percentagem. Por exemplo, uma sequência problema de 20 resíduos de aminoácido pode ser alinhada com uma região de 20 aminoácidos de uma sequência individual. A sequência individual pode ser idêntica aos resíduos de aminoácido 5, 9-15, e 17-19 da sequência problema. A região de alinhamento mais forte é por isso a região de alongamento dos resíduos 9-19, um alongamento de 11 aminoácido. A percentagem da região de alinhamento é: 11 (comprimento da região de alinhamento mais forte) dividido por (comprimento da sequência problema) 20 ou 55%.

A percentagem da identidade das sequências calcula-se por contagem do número de aminoácido emparelhados entre as sequências problema e individual e dividindo o número total de emparelhamentos pelo número de resíduos das sequências individuais encontradas na região de mais forte alinhamento. Por conseguinte, a percentagem de identidade no exemplo acima seria de 10 emparelhamentos dividido por 11 aminoácidos, ou aproximadamente, 90,9%.

O valor de P é a probabilidade de o alinhamento ter sido produzido por acaso. Para um alinhamento simples, o valor de p pode ser calculado de acordo com Karlin et al., Proc. Natl. Acad. Sci. (1990) 87:2264 e Karlin et al., Proc. Natl. Acad. Sci. (1993) 90. O valor de p de alinhamentos múltiplos utilizando a mesma sequência problema pode ser calculado usando uma abordagem heurística descrita em Altschul et al., Nat. Genet. (1994) 6:119. Programas de alinhamento tais como o BLAST conseguem calcular o valor de p . Ver também Altschul et al., Nucleic Acid Res. (1997) 25:3389-3402.

Outro factor a ter em consideração na determinação da identidade ou semelhança é a localização da semelhança ou identidade. Um alinhamento local forte pode indicar semelhança mesmo se o comprimento do alinhamento fôr curto. A identidade de sequências espalhada ao longo do comprimento da sequência problema pode também indicar uma semelhança entre as sequências problema e perfil. As fronteiras da região onde as sequências alinham pode ser determinada de acordo com Doolittle, supra; BLAST 2.0 (ver, p. ex., Altschul, et al. Nucleic Acids Res. (1997) 25:3389-3402) ou programas FAST; ou por determinação da área onde a identidade de sequências é mais elevada.

Semelhança Elevada, Geralmente, nos resultados de alinhamento considerados de semelhança elevada, a percentagem da região de alinhamento é tipicamente pelo menos cerca de 55% do comprimento total da sequência problema; mais comumente, pelo menos cerca de 58%; ainda mais frequentemente, pelo menos cerca de 60% do comprimento total dos resíduos da sequência problema. Normalmente, a percentagem da região de alinhamento pode ser até cerca de 62%; comumente, até cerca de 64%; frequentemente, até cerca de 66%. Para além disso, numa semelhança elevada, a região

de alinhamento, tipicamente, exibe pelo menos cerca de 75% da identidade de sequências; comumente, pelo menos cerca de 78%; frequentemente, pelo menos cerca de 80% da identidade de sequências. Normalmente, a percentagem da identidade de sequências pode ser até cerca de 82%; comumente, até cerca de 84%; frequentemente, até cerca de 86%.

O valor de p é usado em conjunto com esses métodos. Se for encontrada uma semelhança elevada, considera-se que existe uma semelhança elevada entre a sequência problema e uma sequência perfil quando o valor de p é menor ou igual a cerca de 10^{-2} ; comumente, menor ou igual a cerca de 10^{-3} ; frequentemente, menor ou igual a cerca de 10^{-4} . Tipicamente, o valor de p não é superior a cerca de 10^{-5} ; comumente, não é maior ou igual a cerca de 10^{-10} ; frequentemente, não é maior ou igual a cerca de 10^{-15} para que a sequência problema seja considerada de semelhança elevada.

Baixa Semelhança. Geralmente, quando os resultados de alinhamento são considerados de baixa semelhança, não existe uma percentagem mínima da região de alinhamento nem um comprimento mínimo de alinhamento. Um melhor exemplo de baixa semelhança é considerado quando a região de alinhamento tem, tipicamente, pelo menos cerca de 15 resíduos de aminoácido de comprimento; comumente, pelo menos cerca de 20; frequentemente, pelo menos cerca de 25 resíduos de aminoácido de comprimento. Normalmente, o comprimento da região de alinhamento pode ser até cerca de 30 resíduos de aminoácido; comumente, até cerca de 40; frequentemente, até cerca de 60 resíduos de aminoácido. Além disso, para uma baixa semelhança, a região de alinhamento, tipicamente, exibe pelo menos cerca de 35% de identidade de sequências; comumente, pelo menos cerca de 40%; frequentemente,

pelo menos cerca de 45% de identidade de sequências. Normalmente, a percentagem de identidade de sequências pode ser até cerca de 50%; comumente, até cerca de 55%; frequentemente, até cerca de 60%.

Se fôr encontrada uma semelhança, considera-se que a sequência problema tem uma baixa semelhança com uma sequência perfil quando o valor de p é normalmente menor ou igual a cerca de 10^{-2} ; comumente, menor ou igual a cerca de 10^{-3} ; frequentemente, menor ou igual a cerca de 10^{-4} . Tipicamente, o valor de p não é superior a cerca de 10^{-5} ; comumente, não é maior ou igual a cerca de 10^{-10} ; frequentemente, não é maior ou igual a cerca de 10^{-15} para que a sequência problema seja considerada de baixa semelhança.

Semelhança Determinada Apenas por Identidade de Sequências.

A identidade de sequências pode ser utilizada sózinha para determinar a semelhança de uma sequência problema com uma sequência individual e pode indicar a actividade da sequência. Tal alinhamento, de preferência, permite **intervalos** no alinhamento de sequências. Tipicamente, a sequência problema está relacionada com a sequência perfil se a identidade de sequências na totalidade da sequência problema fôr pelo menos cerca de 15%; normalmente, pelo menos cerca de 20%; frequentemente, pelo menos cerca de 25%; mais frequentemente, pelo menos cerca de 50%. A identidade de sequência por si só como medida da semelhança é mais útil quando a sequência problema tem normalmente, pelo menos 80 resíduos de comprimento; comumente, 90 resíduos; frequentemente, pelo menos 95 resíduos de aminoácido de comprimento. Normalmente, pode concluir-se sobre a semelhança baseando-se na apenas na identidade de sequências quando a sequência problema tem de preferência 100 resíduos de

comprimento; desejavelmente, 120 resíduos de comprimento; desejavelmente, 150 resíduos de aminoácido de comprimento.

Alinhamentos com Perfil e Alinhamento Múltiplo de Sequências. As traduções dos polinucleotídeos fornecidos podem ser alinhadas com perfis de aminoácido que definem famílias de proteína ou motivos comuns. Além disso, as traduções dos polinucleotídeos fornecidos podem ser alinhadas em alinhamentos múltiplos de sequências (MSA) contendo sequências polipeptídicas de membros das famílias de proteínas ou motivos. A semelhança ou identidade com sequências perfil ou MSAs pode ser utilizada para determinar a actividade dos produtos genéticos (p. ex., polipeptídeos) codificados pelos polinucleotídeos fornecidos ou correspondente cDNA ou genes. Por exemplo, sequências que apresentam uma identidade ou semelhança com um perfil quimiocina ou MSA podem exhibir actividades quimiocina.

Os perfis podem ser concebidos manualmente por (1) criação de uma MSA, a qual é um alinhamento da sequência de aminoácidos de membros da família ou (2) construção de uma representação estatística do alinhamento. Tais métodos encontram-se descritos, por exemplo, em Birney et al., Nucl. Acid Res. (1996) 24(14):2730-2739. MSAs de algumas famílias de proteínas e motivos são do conhecimento público. Por exemplo, <http://genome.wustl.edu/Pfam/> inclui MSAs de 547 famílias e motivos diferentes. Essas MSAs encontram-se também descritas em Sonnhammer et al., Proteins (1997) 28: 405-420. Outras fontes da world wide web incluem o site <http://www.embl-heidelberg.de/argos/ali/ali.html>; alternativamente, pode enviar-se uma mensagem para ALI@EMBL-HEIDELBERG.DE para obter a informação. Uma breve descrição dessas MSAs encontra-se descrita em Pascarella et al., Prot. Eng. (1996) 9(3):249-251. Técnicas de

construção de s a partir de MSAs são descritas em Sonnhammer et al., supra; Birney et al., supra; and " Computer Methods for Macromolecular Sequence Analysis,"Methods in Enzymology (1996) 266, Doolittle, Academic Press, Inc., San Diego, California, USA.

A semelhança entre uma sequência problema e uma família de proteínas ou motivo pode ser determinada por (a) comparação da sequência problema com o perfil e/ou (b) alinhamento da sequência problema com os membros da família ou motivo. Tipicamente, utiliza-se um programa como Searchwise para comparar a sequência problema com a representação estatística do alinhamento múltiplo, também designado perfil (ver Birney et al., supra). Outras técnicas de comparação da sequência com o perfil são descritas em Sonnhammer et al., supra and Doolittle, supra.

De seguida, podem usar-se os métodos descritos por Feng et al., J. Mol. Evol. (1987) 25:351 e Higgins et al., CABIOS (1989) 5:151 para alinhar a sequência problema com os membros de uma família ou motivo, também designados MSA. Os alinhamentos de sequência podem ser gerados recorrendo a uma série de ferramentas de software. Exemplos incluem PileUp, que cria um alinhamento múltiplo de sequências, e está descrita em Feng et al., J. Mol. Evol. (1987) 25:351. Outro método, INTERVALO, utiliza o método de alinhamento de Needleman et al., J. Mol. Biol. (1970) 48:443. INTERVALO é mais adequado para alinhamento global de sequências. Um terceiro método, BestFit, funciona por inserção de intervalos para maximizar o número de emparelhamentos recorrendo ao algoritmo de homologia local de Smith et al., Adv. Appl. Math. (1981) 2:482. Geralmente, utilizam-se os seguintes factores para determinar se existe semelhança entre a sequência problema e um perfil ou MSA: (1) número de resíduos conservados encontrados na sequência problema, (2) percentagem de resíduos conservados

encontrados na sequência problema, (3) número de *frameshifts*, e (4) espaçamento entre resíduos conservados.

Alguns programas de alinhamento que traduzem e alinham sequências conseguem produzir qualquer número de *frameshifts* durante a tradução da sequência nucleotídica para conseguir o melhor alinhamento. Quanto menos *frameshifts* forem necessários para a produção de um alinhamento, mais forte a semelhança ou identidade entre a sequência problema e o perfil ou as MSAs. Por exemplo, a baixa semelhança resultante da inexistência de *frameshifts* pode ser uma melhor indicação da actividade ou estrutura de uma sequência problema, do que uma forte semelhança resultante de dois *frameshifts*. De preferência, encontram-se três ou menos *frameshifts* num alinhamento; comumente dois ou menos *frameshifts*; desejavelmente, um ou menos *frameshifts*; melhor ainda será não existirem *frameshifts* num alinhamento de problema e perfil ou MSAs.

Resíduos conservados são os aminoácidos numa posição particular em toda ou em parte dos membros da família ou sequências repetitivas padrão. Alternativamente, uma posição é considerada conservada se apenas se encontram aminoácidos de uma certa classe numa posição particular em todos ou alguns dos membros da família. Por exemplo, a posição N-terminal pode conter um aminoácido carregado positivamente, como a lisina, arginina, ou histidina.

Tipicamente, um resíduo de um polipeptídeo é conservado quando uma classe de aminoácidos ou um só aminoácido se encontra numa posição particular em pelo menos cerca de 40% dos membros da classe; normalmente, pelo menos cerca de 50%; comumente, em pelo menos cerca de 60% dos membros. Normalmente, um resíduo é conservado quando uma classe ou um só aminoácido se encontra pelo

menos em cerca de 70% dos membros de uma família ou motivo; comumente, pelo menos cerca de 80%; frequentemente, pelo menos cerca de 90%; mais frequentemente, pelo menos cerca de 95%.

Um resíduo é considerado conservado quando três aminoácidos não relacionados se encontram numa posição particular em alguns ou em todos os membros; comumente, dois aminoácidos não relacionados. Esses resíduos são conservados quando os aminoácidos não relacionados se encontram em posições particulares em pelo menos cerca de 40% dos membros da classe; normalmente, pelo menos cerca de 50%; frequentemente, pelo menos cerca de 60% dos membros. Normalmente, um resíduo é conservado quando uma classe ou um só aminoácido se encontra em pelo menos cerca de 70% dos membros de uma família ou motivo; comumente, pelo menos cerca de 80%; mais comumente, pelo menos cerca de 90%; frequentemente, pelo menos cerca de 95%.

Uma sequência problema tem semelhança com um perfil ou MSA quando a sequência problema contém pelo menos cerca de 25% dos resíduos conservados do perfil ou MSA; comumente, pelo menos cerca de 30%; frequentemente, pelo menos cerca de 40%. Tipicamente, a sequência problema tem uma semelhança mais forte com uma sequência perfil ou MSA quando a sequência problema contém pelo menos cerca de 45% dos resíduos conservados do perfil ou MSA; normalmente, pelo menos cerca de 50%; frequentemente, pelo menos cerca de 55%.

Identificação de Polipeptídeos Segregados e Ligados a Membranas

Tanto os polipeptídeos segregados como os ligados a membranas da presente invenção são de particular interesse. Por exemplo, os níveis de polipeptídeos segregados podem ser medidos em fluidos corporais convenientes, tais como sangue, plasma,

soro, e outros fluidos corporais como a urina, fluido prostático e sémen. Os polipeptídeos ligados a membranas são úteis na criação de vacinas de antígenos ou na indução de uma resposta imune. Tais antígenos conteriam a totalidade ou parte da região extracelular dos polipeptídeos ligados à membrana. Porque tanto os polipeptídeos segregados como os ligados a membranas contêm um fragmento de amino ácidos hidrofóbicos consecutivos, pode recorrer-se a algoritmos capazes de prever a hidrofobicidade para identificar tais polipeptídeos.

Uma sequência de sinal é normalmente codificada pelos genes dos polipeptídeos segregados e dos ligados a membranas para enviar um polipeptídeo à superfície da célula. A sequência sinal normalmente contém um alongamento de resíduos hidrofóbicos. Essa sequência sinal pode dobrar-se em estruturas helicoidais. Os polipeptídeos ligados a membranas tipicamente contêm, pelo menos, uma região transmembrana que possui um alongamento de aminoácidos hidrofóbicos que conseguem atravessar a membrana. Algumas regiões transmembranares também exibem uma estrutura helicoidal. Os fragmentos hidrofóbicos no seio de um polipeptídeo podem ser identificados recorrendo a algoritmos computacionais. Tais algoritmos incluem Hopp & Woods, *Proc. Natl. Acad. Sci. USA* (1981) 78:3824-3828; Kyte & Doolittle, *J. Mol. Biol.* (1982) 157: 105-132; e algoritmo RAOAR, Degli Esposti et al., *Eur. J. Biochem.* (1990) 190: 207-219.

Outro método de identificação de polipeptídeos segregados e ligados a membranas é a tradução dos polinucleotídeos da invenção nas seis grelhas de leitura e determinação de se pelo menos 8 aminoácidos hidrofóbicos consecutivos se encontram presentes. Os polipeptídeos traduzidos com pelo menos 8; normalmente, 10; frequentemente, 12 aminoácidos hidrofóbicos consecutivos são

considerados polipeptídeos putativos segregados ou ligados a membranas. Os aminoácidos hidrofóbicos incluem a alanina, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, treonina, triptofano, tirosina, e valina.

Identificação da Função de um Produto da Expressão de um Gene Completo

Podem utilizar-se ribozimas, construtos *antisense*, e mutantes negativos dominantes para determinar a função do produto da expressão de um gene correspondente a um polinucleotídeo aqui fornecido, e podem ainda ser utilizados na inibição da produção de produtos genéticos funcionais, codificados por um gene correspondente a um polinucleotídeo aqui descrito. No contexto da caracterização funcional do produto genético codificado, o recurso a *antisense*, ribozimas, e/ou mutantes negativos dominantes é particularmente útil quando os polinucleotídeos fornecidos não exibem homologia significativa ou substancial com uma sequência codificadora de um gene de função conhecida.

Podem formar-se moléculas *antisense* e ribozimas a partir de polinucleotídeos sintéticos. Tipicamente, utiliza-se o método da fosforamidite de síntese de oligonucleotídeos. Ver Beaucage et al., Tet. Lett. (1981) 22:1859 and USPN 4,668,777. Encontram-se disponíveis mecanismos automáticos de síntese para formar oligonucleotídeos recorrendo a este tipo de química. Exemplos de tais mecanismos incluem Biosearch 8600, Modelos 392 e 394 da Applied Biosystems, uma divisão da Perkin-Elmer Corp., Foster City, California, USA; e Expedite by Perceptive Biosystems, Framingham, Massachusetts, USA. Pode ainda produzir-se RNA sintético, análogos fosfatados dos oligonucleotídeos, e oligonucleotídeos derivatizados quimicamente, e podem ser ligados

covalentemente a outras moléculas. Podem ser sintetizados oligonucleotídeos de RNA, por exemplo, recorrendo a fosforamidites de RNA. Este método pode ser levado a cabo num sintetizador automático, tal como Applied Biosystems, Modelos 392 e 394, Foster City, California, USA.

Também se podem sintetizar oligonucleotídeos de fosforotioato para a produção de construção *antisense*. Pode recorrer-se a um reagente sulfurizador, tal como o dissulfureto de tetraetiltiruoamo (tetraethylthiruam disulfide-TETD) em acetonitrilo para converter o internucleotídeo fosfito de cianoetilo no triéster de fosforotioato em cerca de 15 minutos à temperatura ambiente. O TETD substitui o iodo como reagente, enquanto que os restantes reagentes utilizados na química de fosforamidite padrão permanecem os mesmos. Tal método de síntese pode ser automatizado utilizando os Modelos 392 e 394 da Applied Biosystems, por exemplo.

Podem ser sintetizados oligonucleotídeos de até 200 nt, normalmente, 100 nt, comummente 50 nt; frequentemente 30 a 40 nt. Esses fragmentos sintéticos podem ser anelados e ligados para produzir fragmentos maiores. Ver, por exemplo, Sambrook et al., supra. RNAs catalíticos de clivagem trans (ribozimas) são moléculas de RNA com actividade de endoribonuclease. As ribozimas são concebidas especificamente para um alvo particular, e a mensagem do alvo deve conter uma sequência nucleotídica específica. São concebidos por engenharia para quebrar qualquer espécie de RNA num local específico no conteúdo total do RNA celular. A quebra torna o mRNA instável e impede expressão de proteínas. É importante o facto de as ribozimas poderem ser usadas para inibir a expressão de um gene de função desconhecida com o objectivo de determinação da sua função *in vitro* ou *in*

vivo, através da detecção do efeito fenotípico. Um ribozima motivo de uso comum é a "cabeça de martelo" (*hammerhead*), para o qual os requisitos de sequência do substrato são mínimos. A concepção da ribozima "cabeça de martelo", assim como as utilizações terapêuticas de ribozimas, são apresentadas em Usman et al., *Current Opin. Struct. Biol.* (1996) 6:527. Métodos de produção de ribozimas, incluindo fragmentos de ribozima com estrutura de ansa de emparelhamento, métodos de aumento de especificidade de ribozimas, e outros bem conhecidos na especialidade.

A região de hibridização da ribozima pode ser modificada ou pode ser preparada como uma estrutura ramificada tal como descritas em Horn e Urdea, *Nucleic Acids Res.* (1989) 17:6959. A estrutura básica das ribozimas pode também ser alterada quimicamente de formas familiares às pessoas da especialidade, e ribozimas quimicamente sintetizadas podem ser administradas como derivados sintéticos de oligonucleotídeos modificados por unidades monoméricas. Num contexto terapêutico, distribuição de ribozimas por liposomas melhora a absorção celular, tal como descrito em Birikh et al., *Eur. J. Biochem.* (1997) 245:1.

Os ácido nucleicos *antisense* são concebidos para se ligarem especificamente ao RNA, resultando na formação de híbridos de RNA-DNA ou RNA-RNA, com uma detenção da replicação do DNA, transcrição reversa ou tradução de RNA mensageiro. Os polinucleotídeos *antisense* baseados numa sequência polinucleotídica seleccionada pode interferir com a expressão do correspondente gene. Os polinucleotídeos *antisense* são tipicamente gerados na célula por expressão de constructos *antisense* que contêm a cadeia *antisense* como cadeia transcrita. Os polinucleotídeos *antisense* baseados nos polinucleotídeos

apresentados irão ligar-se e/ou interferir com a tradução de mRNA contendo a sequência complementar do polinucleotídeo *antisense*. Os produtos de expressão de células de controlo e células tratadas com o constructo *antisense* são comparados para detectar a proteína produzida pelo gene correspondente ao polinucleotídeo no qual o constructo *antisense* é baseado. A proteína é isolada e identificada por recurso a métodos bioquímicos de rotina.

Dada a extensa bibliografia de base e a experiência clínica em terapia *antisense*, um especialista na matéria pode utilizar polinucleotídeos seleccionados da invenção como potenciais terapêuticos adicionais. A escolha de um polinucleotídeo pode ser reduzida pelo primeiro teste da sua ligação a regiões "hot spot" do genoma das células cancerosas do cólon. Caso seja identificado como um polinucleotídeo que se ligue a um "hot spot", o ensaio do polinucleotídeo como um composto de sentido inverso em células do cancro do cólon correspondente está garantido.

Como método alternativo para a identificação da função do gene que corresponde a um polinucleotídeo, divulgada neste documento, são facilmente geradas mutações dominantes negativas para as proteínas correspondente que estão ativas como homomultímeros. Um polipeptídeo mutante irá interagir com polipeptídeos de tipo selvagem (feitas a partir do outro alelo) e formar multímeros não-funcionais. Deste modo, uma mutação encontra-se num domínio de ligação do substrato, num domínio catalítico, ou num domínio de localização celular. De preferência, haverá sobreprodução de polipeptídeo mutante. São feitas mutações pontuais que têm esse efeito. Além disso, a fusão de diferentes polipeptídeos de vários comprimentos à região terminal de uma proteína pode produzir mutantes negativos dominantes. Existem estratégias genéricas para produzir mutantes

negativos dominantes (ver, por exemplo, Herskowitz, Nature (1987) 329:219).

Estas técnicas podem ser usadas para criar mutações de perda de função, que são úteis para determinar a função proteica.

Polipeptídeos e Variantes dos mesmos

Os polipeptídeos da invenção incluem aqueles codificados pelos polinucleótidos referidos, assim como ácidos nucleicos que, em virtude da degradação do código genético, não são idênticos em sequência aos polinucleótidos referidos. Assim, a invenção inclui no seu âmbito um polipeptídeo codificado por um polinucleotídeo contendo a sequência de SEQ ID NO:_1 ou de uma variante do mesmo. Exemplos de polipeptídeos codificados por uma janela aberta de leitura (*open reading frame*) de um polinucleotídeo aqui descrito incluem os SEQ ID NOS:_2.

Em geral, o termo "polipeptídeo", como é aqui usado refere-se tanto à totalidade do polipeptídeo codificado pelo referido polinucleotídeo, ao polipeptídeo codificado pelo gene representado pelo referido polinucleotídicas, bem como o uso de fragmentos ou porções dos mesmos. "Polipeptídeos" também inclui variantes das proteínas que existem na natureza, em que as variantes são homólogas ou substancialmente semelhantes às que ocorrem nas proteínas que existem na natureza, e podem ser obtidas na mesma espécie ou noutra diferente daquela que é a da proteína que existe na natureza (por exemplo, seres humanos, roedores, ou algumas outras espécies que naturalmente exprimem o referido polipeptídeo, geralmente uma espécie de mamífero). Em geral, as variantes de polipeptídeos têm uma sequência que tem, pelo menos, cerca de 80%, normalmente, pelo menos, cerca de 90%, e, mais geralmente, pelo menos, cerca de 98% de semelhança na

sequência com um polipeptídeo expresso diferencialmente da invenção, medido pela BLAST 2,0 utilizando os parâmetros descritos acima. A variante dos polipeptídeos pode ser ou não glicosilada naturalmente, ou seja, o padrão glicosilação do polipeptídeo difere do padrão encontrado na glicosilação que ocorre nas proteínas naturais correspondentes.

A invenção também engloba homólogos dos polipeptídeos referidos (ou fragmentos destes) em que os homólogos são isolados a partir de outras espécies, ou seja, outras espécies animais ou vegetais, onde tais homólogos, geralmente espécies de mamíferos, por exemplo, roedores, como ratos, ratazanas, os animais domésticos, por exemplo, cavalos, vacas, cães, gatos e seres humanos. Por "homólogo" entende-se um polipeptídeo tendo pelo menos cerca de 35%, normalmente, pelo menos, cerca de 40% e, mais geralmente, pelo menos, cerca de 60% da sequência de aminoácido idêntica a uma proteínas específica expressa diferencialmente como identificada acima, onde identidade da sequência é determinada utilizando o algoritmo BLAST 2,0, com os parâmetros descritos *supra*.

Em geral, os polipeptídeos que são o objecto da invenção são produzidos num ambiente não ocorrente na natureza, por exemplo, estão separadas do seu ambiente natural. Em certas modalidades, a proteína está presente numa composição que é enriquecida em proteína em comparação com um controlo. Como tal, é fornecido polipeptídeo purificado, onde por purificado por se entende que a proteína está presente numa composição que se encontra substancialmente livre de polipeptídeos expressos de modo não diferencial, onde por substancialmente livre se entende que menos de 90%, geralmente menos de 60% e, mais geralmente menos de 50%

da composição é composta de expressos polipeptídeos não-diferencialmente.

Também as variantes são do âmbito de aplicação da invenção; variantes de polipeptídeos incluem mutantes, fragmentos e fusões. Os mutantes podem incluir substituições de aminoácido, adições ou supressões. As substituições de aminoácidos podem ser substituições aminoácido conservativas ou substituições para eliminar aminoácidos não-essenciais, como modificar um local de glicosilação, de fosforilação ou de acetilação, ou para minimizar a desdobragem que possa ocorrer por substituição ou supressão de um ou mais resíduos cisteína que não são necessários para a função. As substituições aminoácido conservativas são aquelas que preservam a carga geral; hidrofobicidade/hidrofilicidade e/ou impedimento estereoquímico do aminoácido substituído.

As variantes podem ser concebidas de modo a manter ou reforçar a actividade biológica de uma determinada região da proteína (por exemplo, um domínio funcional e/ou, quando o polipeptídeo for um membro de uma família de proteínas, uma região associada a uma sequência de consenso). A seleção das alterações de aminoácidos para a produção de variantes podem ser baseada na acessibilidade (interior versus exterior) do aminoácido (ver, por exemplo, Go et al, Int. Péptido J. Protein Res. (1980) 15: 211), na termostabilidade da variante do polipeptídeo (ver, por exemplo, Querol et al., Prot. Eng. (1996) 9: 265), nos locais de glicosilação desejados (ver, por exemplo, Olsenn e Thomsen, J. Gen. Microbiol. (1991) 137: 579), nas pontes de dissulfeto desejadas (ver, por exemplo, Clarke et al., Bioquímica (1993) 32: 4322; e Wakarchuk et al., Protein Eng. (1994) 7: 1379), nos locais de ligação a metais desejados (v., por exemplo, toma et al., Bioquímica (1991) 30: 97, e

Haezebrouck et al., Protein Eng. (1993) 6: 643), e substituições desejadas nas ansas "loops" de prolina (ver, por exemplo, Masul et al., Appl. Env Microbiol . (1994) 60: 3579). Muteínas empobrecidas em cisteína podem ser produzidas como apresentado em USPN 4959314.

As variantes incluem também os fragmentos de polipeptídeos divulgadas neste documento, nomeadamente fragmentos biologicamente activos e/ou fragmentos correspondentes a domínios funcionais. Fragmentos de interesse irão normalmente possuir, pelo menos, cerca de 10 aa até, pelo menos, cerca de 15 aa de comprimento, normalmente, pelo menos, até cerca de 50 aa de comprimento, e podem atingir comprimentos de 300 aa ou mais, mas que normalmente não excedem cerca de 1000 aa de comprimento, onde o fragmento terá uma sequência de aminoácidos que é idêntica a um polipeptídeo codificado por um polinucleotídeo que contém uma sequência de SEQ ID NO:_1 ou um homólogo do mesmo. As variantes descritas neste documento são proteínas codificadas por polinucleótidos que estão dentro do âmbito de aplicação da invenção. O código genético pode ser usado para seleccionar os codões adequados para a construção das variantes correspondentes.

Modalidades computacionalmente relacionadas

Em geral, uma biblioteca de polinucleótidos é uma coleção de informações sequenciais, cuja informação é fornecida, quer na forma bioquímica (por exemplo, como uma coleção de moléculas polinucleotídicas), ou em formato electrónico (por exemplo, como uma colecção de sequências polinucleotídicas armazenadas de forma computacionalmente legível, integradas num sistema de computador e/ou como parte de um programa de computador). A sequência de informação dos polinucleótidos pode ser usada em uma variedade de

maneiras, por exemplo, como um recurso para a descoberta de genes, como uma representação de sequências expressas num tipo seleccionado de células (por exemplo, tipo de células marcadores), e/ou como marcadores de uma dada doença ou estado de doença. Em geral, um marcador doença é uma representação de um produto genético que está presente em todas as células afectadas pela doença, seja através de um aumento ou através de uma diminuição em relação a um nível normal nas células (por exemplo, uma célula do mesmo tipo ou similar que não esteja substancialmente afectada pela doença). Por exemplo, uma sequência polinucleotídica de uma biblioteca pode ser um polinucleotídeo que representa um mRNA, um polipeptídeo, ou outro produtos genético codificado pelo polinucleotídeo que seja sobre-expresso ou sub-expresso numa célula do cólon afectada pelo cancro em relação a uma célula do cólon normal (ou seja, substancialmente livre de doença).

A informação de sequência nucleotídica da biblioteca pode ser construída de qualquer forma adequada, por exemplo, sob formas electrónicas ou bioquímicas. Por exemplo, uma biblioteca de informação de sequência consubstanciada em formato electrónico compreende um ficheiro de arquivo de dados acessível no computador (ou, na forma bioquímica, uma coleção de moléculas de ácidos nucleicos), que contém as sequências de nucleótidos representativas dos genes que são expressos diferencialmente (por exemplo, sobre-expressos ou sub-expressos) entre, por exemplo, i) uma célula do cólon cancerosa e uma célula do cólon normal, ii) uma célula do cólon cancerosa e uma célula cólon displática, iii) uma célula do cólon cancerosa e uma célula cólon afectada por uma doença ou condição, com excepção do cancro; iv) uma célula do cólon cancerosa metástica e uma célula cólon normal e/ou uma

célula do cólon cancerosa não-metastática; v) um célula do cólon cancerosa maligna e uma célula do cólon cancerosa não-maligna (ou uma célula normal do cólon) e/ou vi) uma célula do cólon displática relativamente a uma célula do cólon normal. Outras combinações e comparações de células do cólon afectadas por doenças ou várias fases de doenças serão facilmente perceptíveis para os especialistas no ramo. Modalidades bioquímicas da biblioteca incluem uma colecção de ácidos nucleicos que têm as sequências dos genes na biblioteca, onde os ácidos nucleicos podem corresponder à totalidade do gene na biblioteca ou a um fragmento do mesmo, conforme descrito em maior detalhe a seguir.

As bibliotecas polinucleotídicas que são objecto da invenção geralmente incluem informação de uma pluralidade de sequências polinucleotídicas, em que, pelo menos um dos polinucleótidos contém uma sequência de SEQ ID NO: 1. Por pluralidade entende-se, no mínimo, 2, normalmente, pelo menos 3 e pode incluir todos os menos 3 e pode incluir todos os polinucleótidos de SEQ ID NOS: 1, 3, 5, 7, 9, 11-13, 15, 16, 18, 20, 22, 24, 26, 27 e 29. O comprimento e o número de polinucleótidos na biblioteca variará em função da natureza da biblioteca, por exemplo, se a biblioteca é uma matriz de oligonucleótidos, uma matriz de cDNA, uma base de dados computacional da sequência informações, etc.

Quando a biblioteca é uma biblioteca electrónica, a informação das sequências de ácidos nucleicos pode estar presente sob uma variedade de suportes. O termo suporte refere-se a um dispositivo, que não seja uma molécula isolada de ácido nucleico, que contém a informação da sequência da presente invenção. Tal dispositivo proporciona a sequência do genoma ou um subconjunto dele sob uma forma que pode ser examinada por meios não directamente aplicáveis à sequência, tal como existe num ácido

nucleico. Por exemplo, a sequência de nucleótidos da presente invenção, e.g., as seqüências de ácidos nucleicos dos polinucleótidos de SEQ ID NO: 1, podem ser gravados em suportes computacionais, por exemplo, em qualquer meio que possa ser lido e acedido directamente usando um computador. Os meios referidos incluem, mas não estão limitados a: suportes magnéticos, tais como uma disquete, um disco duro e uma fita magnética; meios de armazenamento óptico, como os CD-ROM; meios de armazenamento eléctricos como a RAM e ROM; e híbridos destas categorias, tais como magnéticos/ópticos.

Um especialista pode facilmente apreciar o modo como qualquer um dos meios legíveis no computador actualmente conhecidos pode ser usado para criar um dispositivo contendo a gravação da presente informação da sequência. "Gravado" refere-se a um processo para armazenar informações sobre suporte informático legível, utilizando qualquer dos métodos que são conhecidos no sector. Qualquer estrutura conveniente de armazenamento de dados pode ser escolhida com base nos meios utilizados para o acesso às informações armazenadas. Uma variedade de programas e formatos de processamento de dados pode ser usada para o armazenamento, por exemplo, ficheiro de texto num processador de texto, formato banco de dados, etc. Para além das informações da sequência, as versões electrónicas de bibliotecas da invenção podem ser produzidas em conjunto ou em ligação com outras informações legíveis no computador e/ou outros tipos de arquivos legíveis no computador (por exemplo: ficheiros pesquisáveis, ficheiros executáveis, etc, incluindo mas não limitado a, por exemplo, software de programas de pesquisa, etc.).

Ao fornecer a sequência nucleotídica em suporte informático, as informações podem ser acedidas para uma variedade de finalidades. As aplicações informáticas para aceder a informação da sequência encontram-se publicamente disponíveis. Por exemplo, os algoritmos de busca *Intervaloped BLAST* (Altschul et al. *Ácidos Nucleicos Res.* (1997) 25: 3389-3402) e *Blaze* (Brutlag et al. *Comp. Chem.* (1993) 17: 203) podem ser usados num sistema Sybase para identificar janelas abertas de leitura (*open reading frames*) (ORFs) no genoma que contém homologia com ORFs de outros organismos.

Tal como aqui utilizado, "um sistema computacional" refere-se ao *hardware*, *software*, meios de armazenamento de dados e meios utilizados para analisar a informação da sequência nucleotídica da presente invenção. O mínimo de hardware dos sistemas computacionais da presente invenção compreende uma unidade central de processamento (CPU), meios de entrada (input), meios de saída (output), e meios de armazenamento de dados. Um especialista qualificado pode facilmente compreender que qualquer um dos sistemas computacionais actualmente disponíveis é adequado para ser utilizado na presente invenção. Os meios de armazenamento de dados podem incluir qualquer dispositivo que contenha um registo da presente informação da sequência, como descrita acima, ou que possua uma memória de acesso que possa aceder a um dispositivo desse tipo.

"Meio de Pesquisa" refere-se a um ou vários programas implementados no sistema computacional, para comparar uma sequência alvo ou região estrutural alvo, ou os níveis de expressão do polinucleotídeo numa amostra, com as informações da sequência armazenadas. Os meios de pesquisa podem ser utilizados para identificar fragmentos ou regiões do genoma que correspondem

a uma determinada sequência alvo ou região alvo. São conhecidos publicamente conhecidos e encontram-se comercialmente disponíveis uma grande variedade de algoritmos como, por exemplo, MacPattern (EMBL), BLASTN e BLASTX (NCBI). A "sequência alvo" pode ser qualquer polinucleotídeo ou aminoácido com uma sequência de seis ou mais nucleótidos contíguos ou dois ou mais aminoácidos, de preferência com cerca de 10 a 100 aminoácidos ou com cerca de 30 a 300 nt. Uma grande variedade de meios de comparação pode ser utilizada para realizar a comparação das informações sequenciais a partir de uma amostra (por exemplo, para análise de sequências-alvo, regiões alvo, ou níveis de expressão relativos) com os meios de armazenamento de dados. Um especialista qualificado pode facilmente reconhecer que qualquer dos programas de pesquisa de homologia publicamente disponíveis podem ser usados como meio de pesquisa para os sistemas computacionais da presente invenção para realizar a comparação de sequências-alvo e regiões. Programas de computador para analisar os níveis de expressão numa amostra e nos controlos são também bem conhecidos no ramo.

A "região estrutural alvo", ou "região alvo" refere-se a qualquer sequência ou combinação de sequências racionalmente selecionada em que as sequências (s) são escolhidas com base numa configuração tridimensional que é formada após a dobrar a região alvo, ou em sequências consensuais de centros activos ou de regulação. Há uma variedade de regiões alvo conhecidas na especialidade das regiões alvo de proteínas que incluem, mas não que não são limitadas a, centros activos de enzimas e sequências de sinal. As regiões alvo de ácidos nucleicos incluem, mas não se limitam a, estruturas em ansas de emparelhamento, sequências promotoras e outros elementos de expressão como locais de ligação para factores de transcrição.

Pode ser usada uma variedade de formatos estruturais para os meios de entrada (input media) e de saída (output media) para conseguir importar e exportar a informação do sistema computacional da presente invenção. Um formato de exportação classifica a expressão relativa dos níveis diferentes polinucleótidos. Essa apresentação proporciona a um especialista um ranking de níveis de expressão relativa para determinar um perfil de expressão genética.

Como discutido anteriormente, a "biblioteca" da invenção abrange também as bibliotecas bioquímicas de polinucleotídeo de SEQ ID NO: 1, por exemplo, coleções de ácidos nucleicos que representam os polinucleótidos fornecidos. As bibliotecas bioquímicas podem assumir diversas formas, por exemplo, uma solução de cDNAs, um padrão de ácidos nucleicos sonda estabilizadamente associado a uma superfície de um sólido de suporte (ou seja, uma matriz) e similares. De particular interesse são matrizes de ácidos nucleicos em que a SEQ ID NO: 1 está representada na matriz. Por matriz entende-se um artigo de um dispositivo que tenha, pelo menos, um substrato com pelo menos dois alvos de ácidos nucleicos diferentes numa das suas superfícies, onde o número de ácidos nucleicos diferentes pode ser consideravelmente mais elevado, sendo tipicamente, pelo menos, 10 nt, geralmente pelo menos 20 nt e, muitas vezes, pelo menos 25 nt. Uma panóplia de diferentes formatos de matriz tem sido desenvolvida e é conhecida pelos especialistas. As matrizes do assunto desta invenção têm utilidade numa variedade de aplicações, incluindo a análise da expressão genética, rastreio de droga, análise de mutação e similares, como divulgado nos documentos de patente exemplificativos supralistados.

Além das referidas bibliotecas de ácidos nucleicos, são também fornecidas bibliotecas análogas de polipeptídeos, onde os polipeptídeos da biblioteca representam, no mínimo, uma porção dos polipeptídeos codificados por SEQ ID NO:_1.

Utilidades

Sondas polinucleotídicas, geralmente incluindo pelo menos 12 nt contíguos de um dos polinucleotídeos representados na Listagem de Sequências, são utilizadas para uma variedade de fins, como o mapeamento cromossómico dos polinucleotídeos e detecção dos níveis transcrição. Informações complementares sobre as regiões preferidas das sequências polinucleotídicas discutidas são encontradas nos Exemplos. Uma sonda que se hibridiza especificamente para um dos polinucleotídeos aqui referidos deve proporcionar um sinal detecção que seja, pelo menos, 5, 10, ou 20 vezes maior do que a hibridização de fundo devida a outras sequências não relacionadas.

Detecção de Níveis de Expressão. São utilizadas sondas de nucleótídeos para detectar a expressão de um gene correspondente a um polinucleotídeo seleccionado. Na técnica Northern blots, o mRNA é separado por electroforese e depois sondado. Uma sonda é detectada como uma hibridização a uma espécie de mRNA de dimensão específica. A extensão da hibridização é quantificada para determinar as quantidades relativas de expressão, por exemplo, ao abrigo de uma condição particular. As sondas são utilizadas para a hibridização "*in situ*" de células de modo a detectar a expressão. As sondas podem também ser usadas para detecção diagnóstica "*in vivo*" de sequências de hibridização. As sondas são normalmente marcadas com um isótopo radioactivo. Podem ser usados outros tipos de marcadores detectáveis como cromóforos,

flúores e enzimas. Outros exemplos de testes de hibridização de nucleótidos são descritos em WO92/02526 e USPN 5124246.

Alternativamente, outro meio para a detecção de pequenas quantidades de ácidos nucleicos alvo é a reacção em cadeia da polimerase (Polymerase Chain Reaction -PCR) (ver, por exemplo, Mullis et al., Meth. Enzymol. (1987) 155: 335; USPN 4683195; e USPN 4.683.202). Dois primers nucleotídicos polinucleotídicos que originarão híbridos com os ácidos nucleicos alvo são usados para iniciar a reacção. Os primers podem ser compostos de sequência no interior ou em 3' e 5' dos polinucleótidos da Listagem de Sequências. Alternativamente, se os iniciadores são 3' e 5' para estes polinucleótidos, não necessitam de hibridizar para eles ou para os seus complementares. Após a amplificação do alvo com uma polimerase termoestável, os ácidos nucleicos alvo amplificados podem ser detectados através de métodos conhecidos na especialidade como, por exemplo, Southern blot. O mRNA ou cDNA também pode ser detectada por técnicas de blotting tradicionais (por exemplo, Southern blot, Northern blot, etc) descritas em Sambrook et al., "Molecular Cloning: A Laboratory Manual" (New York, Cold Spring Harbor Laboratory, 1989) (por exemplo, sem PCR). Em geral, o mRNA ou cDNA produzidos a partir de mRNA utilizando uma enzima polimerase pode ser purificado e separado usando eletroforese em gel, e transferido para um suporte sólido, como a nitrocelulose. O suporte sólido é exposto a uma sonda marcada, lavada para remover qualquer sonda não hibridizada, e os duplexes contendo a sonda marcada são detectados.

Mapeamento. Os polinucleotídeos da presente invenção podem ser usados para identificar um cromossoma no qual reside o gene correspondente. Esse mapeamento pode ser útil na identificação da função do gene relacionado com o polinucleotídeo, pela sua

proximidade com outros genes de função conhecida. A função pode também ser atribuída ao gene relacionado com o polinucleotídeo quando síndromas ou doenças específicas são mapeadas no mesmo cromossoma. Por exemplo, a utilização de sondas polinucleotídicas na identificação e quantificação de aberrações de sequência dos ácidos nucleicos é descrita na USPN 5783387. Um método exemplificativo de mapeamento é a hibridização *in situ* de fluorescência (FISH), o que facilita a hibridação genômica comparativa para permitir a total avaliação de alterações do genoma relativamente número de cópias de sequências de DNA (ver, por exemplo, Valdes et al., *Methods in Molecular Biology* (1997) 68:1). Os polinucleótidos podem também ser mapeados a cromossomas específicos, utilizando, por exemplo, híbridos de radiação ou painéis de híbridos cromossomo-específicos. Ver Leach et al., *Advances in Genetics*, (1995) 33: 63-99; Walter et al., *Nature Genetics* (1994) 7: 22; Walter e Goodfellow, *Trends in Genetics* (1992) 9: 352. Painéis para o mapeamento com híbridos de radiação estão disponíveis em Research Genetics, Inc., Huntsville, Alabama, E.U.A. Bases de dados de marcadores usando vários painéis estão disponíveis através da World Wide Web em <http://F/shgo-www.stanford.edu> e em <http://www-genome.wi.mit.edu/lcgin/contig/rhmapper.pl>. O programa estatístico RHIVIAP pode ser usado para construir um mapa com base nos dados de hibridização por radiação com uma medida relativa da probabilidade de um modo versus outro. RHMAP está disponível através da World Wide Web, no grupo <http://www.sph.umich.edu/statgen/software>. Além disso, existem programas comerciais para a identificação de regiões dos cromossomas vulgarmente associadas a doenças, tais como o cancro.

Tipagem de Tecidos ou Profiling. A expressão de mRNA específico correspondente à dos polinucleótidos fornecidos pode

variar consoante os diferentes tipos de células e pode ser específica consoante os diferentes tipos de tecidos. Esta variação nos níveis de Mrna em diferentes tipos de células pode ser explorada com testes com sonda de ácidos nucleicos para determinar os tipos de tecido. Por exemplo, análises usando PCR de sonda ramificada de ADN, ou técnicas blotting utilizando sondas de ácidos nucleicos substancialmente idênticas ou complementares aos polinucleótidos listados na Sequência de Listagem pode determinar a presença ou ausência do cDNA ou mRNA correspondente.

A tipagem de tecidos pode ser usada para identificar a fonte de desenvolvimento do órgão ou tecido de uma lesão metastática através da identificação da expressão de um marcador específico para esse órgão ou tecido. Se um polinucleotídeo é expresso apenas num determinado tipo de tecido, e se descobre que uma lesão metastática expressa esse polinucleotídeo, então a fonte de desenvolvimento da lesão foi identificada. A expressão de um determinado polinucleotídeo podem ser analisada pela detecção quer do mRNA quer da proteína produzida correspondente. Como seria facilmente perceptível para qualquer cientista forense, as sequências aqui divulgadas são úteis na diferenciação entre tecidos humanos e tecidos não-humanos. Em particular, essas sequências são úteis para diferenciar tecido humano dos tecidos de aves, répteis e anfíbios, por exemplo.

Uso de polimorfismos. Um polinucleotídeo da invenção pode ser usada em ciências forenses, análise genética, mapeamento e aplicações diagnósticas onde o correspondente a uma região de um gene é polimórfico na população humana. Qualquer meio de detecção de um polimorfismo num gene pode ser usado, incluindo, mas não limitando a, eletroforese de variantes polimórficas de proteínas,

sensibilidade diferencial à clivagem por enzimas de restrição e hibridação por sondas alelo-específicas.

Produção de Anticorpos

Os produtos de expressão de um polinucleotídeo da invenção, bem como o correspondente mRNA, cDNA, ou gene completo, podem ser preparados e utilizados para elevar a produção de anticorpos para fins experimentais, de diagnóstico e terapêuticos. Para polinucleótidos para os quais não foi ainda atribuído um gene correspondente isto fornece um método adicional para identificar o gene correspondente. O polinucleotídeo ou cDNA relacionado é expresso como descrito acima, e os anticorpos são preparados. Esses anticorpos são específicos para um epítopo no polipeptídeo codificado pelo polinucleotídeo e podem precipitar ou ligar-se às proteínas nativas correspondentes numa preparação de células ou tecido ou num extracto de células livres de um sistema de expressão *in vitro*.

Métodos para a produção de anticorpos que se ligam especificamente a um antígeno seleccionado são bem conhecidos na especialidade. Imunógenos para aumentar os anticorpos podem ser preparados misturando polipeptídeo codificado por um polinucleotídeo da invenção com um adjuvante, e/ou fazendo a fusão de proteínas com proteínas imunogénicas maiores. Os polipeptídeos também podem ser covalentemente ligados a outras proteínas imunogénicas maiores, como, por exemplo, a hemocianina de caramujo (keyhole limpet hemocyanin). Os imunógenos são normalmente administradas intradérmicamente, subcutaneamente ou intramuscularmente a animais de experimentação como coelhos, ovinos e ratinhos, para produzir anticorpos. Os anticorpos monoclonais podem ser gerados isolando células do baço e pela

fusão de células de mieloma para formação de hibridomas. Alternativamente, o polinucleotídeo selecionado é administrado directamente, por exemplo por injeção intramuscular, e expresso *in vivo*. A proteína expressa gera uma variedade de respostas imunitárias a proteínas específicas, incluindo a produção de anticorpos, o que é comparável à administração da proteína.

As preparações de anticorpos monoclonais e policlonais específicos para polipeptídeos codificados por um polinucleotídeo selecionado são feitos usando métodos padrão conhecidos na especialidade. Os anticorpos ligam-se especificamente aos epítopos presentes nos polipeptídeos codificados por polinucleotídeos divulgados na Listagem de Sequências. Tipicamente, são necessários no mínimo, 6, 8, 10 ou 12 aminoácidos contíguos para formar um epítipo. Epítopos que envolvem aminoácidos não contíguos podem exigir um polipeptídeo mais longo, por exemplo, no mínimo, 15, 25, ou 50 aminoácidos. Os anticorpos que se ligam especificamente a polipeptídeos humanos codificados pelos polipeptídeos fornecidos deve proporcionar uma detecção com um sinal pelo menos 5, 10 ou 20 vezes maior do que um sinal de detecção obtido com outras proteínas, quando utilizadas em Western blots ou outros ensaios imunoquímicos. De preferência, os anticorpos que se ligam especificamente a polipeptídeos da invenção não se ligam, em níveis detectáveis, a outras proteínas nos ensaios imunoquímicos e que podem imunoprecipitar o polipeptídeo específico da solução.

A invenção também contempla anticorpos naturalmente presentes que sejam específicos para um polipeptídeo da invenção. Por exemplo, anticorpos séricos de um polipeptídeo da invenção numa população humana podem ser purificados por métodos bem conhecidos na especialidade, por exemplo, passando o antissoro

por uma coluna à qual o polipeptídeo correspondente selecionado ou a proteína de fusão esteja ligado. Os anticorpos podem então ser eluídos da coluna, por exemplo, usando um tampão com uma alta concentração salina.

Para além dos anticorpos discutidos anteriormente, a invenção também contempla anticorpos obtidos por engenharia genética, anticorpos derivados (por exemplo, anticorpos de cadeia única, fragmentos de anticorpos (por exemplo, Fab, etc)), de acordo com métodos bem conhecidos na especialidade.

Diagnóstico e Outros Métodos Envolvendo a Detecção de Produtos de Genes Expressos Diferencialmente

A presente invenção fornece métodos de usar os polinucleótidos aqui descritos. Em modalidades específicas não-limitante, os métodos são úteis para a detecção de células de cancro do cólon, facilitando o diagnóstico do cancro e da severidade um cancro (por exemplo, o grau do tumor, a carga tumoral, e similares) num indivíduo, facilitando uma determinação do prognóstico do indivíduo e a avaliação da capacidade de resposta do sujeito à terapêutica (por exemplo, fornecendo uma medida de efeito terapêutico, por exemplo, avaliação da carga tumoral durante ou após um regime quimioterapêutico). A detecção pode ser baseada na detecção de um polinucleotídeo que é expresso diferencialmente numa célula do cólon cancerosa, e/ou detecção de um polipeptídeo codificado por um polinucleotídeo que é expresso diferencialmente numa célula do cólon cancerosa ("um polipeptídeo associado com o cancro do cólon"). Os métodos de detecção da invenção podem ser realizados *in vitro* ou *in vivo*, em células isoladas, na totalidade de tecidos corporais ou num fluido corporal, por exemplo, sangue, plasma, soro, urina e afins).

Em geral, os métodos da invenção envolvendo a detecção de um produto genético (por exemplo, mRNA, cDNA gerado a partir desse mRNA e polipeptídeos) envolvem o contacto de uma amostra com uma sonda específica para o produto genético de interesse. O termo "sonda", como aqui usado em tais métodos, é utilizado para referir-se a uma molécula que se liga especificamente a um produto genético de interesse (por exemplo, a sonda liga-se ao produto genético alvo com uma especificidade suficiente para distinguir a ligação ao alvo sobre a ligação não específica a moléculas não-alvo (de fundo)). As "sondas" incluem, mas não estão limitadas a, sondas de ácidos nucleicos (por exemplo, DNA, RNA, ácidos nucleicos modificados e similares), anticorpos (por exemplo, anticorpos, fragmentos de anticorpos que mantenham a ligação a um epítipo-alvo, anticorpos de cadeia única e similares), ou outro polipeptídeo, peptídeo ou molécula (por exemplo, ligando receptor), que se liga especificamente a um produto genético alvo de interesse.

A sonda e a amostra suspeita de ter o produto genético de interesse são postos em contacto em condições adequadas para a ligação da sonda ao produto genético. Por exemplo, o contacto é geralmente por um tempo suficiente para permitir a ligação da sonda ao produto genético (por exemplo, a partir de vários minutos a poucas horas), e a uma temperatura e condições de osmolaridade e similares que garantem a ligação da sonda ao produto genético a um nível que seja suficientemente diferenciado das ligações de fundo da sonda (por exemplo, em condições que minimizam as ligações não específicas). Condições adequadas para a ligação entre a sonda e o produto genético alvo podem ser facilmente determinadas utilizando técnicas de controlo e outras

disponíveis e conhecidas daqueles que tenham um conhecimento regular da especialidade.

Nesta modalidade, a sonda pode ser um anticorpo ou outro polipeptídeo, peptídeo ou molécula (p. ex., ligando receptor) que se ligue especificamente a um polipeptídeo alvo de interesse.

Os métodos de detecção podem ser fornecidos como parte de um kit. Assim, a invenção prevê ainda kits para a detecção da presença e/ou de um nível de um polinucleotídeo que é expresso diferencialmente numa célula de cancro do cólon (por exemplo, a detecção de um mRNA codificado pelo gene expresso diferencialmente de interesse), e/ou um polipeptídeo codificado deste modo numa amostra biológica. Os procedimentos usando estes kits podem ser realizados por laboratórios clínicos, por laboratórios experimentais, por médicos ou por particulares. Os kits da invenção para a detecção de um polipeptídeo codificado por um polinucleotídeo que é expresso diferencialmente numa célula de cancro do cólon compreendem um meio que liga especificamente o polipeptídeo, que pode ser um anticorpo específico. Os kits da invenção para a detecção de um polinucleotídeo que é expresso diferencialmente numa célula de cancro do cólon compreende um meio que hibridiza especificamente para esses polinucleotídeos. O kit pode opcionalmente fornecer componentes adicionais que são úteis no processo, incluindo, mas não limitado a, soluções tampão, reagentes de desenvolvimento, marcadores, superfícies de reacção, meios de detecção, amostras de controlo, padrões, instruções e informações interpretativas.

Detecção de um polipeptídeo codificado por um polinucleotídeo que é expresso diferencialmente numa célula de cancro do cólon

Em algumas modalidades, os métodos são fornecidos para uma célula de cancro do cólon através da detecção na célula de um polipeptídeo codificado por um gene expresso diferencialmente numa célula de cancro do cólon. Qualquer método de uma variedade de métodos conhecidos pode ser usado para a detecção, incluindo, mas não limitado a, imunoensaio, utilização de anticorpos específicos para o polipeptídeo codificado, por exemplo, ensaio imunoenzimático da enzima ligada (ELISA), radioimunoensaio (RIA) ou similares, e ensaios funcionais para o polipeptídeo codificado, por exemplo, actividade de ligação ou actividade enzimática.

Por exemplo, um teste de imunofluorescência pode ser facilmente realizado em células sem primeiro isolar o polipeptídeo codificado. As células são primeiro fixadas num suporte sólido, como uma lâmina de microscópio ou num poço (célula) de uma placa de titulação (microplaca). Essa etapa de fixação pode permeabilizar a membrana celular. A permeabilização da membrana celular permite que a sonda polipeptídeo-específica (por exemplo., anticorpo) se ligue. Alternativamente, quando o polipeptídeo é segregado ou ligado à membrana ou que se encontra de outro modo acessível na superfície celular (por exemplo, através de receptores e outras moléculas estavelmente associadas com a membrana celular exterior ou de outra forma estavelmente associadas à membrana celular), esta permeabilização pode não ser necessária.

Seguidamente, as células fixadas são expostas a um anticorpo específico para o polipeptídeo codificado. Para aumentar a sensibilidade do ensaio, as células fixadas podem ainda ser

expostas a um segundo anticorpo, que é marcado e se liga ao primeiro anticorpo, que é específico para o polipeptídeo codificado. Tipicamente, o anticorpo secundário está marcado de modo detectável, por exemplo, com um marcador fluorescente. As células que expressam o polipeptídeo codificado serão fluorescentemente marcadas e facilmente visualizadas ao microscópio. Ver, por exemplo, Hashido et al. (1992) Biochem. Biophys. Res. Comm. 187: 1241-1248.

Como será facilmente perceptível, para o especialista regularmente qualificado, mediante a leitura da presente especificação, os métodos de detecção e outros métodos descritos neste documento podem ser facilmente alterados. Estas alterações estão dentro do âmbito de aplicação previsto da invenção. Por exemplo, no sistema de detecção acima, a sonda utilizada na detecção pode ser imobilizada sobre um suporte sólido, e a amostra do ensaio posta em contacto com a sonda imobilizada. A ligação da amostra do ensaio com a sonda poderá ser detectada através de diversas formas, por exemplo, a detecção de um marcador identificável ligado à amostra do ensaio para facilitar a detecção de complexos da sonda com a amostra de teste imobilizada.

A presente invenção prevê ainda métodos para detectar a presença de e/ou medir o nível de um polipeptídeo numa amostra biológica, polipeptídeo esse que é codificado por um polinucleotídeo que representa um gene expresso diferencialmente no cancro, em particular, numa célula de cancro do cólon, utilizando um sonda específica para o polipeptídeo codificado. Nesta modalidade, a sonda pode ser um anticorpo ou outro polipeptídeo, peptídeo ou molécula (por exemplo, ligando

receptor), que se liga especificamente a um polipeptídeo alvo de interesse.

Os métodos geralmente incluem: a) colocar a amostra em contacto com um anticorpo específico para um polipeptídeo expresso diferencialmente numa célula de teste; e b) detectar ligações entre anticorpos e as moléculas da amostra. O nível de ligações dos anticorpos (seja qualitativo ou quantitativo) indica o estado canceroso das células. Por exemplo, quando o gene expresso diferencialmente é aumentado nas células cancerosas, a detecção de um aumento do nível de ligação dos anticorpos à amostra do ensaio, em relação ao nível de ligação dos anticorpos a células normais, indica que a célula de teste é cancerosa.

Controlos adequados incluem uma amostra conhecida por conter o polipeptídeo codificado, e uma amostra que contactou com um anticorpo não específico para o polipeptídeo codificado, por exemplo, um anticorpo anti-idiotipo. Uma variedade de métodos para a detecção de interações específicas anticorpo-antígeno são conhecidas na especialidade e pode ser usada no método, incluindo, mas não se limitando a, métodos imunohistológicos padrão, imunoprecipitação, um imunoensaio enzimático e um rádio-imunoensaio.

Em geral, os anticorpos específicos serão marcados por detecção, quer directa quer indirectamente. Marcadores directos incluem: radioisótopos; enzimas cujos produtos são detectáveis (por exemplo, Luciferasa, b-galactosidase, e similares); marcadores fluorescentes (por exemplo, fluoresceína isotiocianato, rhodamine, ficoeritrina, e similares); metais emissores de fluorescência, por exemplo, ^{152}Eu , ou outros da série dos lantanídeos, ligado ao anticorpo através de grupos quelantes do metal como o EDTA; compostos quimioluminescentes,

por exemplo, luminol, isoluminol, sais de acridinio e similares; compostos bioluminescentes, por exemplo, luciferina, aequorina (proteína verde fluorescente) e similares.

Os anticorpos podem estar associadas (acoplados) a um suporte insolúvel, como uma placa de poliestireno ou um leito. Marcadores indirectos incluem: segundos anticorpos específicos para os anticorpos específicos para o polipeptídeo codificado ("primeiros anticorpos específicos"), onde o segundo anticorpo é marcado como descrito acima; e membros dos pares das ligações específicas, por exemplo, biotina-avidina, e similares. A amostra biológica pode ser posta em contacto com e imobilizada num suporte sólido ou transportador, como a nitrocelulose, que é capaz de imobilizar células, partículas de células, ou proteínas solúveis. Este suporte pode então ser lavado com as soluções tampão adequadas, seguido pelo contacto com um primeiro anticorpo específico marcado para detecção. Métodos de detecção são conhecidos na especialidade e serão escolhidos de acordo com o sinal emitido pelo marcador detectável. A detecção é conseguida, geralmente, por comparação com um controlo adequado e de padrões adequados.

Em algumas modalidades, os métodos são adaptados para utilização *in vivo*, por exemplo, para localizar e identificar locais onde estão presentes células de cancro do cólon. Nestas modalidades, um meio marcador detectável, por exemplo, um anticorpo que é específico para um cancro do cólon associado a polipeptídeos é administrado a um indivíduo (por exemplo, por injeção) e as células marcadas são localizadas utilizando técnicas de imagiologia padrão, incluindo, mas não se limitando a, imagiologia de ressonância magnética, tomografia

computadorizada de varrimento, e afins. Desta forma, as células de cancro do cólon são diferencialmente marcadas.

Detecção de um polinucleotídeo que representa um gene expresso diferencialmente numa célula de cancro do cólon

Em algumas modalidades, os métodos são fornecidos para a detecção de uma célula de cancro do cólon por detecção de uma expressão na célula de uma transcrição ou que é expressa diferencialmente numa célula de cancro do cólon. Qualquer um de uma variedade de métodos conhecidos pode ser usado na detecção, incluindo, mas não se limitando a: detecção de uma transcrição por hibridização com um polinucleotídeo que hibridiza com um polinucleotídeo que é expresso diferencialmente numa célula de cancro do cólon; detecção de uma transcrição de uma reacção em cadeia de uma polimerase utilizando iniciadores de oligonucleótídeos específicos; hibridização *in situ* de uma célula usando como uma sonda um polinucleotídeo que hibridiza com um gene que é expresso diferencialmente numa célula de cancro do cólon.

Os métodos podem ser usados para detectar e/ou medir os níveis de mRNA de um gene que é expresso diferencialmente numa célula de cancro do cólon. Em algumas modalidades, os métodos incluem: a) colocar uma amostra em contacto com um polinucleotídeo que corresponde a um gene diferencialmente expresso aqui descrito em condições que permitam a hibridação; e b) detecção de hibridização, se fôr o caso. A detecção de hibridização diferencial, quando feita por comparação com um controlo adequado, é uma indicação da presença na amostra de um polinucleotídeo que é expresso diferencialmente numa célula de cancro do cólon. Controlos adequados incluem, por exemplo, uma

amostra que seja conhecida por não conter um polinucleotídeo que seja expresso diferencialmente numa célula de cancro do cólon, e da utilização de um polinucleotídeo marcado da mesma "sensibilidade" que o polinucleotídeo que é expresso diferencialmente numa célula de cancro do cólon. As condições que permitem efectuar a hibridização são conhecidas na especialidade, e já foram descritas com maior detalhe acima.

A detecção também pode ser conseguida através de qualquer método conhecido, incluindo mas não limitado a, hibridização *in situ*, PCR (reacção em cadeia de polimerase), RT-PCR (transcrição reversa-PCR), e "Northern" blotting (ou RNA blotting), ou combinações de tais técnicas, utilizando um polinucleotídeo devidamente marcado. Uma variedade de marcadores e de métodos de marcação de polinucleótidos são conhecidos na especialidade e podem ser utilizados nos métodos de ensaio da invenção. A hibridização específica pode ser determinada por comparação com os controlos adequados.

Um polinucleotídeo geralmente composto de pelo menos 12 nt contíguos de um polinucleotídeo aqui previsto, como mostrado na Listagem de Sequências ou da sequências de genes que correspondem aos polinucleótidos da Listagem de Sequências, são utilizadas para uma grande variedade de finalidades, tais como sondas para a detecção de e/ou medição de níveis de transcrição de um polinucleotídicas que é diferencialmente expresso numa célula de cancro do cólon. A descrição adicional das regiões preferidas das sequências polinucleotídicas descritas é encontrada nos exemplos. Uma sonda que hibridiza especificamente para um polinucleotídeo aqui divulgado deve proporcionar um sinal de detecção de, pelo menos 5, 10, ou 20 vezes maior do que a hibridização de fundo com outras sequências independentes. Note-se que o termo "sonda" como

é utilizado no presente contexto de detecção de ácidos nucleicos é utilizado para referir-se a uma sequência polinucleotídica utilizada para a detecção de um produto genético expresso diferencialmente numa amostra de teste. Como será facilmente reconhecido pelo especialista regularmente qualificado, a sonda pode ser marcada para detecção e contacta com, por exemplo, uma matriz contendo polinucleótídeos imobilizados obtidos a partir de uma amostra de teste (por exemplo, mRNA). Alternativamente, a sonda pode ser imobilizada numa matriz e a amostra do ensaio marcada para detecção. Estas e outras variações dos métodos da invenção encontram-se claramente dentro das práticas da especialidade e estão dentro do âmbito de aplicação da invenção.

São utilizadas sondas de nucleotídeos para detectar a expressão de um gene correspondente ao polinucleotídeo fornecido. Nas Northern blots, o mRNA é separado por electroforese e colocado em contacto com uma sonda. A sonda é detectada por hibridização a uma espécie de mRNA de dimensão específica. A quantidade de hibridização pode ser quantificada para determinar quantidades relativas de expressão, por exemplo, com uma condição particular. As sondas são utilizadas para a hibridização *in situ* de células para detectar a expressão. As sondas podem também ser usadas para a detecção diagnóstica *in vivo* de sequências de hibridização. As sondas são normalmente marcadas com um isótopo radioactivo. Podem ser usados outros tipos de marcadores detectáveis como cromóforos, fluoróforos, e enzimas. Outros exemplos de testes de hibridização de nucleotídeos são descritos na WO92/02526 e USPN 5,124,246.

A PCR é outro meio para a detecção de pequenas quantidades de ácidos nucleicos alvo (ver, por exemplo, Mullis et al., Meth. Enzymol. (1987) 155: 335; USPN 4683195; e USPN 4.683.202). Dois

nucleótideos iniciadores de polinucleótidos que originarão híbridos com os ácidos nucleicos alvo são usados para iniciar a reacção. Os iniciadores podem ser compostos de sequência no interior de 3' e 5' nos polinucleótidos da Listagem de Sequências. Alternativamente, se os iniciadores são 3' e 5' para estes polinucleótideos, eles não hibridarão para eles ou para os seus complementares. Após a amplificação do alvo com uma polimerase termoestável, os ácidos nucleicos alvo amplificados podem ser detectados através de métodos conhecidos na especialidade como, por exemplo, Southern blot. O mRNA ou o cDNA podem também ser detectados por técnicas de blotting tradicionais (por exemplo, Southern blot, Northern blot, etc) descritas em Sambrook et al., " Molecular Cloning: A Laboratory Manual " (New York, Cold Spring Harbor Laboratory, 1989) (por exemplo, sem amplificação PCR). Em geral, o mRNA ou o cDNA, gerado a partir de mRNA utilizando uma enzima polimerase, pode ser purificado e separado usando eletroforese em gel, e transferidos para um suporte sólido, como, por exemplo, a nitrocelulose. O suporte sólido é exposto a uma sonda marcada, lavado para remover eventuais resíduos de sonda não hibridizada, e os duplexes contendo a sonda marcados são detectados.

Podem ser efectuados métodos utilizando a amplificação por PCR no DNA a partir de uma única célula, embora seja conveniente para a utilização, pelo menos, cerca de 10^5 células. A utilização da reacção em cadeia da polimerase é descrita em Saiki et al. (1985) Science 239: 487, e uma revisão das técnicas actuais pode ser encontrada em Sambrook, et al. Molecular Cloning: A Laboratory Manual, CSH Press 1989, pp. 14.2-14.33. Um marcador detectável pode ser incluído na reacção de amplificação. Marcadores detectáveis adequados incluem fluorocromos, (por

exemplo, isotiocianato de fluoresceína (FITC), rodamina, vermelho do Texas, ficoeritrina, aloficocianina, 6 - de carboxi-fluoresceína (6 - FAM), 2 ',7'-dimethoxi-4',5'- dicloro-6-carboxifluoresceína, 6 - Carboximetilcelulose-X-rodamina (ROX), 6 - Carboximetilcelulose-2 ', 4', 7 ', 4,7 - hexaclorofluoresceína (HEX), 5-carboxifluoresceína (5 - FAM) ou N, N, N',N'-tetrametil-6-carboxi-rodamina (TAMRA)), marcadores radioactivos, (por exemplo, ^{32}P , ^{35}S , ^3H , etc), e similares. O marcador pode ser um sistema de duas fases, onde o polinucleótido é conjugado com biotina, haptenos, etc, com uma elevada afinidade com o parceiro da ligação, por exemplo, avidina, anticorpos específicos, etc, onde o parceiro de ligação é associado a um marcador detectável. O marcador pode estar associado a um ou ambos os iniciadores. Alternativamente, o conjunto de nucleótidos utilizados na amplificação é marcado, de modo a incorporar o marcador no produto de amplificação.

Matrizes

Matrizes polinucleotídicas proporcionam uma técnica de produção elevada que pode permitir o ensaio de um grande número de polinucleótidos ou polipeptídeos numa amostra. Esta tecnologia pode ser utilizada como uma ferramenta de teste para a expressão diferencial.

Uma variedade de métodos de produção de matrizes, bem como as variações destes métodos, são conhecidos na especialidade e a sua utilização está prevista na invenção. Por exemplo, podem ser criadas matrizes por distribuição de sondas polinucleotídicas sobre um substrato (por exemplo, vidro, nitrocelulose, etc) numa matriz ou distribuição bidimensional com sondas ligadas. As sondas podem ser ligadas ao substrato por ligações covalentes ou

por interações não específicas, como, por exemplo, interações hidrofóbicas.

As amostras de polinucleótidos podem ser marcadas de modo detectável (por exemplo, usando marcadores radioactivos ou fluorescentes) e, depois, hibridizadas para as sondas. Os polinucleótidos de cadeia dupla, incluindo a amostra de polinucleótidos marcados ligados aos polinucleótidos sonda, podem ser detectados depois da parte do material não ligado da amostra ter sido removida por lavagem. Alternativamente, os polinucleótidos de amostra de teste podem ser imobilizados sobre a matriz, e as sondas marcadas de modo detectável. As técnicas para a construção de arranjos e os métodos de utilização dos mesmos são descritos em, por exemplo, Schena et al. (1996) Proc Natl Acad Sci E.U.A.. 93 (20): 10614-9; Schena et al. (1995) Science 270 (5235): 467-70; Shalon et al. (1996) Genome Res. 6 (7): 639-45, USPN 5807522, EP 799 897; WO 97/29212, WO 97/27317; EP 785 280; WO 97/02357; USPN 5593839; USPN 5578832; EP 728 520; USPN 5599695; EP 721 016; USPN 5556752; WO 95/22058, e USPN 5631734.

As matrizes podem ser usadas para, por exemplo, analisar a expressão diferencial de genes e pode ser usado para determinar a função dos genes. Por exemplo, podem ser utilizados arranjos para detectar a expressão diferencial de um gene correspondente a um dos polinucleotídeos aqui descritos, sendo a expressão comparada entre uma célula de teste e uma célula de controlo (por exemplo, células cancerosas e células normais). Por exemplo, a expressão elevada de uma determinada mensagem numa célula de cancro que não é observada numa célula normal correspondente, pode indicar um determinado produto específico do gene do cancro. Exemplos de utilizações de matrizes são discutidos com mais

pormenorizadamente em, por exemplo, em Pappalarado et al., Sem. Radiation Oncol. (1998) 8: 217; e Ramsay Nature Biotechnol. (1998) 16: 40. Além disso, muitas das variações nos métodos de detecção usando matrizes são bem conhecidos dos especialistas no ramo e estão dentro do âmbito de aplicação da presente invenção. Por exemplo, em vez de imobilizar a sonda num suporte sólido, a amostra pode ser imobilizada num suporte sólido que é então contactado pela sonda.

Diagnóstico de Prognóstico, Avaliação de Terapia (Therametrics), e Gestão do Cancro

Os polinucleótidos aqui descritos, assim como os seus produtos genéticos e correspondentes genes e produtos genéticos, são de especial interesse quer como marcadores genéticos ou bioquímicos (por exemplo, no sangue ou tecidos) que irão detectar as primeiras alterações ao longo do percurso carcinogénico e/ou para monitorar a eficácia de várias terapias e intervenções preventivas.

Por exemplo, o nível de expressão de alguns polinucleótidos pode ser indicativo de um pior prognóstico, e, por isso, garante a existência de quimio ou radioterapia mais agressiva para um paciente ou vice-versa. A correlação de características específicas alternativas do tumor com resposta ao tratamento e os resultados nos doentes podem definir indicadores de prognóstico que permitem o design de terapia adaptada baseada no perfil molecular do tumor. Estas terapias incluem antagonistas para detecção de anticorpos, (por exemplo, pequenas moléculas), e terapia genética.

A determinação da expressão de alguns polinucleótidos e a comparação de um perfil de paciente de expressão conhecida em

tecidos normais e variantes da doença permite uma determinação do melhor tratamento possível para um paciente, tanto em termos de especificidade de tratamento como em termos do nível de conforto do doente . Marcadores tumorais alternativos, como a expressão de polinucleotídeos, podem também ser utilizados para melhorar a classificação e, assim, diagnosticar e tratar as diferentes formas e estados de doença do cancro. Duas classificações amplamente utilizadas em oncologia que podem beneficiar da identificação dos níveis de expressão dos genes correspondentes aos polinucleótídeos descritos neste documento são o estadiamento da desordem do cancro, e a graduação da natureza do tecido canceroso.

Os polinucleotídeos que correspondem a genes expressos diferencialmente, bem como os seus produtos genéticos codificados, podem ser úteis para monitorar pacientes tendo ou suscetíveis a ter cancro, para detectar ocorrências potencialmente malignas a um nível molecular, antes de serem detectáveis a um nível morfológico bruto. Além disso, os polinucleótidos aqui descritos, assim como os genes correspondentes a esses polinucleótidos, podem ser úteis como indicadores de medida da terapia (*therametrics*), por exemplo, para avaliar a eficácia da terapia através da utilização dos polinucleótidos ou dos seus produtos genéticos codificados, a fim de avaliar, por exemplo, a extensão do dano tumoral no paciente antes, durante e após a terapia.

Além disso, um polinucleotídeo identificado como o que corresponde a um gene que é expresso diferencialmente e, portanto, importante para um tipo de cancro, pode também ter implicações para o desenvolvimento ou o risco de desenvolvimento de outros tipos de cancro, por exemplo, quando um polinucleotídeo

representa um gene expresso diferencialmente em vários tipos de cancro. Assim, por exemplo, a expressão de um polinucleotídeo correspondendo a um gene que tem implicações clínicas no cancro do cólon metastático pode também ter implicações clínicas para o cancro de mama ou para o cancro do ovário.

Estadiamento. O estadiamento é um método utilizado pelos médicos para descrever o avanço do estado canceroso num paciente. O estadiamento ajuda o médico na definição de um prognóstico, no planeamento do tratamento e na avaliação dos resultados desse tratamento. Os sistemas de estadiamento variam de acordo com os tipos de cancro, mas geralmente envolvem a prossecução do sistema "TNM": o tipo de tumor, indicado por T; se o cancro está metastizado nos gânglios linfáticos próximos, indicado por N; e se o cancro está metastizado até partes mais distantes indicada por M. Geralmente, quando um cancro é apenas detectável na zona da lesão primária sem se ter propagado a qualquer dos gânglios linfáticos é designado de Estágio I. Se ele apenas se espalhou para os gânglios linfáticos mais próximos, é designado de Estágio II. N Estágio II, o cancro geralmente espalhou-se para os gânglios linfáticos na proximidade próximo ao local da lesão primária. Cancros que se propagaram a uma parte distante do corpo, tais como o fígado, ossos, cérebro ou outros locais, estão no Estágio IV, o mais avançado.

Os polinucleótidos e genes e produtos genéticos correspondentes aqui descritos podem facilitar o ajuste fino do processo de estadiamento através da identificação de marcadores para a agressividade de um cancro, por exemplo, o potencial metastático, bem como a presença em diferentes áreas do corpo. Assim, um cancro de Estágio II com um polinucleotídeo significando um elevado potencial de cancro metastático pode ser

usado para alterar a classificação de um tumor no limite do Estágio II para tumor do Estágio III, justificando uma terapêutica mais agressiva. Inversamente, a presença de um polinucleotídeo significando um menor potencial metastático permite um estagiamento mais conservador do tumor.

Graduação dos cancros. Grau é um termo usado para descrever o quanto um tumor se assemelha de perto ao tecido normal do mesmo tipo. A aparência microscópica de um tumor é usada para identificar o grau do tumor com base em parâmetros como a morfologia celular, a organização celular e outros marcadores de diferenciação. Como regra geral, o grau de um tumor corresponde à sua taxa de crescimento ou de agressividade, com os tumores indiferenciados ou de grau elevado geralmente a ser mais agressivos do que os tumores bem diferenciados ou de baixo grau. As seguintes directrizes são geralmente utilizadas na graduação de tumores: 1) GX - O grau não pode ser avaliado; 2) G1 - Bem diferenciado; G2 - Moderadamente bem diferenciado; 3) G3 - pouco diferenciado; 4) G4 - indiferenciados. O polinucleótidos da Listagem de Sequências, e os seus genes e produtos genéticos correspondentes, podem ser particularmente valiosos para a determinação do grau do tumor, pois eles não só ajudam à determinação do estado de diferenciação das células de um tumor, como também podem identificar outros factores que não a diferenciação que são valiosos para determinar a agressividade de um tumor, como o potencial metastático.

A detecção de cancro do cólon. Os polinucleótidos que correspondem a genes que exibem o padrão adequado de expressão podem ser usados para detectar cancro do cólon num indivíduo. O cancro colorectal é uma das neoplasias mais comuns em seres humanos e, talvez a forma mais frequente de neoplasia

hereditária. A prevenção e detecção precoce são factores-chave para o controlo e a cura do cancro colorrectal. O cancro colorrectal começa sob a forma de pólipos, que são pequenos crescimentos benignos de células que se formam sobre o revestimento interior do cólon. Ao longo de um período de vários anos, alguns destes pólipos acumulam mutações adicionais e tornam-se cancerosos. Têm sido identificadas múltiplas famílias de distúrbios de cancro colorrectal, as quais podem ser resumidas como segue: 1) Polipose adenomatosa familiar (FAP); 2) Síndrome de Gardner; 3) Não-polipose hereditária do cancro do cólon (HNPCC); e 4) Cancro colo-rectal familiar nos Judeus Asquenazi.

A expressão de polinucleótídeos adequados pode ser utilizada no diagnóstico, prognóstico e gestão do cancro colorrectal. A detecção de cancro do cólon pode ser realizada usando os níveis de expressão de qualquer uma destas sequências isoladamente ou em combinação com os níveis de expressão. A determinação da natureza agressiva e/ou o potencial metastático de um cancro do cólon pode ser realizada por comparação dos níveis de um ou mais produtos genéticos dos genes correspondentes aos polinucleótídeos aqui descritos, e comparando com os níveis totais de outra sequência conhecida por variar em tecido canceroso, por exemplo, a expressão de p53, DCC, ras, FAP (ver, por exemplo, Fearon ER, et al., Cell (1990) 61 (5): 759; Hamilton SR et al., Cancer (1993) 72: 957; Bodmer W, et al., Nat Genet. (1994) 4 (3): 217; Fearon ER, Ann NY Acad Sci. (1995) 768: 101).

Por exemplo, o desenvolvimento do cancro do cólon pode ser detectado através da análise do nível de expressão de um gene correspondente a um polinucleótídeo aqui descrito para os níveis de oncogenes (por exemplo ras), ou de genes supressores tumorais (por exemplo, FAP ou p53). Assim, a expressão de polinucleótídeos

marcadores específicos pode ser usada para distinguir entre os tecidos de cólon normal e cancerosos, para distinguir entre cancros do cólon com diferentes células de origem, para distinguir entre cancros do cólon com diferentes taxas de potencial metastáticos, etc. Para uma revisão dos marcadores de cancro, ver, por exemplo, Hanahan *et al.* (2000) *Cell* 100: 57-70.

Tratamento do Cancro do Cólon

A invenção prevê ainda métodos para reduzir o crescimento de células de cancro do cólon. Os métodos permitem diminuir a expressão de um gene que é diferencialmente expresso numa célula de cancro do cólon ou diminuir o nível de e/ou diminuindo uma actividade de um polipeptídeo associado ao cancro do cólon. Em geral, os métodos consistem de um contacto com células de cancro do cólon com uma substância que modula (1) a expressão de um gene que é expresso diferencialmente no cancro do cólon ou (2) um nível de e/ou uma actividade de um polipeptídeo associado ao cancro do cólon. "Reduzir o crescimento de células de cancro do cólon" inclui, mas não se limita a, reduzir a proliferação de células de cancro do cólon e reduzir a incidência da transformação de células cólon não-cancerosas em células do cólon cancerosas. Se uma redução do crescimento celular do cancro do cólon foi conseguida, pode ser facilmente determinado utilizando qualquer ensaio conhecido incluindo, mas não se limitando a: incorporação de [³H]-timidina; contagem do número de células durante um período de tempo; detecção e/ou medição de um marcador associado ao cancro do cólon (por exemplo, CEA, CA 19-9, e LASA).

A presente invenção fornece métodos para o tratamento do cancro do cólon, geralmente compreendendo a administração, a um indivíduo necessitado, de uma substância que reduz o crescimento

celular do cancro do cólon, numa quantidade suficiente para reduzir o crescimento de células do cancro do cólon e tratar o cancro do cólon. Pode ser avaliado se uma substância, ou uma quantidade específica da substância, é eficaz no tratamento do cancro do cólon usando qualquer um de uma série de testes para diagnóstico conhecidos de cancro do cólon, incluindo, mas não se limitando a, sigmoidoscopia, protoscopia, exame rectal, colonoscopia com biópsia, estudos de contraste radiográfico, varrimentos CAT, angiografia e detecção de um marcador tumoral associado com o cancro do cólon no sangue do indivíduo. A substância pode ser administrada sistemicamente ou localmente. Assim, em algumas modalidades, a substância é administrada localmente, e o crescimento do cancro do cólon diminui no local da administração. A administração local pode ser útil no tratamento, por exemplo, de um tumor sólido.

[0179] Uma substância que reduz o crescimento celular do cancro do cólon pode ser aplicada a uma célula de cancro do cólon. Assim, em algumas modalidades, a invenção proporciona um método de aplicação a uma célula de cancro do cólon, que consiste na administração de um complexo fármaco-anticorpo a um indivíduo, sendo o anticorpo específico para polipeptídeos associados ao cancro do cólon, e este fármaco reduzindo o crescimento celular do cancro do cólon, dos quais são conhecidos uma grande variedade na especialidade. A aplicação dirigida pode ser obtida por acoplamento (por exemplo, ligando, directamente ou através de uma molécula de ligação, quer covalente ou não-covalentemente, de modo a formar um complexo fármaco-anticorpo) de um fármaco a um anticorpo específico para um polipeptídeo associado ao cancro do cólon. Métodos de acoplamento de um fármaco a um anticorpo são

bem conhecidos na especialidade e não precisam de ser aqui desenvolvidos.

Identificação de Alvos Terapêutico e Agentes Terapêuticos Anti-Cancro

A presente invenção, abrange também os métodos de identificação dos agentes que tenham a capacidade de modular a actividade de um produto genético expresso diferencialmente, bem como métodos para a identificação de um produto genético expresso diferencialmente como um alvo terapêutico para o tratamento do cancro, sobretudo cancro do cólon.

Agentes Candidatos

A identificação de compostos que modulam a actividade de um produto genético expresso diferencialmente pode ser realizada usando qualquer uma de uma variedade de técnicas de rastreio de fármacos. Esses agentes são candidatos ao desenvolvimento de terapias do cancro. De particular interesse são os ensaios de despistagem de agentes que possuem toxicidade tolerável para as células humanas normais, não-cancerosas. Os testes de despistagem da invenção geralmente são baseados na capacidade do agente de modular uma actividade de um produto genético expresso diferencialmente e/ou no fenómeno de inibição ou supressão associado ao cancro (por exemplo, a proliferação celular, formação de colónias, detenção do ciclo celular, metástase e similares).

O termo "agente", como utilizado neste documento descreve qualquer molécula, por exemplo, proteína ou fármaco, com a capacidade de modular uma actividade biológica de um produto genético de um gene diferencialmente expresso. Geralmente é

realizada em paralelo uma pluralidade de misturas de ensaio com diferentes concentrações de agente para obter respostas diferenciadas às diferentes concentrações. Tipicamente, uma dessas concentrações serve como controle negativo, ou seja, à concentração zero ou abaixo do nível de detecção.

Os candidatos a agentes abrangem inúmeras categorias químicas, embora normalmente sejam moléculas orgânicas, de preferência, compostos orgânicos pequenos com um peso molecular de mais de 50 e menos de cerca de 2500 daltons. Os candidatos a agentes consistem de grupos funcionais necessários para a interacção com proteínas estruturais, particularmente por pontes de hidrogénio, e tipicamente incluem pelo menos uma amina, grupo carbonilo, hidroxilo ou carboxilo, de preferência, pelo menos dois dos grupos funcionais químicos. Os candidatos a agentes muitas vezes incluem estruturas cíclicas ou heterocíclicas de carbono e/ou policíclicos aromáticos ou estruturas substituídas com um ou mais dos referidos grupos funcionais. Os candidatos a agentes também são encontrados entre as biomoléculas incluindo, mas não limitados a: peptídeos, sacarídeos, ácidos gordos, esteróides, purinas, pirimidinas, derivados, análogos estruturais ou combinações dos mesmos.

Os candidatos a agentes são obtidos a partir de uma ampla variedade de fontes, incluindo bibliotecas de compostos naturais ou sintéticos. Por exemplo, inúmeros meios estão disponíveis para a síntese aleatória e dirigida de uma ampla variedade de compostos orgânicos e biomoléculas, incluindo a expressão de oligonucleotídeos e oligopeptídeos aleatórios. Alternativamente, bibliotecas de compostos naturais, sob a forma de extractos de bactérias, de fungos, de vegetais e de animais (incluindo os extractos de tecido humano para identificar factores endógenos

que afectam produtos genéticos expressos diferencialmente) estão disponíveis ou são facilmente produzidos. Adicionalmente, bibliotecas e compostos produzidos de forma natural ou sintética são facilmente modificados através de processos usando agentes químicos, físicos e bioquímicos convencionais, e podem ser utilizados para produzir bibliotecas combinatórias. Agentes farmacológicos conhecidos podem ser submetidos a modificações químicas dirigidas ou aleatórias, tais como acilação, alquilação, esterificação, amidificação, etc. para produzir análogos estruturais.

Exemplos de candidatos a agentes de particular interesse incluem, mas não se limitam a, polinucleótidos *antisense* e anticorpos, receptores solúveis, e similares. Anticorpos e receptores solúveis são de especial interesse como candidatos a agentes onde os produtos genéticos expressos diferencialmente pretendidos são segregados ou acessíveis na superfície de células (por exemplo, receptores e outra molécula associados de modo estável com a membrana celular exterior.

Rastreo dos candidatos a agentes

Os ensaios de rastreio podem ser baseados em qualquer de uma variedade de técnicas facilmente disponíveis e conhecidas para os especialistas na área. Em geral, os ensaios de rastreio envolvem o contacto de uma célula cancerosa (de preferência uma célula cancerosa do cólon) com um agente candidato, e avaliando o efeito sobre a actividade biológica de um produto genético expresso diferencialmente. O efeito sobre uma actividade biológica pode ser detectado através, por exemplo, da detecção da expressão de um produto genético de um gene expresso diferencialmente (por exemplo, uma diminuição de mRNA ou nos níveis de polipeptídeo

iria, por sua vez, causar uma diminuição da actividade biológica do produto genético). Em alternativa ou em complemento, o efeito do agente candidato pode ser avaliado examinando os efeitos dos candidatos a agente numa análise funcional. Por exemplo, quando o produto genético expresso diferencialmente é uma enzima, então o efeito sobre a actividade biológica pode ser avaliado através da detecção do nível de actividade enzimática associada ao produto genético expresso diferencialmente. A análise funcional será seleccionada de acordo com o produto genético expresso diferencialmente. Em geral, quando o gene expresso diferencialmente é aumentado em expressão numa célula cancerosa, os agentes de interesse são aqueles que provocam a diminuição da actividade do produto genético diferencialmente expresso.

Os ensaios descritos abaixo podem ser facilmente adaptados em modalidades de ensaios de rastreio da invenção. Exemplos de ensaios úteis em termos de despistagem de candidatos a agentes, incluem, mas não se limitam a, ensaios à base de hibridização (por exemplo, o uso de sondas de ácidos nucleicos ou iniciadores para avaliar os níveis de expressão), com ensaios de anticorpos (por exemplo, para avaliar os níveis de produtos genéticos polipeptídicos), ensaios de ligação (por exemplo, para detectar a interacção de um candidato a agente com um polipeptídeo expresso diferencialmente, podendo estes ensaios ser ensaios competitivos onde um ligando natural ou sintético do polipeptídeo está disponível), e similares. Exemplos adicionais de testes incluem, mas não estão limitados a, testes de proliferação celular, testes de antisense de eliminação, testes para detectar a inibição do ciclo celular, testes de indução de morte celular/apoptose, e afins. Geralmente esses ensaios são realizados *in vitro*, mas

muitos ensaios podem ser adaptados para análises in vivo, por exemplo, num modelo animal do cancro.

Identificação dos alvos terapêuticos

Noutra modalidade, a invenção contempla a identificação de genes expressos diferencialmente e dos produtos genéticos como alvos terapêuticos. Em alguns aspectos, este é o inverso dos ensaios acima descritos para a identificação dos agentes com actividade na modulação (por exemplo, diminuindo ou aumentando) da actividade de um produto genético diferencialmente expresso.

Nesta modalidade, os alvos terapêuticos são identificados através da análise dos efeitos de um agente que podem ser demonstrados ou tenham sido demonstrados a modular um fenótipo canceroso (por exemplo, repressão de inibidor ou evitar o desenvolvimento de um fenótipo canceroso). Esses agentes são geralmente mencionados neste documento como um "agente anti-cancerígeno", que englobam agentes quimioterápicos. Por exemplo, o agente pode ser um oligonucleótido *antisense* que é específico para uma determinada transcrição genética. Por exemplo, o oligonucleótido *antisense* pode ter uma sequência que corresponde a uma sequência de um gene diferencialmente expresso aqui descrito, por exemplo, uma sequência de uma dos SEQ ID NO : 1.

Os ensaios para a identificação de alvos terapêuticos podem ser realizados através de uma variedade de formas usando métodos que são bem conhecidos dos especialistas no ramo. Por exemplo, um teste de células cancerosas que expressa ou sobre-expressa um gene expresso diferencialmente é posto em contacto com um agente anti-cancerígeno, sendo o efeito sobre um fenótipo canceroso e sobre a actividade biológica do produto genético candidato avaliado. A actividade biológica dos produtos genéticos

candidatos pode ser analisada examinando, por exemplo, a modulação da expressão de um gene que codifica o produto genético candidato (por exemplo, como é detectado, por exemplo, um aumento ou uma diminuição nos níveis de transcrição ou nos níveis de polipeptídeo), ou a modulação de uma actividade enzimática ou outra do produto genético. O fenótipo cancerígeno pode ser, por exemplo, a proliferação celular, a perda de inibição de contacto do crescimento (por exemplo, formação de colónias), o crescimento tumoral (*in vitro* ou *in vivo*) e similares. Em alternativa ou em complemento, o efeito de modulação de uma actividade biológica do gene alvo candidato, após a morte celular/apoptose ou a regulação ciclo celular, podem ser avaliados.

A inibição ou supressão de um fenótipo canceroso, ou um aumento na morte celular/apoptose, como resultado da modulação da actividade biológica de um produto genético candidato, indica que o produto genético candidato é alvo adequado para a terapia do cancro. Os ensaios descritos abaixo podem ser facilmente adaptados para ensaios para a identificação de alvos terapêuticos. Geralmente esses ensaios são realizados *in vitro*, mas muitos ensaios podem ser adaptados para análises *in vivo*, por exemplo, num modelo animal de cancro adequado e aceite na especialidade.

Identificação dos Peptídeos Análogos e Antagonistas

Polipeptídeos codificados pelos genes expressos diferencialmente, identificados neste documento, podem ser usados para rastrear bibliotecas de peptídeos para identificar parceiros para ligação, como receptores, de entre os polipeptídeos codificados. As bibliotecas de peptídeos podem ser sintetizadas

de acordo com métodos conhecidos na especialidade (ver, por exemplo, USPN 5010175, e WO 91/17823).

Os agonistas ou antagonistas dos polipeptídeos da invenção podem ser rastreados usando qualquer método disponível conhecido na especialidade, como sinal de transdução, ligação a anticorpos, ligação a receptores, ensaios mitogénicos, ensaios quimiotáxicos, etc. As condições ideais do ensaio devem ser semelhantes às condições sob as quais a actividade nativa é exibida *in vivo*, ou seja, nas condições do pH, temperatura e força iónica fisiológicas. Agonistas ou antagonistas adequados irão apresentar forte inibição ou reforço da actividade nativa em concentrações que não causam efeitos colaterais tóxicos no indivíduo. Agonistas ou antagonistas que concorrem para a ligação dos polipeptídeos nativos podem exigir concentrações que sejam iguais ou superiores à concentração nativa, ao mesmo tempo que inibidores capazes de ligar irreversivelmente o polipeptídeo podem ser adicionados em concentrações da ordem da concentração nativa.

Essa triagem e experimentação pode levar à identificação de um parceiro de ligação do polipeptídeo, como um receptor, codificado por um gene ou um cDNA correspondente a um polinucleotídeo aqui descrito, e, pelo menos, um peptídeo agonista ou antagonista do parceiro de ligação. Tais agonistas e antagonistas podem ser usados para modular, incrementar ou inibir a função de receptor nas células para as quais o receptor é nativo, ou em células que possuem o receptor como resultado da engenharia genética. Além disso, se o receptor partilha características biologicamente importantes com um receptor conhecido, informações sobre as ligações do agonista/ antagonista podem facilitar o desenvolvimento de melhores agonistas/ antagonistas do receptor conhecido.

Composições Farmacêuticas e Usos Terapêuticos

As composições Farmacêuticas da invenção podem incluir polipeptídeos, anticorpos, ou polinucleótidos (incluindo nucleotídeos antisense e ribozimas) da invenção reivindicada numa quantidade terapeuticamente eficaz. O termo "quantidade terapeuticamente eficaz", como é usado neste documento, refere-se a uma quantidade de um agente terapêutico adequada para tratamento, melhoria, ou para impedir uma doença ou condição de interesse, ou para exibir um efeito terapêutico ou preventivo detectável. O efeito pode ser detectado através de, por exemplo, marcadores químicos ou níveis de antigénio. Efeitos terapêuticos incluem também a redução de sintomas físicos, tais como a diminuição da temperatura corporal. A quantidade eficaz exacta para um indivíduo vai depender da dimensão e da saúde do indivíduo, da natureza e extensão do estado de doença, e da terapêutica ou da combinação de terapias seleccionados para serem administradas. Assim, não é útil especificar *a priori* uma quantidade exacta eficaz. No entanto, a quantidade eficaz para uma dada situação é determinada por experimentação de rotina e pertence ao âmbito do julgamento do médico. Para efeitos da presente invenção, uma dose eficaz terá geralmente a partir de cerca de 0,01 mg/kg até 50 mg/kg ou 0,05 mg/kg até cerca de 10 mg/kg dos constructos de DNA no indivíduo a que é administrada.

A composição farmacêutica também pode conter um transportador farmacêuticamente aceitável. O termo "transportador farmacêuticamente aceitável" refere-se a um transportador para a administração de um agente terapêutico, como sejam anticorpos ou um polipeptídeo, genes, e de outros agentes terapêuticos. O termo refere-se a qualquer transportador farmacêutico que não possa induzir por si mesmo a produção de anticorpos prejudiciais para o

indivíduo que recebe a composição, e que pode ser administrado sem toxicidade. Os transportadores adequados podem ser macromoléculas grandes, metabolizadas lentamente como as proteínas, os polissacarídeos, os ácidos polilácticos, os ácidos poliglicólicos, os aminoácidos poliméricos, os copolímeros de aminoácidos, e partículas de vírus inativos. Esses transportadores são bem conhecidos dos especialistas no ramo. Transportadores farmacologicamente aceitáveis em composições terapêuticas podem incluir líquidos como a água, o soro fisiológico, o glicerol e o etanol. Substâncias auxiliares, tais como agentes humidificantes ou emulsionantes, substâncias tampão de pH e similares podem também estar presentes nesses veículos.

Normalmente, as composições terapêuticas são preparadas como injectáveis, quer como soluções líquidas ou suspensões; Também podem ser preparadas formas sólidas adequadas para preparar a solução líquida, ou a suspensão líquida, antes da injeção. Os lipossomas estão incluídos na definição de um transportador farmacologicamente aceitável. Sais farmacologicamente aceitáveis também podem estar presentes na composição dos produtos farmacêuticos, por exemplo, sais minerais ácidos, tais como hidrocloreto, hidrobrometo, fosfatos, sulfatos e afins; e os sais de ácidos orgânicos, tais como acetatos, propionatos, malonatos, benzoatos e afins. Uma discussão aprofundada de excipientes farmacologicamente aceitáveis está disponível na Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Métodos de administração. Uma vez formuladas, as composições da invenção podem ser (1) administradas directamente ao indivíduo (por exemplo, como polinucleotídicas ou polipeptídeos); ou (2) fornecidas ex vivo, para células derivadas do indivíduo (por exemplo, como na terapia genética ex vivo). A administração

directa das composições será em geral conseguida por injeção parenteral, por exemplo, por via subcutânea, intraperitoneal, intravenosa ou intramuscular, intratumoral ou para o espaço intersticial de um tecido. Outros modos de administração incluem a administração oral e pulmonar, supositórios, e aplicações transdérmicas, agulhas, e *hiposprays* (injectores sem agulha) ou pistolas de genes. A dosagem do tratamento pode ser através de uma dose única ou ser agendado um calendário de doses múltiplas.

[0199] Os métodos *ex vivo* para o fornecimento e reimplante das células transformadas num indivíduo são conhecidos na especialidade, e descritos, por exemplo, em International Publication No. 93/14778 WO. Exemplos de células úteis em aplicações *ex vivo* incluem, por exemplo, as células estaminais, nomeadamente as hematopoéticas, as células linfáticas, os macrófagos, as células dendríticas ou as células tumorais. Geralmente, a administração dos ácidos nucleicos para ambas as aplicações *ex vivo* e *in vitro* pode ser conseguida através de, por exemplo, transfecção mediada por dextrano, precipitação de fosfato de cálcio, transfecção mediada por polibreno, por fusão protoplástica, por eletroporação, por encapsulamento do(s) polinucleotídeo(s) em lipossomas, e por microinjecção directa de DNA em núcleos, que são bem conhecidos dos especialistas.

Quando um gene correspondente a um polinucleotídeo da invenção se correlaciona com uma disfunção proliferativa, como neoplasia, displasia, e hiperplasia, a disfunção pode ser passível de tratamento por administração de um agente terapêutico baseado nos polinucleotídeos fornecidos, polipeptídeo correspondente ou outra molécula correspondente (p. ex., *antisense*, ribozima, etc.).

A dose e o meio de administração das composições farmacêuticas inovadores são determinados com base nas qualidades específicas da composição terapêutica, na condição, idade, e peso do paciente, a evolução da doença, e outros factores relevantes. Por exemplo, a administração de composições de agentes polinucleotídicos terapêuticos da invenção inclui a administração local ou sistémica, incluindo injeção, administração oral, pistola de partículas ou administração por catéter, e administração tópica. De preferência, a composição terapêutica polinucleotídica contém um constructo de expressão contendo um promotor operacionalmente ligado a um polinucleotídeo de, pelo menos, 12, 22, 25, 30, or 35 nt consecutivos do polinucleotídeo aqui referido.

Podem ser usados vários métodos para administrar a composição terapêutica directamente num local específico do corpo. Por exemplo, uma pequena lesão metastática é localizada e a composição terapêutica injectada várias vezes em vários locais diferentes no seio do tumor. Alternativamente, identifica-se as artérias que irrigam o tumor, e a composição terapêutica é injectada nessas artérias, de modo a enviar a composição directamente para o tumor. Um tumor que tenha um centro necrótico é aspirado e a composição injectada directamente no agora vazio centro do tumor. A composição *antisense* é directamente administrada na superfície do tumor, por exemplo, por aplicação tópica da composição. Recorre-se à imagiologia de raios-X como método de auxílio em alguns dos métodos de administração acima descritos.

Também se pode recorrer à aplicação orientada mediada pelo receptor das composições terapêuticas contendo polinucleotídeo *antisense*, polinucleotídeos subgenómicos, ou anticorpos em

tecidos específicos. As técnicas de aplicação orientada mediada pelo receptor de DNA encontram-se descritas em, por exemplo, Findeis et al., Trends Biotechnol. (1993) 11:202; Chiou et al., Gene Therapeutics: Methods And Applications Of Direct Gene Transfer (J.A. Wolff, ed.) (1994); Wu et al., J. Biol. Chem. (1988) 263:621; Wu et al., J. Biol. Chem. (1994) 269:542; Zenke et al., Proc. Natl. Acad. Sci. (USA) (1990) 87:3655; Wu et al., J. Biol. Chem. (1991) 266:338. As composições terapêuticas contendo um polinucleotídeo são administradas numa gama de cerca de 100 ng a cerca de 200 mg de DNA na administração local num protocolo de terapia genética. Podem ainda usar-se gamas de concentração de cerca de 500 ng a cerca de 50 mg, cerca de 1 µg a cerca de 2 mg, cerca de 5 µg a cerca de 500 µg, e cerca de 20 µg a cerca de 100 µg de DNA num protocolo de terapia genética. Factores como o método de acção (p. ex., para aumentar or inibir níveis do produto genético codificado) e a eficácia de transformação e expressão são considerações que irão afectar a dosagem requerida para uma total eficácia dos polinucleotídeos *antisense* subgenómicos. Quando se pretende uma maior expressão numa área maior de tecido, pode ser necessário utilizar maiores quantidades de polinucleotídeos *antisense* subgenómicos ou as mesmas quantidades readministradas num protocolo de administrações sucessivas, ou várias administrações em porções de tecido próximas ou adjacentes, por exemplo, um local do tumor, para se obter um resultado terapêutico positivo. Em todos os casos, experiências de rotina em testes clínicos irão determinar as gamas específicas para a optimização do efeito terapêutico. Para genes relacionados com polinucleotídeos que codificam polipeptídeos ou proteínas com actividade anti-inflamatória, a

utilização apropriada, doses, e administração encontram-se descritas na USPN 5,654,173.

Os polinucleotídeos e polipeptídeos terapêuticos da presente invenção podem ser administrados por recurso a veículo de entrega de genes. O veículo de entrega de genes pode ser de origem viral ou não (ver geralmente, Jolly, *Cancer Gene Therapy* (1994) 1:51; Kimura, *Human Gene Therapy* (1994) 5:845; Connelly, *Human Gene Therapy* (1995) 1:185; and Kaplitt, *Nature Genetics* (1994) 6:148). A expressão de tais sequências codificadoras pode ser induzida por recurso a promotores endógenos de mamíferos ou heterólogos. A expressão da sequência codificadora pode ser constitutiva ou regulada.

Vectores de base viral para administração de um polinucleotídeo desejado e expressão numa célula desejada são bem conhecidos na especialidade. Exemplos de veículos de base viral incluem, mas não se limitam a retrovírus recombinantes (ver, p. ex., WO 90/07936; WO 94/03622; WO 93/25698; WO 93/25234; USPN 5,219,740; WO 93/11230; WO 93/10218; USPN 4,777,127; Patente na Grã-Bretanha No. 2,200,651; EP 0 345 242; e WO 91/02805), vectores baseado em alfavírus (p. ex., vectores do vírus de Sindbis, vírus da floresta de Semliki (ATCC VR-67; ATCC VR-1247), vírus de Ross River (ATCC VR-373; ATCC VR-1246) e vírus da encefalite equina Venezuelana (ATCC VR-923; ATCC VR-1250; ATCC VR 1249; ATCC VR-532), e vectores de vírus adeno-associados (AAV) (ver, p. ex., WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 and WO 95/00655). Também se pode empregar a administração de DNA ligado a adenovírus mortos, tal como descrito em Curiel, *Hum. Gene Ther.* (1992) 3:147.

[0206] Também podem ser empregues veículos e métodos não virais de distribuição, incluindo, mas não se limitando a DNA

policatiónico condensado ligado ou não a adenovírus mortos (ver, p. ex., Curiel, Hum. Gene Ther. (1992) 3:147); DNA ligado a ligandos (ver, p. ex., Wu, J Biol. Chem. (1989) 264:16985); células veículo de distribuição de células eucarióticas (ver, p. ex., USPN 5,814,482; WO 95/07994; WO 96/17072; WO 95/30763; e WO 97/42338) e neutralização de carga nucleica ou fusão com membranas celulares. Também se pode empregar DNA desprotegido. Exemplos de métodos de introdução de DNA desprotegido estão descritos na WO 90/11092 e USPN 5,580,859. Lipossomas que podem actuar como veículos de distribuição de genes estão descritos na USPN 5,422,120; WO 95/13796; WO 94/23697; WO 91/14445; e EP 0524968. Abordagens adicionais estão descritas em Philip, Mol. Cell Biol. (1994) 14:2411, e em Woffendin, Proc. Natl. Acad. Sci. (1994) 91:1581.

Outros métodos de administração não viral adequados incluem sistemas de distribuição mecânica tais como as abordagens descritas em Woffendin et al., Proc. Natl. Acad. Sci. USA (1994) 91 (24):11581. A sequência codificadora e o produto de expressão da mesma podem ainda ser administrados através da deposição de hidrogéis fotopolimerizados ou por utilização de radiação ionizante (ver, p. ex., USPN 5,206,152 e WO 92/11033). Outro método convencional de distribuição de genes que pode ser usado para administração da sequência codificadora inclui, por exemplo, utilização de uma pistola de partículas de transferência de genes manual (ver, p. ex., USPN 5,149,655); utilização de radiação ionizante para activar o gene transferido (ver, p. ex., USPN 5,206,152 e WO 92/11033).

A presente invenção será agora ilustrada por referência aos seguintes exemplos os quais constituem modalidades particularmente vantajosas. Contudo, deve ser notado que essas

modalidades são ilustrativas e não devem ser entendidas como limitativas da invenção de qualquer modo.

EXEMPLOS

Os seguintes exemplos são apresentados principalmente com o objectivo de ilustração. Será evidente para os especialistas que as formulações, dosagens, métodos de administração, e outros parâmetros desta invenção podem ser modificados ou substituídos de várias maneiras sem se sair do espírito e âmbito da invenção.

Exemplo 1: Fontes de Materiais Biológicos e Panorâmica de Polinucleotídeos Expressos pelos Materiais Biológicos

De forma a identificar genes que são expressos diferencialmente no cancro do cólon, prepararam-se bibliotecas de cDNA a partir de várias linhagens celulares diferentes e fontes de tecido. A Tabela 1 fornece um sumário dessas bibliotecas, incluindo o diminutivo do nome da biblioteca (usado daqui em diante), a fonte de mRNA usado para preparar a biblioteca de cDNA, a "alcunha" da biblioteca que é usada nas tabelas abaixo (entreaspas), e o número clones aproximado na biblioteca. As bibliotecas de cDNA foram preparadas de acordo com métodos bem conhecidos na especialidade, e as sequências das inserções de cDNA foram determinadas utilizando métodos bem conhecidos.

Tabela 1. Descrição de Bibliotecas de cDNA		
Biblioteca	Descrição	Número de Clones
1	Linhagem celular Km12 L4 do Cólon Humano: Elevado Potencial Metastático (derivada da Km12C)	308731
2	Linhagem celular Km12C do Cólon Humano: Baixo Potencial Metastático	284771
3	Linhagem celular MDA-MB-231 do Cancro Humano da Mama: Elevado Potencial Metastático; micromets no pulmão.	326937
4	Linhagem celular MCF7 do Cancro Humano da Mama: Não Metastático	318979
8	Linhagem celular MV-522 do Cancro do Humano Pulmão: Elevado Potencial Metastático	223620
9	Linhagem celular UCP-3 do Cancro do Humano Pulmão: Baixo Potencial Metastático	312503
12	Células microvasculares endoteliais humanas (HMEC) - NÃO TRATADAS (biblioteca de cDNA PCR (OligodT))	41938
13	Células microvasculares endoteliais humanas (HMEC) - TRATADAS com bFGF (biblioteca de cDNA PCR (OligodT))	42100
14	Células microvasculares endoteliais humanas (HMEC) - TRATADAS com VEGF (biblioteca de cDNA PCR (OligodT))	42825
15	Cólon Normal - Paciente UC#2 (biblioteca de PCR cDNA MICRODISSECADO (OligodT))	282718
16	Tumor do Cólon - Paciente UC#2 (biblioteca de PCR cDNA MICRODISSECADO (OligodT))	298829
17	Metástase no Fígado do Tumor do Cólon do Paciente UC#2 (biblioteca de PCR cDNA MICRODISSECADO (OligodT))	303462
18	Cólon Normal - Paciente UC#3 (biblioteca de PCR cDNA MICRODISSECADO (OligodT))	36216
19	Tumor do Cólon - Paciente UC#3 (biblioteca de PCR cDNA MICRODISSECADO (OligodT))	41388
20	Metástase no Fígado do Tumor do Cólon do Paciente UC#3 (biblioteca de PCR cDNA MICRODISSECADO (OligodT))	30956
21	Células GRRpz obtidas de epitélio da próstata normal	164801
22	Células W0ca obtidas do epitélio de um cancro da próstata de Grau 4 de Gleason	162088
23	Epitélio Normal do Pulmão do Paciente #1006 (biblioteca de PCR cDNA MICRODISSECADO (OligodT))	309349
24	Tumor Primário, Carcinoma das Células Grandes do Paciente #1006 (biblioteca de PCR cDNA MICRODISSECADO (OligodT))	309349
25	Epitélio Normal da Próstata do Paciente IF97-26811	279437
26	Epitélio do Cancro da Próstata Gleason 3+3 Paciente IF97-26811	269366

A linhagem celular KM12L4 é obtida da linhagem celular KM12C (Morikawa, et al, Cancer Research (1988) 48:6863). A linhagem

celular KM12C, que é fracamente metastática (pouco metastática) foi estabelecida em cultura a partir de um espécime cirúrgico de estadio B2 de Dukes (Morikawa et al. Cancer Res. (1988) 48:6863). A KML4-A é uma sublinhagem altamente metastática obtida da KM12C (Yeatman et al. Nucl. Acids Res. (1995) 23:4007; Bao-Ling et al. Proc. Annu. Meet. Am. Assoc. Cancer Res. (1995) 21:3269). As linhagens celulares (KM12C e obtidas da KM12C p. ex., KM12L4, KM12L4-A, etc.) são reconhecidas na especialidade como modelo de linhagem celular no estudo do cancro do cólon (ver, p. ex., Moriakawa et al., supra; Radinsky et al. Clin. Cancer Res. (1995) 1:19; Yeatman et al., (1995) supra; Yeatman et al. Clin. Exp. Metastasis (1996) 14:246).

A linhagem celular MDA-MB-231 foi originalmente isolada de effusões pleurais (Cailleau, J. Natl. Cancer. Inst. (1974) 53:661), é de alto potencial metastático, e forma adenocarcinomas pouco diferenciados de grau II em ratos carecas consistentes com carcinoma da mama. A linhagem celular MCF7 foi obtida de uma effusão pleural de um adenocarcinoma da mama e não é metastática. Estas linhagens celulares são reconhecidas na especialidade como modelos para o estudo dos cancros humanos da mama e do pulmão (ver, p. ex., Chandrasekaran et al., Cancer Res. (1979) 39:870; Gastpar et al., J Med Chem (1998) 41:4965; Ranson et al., Br J Cancer (1998) 77:1586; Kuang et al., Nucleic Acid Res (1998) 26:1116. As amostras de bibliotecas 15-20 são obtidas de dois pacientes diferentes (UC#2 e UC#3). As linhagens celulares GRRpz e WOca foram fornecidas pela Dr. Donna M. Peehl, Departamento de Medicina, Stanford University School of Medicine. A GRRpz foi obtida do epitélio normal da próstata. A linhagem celular WOca é uma linhagem celular de Grau 4 de Gleason.

Cada uma das bibliotecas é composta por uma colecção de clones de cDNA que por sua vez são representativos dos mRNAs expressos na fonte de mRNA indicada. De forma a facilitar a análise dos milhões de sequências em cada biblioteca, as sequências foram atribuídas a *clusters*. O conceito de "*cluster* de clones" é obtido a partir da escolha/agrupamento de clones de cDNA baseada no seu padrão de hibridação com um painel de sondas de oligonucleotídeos de aproximadamente 300 7bp (ver Drmanac et al., Genomics (1996) 37(1):29). Clones aleatórios de cDNA de uma biblioteca de tecidos são hibridados em condições de restringência moderada com 300 7bp oligonucleotídeos. Cada oligonucleotídeo tem alguma hibridação específica com um clone específico. A combinação de 300 dessas medidas de hibridação para 300 sondas dá a "assinatura de hibridação" para um clone específico. Clones com sequências semelhantes terão assinaturas de hibridação semelhantes. Ao desenvolver um algoritmo escolha/agrupamento para analisar essas assinaturas, grupos de clones numa biblioteca podem ser identificados e agrupados computacionalmente. Esses grupos de clones são designados por "*clusters*".

Dependendo da restringência da selecção do algoritmo (semelhante à restringência da hibridação no protocolo de rastreio de uma biblioteca de cDNA clássica), a "pureza" de cada *cluster* pode ser controlada. Por exemplo, artefactos de *clustering* podem ocorrer no *clustering* computacional tal como artefactos podem ocorrer num rastreio em "*wet-lab*" de uma biblioteca de cDNA com fragmentos de cDNA com 400 bp, até à mais elevada restringência. A restringência aqui utilizada na implementação de *clusters* fornece grupos de clones que são geralmente do mesmo cDNA ou de cDNAs proximamente relacionados.

Clones proximamente relacionados podem ser um resultado de clones de diferentes tamanhos do mesmo cDNA, proximamente relacionados clones de famílias de gene muito relacionadas, ou variantes de junção do mesmo cDNA.

A expressão diferencial para um *cluster* seleccionado foi conseguida determinando o número de clones de cDNA correspondente ao *cluster* seleccionado na primeira biblioteca (Clones na primeira), e determinando o número de clones de cDNA correspondente ao *cluster* seleccionado na segunda biblioteca (Clones na segunda). A expressão diferencial do *cluster* seleccionado na primeira biblioteca, relativamente à segunda biblioteca, é expressa como a "razão" de percentagem de expressão entre as duas bibliotecas. Geralmente, a "razão" é determinada por: 1) cálculo da percentagem de expressão do *cluster* seleccionado na primeira biblioteca por divisão do número de clones correspondente a um *cluster* seleccionado na primeira biblioteca pelo número total de clones analisado da primeira biblioteca; 2) cálculo da percentagem de expressão do *cluster* seleccionado na segunda biblioteca por divisão do número de clones correspondente a um *cluster* seleccionado na segunda biblioteca pelo número total de clones analisados da segunda biblioteca; 3) divisão da percentagem de expressão calculada da primeira biblioteca pela percentagem de expressão calculada da segunda biblioteca. Se o "número de clones" correspondente ao *cluster* seleccionado numa biblioteca é zero, o valor é ajustado a 1 para ajudar o cálculo. A fórmula utilizada no cálculo da razão entra em consideração com a "profundidade" de cada uma das bibliotecas em comparação, i.e., o número total de clones analisado em cada biblioteca.

Como resultado desta comparação de bibliotecas, foram identificados 17 polinucleotídeos, listados como SEQ ID NOS: 1, 3, 5, 7, 9, 11-13, 15, 16,18, 20, 22, 24, 26, 27 e 29 na Lista de Sequências anexa e sumariados na Tabela 2, como correspondentes a genes expressos diferencialmente em tecidos de pacientes de cancro do cólon. A Tabela 2 fornece: 1) o número de identificação da sequência ("SEQ ID NO do polinucleotídeo") atribuído a cada sequência para uso na presente especificação; 2) o número de identificação do *cluster* ("CLUSTER"); 3) o número de Identificação do Candidato; 4) o número CHIR (que serve de referência cruzada do oligos *antisense* discutidos abaixo), com, por exemplo CHIR7 tendo correspondência com os oligos CHIR7-2AS (*antisense*) e CHIR7-RC (controlo reverso); 5) o nome da sequência ("SEQ NOME") usado como identificador interno da sequência; 6) o nome atribuído ao clone do qual as sequências foram isoladas ("CLONE ID"); 7) o primeiro nucleotídeo dos codões de iniciação e de término de janelas de leitura aberta identificadas ("ORF de iniciação" e "ORF de término");; e 8) o número de identificação da sequência ("SEQ ID NO do polipeptídeo codificado") atribuído ao polipeptídeo codificado, quando apropriado. Porque os polinucleotídeos fornecidos representam transcrições parciais de mRNA, dois ou mais polinucleotídeos da invenção podem representar diferentes regiões da mesma transcrição de mRNA e o mesmo gene. Assim, se duas ou mais sequências são identificadas como pertencentes ao mesmo clone, qualquer dessas sequências pode ser usada para obter o mRNA ou gene completo.

Tabela 2. Identificação e Caracterização de Sequências Polinucleotídicas							
SEQ ID NO	CLUSTER	ID do Candidato	CHIR	NOME SEQ	ORF		SEQ ID NO do polipeptídeo codificado
					start	stop	
1	719	196	CHIR-7	SK1	21	396	2
3	9083	181	CHIR-8	SK2	219	693	4
5	115762	188	CHIR-16	SK5	5	1760	6
7	1665	195	CHIR-9	1665 longo	78	642	8
9	1665	195	CHIR-9	1665 curto	79	232	10
11	2334			SK8 parcial			
12	2334			SK8 completo			
13	3376	118	CHIR-11	SK19	79	376	14
15	376130			Junc2	181, 363, 731	361, 542, 911	
16	402380	202	CHIR-33	XD4	16	538	17
18	726682	198	CHIR-43	XD1	2	551	19
20	552930	174	CHIR-42	XD7	240	585	21
22	454001	161	CHIR-29	XD10	53	1700	23
24	378805	163	CHIR-31	XD11	10	400	25
26	374641	160	CHIR-32	374641 longo (Junc4)	33, 420	183, 615	
27	374641	160	CHIR-32	374641 curto (XD6)	324	519	28
29	374641	160	CHIR-32	374641 electrónico	40, 388	190, 583	

A Tabela 3 sumariza os polinucleotídeos que correspondem a genes expressos diferencialmente em tecido do cólon de um único paciente.

Tabela 3							
SEQ ID NO	CLUSTER	Clones (Bibl 15) Normais	Clones (Bibl 16) Tumoraís	Clones (Bibl 17) Alto Met	Tumoraís/Normais (Bibl 15/Bibl 16)	Alto Met/Normal (Bibl 17/Bibl 15)	Alto Met/Tumoraís (Bibl17/Bibl16)
1	719	0	20	27	20	27	1
3	9083	0	10	14	10	14	1
5	115662	0	6	7	6	7	1
7	1665	4	14	20	3.5	5	1
12	2334	0	6	1	6	1	0
13	3376	3	20	19	7	6	1
15	376130	0	9	15	9	15	2
16	402380	0	15	2	15	2	0

(continuação)

18	726682	0	52	0	52	0	0
20	552930	1	14	2	14	2	0
22	454001	0	8	13	8	13	2
26	378805	1	12	12	12	12	1
	374641	9	47	129	5	14	3

Exemplo 2: Análise e Caracterização dos Polinucleotídeos da Invenção

Vários dos polinucleotídeos fornecidos contêm uma ou mais janelas de leitura aberta (ORFs) putativas que codificam um produto genético. Os locais de iniciação e de término para essas ORFs encontram-se listados na Tabela 2.

SEQ ID NO:15 contém três ORFs. A primeira ORF compreende do nucleotídeo 181 ao nucleotídeo 361. A segunda ORF compreende do nucleotídeo 363 ao nucleotídeo 542. A terceira ORF compreende do nucleotídeo 731 ao nucleotídeo 911.

SEQ ID NO:26 contém uma sequência de inserção de 39 nucleotídeos (do nucleotídeo 269 ao nucleotídeo 307) e duas ORFs. A primeira ORF compreende do nucleotídeo 33 ao nucleotídeo 183. A segunda ORF compreende do nucleotídeo 420 ao nucleotídeo 615.

SEQ ID NO:29 é uma sequência electrónica de acordo com o resultado do 5'-RACE e contém duas ORFs. A primeira ORF compreende do nucleotídeo 40 ao nucleotídeo 190. A segunda ORF compreende do nucleotídeo 388 ao nucleotídeo 583.

Exemplo 3: Membros das Famílias de Proteínas

As traduções dos polinucleotídeos fornecidos foram alinhadas com perfis de aminoácidos que definem famílias de proteínas ou motivos comuns. Descobriu-se que vários dos polinucleotídeos da invenção codificam polipeptídeos com características de um polipeptídeo pertencente a uma família de proteínas conhecida (e

portanto representa novos membros dessas famílias de proteínas) e/ou contendo um domínio funcional conhecido. A semelhança entre a sequência problema e uma família de proteínas ou motivo foi determinada por (a) comparação da sequência problema como perfil e/ou (b) alinhamento da sequência problema com os membros da família ou motivo.

Cada um dos perfis de sucesso encontra-se descrito mais pormenorizadamente abaixo. A Tabela 4 fornece a SEQ ID NO correspondente aos polinucleotídeos fornecidos que codificam produtos genéticos com semelhança ou identidade com as sequências de perfil. A semelhança (forte ou fraca) também se encontra patente na Tabela 4. Os acrónimos dos perfis (fornecidos em parêntesis) são os utilizados na identificação do perfil nas bases de dados Pfam e Prosite. A base de dados Pfam pode ser acedida através de qualquer das seguintes URLs: <http://pfam.wustl.edu/index.html>; <http://www.sanger.ac.uk/Software/Pfam/>; e <http://www.cgr.ki.se/Pfam/>. A base de dados Prosite pode ser acedida em <http://www.expasy.ch/prosite/>.

Tabela 4. Perfis de Sucesso

SEQ ID NO	CLUSTER	Perfil	Descrição	Semelhança
1	719		Glicosil hidrolase	fraca
3	9083	ANK	<i>Ankyrin repeats</i>	forte
5	115762	7tm_1	7 receptor transmembranar (família da rodopsina)	fraca
11	2334	EFhand	EF-hand	forte
12	2334	Efhand	EF-hand	forte
15	376130		Protease/integrase endógena retrógrada	
16	402380	Rrm	Motivo de reconhecimento de RNA (domínio aka RRM, RBD ou RNP)	

Família da Glicosil hidrolase 5 (GLICOSIL HIDROL F5; Pfam Adesão No.

PS00659; PDOC00565). SEQ ID NO: 1 corresponde a um gene que codifica um polipeptídeo homólogo a polipeptídeos da família da glicosil hidrolase 5 (Henrissat Biochem. J. (1991) 280:309-316) (também como família da celulase A (Henrissat et al. Gene (1989) 81:83-95)). Os membros desta família participam na degradação da celulase e xilanos, e encontram-se geralmente em bactérias, fungos, e leveduras. O padrão de consenso para membros desta família é: [LIV]-[LIVMFYWGA](2)-[DNEQG]-[LIVMGST]-x-N-E-[PV]-[RHDNSTLIVFY] (onde E é um resíduo de centro activo putativo).

SEQ ID NO: 1 corresponde a um gene que codifica um membro de uma das famílias das glicosil hidrolases (Henrissat et al. Biochem. J. (1993) 293:781-788). Essas enzimas contêm pelo menos um resíduo conservado de ácido glutâmico (ou resíduo de ácido aspártico) o qual já se provou estar directamente envolvido na quebra da ligação glicosídica ao actuar como nucleófilo.

Repetições Ank (ANK; Pfam Adesão No. PF0023). SEQ ID NO:3 corresponde a um gene que codifica uma proteína contendo repetições Ank. O motivo anquirina é uma sequência de 33 aminoácidos que recebeu o nome da proteína *anquirina* que tem 24 motivos de 33 aminoácidos em paralelo. As repetições *ank* foram originalmente identificadas na proteína de controlo do ciclo celular *cdc10* (Breedon et al., Nature (1987) 329:651). Proteínas contendo repetições anquirina incluem anquirina, miotropina, proteínas *I-kappaB*, proteínas do ciclo celular *cdc10*, o receptor de Notch (Matsuno et al., Development (1997) 124 (21):4265); G9a (ou BAT8) da região de classe III do complexo de histocompatibilidade principal (Biochem J. 290:811-818, 1993), FABP, GABP, 53BP2, Lin12, glp-1, SW14, e SW16. As funções das

repetições anquirina são compatíveis com um papel nas interações proteína-proteína (Bork, família da glicosil hidrolase 5 (GLICOSIL HYDROL F5; Pfam Adesão No.

Sete Proteínas de Membrana Integral Transmembranares -- Família da Rodopsina (7tm 1: Pfam Adesão No. PF00001). SEQ ID NO:3 corresponde a um gene que codifica um polipeptídeo do receptor da rodopsina membro da família das sete proteínas transmembranares (7tm). Os receptores da (7tm) família da rodopsina acoplados à proteína G (também designados R7G) são um extenso grupo de hormonas, neurotransmissores, e receptores de luz que convertem sinais extracelulares por interacção com proteínas (G) de ligação ao nucleotídeo guanina (Strosberg A.D. Eur. J. Biochem. (1991) 196:1, Kerlavage A.R Curr. Opin. Struct. Biol. (1991) 1:394, Probst, et al., DNA Cell Biol. (1992) 11:1, Savarese, et al., Biochem. J. (1992) 283:1, <http://www.gcrdb.uthscsa.edu/>, <http://swift.embl-heidelberg.de/7tm/>. O padrão de consenso que contém o triplete conservado e que também abarca a maior parte da terceira hélice transmembranar é usado para detectar esta vasta família de proteínas: [GSTALIVMFYWC]-[GSTANCPDE]-{EDPKRH}-x(2)-[LIVMNQGA]-x(2)-[LIVMFT]-[GSTANC]-[LIVMFYWSTAC]-[DENH]-R-[FYWCSH]-x(2)-[LIVM].

EF Hand (EFhand: Pfam Adesão No. PF00036). SEQ ID NOS: 11 e 12 correspondem a genes que codificam uma proteína da família das proteínas EF-hand. Muitas proteínas ligadoras de cálcio pertencem à mesma família evolucionária e partilham um tipo de domínio de ligadores de cálcio conhecido por EF-hand (Kawasaki et al., Protein. Prof (1995) 2:305-490). Este tipo de domínio consiste numa ansa de doze resíduos ladeado em ambos os lados por um domínio de hélice alfa de doze resíduos. Numa ansa, o ião cálcio encontra-se coordenado numa configuração bipiramidal pentagonal.

Os seis resíduos envolvidos na ligação encontram-se nas posições 1, 3, 5, 7, 9 e 12; esses resíduos são designados por X, Y, Z, -Y, -X and -Z. A invariante Glu ou Asp na posição 12 fornece dois oxigénios para ligação ao Ca (ligando bidentado). O padrão de consenso inclui a ansa EF-hand completo bem como o primeiro resíduo que se segue à ansa e que parece ser sempre hidrofóbico: D-x-[DNS]-{ILVFYW}-[DENSTG]-[DNQGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-x(2)-[DE]-[LIVMFYW].

Protease/integrase retroviral Endógena. SEQ ID NO:15 corresponde a um gene que codifica um polipeptídeo contendo um domínio homólogo de um domínio de protease/integrase retroviral endógena humana de uma proteína pol retroviral.

Motivo de Reconhecimento de RNA (rrm; Pfam Adesão No. PF00076). SEQ ID NO:16 corresponde a um gene que codifica um motivo de reconhecimento de RNA, também designado um domínio RRM, RBD, ou RNP. Este domínio, com cerca de 90 aminoácidos de comprimento, está contido em proteínas eucarióticas que se ligam a RNA de cadeia simples (Bandziulis et al. Genes Dev. (1989) 3:431-437; Dreyfuss et al. Trends Biochem. Sci. (1988) 13:86-91). Duas regiões no seio do domínio de ligação ao RNA são altamente conservadas:

O primeiro é um segmento hidrofóbico de seis resíduos (designado motivo RNP-2), o segundo é um motivo octapeptídico (designado RNP-1 ou RNP-CS). O padrão de consenso é: [RK]-G-{EDRKHPCG}-[AGSCI]-[FY]-[LIVA]-x-[FYLM].

Exemplo 4: Detecção e Quantificação dos Polinucleotídeos da Invenção

Os polinucleotídeos da invenção foram detectados e quantificados em amostras de tecidos de pacientes por

transcriptase reversa PCR (RT-PCR). As amplificações de RNA completo foram feitas recorrendo ao sistema ciclotérmico LightCycler™ (Roche Diagnostics) numa reacção de PCR padrão contendo os *primers* fornecidos e o pigmento de ligação ao dsDNA SYBR Green I. A amplificação por PCR foi seguida por fluorescência do pigmento SYBR Green I, que fluoresce apenas quando ligado a DNA de cadeia dupla. A especificação dos produtos foi verificada por análise da curva de fusão.

Preparação Padrão. 1 µg de RNA completo de placenta humana (Clontech, Palo Alto, CA) foi transcrito de modo reverso a 42°C durante 1 hora e depois aquecido a 94°C durante 5 minutos num volume total de reacção de 20 µl (1st-Strand™ cDNA Synthesis Kit, Clontech). A mistura reaccional foi usada como um 1x matriz padrão. Foram então feitas diluições em série do 1x matriz padrão: $10^{-1}x$, $10^{-2}x$, $10^{-3}x$, $10^{-4}x$, $10^{-5}x$, $10^{-6}x$ template padrões.

Preparação de Amostras de RNA Completo. As amostras de tecido dos pacientes foram transportadas em reagente TRIZOL congelado. As amostras foram homogeneizadas em reagente TRIZOL. Adicionou-se então clorofórmio para isolar o RNA, seguido de precipitação do RNA com isopropanol. Os precipitados de RNA foram lavados com etanol a 75%, secos ao ar, depois dissolvidos em água destilada isenta de RNase. Antes da transcrição reversa, as amostras de RNA foram tratadas com DNase I (isenta de RNase) (2 U/µl, Ambion, Austin, TX) e limpas com o RNeasy Mini Kit (Qiagen, Santa Clarita, CA).

RT-PCR As amostras de RNA completo foram foi transcritas de modo reverso com *primer* oligo-dT18 (1st-Strand™ cDNA Synthesis Kit, Clontech). A PCR foi feita utilizando os seguintes *primers* específicos de genes:

SK1:	primer directo	5'- AGGAGTTTCTGAGGACCATGCAC -3'	(SEQ ID NO:30)
	primer reverso	5'- TCAAGGGTTGGGGATACACACG -3'	(SEQ ID NO:31)
SK2:	primer directo	5'- CTTGCTTGCTTTCTTCTCTGGC -3'	(SEQ ID NO:32)
	primer reverso	5'- AGTCTGGAAATCCACATGACCAAG -3'	(SEQ ID NO:33)
SK5:	primer directo	5'- CCCAATGAGGAACCTAAAGTTGC -3'	(SEQ ID NO:34)
	primer reverso	5'- GGTGCCAAATCTGGACTCTTGTC -3'	(SEQ ID NO:35)

(continuação)

1685:	primer directo	5'- GATCCATTTTCAGCAGTGCTCTG -3'	(SEQ ID NO:36)
	primer reverso	5'- CAGTGTTCACAGAAGGGGTACTCAC -3'	(SEQ ID NO:37)
SK8:	primer directo	5'- ACGAGAGCGACACGGACAAG -3'	(SEQ ID NO:38)
	primer reverso	5'- TCTGAGGCTGTGGCAGGTGC -3'	(SEQ ID NO:39)
SK19:	primer directo	5'- CCAGTCTTTGCCAACTCGTGC -3'	(SEQ ID NO:40)
	primer reverso	5'- TTCGATCTTCAAACCTGTGCCTTG -3'	(SEQ ID NO:41)
Junc2:	primer directo	5'- TTGGCAACCAGACCAGCATC -3'	(SEQ ID NO:42)
	primer reverso	5'- TTTCCCATAGGTGTGAGTGGCG -3'	(SEQ ID NO:43)
XD4:	primer directo	5'- GACTGGTGTTTTGTTCGGGGTC -3'	(SEQ ID NO:44)
	primer reverso	5'- TTTGTCCAAGGCTGCATGGTC -3'	(SEQ ID NO:45)
XD1:	primer directo	5'- TGCCCTGGTTAAGCCAGAAGTC -3'	(SEQ ID NO:46)
	primer reverso	5'- AGCTTCACTTTGGTCTTGACGG -3'	(SEQ ID NO:47)
XD7:	primer directo	5'- GGTCATCTGCATCAAGGTTGGC -3'	(SEQ ID NO:48)
	primer reverso	5'- GGTTCGTAACCGTGACTTCAGG -3'	(SEQ ID NO:49)
XD10:	primer directo	5'- GCATCCTTTTCCAGTCTTCCG -3'	(SEQ ID NO:50)
	primer reverso	5'- TGCAGCAAACATGCCTGAGC -3'	(SEQ ID NO:51)
XD11:	primer directo	5'- TGTTCACGAGCAAAGCATGTG -3'	(SEQ ID NO:52)
	primer reverso	5'- ATCCTTCTTCCACTCCCGCTTC -3'	(SEQ ID NO:53)
37841:	primer directo	5'- TCGGCTTGACTACACTGTGTGG -3'	(SEQ ID NO:54)
	primer reverso	5'- TACAAAGACCACTGGGAGGCTG -3'	(SEQ ID NO:55)
β -actin:	primer directo	5'- CGGGAAATCGTGCGTGACATTAAG -3'	(SEQ ID NO:56)
	primer reverso	5'- TGATCTCCTTCTGCATCCTGTCCG -3'	(SEQ ID NO:57)
GAPDH:	primer directo	5'- TTTGGCTACAGCAACAGGGTG -3'	(SEQ ID NO:58)
	primer reverso	5'- TGTGAGGAGGGGAGATTCAGTG -3'	(SEQ ID NO:59)

Foram utilizadas β -actina e INTERVALODH como controlos positivos. Todos os produtos da PCR são de 150-250 pb. Os 20 μ l

de mistura reaccional PCR em cada LightCycler™ capilar continham 2 µl de 10x tampão PCR 11, 3 mM de MgCl₂ (Perkin-Elmer, Foster City, Califórnia), 140 µM de dNTP, 1: 50000 de SYBR Green 1, 0.25 mg / ml de BSA, 1 unidade de polimerase Taq (Boehringer Mannheim, Indianapolis, IN), 0,175 µM de cada *primer*, 2 µl de mistura reaccional RT. A amplificação por PCR começou com 20 segundos de desnaturação a 95°C, seguidos de 45 ciclos de desnaturação a 95°C durante 5 segundos, emparelhamento a 60°C durante 1 segundo e extensão a 72°C durante 30 segundos. No fim do último ciclo, os produtos de PCR foram emparelhados a 60°C durante 5 segundos, depois lentamente aquecidos até 95°C a 0,2°C/segundo, para medir a curva de fusão dos produtos específicos da PCR. Todas as experiências foram realizadas em duplicado.

A análise de dados foi levada a cabo utilizando software LightCycler™ (Roche Diagnostics) com opções de quantificação e curva de fusão. A fluorescência é normalizada em relação a controlos positivos and negativos.

Sobreexpressão de genes em tecido completo de paciente com cancro do cólon. Os resultados fornecidos nas tabelas abaixo incluem dados de fluorescência para polinucleotídeos isolados de amostras de tecido do cólon colhidas directamente, não microdissecadas (i.e., tecido completo), e amplificadas por recurso aos *primers* indicados. Células de tipo normal, de tumor primário e metastáticas são designadas por N, PT e Met, respectivamente. A sobreexpressão foi determinada por comparação das células metastáticas ou de tumor primário, ou ambas, com células normais. Os resultados para cada gene correspondente aos *clusters* indicados na amostra em cada paciente encontram-se sumarizados nas tabelas abaixo. Todos os valores foram ajustados aos níveis relativos ao controlo de beta-actina.

Cluster#719 (SK1): Sobreexpressão detectada em 4 de 5 pacientes (67%)			
Pacientes	N	PT	MET
UC#1	0.022	0.117	0.364
UC#2	0.121	0.109	0.142
UC#4	0.083	0.053	0.078
UC#7	0.042	0.199	0.145
UC#8	0.215	0.515	0.794
UC#9	0.233	0.565	0.613

Cluster#9083 (SK2): Sobreexpressão em 3 de 4 pacientes (75%)			
Pacientes	N	PT	MET
UC#1	0.0021	0.0013	0.0078
UC#2	0.008	0.012	0.014
UC#4	0.0021	0.0022	0.0026
UC#7	0.0009	0.0021	0.0039

Cluster#115782 (SK5): Sobreexpressão em 5 de 6 pacientes (83%)			
Pacientes	N	PT	MET
UC#1	0.0053	0.0159	0.044
UC#2	0.0185	0.0174	0.0269
UC#4	0.022	0.033	0.034
UC#7	0.013	0.028	0.025
UC#8	0.0275	0.105	0.143
UC#9	0.0336	0.0595	0.0541

Cluster#1685: Sobreexpressão em 4 de 6 pacientes (67%)			
Pacientes	N	PT	MET
UC# 1	0.00006	0.0003	0.002
UC#2	0.0015	0.001	0.0012
UC#4	0.0016	0.0013	0.0016
UC#7	0.00003	0.0003	0.0012
UC#8	0.0016	0.0122	0.0154
UC#9	0.006	0.057	0.097

Cluster#2334 (SK8): Sobreexpressão em 4 de 6 pacientes (67%)			
Pacientes	N	PT	MET
UC# 1	0.011	0.022	0.017

(continuação)

Cluster#2334 (SK8): sobreexpressão em 4 de 6 pacientes (67%)			
Pacientes	N	PT	MET
UC#2	0.0266	0.0317	0.026
UC#4	0.02	0.006	0.01
UC#7	0.046	0.093	0.042
UC#8	0.042	0.168	0.472
UC#9	0.208	0.322	0.29

Cluster#3376 (SK19): sobreexpressão em 4 de 6 pacientes (67%)			
Pacientes	N	PT	MET
UC#1	0.00018	0.00042	0.0012
UC#2	0.002	0.0025	0.0016
UC#4	0.0013	0.0012	0.002
UC#7	0.00024	0.00055	0.00062
UC#8	0.0003	0.00127	0.0023
UC#9	0.001	0.0075	0.009

Cluster#376130 (Junc2): sobreexpressão em 3 de 4 pacientes (75%)			
Pacientes	N	PT	MET
UC#1	0.00871	0.0111	0.0142
UC#2	0.000567	0.00663	0.0163
UC#4	0.000107	0.00048	0.000237
UC#7	0.0000401	0.000259	0.00159

Cluster#402380 (XD4): sobreexpressão em 2 de 4 pacientes (50%)			
Pacientes	N	PT	MET
UC# 1	0.0763	0.123	0.2
UC#2	0.0867	0.0629	0.069
UC#4	0.0735	0.0672	0.0664
UC#7	0.0559	0.112	0.139

Cluster#726682 (XD1): sobreexpressão em 0 de 4 pacientes (0%)			
Pacientes	N	PT	MET
UC#1	0.0879	0.0822	0.136
UC#2	0.175	0.124	0.147
UC#4	0.2	0.145	0.145

(continuação)

Cluster#726682 (XD1): sobreexpressão em 0 de 4 pacientes			
Pacientes	N	PT	MET
UC#7	0.108	0.144	0.114

Cluster#552930 (XD7): sobreexpressão em 1 de 4 pacientes (25%)			
Pacientes	N	PT	MET
UC# 1	0.018	0.019	0.0902
UC#2	0.204	0.161	0.212
UC#4	0.299	0.25	0.238
UC#7	0.246	0.409	0.248

Cluster#454001 (XD10): sobreexpressão em 2 de 4 pacientes)			
Pacientes	N	PT	MET
UC#1	0.0197	0.0363	0.0587
UC#2	0.0514	0.0451	0.069
UC#4	0.0587	0.0889	0.096
UC#7	0.0342	0.1	0.0705

Cluster#378805 (XD11): sobreexpressão em 1 de 4 pacientes)			
Pacientes	N	PT	MET
UC#1	0.00117	0.00269	0.00697
UC#2	0.00864	0.00371	0.00672
UC#4	0.0098	0.00525	0.00497
UC#7	0.00912	0.00989	0.0127
Cluster#374641: sobreexpressão em 3 de 4 pacientes (75%)			
Pacientes	N	PT	MET
UC# 1	0.0124	0.163	0.0947
UC#2	0.28	0.317	0.544
UC#4	0.685	1.809	1.996
UC#7	0.569	1.714	1.073

Sobreexpressão de genes no epitélio de pacientes com cancro do cólon. Os resultados fornecidos nas tabelas abaixo incluem dados de fluorescência para polinucleotídeos isolados de células epiteliais do cólon preparadas pelo método de sacudir o epitélio para obter epitélio com > 97% de pureza sem estroma. As células de tipo normal, pré-canceroso (pólipo adenomatoso), e de tumor primário são designadas por N, polyp e PT, respectivamente. A sobreexpressão foi determinada por comparação com células de tumor primário ou células pré-cancerosas, ou ambas, com células normais. Todos os valores foram ajustados aos níveis relativos ao controlo de beta-actina.

Cluster#719 (SK1): sobreexpressão em 4 de 4 pacientes (100%)			
Pacientes	N	Polyp	PT
UW#17	0.0924	0.117	N/A
UW#18	0.0864	N/A	0.327
UW#19	0.151	N/A	0.227
UW#20	0.0624	0.162	0.164

Cluster#115762 (SK5): sobreexpressão em 4 de 4 pacientes (100%).			
Pacientes	N	Polyp	PT
UW#17	0.00724	0.0122	N/A
UW#18	0.0156	N/A	0.111
UW#19	0.0158	N/A	0.0461
UW#20	0.00728	0.0187	0.0306

Cluster#1665: sobreexpressão em 4 de 4 pacientes (100%)			
Pacientes	N	Polyp	PT
UW#17	0.0041	0.0306	N/A
UW#18	0.0029	N/A	0.0357
UW#19	0.0045	N/A	0.0357
UW#20	0.0028	0.025	0.047

Cluster#2334 (SK8) sobreexpressão em 1 de 4 pacientes (25%)			
Pacientes	N	Polyp	PT
UW#17	0.1835	0.041	N/A
UW#18	0.0638	N/A	0.0927
UW#19	0.04	N/A	0.04
UW#20	0.2236	0.0576	0.0454

Cluster#3376 (SK19) sobreexpressão em 4 de 4 pacientes (100%)			
Pacientes	N	Polyp	PT
UW#17	0.0053	0.012	N/A
UW#18	0.0028	N/A	0.0084
UW#19	0.003	N/A	0.0135
UW#20	10.0023	10.023	10.012

Exemplo 5: Análise de Northern Blot

A expressão genética diferencial nas células cancerosas do cólon pode ser ainda confirmada por outras técnicas, tal como a análise *Northern blot*. A análise *Northern* pode ser feita por métodos bem conhecidos na especialidade. Em suma, 5 mg/ml de DNA de cadeia simples do esperma desnaturado em tampão *rapid-Hyb* (Amersham Life Science, Little Chalfont, England) é pré-aquecido a 65°C e *blots* de RNA completo (Invitrogen, Carlsbad, CA) de tumor do cólon humano são pré-hibridados no tampão com agitação a 65°C durante 30 minutos. Sondas de DNA (50 ng por reacção) específicas para genes marcadas com [α -32P]dCTP (3000Ci/mmol, Amersham Pharmacia Biotech Inc., Piscataway, NJ) (Prime-It RmT Kit, Stratidadene, La Jolla, CA) e purificadas com Micro Colunas SondaQuant™ G-50 (Amersham Pharmacia Biotech Inc.) são adicionadas e hibridadas com os *blots* sob agitação a 65°C durante a noite. Os *blots* são lavados em 2x SSC, 0.1%(w/v) SDS à temperatura ambiente durante 20 minutos, duas vezes em 1x SSC, 0.1%(w/v) SDS a 65°C durante 15 minutos, e depois expostos a *Hyperfilms* (Amersham Life Science).

Exemplo 6: Análise da expressão do gene correspondente a SK2 (*cluster* 9083 (c9083)) (SEQ ID NO:3) no Carcinoma colo-rectal

A expressão do gene contendo a sequência de SK2, cujos *clusters* do *cluster* da i.d. no. 9083, foi examinado por PCR quantitativo em várias linhagens celulares cancerosas, incluindo algumas linhagens celulares de carcinoma colo-rectal. As células cuja expressão foi testada encontram-se sumarizadas abaixo.

Linhagem Celular	Fonte do Tecido	Linhagem Celular	Fonte do Tecido
MDA-MB-231	Mama humana; elevado potencial metastático (micromets no pulmão; adenocarcinoma; efusão pleural)	Caco-2	Adenocarcinoma colo-rectal humano
MDA-MB-435	Mama humana; elevado potencial metastático (macrometastases no pulmão)	SW620	Adenocarcinoma colo-rectal humano; de local metastático (nódulo linfático)
MCF-7	Mama humana; não metastático	LS 147T	Adenocarcinoma colo-rectal humano de elevado potencial metastático
MDA-MB-468	Mama humana; adenocarcinoma	LOVO	Adenocarcinoma colo-rectal humano; cólon; de local metastático (cólon)
Alab	Mama humana; metastático	HT29	Adenocarcinoma colo-rectal humano; cólon;
SKOV3	Adenocarcinoma do ovário humano	SW480	Adenocarcinoma colo-rectal humano; cólon;
OVCAR3	Adenocarcinoma do ovário humano	HCT116	Adenocarcinoma colo-rectal humano; cólon;
KM 12C	Cólon humano; baixo potencial metastático	Colo 320DN	Adenocarcinoma colo-rectal humano; cólon;
KM12L4	Cólon humano; elevado potencial metastático (obtido de Km12C)	T84	Adenocarcinoma colo-rectal humano; cólon; de local metastático (pulmão)
DU145	Próstata humana; carcinoma; do local metastático: cérebro	HCT 15	Adenocarcinoma colo-rectal humano; cólon;
HT 1080	Linhagem celular do sarcoma humano	CCD 112	Adenocarcinoma colo-rectal humano, baixo potencial metastático
HMVEC	Células microvasculares endoteliais primárias humanas	DLD1	Cólon humano; Adenocarcinoma colo-rectal
185B4	Células epiteliais normais da mama; quimicamente transformadas	293	Células epiteliais do rim
LNCAP	Carcinoma da próstata; metástase no linfo supraclavicular esquerdo	GRDP	Epitélio primário da próstata
U373MG	Célula de glioblastoma	IMR90	Fibroblasto primário do pulmão
WOCA	Epitélio primário da próstata	PC3	Cancro da próstata; receptor negativo do androgénio

O PCR quantitativo em tempo real foi feito por isolamento de RNA de células recorrendo a um Kit de Isolamento de RNA Roche, de acordo com as instruções do fabricante. Usou-se um micrograma de RNA para sintetizar uma primeira cadeia de cDNA recorrendo a transcriptase reversa (Ambião) MMLV usando o tampão do fabricante e as concentrações recomendadas de oligo dT, nucleotídeos, e *Rnasin*. Esta primeira cadeia de cDNA serviu como matriz para o PCR quantitativo em tempo real por recurso ao por recurso ao Roche *light-cycler*, como recomendado no manual do equipamento. O gene correspondente a SK2 (C9083) (SEQ ID NO:3) foi amplificado com o *primer* directo: 5'-cgctgacctcaaccag-3' (SEQ ID NO:60) *primer* reverso: 5'-ctgtttgccggttcttattac-3' (SEQ ID NO:61). O produto foi quantificado com base no ciclo em que a amplificação entrou na fase linear de amplificação em comparação com um padrão interno e utilizando o software fornecido pelo fabricante. Pequenas diferenças em partes ou na totalidade da matriz na reacção da primeira cadeia de cDNA foram eliminadas por normalização à quantidade de actina amplificada numa reacção quantitativa de PCR separada usando o *primer* directo 5'-CGGGAAATCGTGCGTGACATTAAG-3' (SEQ ID NO:56) e o *primer* reverso: 5'-TGATCTCCTTCTGCATCCTGTCTGG-3' (SEQ ID NO:57). Os resultados são mostrados na Fig. 1.

Exemplo 7: Análise Funcional do gene correspondente a SK2 (c9083) (SEQ ID NO:3)

De modo a melhor compreender o papel do gene correspondente a SK2 (c9083) (SEQ ID NO:3), a informação funcional no gene correspondente a esta sequência foi obtida utilizando tecnologia de eliminação *antisense*. Em suma, o tipo de célula a ser testado, células SW620 ou HT1080, que expressam o polipeptídeo codificado

pelo gene correspondente a c9083, foi plaqueada até aproximadamente 60-80% de confluência numa placa de 6 poços ou, para ensaios de proliferação, de 96 poços. O oligonucleotídeo *antisense* ou reverso de controlo foi diluído a 2 μ M em optiMEM e adicionado ao optiMEM, no qual o veículo de distribuição, lipitóide 116-6 no caso das células SW620 ou 1:1 lipitóide 1:colesteróide 1 no caso de células HT1080, tinha sido diluído. A mistura oligo/veículo de distribuição foi então voltada a diluir no meio com soro por cima das células. A concentração final de oligonucleotídeo em todas as experiências foi de 300 nM, e a razão final de oligo para veículo de distribuição em todas as experiências foi 1.5 nmol lipitóide/ μ g de oligonucleotídeo. As células foram transfectadas durante a noite a 37°C e a mistura de transfecção substituída por meio fresco na manhã seguinte.

Foram testados os seguintes oligonucleotídeos *antisense* relativamente à capacidade de empobrecer RNA c9083 (SEQ ID NO:3):

Nome do Oligon.	Sequência	Nucleótidos
CHIR-8-4AS C9083.P0463	ATTTGGGCATCACTGGCTACAAGCA (SEQ ID NO:64)	25
CHIR-8-4RC C9083.P0463RC	ACGAACATCGGTCCTACTACGGGTTTA (SEQ ID NO:65)	25
CHIR-8-5AS C9083.P0157	CAGAGAGGTGAGACACTCGCCGCA (SEQ ID NO:66)	24
CHIR-8-5RC C9083.P0157RC	ACGCCGCTCACAGAGTGGAGAGAC (SEQ ID NO:67)	24
RC: oligos de controlo reverso (oligos de controlo); AS: Oligos antisense (teste)		

O efeito do oligonucleotídeo nas células foi determinado por quantificação dos níveis de PCR como descrito acima, e em ensaios de proliferação usando quantidades de DNA, tal como quantificado com o Stratidadene Quantos™ kit para determinar o número de células.

Os resultados do nível de quantificação do mRNA são mostrados na Fig. 2. Os efeitos dos oligonucleotídeos sobre a proliferação, durante um período de quatro dias, são mostrados

nas Figs. 3 e 4. Células sem tratamento por oligonucleotídeos (WT) serviram de controlo. O oligo CHIR-8-4AS foi o mais eficiente na diminuição do mRNA para o gene correspondente a 9083c. A transfecção desses oligos em células SW620 resultou numa diminuição da velocidade de proliferação relativamente a oligos de controlo reverso emparelhados, sendo o CHIR 8-4 algo mais eficiente do que o CHIR-8-5 (Fig. 3). Significativamente, o mesmo oligonucleotídeo *antisense* não teve qualquer efeito no crescimento numa linhagem celular de fibrosarcoma, HT1080 (Fig. 4). Isto indica que o papel funcional do gene correspondente a c9083 é específico para tecidos, e ainda que o gene correspondente a c9083 tem um efeito específico no crescimento.

Os oligos foram seguidamente testados relativamente ao efeito na formação colónias num ensaio em agar mole. Os ensaios em agar mole começaram pela formação de uma camada inferior de 2 ml agar a 0.6%, em meio plaqueado de fresco, no espaço de algumas horas de camada sobre as células. A camada celular foi formada na camada inferior por remoção das células transfectadas como descrito acima (ou um oligo *antisense* k-Ras como controlo positivo), CHIR-8-4, CHIR-8-5, CHIR-8-4RC, ou CHIR-8-SRC) de placas usando 0.05% de tripsina e lavando duas vezes com meio. As células foram contadas num contador Coulter, e ressuspensas a 10^6 por ml em meio. Aliquotas de 10 μ l foram colocadas com meio em placas de 96 poços (para verificação da contagem com WST1), ou mais diluídas para ensaios posteriores em agar mole. 2000 células são plaqueadas em 800 μ l de agar a 0.4% em poços duplicados acima da camada do fundo de agar a 0.6%. Após a solidificação do agar da camada celular, 2ml de meio são vertidos por cima e adicionado oligo de controlo *antisense* ou reverso sem veículos de distribuição. Foram adicionados meio fresco e oligos a cada 3-4

dias. Formaram-se colónias entre 10 dias e 3 semanas. Os campos de colónias foram contados a olho. Os valores do metabolismo de WST-1 podem ser usados para compensar pequenas diferenças no número inicial de células.

Os campos maiores podem ser digitalizados para registo visual das diferenças.

Os dois oligos *antisense* CHIR-8-4 e CHIR-8-5 promoveram um decréscimo no tamanho e número de colónias relativamente aos oligos de controlo CHIR-8-4RC e CHIR-8-SRC. Esses resultados validam ainda o gene correspondente a c9083 (SEQ IDNO:3) como alvo para intervenção terapêutica.

Exemplo 8: Efeito dos oligonucleotídeos *antisense* nos níveis de mensagem dos genes alvo

O efeito dos oligonucleotídeos *antisense* sobre os níveis de mensagem dos genes, correspondentes às sequências e *clusters* aqui descritas, foi analisado usando tecnologia de eliminação *antisense*, como descrito para c9083 no Exemplo acima. Especificamente, oligos *antisense* para genes correspondentes a c719, c1665, c3376, c115762, c454001, c3788805, e c776682 foram preparados como descrito acima. Uma vez sintetizados e quantificados, os oligómeros foram rastreados relativamente à eficiência de uma eliminação de transcrito num painel de uma linhagem de células cancerosas. A eficiência da eliminação foi determinada por análise dos níveis de mRNA usando quantificação por *Lightcycler*. Os oligómeros que obtiveram os níveis mais elevados de eliminação de transcrito, nos quais esses níveis foram de pelo menos cerca de 50%, de preferência cerca de 80-90%, até 95% ou mais até uma mensagem não detectável, foram seleccionados para uso num ensaio de proliferação baseado em

célula, um ensaio de ancoragem de crescimento independente, e um ensaio de apoptose.

As células SW620, que expressam o polipeptídeo codificado pelos genes correspondentes em análise, foram plaqueadas a aproximadamente 60-80% confluência numa placa de 6 poços ou, para ensaios de proliferação, 96 poços. Para cada mistura da transfecção, foi preparada uma molécula de transporte, de preferência um lipitóide ou colesteróide, numa concentração de 0.5 mM em água, levada aos ultra-sons para originar uma solução uniforme, e filtrada através de uma membrana de PVDF de 0.45 μ m. O oligonucleotídeo *antisense* ou de controlo foi então preparado numa concentração de trabalho de 100 μ M em água Millipore estéril. O oligonucleotídeo foi ainda diluído em OptiMEM™ (GibcoBRL), num tubo de microcentrífuga, a 2 μ M, ou aproximadamente 20 μ g oligo/ml de OptiMEM™. Num tubo de microcentrífuga separado, o lipitóide ou colesteróide, tipicamente na quantidade de cerca de 1.5-2 nmol lipitóide/ μ g de oligonucleotídeo *antisense*, foi diluído no mesmo volume de OptiMEM™ usado na diluição do oligonucleotídeo. O oligonucleotídeo *antisense* diluído foi imediatamente adicionado ao lipitóide diluído e misturado por pipetagem. O oligonucleotídeo foi adicionado às células numa concentração final de 30 nM.

O nível de mRNA alvo correspondente a um gene alvo de interesse nas transfectadas células foi quantificado nas linhagens de célula cancerosas usando o equipamento de PCR em tempo real Roche LightCycler™. Os valores para o mRNA alvo foram normalizados versus um controlo interno (p. ex., beta-actina). Para cada reacção de 20 μ l, o RNA extraído (normalmente um total de 0.2-1 μ g) foi colocado num tubo de microcentrífuga estéril de

0.5 ou 1.5 ml, e foi adicionada água até um volume total de 12.5 µl. A cada tubo foram adicionados 7.5ml de uma mistura tampão/enzima, preparada por mistura (na ordem listada) de 2.5 µl de H₂O, 2.0 µl de 10X tampão de reacção, 10 µl de oligo dT (20 pmol), 1.0 µl de mistura de dNTP (10 mM cada), 0.5 µl de RNAsin® (20u) (Ambion, Inc., Hialeah, FL), e 0.5 µl de transcriptase reversa MMLV (50u) (Ambion, Inc.). O conteúdo foi misturado por pipetagem, e a mistura reaccional incubada a 42°C durante 1 hora. O conteúdo de cada tubo foi centrifugado antes da amplificação.

Preparou-se uma mistura de amplificação pela seguinte ordem: tampão II 1X PCR, 3 mM em MgCl₂, 140 µM em cada dNTP, 0.175 pmol em cada oligo, dil 1:50,000 de SYBR® Green, 0.25 mg/ml de BSA, 1 unidade de polimerase Taq, e H₂O até 20 µl. (O tampão II PCR encontra-se disponível numa concentração 10X na Perldn-Elmer, Norwalk, CT). Na concentração 1X contém 10 mM de Tris pH 8.3 e 50 mM de KCl. SYBR® Green (Molecular Probes, Eugene, OR) é um pigmento que fluoresce quando ligado a DNA de cadeia dupla. Como o produto de PCR de cadeia dupla é produzido durante a amplificação, a fluorescência do SYBR® Green aumenta. A cada aliquota de 20 µl de mistura de amplificação, foram adicionados 2 µl de matriz RT, e a amplificação levada a cabo de acordo com os protocolos padrão.

Os seguintes oligonucleotídeos *antisense* foram testados relativamente à capacidade de empobrecer os níveis de mensagem do gene correspondente ao *cluster* indicado. Gene Alvo: Localização do Oligo fornece o nome do *cluster* ao qual o gene alvo é atribuído e o nome do oligo usado. AS indica *antisense*; RC indica controlo reverso. Os dados para os genes correspondentes a c9083 são fornecidos para comparação.

Gene Alvo: Localização do Ôlgo	Sequência do Ôlgo	SEQ ID NO:	% KO da Mensagem
c719-1-AS	TTGGTGTCAATTGGGTCAAGGGTTGG	68	88%
C719-1-RC	GGTTGGGAAGTGGTTACTGTGGTT	69	
c719-2-AS	ACAGGGCAGATACGGACCTCGGTG	70	93%
c719-2-RC	GTGGCTCCAGGCATAGACGGGACA	71	
c719-3-AS	TTGTGGGTAAAGCAGTTTCATGTCCG	72	87%
c719-3-RC	CGCTGTACTTTGACGAATGGGTGTT	73	
c719-4-AS	CCTGGATCAGACGCAAGTTATCGCC	74	89%
c719-4-RC	CGGCTATTGAACGCAGACTAGGTCC	75	
C9083-4-AS	ATTGGGCATCACTGGGTACAAGCA	84	83.0
C9083-4-RC	ACGAACATCGGTCACTACGGTTTA	85	
C9083-5-AS	CAGAGAGGTGAGACACTCGCCGCA	86	73.0
C9083-5-RC	ACGCCGCTCACAGAGTGGAGAGAC	87	
C1685-1-AS	CTACTCCGCACACTTCATCGGCAGG	76	73.0
C1685-1-RC	GGACCGCTACTTCACACCCCTCATC	77	
C1685-2-AS	CTCTTGATACTCCAGCGGCAAAACCA	78	81.0
C1685-2-RC	ACCAAAAGGCGACCTCATAGTTCTC	79	
c9378-1-AS	GCGGCCAAGGCGTTCTGTTCTTAAG	80	78.0
c9378-1-RC	GAATTCTTGCCTTGGCGAACCCTCG	81	
c9378-2-AS	CCAGGTAGGCAGGAGTTGGCAAAGA	82	87.0
c9378-2-RC	AGAAACGGTTGAGCACGGATGGACC	83	
c9378-3-AS	GCCATTGAAGATGCCACAGATCCAC	84	99.0
c9378-3-RC	CACCTAGACCCGTAGAAATTACCG	85	
c9378-4-AS	CCTGCGTTTGTCCTCCAGCATCT	86	93.0
c9378-4-RC	TCTACGACCTGCCCTGTTTGGGTCC	87	
c9378-5-AS	AAGTCACAGTCCCGGATACACAGTC	88	86.0
c9378-5-RC	CTGACCATAGGCCCCCTGACACTGAA	89	
c115762-1-AS	TTGTGCGCTTGGCGAGGCATAAAACG	90	97.5
c115762-2-AS	TCTGGTCATCAACTTGGTTTCCGTG	91	99.0
c115762-3-AS	CAGTGTGTTGTGGTGTGCTGTGTGG	92	98.0
c115762-4-AS	GCTCAGCATCGGGGACCAAGCA	93	97.0
c115762-5-AS	TGAGAGACAGTGTGTTGTTGTTGTGC	94	93.0
454001-1-AS	TGCCCTTCACACGCTTGTTATCTTC	95	0
454001-2-AS	GACAACATCGGAGGCTTCAATCACG	96	0
454001-3-AS	GTTGAGGCTCTGAACACCACTGTTG	97	0
454001-4-AS	GTTTGGGAGCAGCTTCAACATTTGG	98	87
454001-5-AS	AGCAGTTTGGCAGCACCTTCAACA	99	92
454001-1-RC	CTTCTATTGGTTGSCACACTTCCGT	100	
454001-2-RC	CCACTAACTTCGGAGGCTACAACA	101	
454001-3-RC	GTTGTACCCACAAGTCTCGGAGTTG	102	
454001-4-RC	GGTTTACAACCTTCAACAGCGTTTG	103	
454001-5-RC	ACAACCTTCCAGGAGGTTTGAACGA	104	
378805-1-AS	ATCTGGCATGGACGGATGAGCGAA	105	41.0
378805-2-AS	GCTGGGTGGTTTCCGAACTCAACG	106	97
378805-3-AS	GTCCCAATCACCTTCCCAACAATCG	107	85.0
378805-4-AS	TCAGATCCTTCTTCCACTCCGCTT	108	100.0
378805-5-AS	TGCTCGTGGAAACAGGTAAAGCTCTG	109	98
378805-1-RC	AAGCGAGTAGGCAGGTACGGTCTA	110	

(continuação)

Gene Alvo: Localização do Oligo	Sequência do Oligo	SEQ ID NO:	% KO da Mensagem
378805:2-RC	GCAACTCAAGCCTTTGGTGGGTCG	111	
378805:3-RC	CCTAACACCCCTTCCACTAACCCTG	112	
378805:4-RC	TTGGCCCTCACCCTTCTTCTAGACT	113	
378805:5-RC	GTCTCGAAATGGACAAGGTGCTCGT	114	
776682:1-AS	AGCTTCACCTTTGGTCTTGACGGCAT	115	81
776682:2-AS	CGGAGGGAAGTCAAGTCAGCCACA	116	80
776682:3-AS	CGGCATTACCCCTCTCCAGCACCT	117	89
776682:4-AS	CCTCCACCTGTTTGCGGGCTTCC	118	81
776682:5-AS	CCACATTGAGGGAGTCCTCTTGCAA	119	80
776682:1-RC	TACGGCAGTTCTGGTTTCACTTCGA	120	
776682:2-RC	ACACCGACTGAAGTGAAGGGAGGC	121	
776682:3-RC	TCCACGACCTCTCCACTTACGGC	122	
776682:5-RC	CCTTCGGGGGTTTGTCCACCTCC	123	
402380:P464:4-AS	CGCCGAACAAAACAGTCAACG	124	94
402380:P464:4-RC	GCAACTGACCACAAAACAAGCCCC	125	
402380:P414:5 AS	GGCCATTGAGTCCCTCCATAGCAGC	126	82
402380:P414:5-RC	CGACGATACCTCCCTGAGTTACGGG	127	

O efeito do oligonucleotídeo nas células foi determinado por quantificação dos níveis de PCR. Os resultados da quantificação do nível de mRNA encontram-se sumarizados na tabela imediatamente acima.

O efeito da perda de mensagem para cada gene acima pode ser determinada ensaios de base celular como descrito no Exemplo 7 acima. Tal utilização do oligonucleotídeo *antisense* descrita por SEQ ID NO: 108 resultou numa inibição da proliferação de células SW620 quando usadas como descrito nos protocolos de ensaios de transfecção e proliferação do Exemplo 7 (Fig. 5).

Exemplo 9: O Efeito da Expressão dos Genes Correspondentes a c3376 e 402380 sobre a Proliferação

O efeito da expressão dos genes correspondentes a c3376 (gene correspondente a SEQ ID NO: 13) e 402380 (gene

correspondente a SEQ ID NO: 16), na inibição da proliferação celular, foi determinado em células SW620 do carcinoma coloretal do cólon.

As células foram plaqueadas a aproximadamente 60-80% confluência em placas de 96 poços. O oligonucleotídeo antisense ou reverso de controlo foi diluído a 2 μ M em OptiMEM™ e adicionado a OptiMEM™ no qual o veículo de distribuição, lipitóide 116-6 no caso de células SW620 ou 1:1 lipitóide 1:colesteróide 1 no caso de células MDA-MB-231, tinha sido diluído. A mistura oligo/veículo de distribuição foi então diluída de novo em meio com soro sobre as células. A concentração final de oligonucleotídeo em todas as experiências foi de 300 nM, e a razão final de oligo para veículo de distribuição para todas as experiências foi de 1.5 nmol de hpitoid/ μ g de oligonucleotídeo.

Os oligonucleotídeos *antisense* foram preparados como descrito acima. As células foram transfectadas durante a noite a 37°C e a mistura de transfecção foi substituída por meio fresco na manhã seguinte. A transfecção foi feita como descrito acima no Exemplo 8. A proliferação foi medida por recurso ao reagente colorimétrico WST-1, de acordo com métodos bem conhecidos na especialidade. Os resultados das experiências *antisense* são mostrados nas Figs. 6-9. Os valores no eixo dos y representam unidades de fluorescência relativa. Os oligos *antisense* e reverso de controlo de K-Ras serviram de controlo para demonstrar que o ensaio decorreu como esperado (Fig. 6).

Exemplo 10: Efeito da Expressão Genética na Formação de Colónias em Agar Mole

O efeito da expressão do gene correspondente a 402380 (gene correspondente a SEQ ID NO: 16), sobre a formação de colónias de células SW620, foi testado num ensaio em agar mole. Os ensaios em agar mole começaram pela formação de uma camada inferior de 2 ml agar a 0.6% em meio plaqueado de fresco no espaço de algumas horas de camada sobre as células. A camada celular foi formada na camada inferior por remoção das células transfectadas como descrito acima de placas usando 0.05% de tripsina e lavando duas vezes com meio. As células foram contadas num contador Coulter, e ressuspensas a 10^6 por ml em meio. Aliquotas de 10 μ l foram colocadas com meio em placas de 96 poços (para verificação da contagem com WST1), ou mais diluídas para ensaios posteriores em agar mole. São plaqueadas 2000 células em 800 μ l de agar a 0.4% em poços duplicados acima da camada do fundo de agar a 0.6%. Após a solidificação do agar da camada celular, 2 ml de meio são vertidos por cima e adicionado oligo de controlo *antisense* ou reverso (produzido como descrito acima) sem veículos de distribuição. Foram adicionados meio fresco e oligos a cada 3-4 dias. Formaram-se colónias entre 10 dias e 3 semanas. Os campos de colónias foram contados a olho. Os valores do metabolismo de WST-1 foram usados para compensar pequenas diferenças no número inicial de células. Os campos maiores podem ser digitalizados para registo visual das diferenças.

Os resultados são mostrados na Fig. 9. O eixo dos y representa o número de células para um sector definido, usando WST-1 para facilitar a contagem celular e normalizada em relação a um controlo. Os oligos *antisense* e reverso de controlo de K-Ras

(kRAS 2576-as e kRAS 2576-rc) serviram de controlos para demonstrar que o ensaio decorreu como esperado.

Exemplo 11: Efeito da Expressão Genética sobre a Morte Celular

O efeito da expressão dos genes correspondentes ao *cluster* 719 (gene correspondente a SEQ ID NO:1, CHIR-7); *cluster* 9083 (gene correspondente a SEQ ID NO:3, CHIR-8); *cluster* 1665 (gene correspondente a SEQ ID NOS:7 e 9, CHIR-9); *cluster* 3376 (gene correspondente a SEQ ID NO: 13, CHIR 11); *cluster* 115762 (gene correspondente a SEQ IDNO:5, CHIR 16); e *cluster* 402380 (gene correspondente a SEQ ID NO:16, CHIR-33) sobre a morte celular num ensaio de citotoxicidade da lactato desidrogenase (LDH) foi examinado em células HT1080 (uma linhagem celular de fibrosarcoma humano), células SW620, e linhagens celulares de células metastáticas de cancro da mama (MDA-MB-231 ("231")). O ensaio de citotoxicidade da lactato desidrogenase (LDH) é essencialmente o seguinte:

O ensaio de citotoxicidade da lactato desidrogenase (LDH) foi levado a cabo essencialmente da forma que se segue:

Dia 1: As células foram cultivadas em 4 placas separadas de 96 poços, tipicamente 5000 células/poço e incubadas a 37°C e 5% de CO₂.

Dia 2: As células foram transfectadas com o *antisense* assim como os complementos de controlo reverso, essencialmente como descrito no Exemplo 4. Uma placa (dia 0) foi deixada não transfectada como controlo de cultura.

A transfecção foi feita usando um veículo lipídico para distribuição como descrito em WO 01/16306. Em suma, a transfecção usou agentes conhecidos "lipitóides" e "colesteróides",

descritos, por exemplo, nas publicações PCT WO 01/16306, WO 98/06437 e WO 99/08711.

Esses conjugados lípido-peptóide catiónico são descritos nessas referências como sendo reagentes eficientes na distribuição de plasmídeos de DNA a células *in vitro*. Qualquer dos transportadores descritos nas aplicações acima referidas é adequado para utilização na transfecção dos oligonucleotídeos aqui descritos.

Esses compostos podem ser preparados por síntese convencional em solução ou em fase sólida. Num procedimento, como descrito em WO 99/08711, citado acima, o N terminal de um peptóide ligado a uma resina é acilado com um espaçador tal como ácido Fmocamino-hexanóico ou Fmoc-3-alanina. Após remoção do grupo Fmoc, o grupo amina primário é feito reagir com chloroformato de colesterol para formar uma ligação carbamato. O produto é então libertado da resina com ácido trifluoroacético e purificado por HPLC de fase reversa. Pode usar-se uma fracção lipídica obtida de ácidos gordos, tal como um fosfolípido, em vez de fracção esteróide. A fracção esteróide ou de outro lípido pode também ser ligada à fracção peptóide por outras ligações, de qualquer comprimento eficiente, facilmente disponível para um especialista.

Dependendo do tipo de célula, foram usados diferentes veículos lipídicos para diferentes tempos de transfecção. Contudo, o tempo de transfecção não excedeu as 24 hrs. A transfecção foi feita em meio completo e a concentração final de oligonucleotídeo *antisense* foi de 300 nM por poço. Nos poços com fármaco, o fármaco foi adicionado à cultura no início da transfecção.

A começar no dia 3: as células foram recuperadas, 1 placa/dia e a libertação de LDH no sobrenadante assim como LDH nas células intactas foi medido usando um kit da Roche de acordo com as instruções do fabricante (Roche Diagnostics, Basel, Switzerland) (dados marcados como dia 1, 2, 3).

Para cada amostra, foi analisada por exame do nível relativo de libertação de LDH em comparação com o LDH total, sendo que um aumento da porção do LDH total significa aumento da morte celular (devido a uma maior proporção de LDH libertado no meio). Os dados foram tratados qualitativamente por comparação com um controlo não tratado (sem oligo). Este ensaio permitiu a determinação de se a perda de mensagem induzida pelo *antisense* para um gene particular causa morte celular quando usada sózinha, ou se esta perda de mensagem torna as células mais sensíveis aos efeitos de um fármaco.

Os resultados são mostrados na tabela imediatamente abaixo.

	HT 1080	SW620	231
chir7-2	negativo	negativo	
	HT 1080	SW620	231
chir8-4	positivo	fracamente positivo	
chir9-5		positivo	
chir1 1-2		negativo	
chir 16-4		negativo	
Chir33-4	positivo muito fraco	positivo forte	positivo muito fraco

Exemplo 12: Detecção de Expressão Diferencial Utilizando Matrizes

O mRNA isolado de amostras de tecido do cólon canceroso e normal obtido de pacientes foi analisado para identificar genes

expressos diferencialmente nas células cancerosas e normais. As células normais e cancerosas recolhidas de tecidos criopreservados de pacientes foram isoladas recorrendo a técnicas de Microdissecação por Captura Laser (LCM), as quais são bem conhecidas na especialidade (ver, p. ex., Ohyama et al. (2000) *Biotechniques* 29:530-6; Curran et al. (2000) *Mol. Pathol.* 53:64-8; Suarez-Quian et al. (1999) *Biotechniques* 26:328-35; Simone et al. (1998) *Trends Genet* 14:272-6; Conia et al. (1997) *J. Clin. Lab. Anal.* 11:28-38; Emmert-Buck et al. (1996) *Science* 274:998-1001).

A Tabela 5 (inserida antes das reivindicações) fornece informação acerca de cada paciente do qual as amostras foram isoladas, incluindo: o "Paciente ID" e "Rel Pat ID", que são números atribuídos ao paciente e aos relatórios de patologia para efeitos de identificação; o "Grupo" ao qual os pacientes foram atribuídos; a localização anatómica do tumor ("Loc Anatom"); o "Tamanho do Tumor Primário"; o "Grau do Tumor Primário"; a identificação do grau histopatológico ("Grau Histopat"); a descrição dos locais invadidos pelo tumor ("Local Invasão"); a presença de metástases em nódulos linfáticos ("Met Nod Linf "); a incidência das metástases nos nódulos linfáticos (fornecidos como o número de nódulos linfáticos positivos para metástases sobre o número de nódulos linfáticos examinados) ("Incidência Met Linfnod "); o "Grau da Região do nódulo linfático"; a identificação ou detecção de metástases em locais distantes do tumor e a sua localização ("Met Distant & Loc"); uma descrição das metástases distantes ("Descri Met Distant "); o grau das metástases distantes ("Grau Met Dist"); e comentários gerais acerca do paciente ou do tumor ("Comentários"). O adenoma não foi descrito em qualquer dos pacientes; uma displasia do adenoma

(descrita como hiperplasia pelo patologista) foi descrita no Paciente ID No. 695. Foram descritas extensões extranodais em duas pacientes, Paciente ID Nos. 784 e 791. Foi descrita invasão linfovascular em sete pacientes, Paciente ID Nos. 128, 278, 517, 534, 784, 786, e 791. Infiltrados de tipo Crohn foram descritos em sete pacientes, Paciente ID Nos. 52, 264, 268, 392, 393, 784, e 791.

Identificação de genes expressos diferencialmente

Foram preparadas sondas de cDNA a partir de RNA completo isolado das células de pacientes descritas acima. Uma vez que a LCM permite o isolamento de tipos de células específicos para se obter uma amostra de células substancialmente homogênea, esta fornece uma amostra de RNA semelhante pura.

O RNA completo foi primeiro transcrito de modo reverso em cDNA usando um *primer* contendo um promotor da RNA polimerase T7, seguido da síntese da segunda cadeia de DNA. O cDNA foi então transcrito *in vitro* para produzir RNA *antisense* usando a expressão mediada pelo promotor T7 (ver, p. ex., Luo et al. (1999) *Nature Med* 5:117-122), e o RNA *antisense* foi então convertido em cDNA. O segundo conjunto de cDNAs foi outra vez transcrito *in vitro*, usando o promotor T7, para fornecer o RNA *antisense*. Opcionalmente, o RNA foi outra vez convertido em cDNA, permitindo uma terceira amplificação mediada por T7 para produzir mais RNA *antisense*. Assim o procedimento forneceu dois ou três ciclos de transcrição *in vitro* para produzir o RNA final usado para marcação fluorescente.

As sondas fluorescentes foram geradas por adição de RNA de controlo à mistura de RNA *antisense*, e produzindo cDNA marcado com fluorescência a partir do RNA de partida. Os cDNAs marcados

com fluorescência preparados a partir da amostra de RNA do tumor foram comparados com cDNAs marcados com fluorescência, preparados a partir de uma amostra de RNA de células normais. Por exemplo, as sondas de cDNA das células normais foram marcadas com pigmento fluorescente Cy3 (verde) e as sondas de cDNA preparadas das células do tumor foram marcadas com pigmento fluorescente Cy5 (vermelho), e vice-versa.

Cada matriz usada tinha um arranjo espacial idêntico e conjunto de manchas de controlo. Cada micro-matriz foi dividido em duas áreas, cada área contendo uma matriz com, em cada metade, doze agrupamentos de 32 x 12 manchas, para um total de cerca de 9,216 manchas em cada matriz. As duas áreas são manchas de forma idêntica o que fornece pelo menos dois duplicados de cada clone por matriz.

Os polinucleotídeos correspondentes aos genes expressos diferencialmente, aqui descritos para utilização nas matrizes, foram obtidos de fontes públicas e de bibliotecas de cDNA geradas a partir da selecção de linhagens celulares e tecidos de pacientes. Os produtos de PCR com cerca de 0.5kb a 2.0 kb amplificadas a partir dessas fontes foram manchas na matriz recorrendo a um visualizador Molecular Dynamics Gen III, de acordo com as instruções do fabricante. A primeira fila de cada uma das 24 regiões da matriz tinha cerca de 32 manchas de controlo, incluindo 4 manchas de controlo negativo e 8 polinucleotídeos de teste. Os polinucleotídeos de teste foram reconstituídas em cada amostra antes da reacção de labeling numa gama de concentrações de 2-600 pg/lâmina e razões de 1:1. Para a concepção de cada matriz, foram hibridadas duas lâminas com o teste das amostras de reversas marcadas na reacção de labeling. Este procedimento forneceu cerca de quatro medições em duplicado

por cada clone, duas de uma côr e duas da outra, para cada amostra.

O ensaio de expressão diferencial foi levado a cabo por mistura de quantidades iguais de sondas de células tumorais e células normais do mesmo paciente. As matrizes foram pré-hibridadas por incubação durante cerca de 2 hrs a 60°C em 5X SSC/0.2% em SDS/1 mM em EDTA, e depois lavadas três vezes com água e duas com isopropanol. Após a pré-hibridação da matriz, a mistura de sondas foi hibridada com a matriz em condições de elevada restringência (durante a noite a 42°C em 50% de formamida, 5X SSC, e 0.2% de SDS. Após a hibridação, a matriz foi lavada a 55°C três vezes do seguinte modo: 1) primeira lavagem com 1X SSC/0.2% em SDS; 2) segunda lavagem em 0.1X SSC/0.2% em SDS; e 3) terceira lavagem em 0.1X SSC.

As matrizes foram então varridas para fluorescência verde e vermelha por recurso a um laser-scanner/detector de cor dupla da Molecular Dynamics Generation III. As imagens foram processadas usando software BioDiscovery Autogene, e os dados de cada conjunto de varrimentos normalizadas para fornecerem a razão de expressão relativamente ao normal. Os dados das experiências de micro-matriz foram analisados de acordo com os algoritmos descritos no Pedido U.S. N°. de Série 60/252,358, depositado em 20 de Novembro de 2000, por E.J. Moler, M.A. Boyle, e F.M. Randazzo, e intitulado "Precisão e exactidão em dados de micro-matriz de cDNA".

A experiência foi repetida, desta vez marcando as duas sondas com a côr oposta de forma a fazer o ensaio em ambas as "direcções da côr." Cada experiência foi por vezes repetida com mais duas lâminas (uma em cada direcção da côr). O nível de fluorescência para cada sequência no sistema expresso como a

razão da média geométrica de manchas de 8 replicados / genes de 4 sistemas ou manchas de 4 replicados / gene de 2 sistemas ou qualquer outra permutação. Os dados foram normalizados usando os controles positivos reconstituídos presentes em cada área duplicada, e a precisão desta normalização foi incluída na determinação final da significância de cada diferencial. A intensidade da fluorescência de cada mancha foi ainda comparada com os controles negativos em cada área duplicada, para determinar que manchas detectaram níveis significativos de expressão em cada amostra.

A análise estatística das intensidades de fluorescência foi aplicada a cada conjunto em duplicado de manchas para determinar a precisão e significância de cada medição diferencial, resultando num valor de p que testa a hipótese nula de não haver diferença no nível de expressão entre as amostras tumoral e normal de cada paciente. Durante a análise inicial das micromatrizes, a hipótese foi aceita se $p > 10^{-3}$, e a razão diferencial foi considerada 1.000 para essas manchas. Todas as outras manchas apresentam uma diferença significativa na expressão entre as amostra tumorais e normais. Se a amostra de tumor apresenta uma expressão detectável e a normal não, a razão é truncada a 1000 uma vez que o valor da expressão na amostra normal seria zero, e a razão não teria um valor matematicamente útil (p. ex., infinito). Se a amostra normal apresenta uma expressão detectável e o tumor não, a razão é truncada a 0.001, uma vez que o valor da expressão na amostra do tumor seria zero e a razão não teria um valor matematicamente útil. Estas duas últimas situações são aqui designadas por "on/off." As tabelas das bases de dados foram preenchidas usando um nível confiança de 95% ($p > 0.05$).

Os resultados são fornecidos na Tabela 6 abaixo. A tabela inclui: 1) a SEQ ID NO; 2) a identificação da amostra (Amostra ID); 3) o número de identificação da mancha ("SpotID"); e 4) a percentagem dos pacientes testados nos quais os níveis de expressão do gene foram pelo menos o dobro no tecido canceroso do que no tecido normal emparelhado ("Pacientes de cólon com valor de p pcorrigido $95 \geq 2x$ "). As razões da expressão diferencial são expressas como o sinal da hibridação normalizado associada à sonda tumoral dividido pelo sinal da hibridação normalizado com a sonda normal. Assim, uma razão maior do que 1 indica que o produto genético vê a sua expressão aumentada em células cancerosas relativamente a células normais, enquanto que uma razão de menos do que 1 indica o oposto.

Tabela 6 SEQ ID NO	Amostra ID	Chip Spot ID	Pacientes Cólon valor p corrigido $95 \geq 2x$
1	RG:727787:Order7TM31:E07	29912	82.14
7	M00055209C:B07	24297	30.30
9	M00056908A:H05	21544	42.42
13	M00057000D:E08	21592	30.30
27	RG:1418951:Order7TM11:D12	33623	78.57
29	RG:1418951:Order7TM11:D12	33623	78.57
22	M00001346C:A05	243	55
22	M00054893C:D03	21952	30

Esses dados constituem uma prova de que os genes representados pelos polinucleotídeos contendo as sequências indicadas são expressos diferencialmente no cancro do cólon.

Os especialistas reconhecerão, ou serão capazes de discernir, sem recurso a nada mais do que experiências de rotina, muitos equivalentes às modalidades específicas da invenção aqui

descrita. Tais modalidades específicas e equivalentes presumem-se abrangidas pelas seguintes reivindicações.

Apesar de a presente invenção ter sido descrita com algum detalhe através de ilustrações e exemplos com a finalidade da clareza e compreensão, é evidente para os especialistas à luz dos ensinamentos desta invenção que certas alterações e modificações podem ser feitas sem que se saia do espírito ou âmbito das reivindicações anexas.

Informação de Depósito. Foi feito um depósito de culturas biologicamente puras dos seguintes vírus na American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209, ao abrigo das disposições do Tratado de Budapeste, na data de submissão da presente candidatura ou antes. O número de adesão indicado foi atribuído após testes de viabilidade bem sucedidos, e pagas as taxas requeridas. O acesso às referidas culturas estará disponível durante pendência da candidatura a patente a quem a Comissão determinar esse direito ao abrigo dos 37 C.F.R § 1.14 e 35 U.S.C. § 122. Todas as restrições de disponibilização das ditas culturas ao público serão irrevogavelmente retiradas após o registo da patente baseada na candidatura. Mais ainda, os depósitos designados serão mantidos por um período de trinta (30) anos a partir da data do depósito, ou por cinco (5) anos após o último pedido de depósito; ou pelo período de vida executória da patente U.S., o qual for mais longo. Se a cultura se tornar inviável ou for destruída inadvertidamente, ou, no caso de estirpes contendo plasmídeos, perda do plasmídeo, será substituída por culturas viáveis da mesma descrição taxonómica.

Esses depósitos são fornecidos meramente para conveniência dos especialistas, e não são uma admissão da necessidade de existência de um depósito. As sequências de ácidos nucleicos

desses plasmídeos, assim como as aminosequências dos polipeptídeos por eles codificados, servirão de controle na eventualidade de qualquer conflito envolvendo a descrição aqui feita. Pode ser exigida uma licença para produzir, usar, ou vender os materiais depositados, não sendo aqui garantida tal licença.

[0286] Para além disso, às misturas de clones seleccionados, assim como às bibliotecas contendo clones específico, foi atribuído um número "ES" (referência interna) e depositados com o ATCC. A Tabela 7 abaixo fornece os N^{os} de Acesso ATCC dos clones depositados, os quais foram todos depositados na data de apresentação do depósito ou antes.

Tabela 7. Misturas de Clones e Bibliotecas Depositados com o ATCC

Nome de Sequência	Clones	CMCC	ATCC
SK1	SK-1	5162	PTA-1360
SK2	SK-2	5163	PTA-1361
SK5	SK-5	5164	PTA-1362
1665 curto	1665 curto	5165	PTA-1363
1665 longo	1665 longo	5166	PTA-1363
sk19	SK-19	5167	PTA-1364
Junc2	Junc2-6	5168	PTA-1365
XD4	XD4b	5169	PTA-1366
XD1	XD1b	5170	PTA-1367
XD7	XD7c	5171	PTA-1368
XD10	XD10b	5172	PTA-1369
XD11	XD11b	5173	PTA-1370
Junc4	Junc4-2	5174	PTA-1371

CMCC refere-se a um N°. de referência interno da Requerente.

Remoção de Clones Individuais do Depósito de Clones Misturados. Sendo o depósito ATCC composto por uma mistura de clones de cDNA ou uma biblioteca de clones de cDNA, o depósito foi preparado por transfecção de cada um dos clones para células bacterianas separadas. Os clones da *pool* ou biblioteca foram então depositados como uma *pool* de misturas iguais no depósito compósito. Podem obter-se clones particulares a partir do depósito compósito recorrendo a métodos bem conhecidos na especialidade. Por exemplo, uma célula bacteriana contendo um

clone particular pode ser identificada por isolamento de colónias simples, e identificação das colónias contendo o clone específico através de técnicas padronizadas de hibridação de colónias, recorrendo a uma sonda ou sondas de oligonucleotídeo concebidas para hibridar especificamente com uma sequência do clone inserção (p. ex., uma sonda baseada numa sequência sem máscara do polinucleotídeo codificado com o SEQ ID NO indicado). A sonda deve ser concebida para ter uma T_m de aproximadamente 80°C (assumindo 2°C por cada A ou T e 4°C por cada G ou C). As colónias positivas podem então ser recolhidas, crescidas em meio de cultura, e os clones recombinantes isolados. Alternativamente, as sondas concebidas deste modo podem ser usadas em PCR para isolar uma molécula de ácido nucleico dos clones misturados, de acordo com métodos bem conhecidos na especialidade, p. ex., por purificação do cDNA a partir da mistura de cultura depositada, e usando as sondas em reacções de PCR para produzir um produto amplificado contendo a correspondente sequência polinucleotídica desejada.

Tabela 5
Dados dos Pacientes

Paciente ID	Idade (anos)	Grupo	Localiz. Anatomico	Tamanho do Tumor Primario	Grav. do Tumor Primario	Grav. de Histopatol	Local. Invasão	Mar. Nodulo Linfat.	Incidência Mar. Nodulo Linfat.	Grav. do Tumor Infiltrado	Mar. & Linf. Infiltrado	Descrição Mar. Ductos	Grav. Mar. Duct	Comentários
125	71	III	Colon Ascendente	4,0	T3	G2	Expositado até ao tecido adiposo subserosal	positivo	3/8	N1	negativo			Adenocarcinoma invasivo moderadamente diferenciado; observada invasão perineural local
126	71	II	Colon Ascendente	9,0	T3	G3	Invasão através da muscular própria; envolvimento subseroso; envolvimento da válvula ileocecal.	negativo	0/12	N0	negativo		M0	Tipo hiperplásico no apêndice
127	140	II	Sigmóide	6	T4	G2	Invasão muscular própria na serosa, envolvendo a submucosa da bexiga	negativo	0/24	N0	negativo		M0	Invasão perineural; coexistência atípica negativa; uma videolidade tubular e um adenoma tubular com displasia de baixo grau
128	144	II	Cego	6	T3	G2	Invasão através da muscular própria até ao tecido adiposo subseroso. Mucosa ileocecal.	negativo	0/19	N0	negativo		M0	Paciente com história de melanoma metastático
129	147	III	Colon transversal	5,0	T3	G2	Invasão da muscular própria	positivo	1/5	N1	negativo		M0	
130	149		Flexura hepática	5,5	T3		Invasão da parede até ao tecido adiposo subcutâneo	positivo	10/24	N2	negativo		M1	
131	152	II	Ceco	5,0	T3	G2	Invasão através da muscular própria até ao tecido peritoneal; a peritonização a configuração visual é normal.	negativo	0/9	N0	negativo		M0	Pequeno adenoma tubular separado (0,4 cm)
141	168	IV	Cego	5,5	T3	G2	Invasão da muscular própria até ao tecido subperitoneal; taxa não através da serosa. Presença de desmoplasia local.	positivo	7/21	N2	positivo (figado)	Adenocarcinoma coexistente com primário	M1	Identificada invasão perineural adjacente a adenocarcinoma metastático

Exercises 2

Indicador ID	Soluções Propostas	Grupo	Localidade	Unidade de Tumor Primária	Grupo de Estadiamento	Local Invasão	Mar. Adiposo Unif.	Indicador Mar. Adiposo Unif.	Gravidade do Tumor Primário	Mar. & Los Foram	Descrição Mar. & Los Foram	Gravidade do Tumor Primário	Comentários
156	175	III	Flórida, República	3,8	G3	Invasão através da membrana própria até ao adiposo subcutâneo pericárdico, sem envolvimento seroso. Configuração nuclear vítreo.	positivo	2/13	R1	negativo		M0	Adesões tumorais e fibroses locais separadas.
228	247	III	Buenos Aires	5,8	G3	Invasão através da membrana própria, envolvendo subserosa, adiposo pericárdico, e seroso.	positivo	1/8	R1	negativo		M0	Pistões hiperplásicos.
263	283	II	Colômbia, departamento	5,5	G2	Invasão através da membrana própria até ao tecido adiposo subcutâneo.	negativo	0/10	R0	negativo		M0	Alto grau de invasão.
266	285	III	Colômbia, departamento	9	G3	Invasão através da membrana própria para o adiposo pericárdico, expandindo-se até à serosa.	negativo	0/15	R1	positivo (dispositivo metastático no)	R a cas, pode representar o tumor.	M0	Alto grau de invasão.
268	287	I	Uruguai	6,5	G2	Invasão total a espessura da membrana própria, mas adiposo metastático livre de malignidade.	negativo	0/12	R0	negativo		M0	
278	287	III	Buenos Aires	4	G3	Invasão através da membrana própria.	positivo	2/13	R2	negativo		M0	Adesões ao tecido conectivo, sem HCTD ou carcinoma identificado.
285	314	I	Colômbia, departamento	2,0	G2	Invasão através da membrana própria até ao tecido adiposo pericárdico.	negativo	0/13	R0	negativo		M0	Adesões ao tecido conectivo, sem HCTD ou carcinoma identificado.

Tabela 5
Dados dos Pacientes

Paciente ID	Idade	Sexo	Grupo	Localiz. Análise de	Tamanho do Tumor Primária	Gravidade de Tumor Primária	Gravidade de Metástase	Local Invasão	Marx. Nodular Unifal.	Incidência Marx. Nodular Unifal.	Gravidade de Tumor Primária	Marx. Nodular Unifal.	Descrição Marx. Nodular	Gravidade de Tumor Primária	Comentários
338	358	F	II	Rectoanal de	6	T3	G2	Expansão até a gordura perirectal mas não atinge a serosa	negativo	0/6	N0	negativo		M0	1 colpox hiperplásico identificado
341	360	F	II	Colón ascendente	2 cm invasivo	T3	G2	Invasão através da muscular própria até envolver a gordura pericolicônica. Proveniente de vilosidade da adenoma.	negativo	0/4	N0	negativo		M0	
350	375	F	II	Sigmoides	6,5	T3	G2	Através da parede do ceco até ao tecido adiposo subseroso. Nódulos espalhados na serosa observados.	negativo	0/4	N0	negativo		M0	
360	412	F	III	Colón ascendente	4,3	T3	G2	Invasão através da muscular própria até a gordura pericolicônica.	positivo	1/5	N1	negativo		M0	Dois polípos na mucosa
392	444	F	IV	Colón ascendente	2	T3	G2	Invasão através da muscular própria até ao tecido adiposo subseroso, não na serosa	positivo	1/6	N1	positivo (Figado)	Estenose muscular vesicular e microvesicular	M1	Tumor proveniente de antes da anastomose cirúrgica ileocolica.
395	445	F	II	Ceco	6,8	T3	G2	Ceco, invaso através da muscular própria para envolver o tecido adiposo subseroso mas não a serosa.	negativo	0/21	N0	negativo		M0	
413	465	F	IV	Colón ascendente	4,8	T3	G2	Invasão através da muscular própria para envolver a gordura pericolicônica; presente na junção ileocecal.	negativo	0/7	N0	positivo (Figado)	Adenocarcinoma em lúmen intestinal	M1	Redesignação da via da anastomose para o ceco de colite metastática.
505	383	F	IV		7,5 cm dim max	T3	G2	Invasão através da muscular própria envolvendo adiposo pericólico, superfície serosa não envilhada.	positivo	2/17	N1	positivo (Figado)	Adenocarcinoma em lúmen intestinal diferenciado, consistentemente com primário	M1	Localização anatomica não mudada em relação evidência de colite crônica.
517	395	F	IV	Sigmoides	3	T3	G2	Serosa a muscular própria; envolver a gordura pericolicônica.	positivo	0/5	N2	negativo		M0	Não são mencionadas Marx. distantes no relatório

Tabela 5
Dados dos Pacientes

Paciente	Relatório Painel T3	Grupo	Localiz. Anatômico	Tamanho do Tumor Primário	Gravidade do Tumor Primário	Gravidade Histopatol.	Local Lesões	Marco Nodal Linfat.	Resposta Marco Linfat.	Resposta Marco Histol.	Uso de Marco Hist.	Comentários
534	553	II	Cérebro anterior	12	T3	G3	Invasão através da meninge própria envolvendo glicoma periclitico. Semelhante de tumor.	negativo			M3	Oncologia com fibrose e exereses da glicoma. Intensificação da resposta aguda e crônica, aumento focal e difuso.
546	565	IV	Cérebro anterior	15,5	T3	G2	Invasão através da meninge própria através da subaracnóide e espalhando-se até a cerebra.	positivo	6/12	Adesão meninge anterior	M3	
577	596	II	Cérebro	11,5	T3	G2	Invasão através da meninge própria através da subaracnóide. Superfície de tumor.	negativo	0/12		M3	Agudamente difusa e focalizada, mas não consistida por tumor

Tabela 5
Dados dos Pacientes

Paciente ID	Relatório Painel II	Grupo	Localiz. Anatom.	Tamanho do Tumor Primário	Gravidade do Tumor Primário	Gravidade Histopatol.	Local Invasão	Marc. Nidulo Linfat.	Incidência Marc. Nidulo Linfat.	Uso da Imunohisto- química	Marc. de Linf. Distant.	Benefício Marc. Distant.	Gravidade Marc. Distant.	Comentários
685	714	II	Cego	14	T3	G2	Expende-se atenção da parte do intestino até à pariete serosa	negativo	G22	N0	negativo		M43	Adenoma tubular e polipos hiperplásicos presentes; adenoma moderadamente diferenciado, com diferenciação oncos (% não referida)
784	803	IV	Cólon ascendente	3,5	T3	G3	Através da muscular própria até aos testículos e nos peritóstios	positivo	5/17	N2	positivo (Figado)		M1	Carcinoma adenocarcinoma linxiva pouco diferenciado
786	805	IV	Cólon descendente	9,5	T3	G2	Através da muscular própria até à gordura peritóstica, mas não na superfície serosa	negativo	G12	N0	positivo (Figado)		M1	Adenocarcinoma parietal moderadamente diferenciado
791	810	IV	Cólon ascendente	5,8	T3	G3	Através da muscular própria até à gordura peritóstica	positivo	13/25	M1	positivo (Figado)		M1	Adenocarcinoma solitário invasivo pouco diferenciado
808	908	IV	Cólon ascendente	2,0	T2	G1	Até à muscular própria	positivo	3/21	N0	positivo (Figado)		M1	Adenocarcinoma moderadamente a muito diferenciado; este portante tem tumores de côlon ascendente e do côlon sigmoide
809	909	IV	Cego	4,8	T3	G1	Através da muscular própria até ao tecido subseroso	positivo	1/4	M1	positivo (Figado)		M1	Adenocarcinoma moderadamente diferenciado

LISTAGEM DAS SEQUÊNCIAS

<110> Kennedy, Giulia C._

Kang, Sanmao

Reinhard, Christoph

Jefferson, Anne Bennett

<120> POLINUCLEOTÍDEOS RELACIONADOS COM O CANCRO DO CÓLON

<130> PP-_1663.003

<140> Unatribuído

<141> 2001-06-15

<150> 60/211,835

<151> 2000-06-15

<160> 127

<170> FastSEQ for Windows Versão 4.0

<210> 1

<211> 564

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (21)..._(396)

<400> 1

ggcgagattt gtgcggcgac atg aaa ctg ctt acc cac aat ctg ctg agc tcg	53
Met Lys Leu Leu Thr His Asn Leu Leu Ser Ser	
1 5 10	
cat gtg cgg ggg gtg ggg tcc cgt ggc ttc ccc ctg cgc ctc cag gcc	101
His Val Arg Gly Val Gly Ser Arg Gly Phe Pro Leu Arg Leu Gln Ala	
15 20 25	
acc gag gtc cgt atc tgc cct gtg gaa ttc aac ccc aac ttc gtg gcg	149
Thr Glu Val Arg Ile Cys Pro Val Glu Phe Asn Pro Asn Phe Val Ala	
30 35 40	
cgt atg ata cct aaa gtg gag tgg tgg gcg ttc ctg gag gcg gcc gat	197
Arg Met Ile Pro Lys Val Glu Trp Ser Ala Phe Leu Glu Ala Ala Asp	
45 50 55	
aac ttg cgt ctg atc cag gtg cgg aaa ggg cgg gtt gag gga tat gag	245
Asn Leu Arg Leu Ile Gln Val Pro Lys Gly Pro Val Glu Gly Tyr Glu	
60 65 70 75	
gag aat gag gag ttt ctg agg acc atg cac cac ctg ctg ctg gag gtg	293
Glu Asn Glu Glu Phe Leu Arg Thr Met His His Leu Leu Leu Glu Val	
80 85 90	
gaa gtg ata gag ggc acc ctg cag tgc cgg gaa tct gga cgt atg ttc	341
Glu Val Ile Glu Gly Thr Leu Gln Cys Pro Glu Ser Gly Arg Met Phe	
95 100 105	
ccc atc agc cgc ggg atc ccc aac atg ctg ctg agt gaa gag gaa act	389
Pro Ile Ser Arg Gly Ile Pro Asn Met Leu Leu Ser Glu Glu Glu Thr	
110 115 120	
gag agt t gattgtgcca ggcgccagtt tttcttgtaa tgactgtgta tttttgttga	446
Glu Ser	
125	
tctataccct gtttcgaat tctgccgtgt gtatcccaa cccttgacc aatgacacca	506
aacacagtgt ttttgagctc ggtattatat atttttttct cattaaaggt ttaaaacc	564

<210> 2

<211> 125

<212> PRT

<213> Homo sapiens

<400> 2

Met	Lys	Leu	Leu	Thr	His	Asn	Leu	Leu	Ser	Ser	His	Val	Arg	Gly	Val	
1				5					10					15		
Gly	Ser	Arg	Gly	Phe	Pro	Leu	Arg	Leu	Gln	Ala	Thr	Glu	Val	Arg	Ile	
			20					25					30			
Cys	Pro	Val	Glu	Phe	Asn	Pro	Asn	Phe	Val	Ala	Arg	Met	Ile	Pro	Lys	
		35					40					45				
Val	Glu	Trp	Ser	Ala	Phe	Leu	Glu	Ala	Ala	Asp	Asn	Leu	Arg	Leu	Ile	
	50					55					60					
Gln	Val	Pro	Lys	Gly	Pro	Val	Glu	Gly	Tyr	Glu	Glu	Asn	Glu	Glu	Phe	
65					70					75					80	
Leu	Arg	Thr	Met	His	His	Leu	Leu	Leu	Glu	Val	Glu	Val	Ile	Glu	Gly	
				85					90					95		
Thr	Leu	Gln	Cys	Pro	Glu	Ser	Gly	Arg	Met	Phe	Pro	Ile	Ser	Arg	Gly	
			100					105					110			
Ile	Pro	Asn	Met	Leu	Leu	Ser	Glu	Glu	Glu	Thr	Glu	Ser				
		115					120					125				

<210> 3

<211> 919

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (219)..._(693)

<400> 3

tggcacgagg	tggcacgagg	gtccgggtcg	ttgaggatta	ggctgctcgg	gcgtaaccgg	60
agctggggcg	cggtgcgcaa	gggcgggccc	ggaagtccca	gcggtcttta	aattctcccg	120
tgctagggcc	agcctgcgca	ttcttacctg	tccgggtgcg	gcgagtgtct	cacctctctg	180
cacttccaag	gactcttgte	atctgcctta	ggcgggaa	atg ctg ttg ctg gat tgc		236
				Met Leu Leu Leu Asp Cys		
				1	5	
aac ccc gag gtg gat ggt ctg aag cat ttg ctg gag aca ggg gcc tcg						284
Asn Pro Glu Val Asp Gly Leu Lys His Leu Leu Glu Thr Gly Ala Ser						
	10		15		20	
gtc aac gca ccc ccg gat ccc tgc aag cag tcg cct gtc cac tta gcc						332
Val Asn Ala Pro Pro Asp Pro Cys Lys Gln Ser Pro Val His Leu Ala						
	25		30		35	
gca gga agc ggc ctt gct tgc ttt ctt ctc tgg cag ctg caa acg ggc						380
Ala Gly Ser Gly Leu Ala Cys Phe Leu Leu Trp Gln Leu Gln Thr Gly						
	40		45		50	
gct gac ctc aac cag cag gat gtt tta gga gaa gct cca cta cac aag						428
Ala Asp Leu Asn Gln Gln Asp Val Leu Gly Glu Ala Pro Leu His Lys						
	55		60		65	70

gca gca aaa gtt gga agc ctg gag tgc cta agc ctg ctt gta gcc agt	476
Ala Ala Lys Val Gly Ser Leu Glu Cys Leu Ser Leu Leu Val Ala Ser	
75 80 85	
gat gcc caa att gat tta tgt aat aag aac ggg caa aca gct gaa gat	524
Asp Ala Gln Ile Asp Leu Cys Asn Lys Asn Gly Gln Thr Ala Glu Asp	
90 95 100	
ctc gct tgg tca tgt gga ttt cca gac tgt gcc aag ttt ctt aca aca	572
Leu Ala Trp Ser Cys Gly Phe Pro Asp Cys Ala Lys Phe Leu Thr Thr	
105 110 115	
att aaa tgt atg cag aca ata aaa gca agt gaa cac cct gac agg aat	620
Ile Lys Cys Met Gln Thr Ile Lys Ala Ser Glu His Pro Asp Arg Asn	
120 125 130	
gat tgt gtt gcc gtg ctc aga cag aaa cgg agt ctc gga agt gta gaa	668
Asp Cys Val Ala Val Leu Arg Gln Lys Arg Ser Leu Gly Ser Val Glu	
135 140 145 150	
aat acc agt ggg aaa agg aag tgc t gatgtcacgt gggttatgaa	713
Asn Thr Ser Gly Lys Arg Lys Cys	
155	
gaagtctgaa gaacgccttc atttcatgca aatctataag ctccctgcttt tggctttacc	773
atatgttgtg tctaattctcc ttctgagaag gacgaaaaac tttcttccaa gtgaagatcc	833
atttaagaac acatgtatctt acatgcctat aatatgctgg ttgtgtatgc tttgtctttt	893
aagttattaa aggaacgtct aaaaaa	919

<210> 4

<211> 158

<212> PRT

<213> Homo sapiens

<400> 4

Met	Leu	Leu	Leu	Asp	Cys	Asn	Pro	Glu	Val	Asp	Gly	Leu	Lys	His	Leu
1				5					10					15	
Leu	Glu	Thr	Gly	Ala	Ser	Val	Asn	Ala	Pro	Pro	Asp	Pro	Cys	Lys	Gln
			20					25					30		
Ser	Pro	Val	His	Leu	Ala	Ala	Gly	Ser	Gly	Leu	Ala	Cys	Phe	Leu	Leu
		35					40					45			
Trp	Gln	Leu	Gln	Thr	Gly	Ala	Asp	Leu	Asn	Gln	Gln	Asp	Val	Leu	Gly
		50				55					60				
Glu	Ala	Pro	Leu	His	Lys	Ala	Ala	Lys	Val	Gly	Ser	Leu	Glu	Cys	Leu
65				70						75				80	
Ser	Leu	Leu	Val	Ala	Ser	Asp	Ala	Gln	Ile	Asp	Leu	Cys	Asn	Lys	Asn
			85						90					95	
Gly	Gln	Thr	Ala	Glu	Asp	Leu	Ala	Trp	Ser	Cys	Gly	Phe	Pro	Asp	Cys
			100				105						110		
Ala	Lys	Phe	Leu	Thr	Thr	Ile	Lys	Cys	Met	Gln	Thr	Ile	Lys	Ala	Ser
		115					120					125			
Glu	His	Pro	Asp	Arg	Asn	Asp	Cys	Val	Ala	Val	Leu	Arg	Gln	Lys	Arg
	130					135					140				
Ser	Leu	Gly	Ser	Val	Glu	Asn	Thr	Ser	Gly	Lys	Arg	Lys	Cys		
145					150				155						

<210> 5
<211> 1949
<212> DNA
<213> Homo sapiens
<220>
<221> CDS
<222> (5)..._(1760)
<400> 5

caac atg gcg ccg tcc acg ccg ctc ttg aca gtc cga gga tca gaa gga	49
Met Ala Pro Ser Thr Pro Leu Leu Thr Val Arg Gly Ser Glu Gly	
1 5 10 15	
ctg tac atg gtg aat gga cca cca cat ttt aca gaa agc aca gtg ttt	97
Leu Tyr Met Val Asn Gly Pro Pro His Phe Thr Glu Ser Thr Val Phe	
20 25 30	
cca agg gaa tct ggg aag aat tgc aaa gtc tgt atc ttt agt aag gat	145
Pro Arg Glu Ser Gly Lys Asn Cys Lys Val Cys Ile Phe Ser Lys Asp	
35 40 45	
ggg acc ttg ttt gcc tgg ggc aat gga gaa aaa gta aat att atc agt	193
Gly Thr Leu Phe Ala Trp Gly Asn Gly Glu Lys Val Asn Ile Ile Ser	
50 55 60	
gtc act aac aag gga cta ctg cac tcc ttc gac ctc ctg aag gca gtt	241
Val Thr Asn Lys Gly Leu Leu His Ser Phe Asp Leu Leu Lys Ala Val	
65 70 75	
tgc ctt gaa ttc tca ccc aaa aat act gtc ctg gca acg tgg cag cct	289
Cys Leu Glu Phe Ser Pro Lys Asn Thr Val Leu Ala Thr Trp Gln Pro	
80 85 90 95	
tac act act tct aaa gat ggc aca gct ggg ata ccc aac cta caa ctt	337
Tyr Thr Thr Ser Lys Asp Gly Thr Ala Gly Ile Pro Asn Leu Gln Leu	
100 105 110	
tat gat gtg aaa act ggg aca tgt ttg aaa tct ttc atc cag aaa aaa	385
Tyr Asp Val Lys Thr Gly Thr Cys Leu Lys Ser Phe Ile Gln Lys Lys	
115 120 125	
atg caa aat tgg tgt cca tcc tgg tca gaa gat gaa act ctt tgt gcc	433
Met Gln Asn Trp Cys Pro Ser Trp Ser Glu Asp Glu Thr Leu Cys Ala	
130 135 140	
cgc aat gtt aac aat gaa gtt cac ttc ttt gaa aac aac aat ttt aac	481
Arg Asn Val Asn Asn Glu Val His Phe Phe Glu Asn Asn Asn Phe Asn	
145 150 155	
aca att gca aat aaa ttg cat ttg caa aaa att aat gac ttt gta tta	529
Thr Ile Ala Asn Lys Leu His Leu Gln Lys Ile Asn Asp Phe Val Leu	
160 165 170 175	
tca cct gga ccc caa cca tac aag gtg gct gtc tat gtt cca gga agt	577
Ser Pro Gly Pro Gln Pro Tyr Lys Val Ala Val Tyr Val Pro Gly Ser	
180 185 190	
aaa ggt gca cct tca ttt gtt aga tta tat cag tac ccc aac ttt gct	625
Lys Gly Ala Pro Ser Phe Val Arg Leu Tyr Gln Tyr Pro Asn Phe Ala	
195 200 205	
gga cct cat gca gct tta gct aat aaa agt ttc ttt aag gca gat aaa	673
Gly Pro His Ala Ala Leu Ala Asn Lys Ser Phe Phe Lys Ala Asp Lys	
210 215 220	
gtt aca atg ctg tgg aat aaa aaa gct act gct gtg ttg gta ata gct	721
Val Thr Met Leu Trp Asn Lys Lys Ala Thr Ala Val Leu Val Ile Ala	
225 230 235	

agc aca gat gtt gac aag aca gga gct tcc tac tat gga gaa caa act	769
Ser Thr Asp Val Asp Lys Thr Gly Ala Ser Tyr Tyr Gly Glu Gln Thr	
240 245 250 255	
cta cac tac att gca aca aat gga gaa agt gct gta gtg caa tta cca	817
Leu His Tyr Ile Ala Thr Asn Gly Glu Ser Ala Val Val Gln Leu Pro	
260 265 270	
aaa aat ggc ccc att tat gat gta gtt tgg aat tct agt tct act gag	865
Lys Asn Gly Pro Ile Tyr Asp Val Val Trp Asn Ser Ser Ser Thr Glu	
275 280 285	
ttt tgt gct gta tat ggt ttt atg cct gcc aaa gcg aca att ttc aac	913
Phe Cys Ala Val Tyr Gly Phe Met Pro Ala Lys Ala Thr Ile Phe Asn	
290 295 300	
ttg aaa tgt gat cct gta ttt gac ttt gga act ggt cct cgt aat gca	961
Leu Lys Cys Asp Pro Val Phe Asp Phe Gly Thr Gly Pro Arg Asn Ala	
305 310 315	
gcc tac tat agc cct cat gga cat ata tta gta tta gct gga ttt gga	1009
Ala Tyr Tyr Ser Pro His Gly His Ile Leu Val Leu Ala Gly Phe Gly	
320 325 330 335	
aat ctg agg gga caa atg gaa gtg tgg gat gtg aaa aac tac aaa ctt	1057
Asn Leu Arg Gly Gln Met Glu Val Trp Asp Val Lys Asn Tyr Lys Leu	
340 345 350	
att tct aaa ccg gtg gct tct gat tct aca tat ttt gct tgg tgc ccg	1105
Ile Ser Lys Pro Val Ala Ser Asp Ser Thr Tyr Phe Ala Trp Cys Pro	
355 360 365	
gat ggt gag cat att tta aca gct aca tgt gct ccc agg tta cgg gtt	1153
Asp Gly Glu His Ile Leu Thr Ala Thr Cys Ala Pro Arg Leu Arg Val	
370 375 380	
aat aat gga tac aaa att tgg cat tat act ggc tct atc ttg cac aag	1201
Asn Asn Gly Tyr Lys Ile Trp His Tyr Thr Gly Ser Ile Leu His Lys	
385 390 395	
tat gat gtg cca tca aat gca gaa tta tgg cag gtt tct tgg cag cca	1249
Tyr Asp Val Pro Ser Asn Ala Glu Leu Trp Gln Val Ser Trp Gln Pro	
400 405 410 415	
ttt ttg gat gga ata ttt cca gca aaa aca ata act tac caa gca gtt	1297
Phe Leu Asp Gly Ile Phe Pro Ala Lys Thr Ile Thr Tyr Gln Ala Val	
420 425 430	
cca agt gaa gta ccc aat gag gaa cct aaa gtt gca aca gct tat aga	1345
Pro Ser Glu Val Pro Asn Glu Glu Pro Lys Val Ala Thr Ala Tyr Arg	
435 440 445	
ccc cca gct tta aga aat aaa cca atc acc aat tcc aaa ttg cat gaa	1393
Pro Pro Ala Leu Arg Asn Lys Pro Ile Thr Asn Ser Lys Leu His Glu	
450 455 460	
gag gaa cca cct cag aat atg aaa cca caa tca gga aac gat aag cca	1441
Glu Glu Pro Pro Gln Asn Met Lys Pro Gln Ser Gly Asn Asp Lys Pro	
465 470 475	
tta tca aaa aca gct ctt aaa aat caa agg aag cat gaa gct aag aaa	1489
Leu Ser Lys Thr Ala Leu Lys Asn Gln Arg Lys His Glu Ala Lys Lys	
480 485 490 495	

gct gca aag cag gaa gca aga agt gac aag agt cca gat ttg gca cct	1537
Ala Ala Lys Gln Glu Ala Arg Ser Asp Lys Ser Pro Asp Leu Ala Pro	
500 505 510	
act cct gcc cca cag agc aca cca cga aac act gtc tct cag tca att	1585
Thr Pro Ala Pro Gln Ser Thr Pro Arg Asn Thr Val Ser Gln Ser Ile	
515 520 525	
tct ggg gac cct gag ata gac aaa aaa atc aag aac cta aag aag aaa	1633
Ser Gly Asp Pro Glu Ile Asp Lys Lys Ile Lys Asn Leu Lys Lys Lys	
530 535 540	
ctg aaa gca atc gaa caa ctg aaa gaa caa gca gca act gga aaa cag	1681
Leu Lys Ala Ile Glu Gln Leu Lys Glu Gln Ala Ala Thr Gly Lys Gln	
545 550 555	
cta gaa aaa aat cag ttg gag aaa att cag aaa gaa aca gcc ctt ctc	1729
Leu Glu Lys Asn Gln Leu Glu Lys Ile Gln Lys Glu Thr Ala Leu Leu	
560 565 570 575	
cag gag ctg gaa gat ttg gaa ttg ggt att t aaagattcac ggaaagcaag	1780
Gln Glu Leu Glu Asp Leu Glu Leu Gly Ile	
580 585	
ttgatgacca gaaatcagtg caaacacatc ttctgttaaa cccattggta tacacagaat	1840
attcctgtgc ccacacttaa tgtcaatcta taattttaac catttatcca agattctact	1900
aagtgtaaaa ttatttaata atgtctatta aattgatatt tatatcttg	1949

<210> 6

<211> 585

<212> PRT

<213> Homo sapiens

<400> 6

Met	Ala	Pro	Ser	Thr	Pro	Leu	Leu	Thr	Val	Arg	Gly	Ser	Glu	Gly	Leu
1				5					10					15	
Tyr	Met	Val	Asn	Gly	Pro	Pro	His	Phe	Thr	Glu	Ser	Thr	Val	Phe	Pro
			20					25					30		
Arg	Glu	Ser	Gly	Lys	Asn	Cys	Lys	Val	Cys	Ile	Phe	Ser	Lys	Asp	Gly
		35					40					45			
Thr	Leu	Phe	Ala	Trp	Gly	Asn	Gly	Glu	Lys	Val	Asn	Ile	Ile	Ser	Val
	50					55					60				
Thr	Asn	Lys	Gly	Leu	Leu	His	Ser	Phe	Asp	Leu	Leu	Lys	Ala	Val	Cys
65				70					75						80
Leu	Glu	Phe	Ser	Pro	Lys	Asn	Thr	Val	Leu	Ala	Thr	Trp	Gln	Pro	Tyr
				85					90					95	
Thr	Thr	Ser	Lys	Asp	Gly	Thr	Ala	Gly	Ile	Pro	Asn	Leu	Gln	Leu	Tyr
			100					105					110		
Asp	Val	Lys	Thr	Gly	Thr	Cys	Leu	Lys	Ser	Phe	Ile	Gln	Lys	Lys	Met
		115					120					125			
Gln	Asn	Trp	Cys	Pro	Ser	Trp	Ser	Glu	Asp	Glu	Thr	Leu	Cys	Ala	Arg
	130					135					140				
Asn	Val	Asn	Asn	Glu	Val	His	Phe	Phe	Glu	Asn	Asn	Asn	Phe	Asn	Thr
145					150				155						160
Ile	Ala	Asn	Lys	Leu	His	Leu	Gln	Lys	Ile	Asn	Asp	Phe	Val	Leu	Ser
				165					170					175	
Pro	Gly	Pro	Gln	Pro	Tyr	Lys	Val	Ala	Val	Tyr	Val	Pro	Gly	Ser	Lys
			180					185					190		
Gly	Ala	Pro	Ser	Phe	Val	Arg	Leu	Tyr	Gln	Tyr	Pro	Asn	Phe	Ala	Gly
		195					200					205			
Pro	His	Ala	Ala	Leu	Ala	Asn	Lys	Ser	Phe	Phe	Lys	Ala	Asp	Lys	Val
	210					215					220				

Thr	Met	Leu	Trp	Asn	Lys	Lys	Ala	Thr	Ala	Val	Leu	Val	Ile	Ala	Ser	225	230	235	240
Thr	Asp	Val	Asp	Lys	Thr	Gly	Ala	Ser	Tyr	Tyr	Gly	Glu	Gln	Thr	Leu	245	250	255	
His	Tyr	Ile	Ala	Thr	Asn	Gly	Glu	Ser	Ala	Val	Val	Gln	Leu	Pro	Lys	260	265	270	
Asn	Gly	Pro	Ile	Tyr	Asp	Val	Val	Trp	Asn	Ser	Ser	Ser	Thr	Glu	Phe	275	280	285	
Cys	Ala	Val	Tyr	Gly	Phe	Met	Pro	Ala	Lys	Ala	Thr	Ile	Phe	Asn	Leu	290	295	300	
Lys	Cys	Asp	Pro	Val	Phe	Asp	Phe	Gly	Thr	Gly	Pro	Arg	Asn	Ala	Ala	305	310	315	320
Tyr	Tyr	Ser	Pro	His	Gly	His	Ile	Leu	Val	Leu	Ala	Gly	Phe	Gly	Asn	325	330	335	
Leu	Arg	Gly	Gln	Met	Glu	Val	Trp	Asp	Val	Lys	Asn	Tyr	Lys	Leu	Ile	340	345	350	
Ser	Lys	Pro	Val	Ala	Ser	Asp	Ser	Thr	Tyr	Phe	Ala	Trp	Cys	Pro	Asp	355	360	365	
Gly	Glu	His	Ile	Leu	Thr	Ala	Thr	Cys	Ala	Pro	Arg	Leu	Arg	Val	Asn	370	375	380	
Asn	Gly	Tyr	Lys	Ile	Trp	His	Tyr	Thr	Gly	Ser	Ile	Leu	His	Lys	Tyr	385	390	395	400
Asp	Val	Pro	Ser	Asn	Ala	Glu	Leu	Trp	Gln	Val	Ser	Trp	Gln	Pro	Phe	405	410	415	
Leu	Asp	Gly	Ile	Phe	Pro	Ala	Lys	Thr	Ile	Thr	Tyr	Gln	Ala	Val	Pro	420	425	430	
Ser	Glu	Val	Pro	Asn	Glu	Glu	Pro	Lys	Val	Ala	Thr	Ala	Tyr	Arg	Pro	435	440	445	
Pro	Ala	Leu	Arg	Asn	Lys	Pro	Ile	Thr	Asn	Ser	Lys	Leu	His	Glu	Glu	450	455	460	
Glu	Pro	Pro	Gln	Asn	Met	Lys	Pro	Gln	Ser	Gly	Asn	Asp	Lys	Pro	Leu	465	470	475	480
Ser	Lys	Thr	Ala	Leu	Lys	Asn	Gln	Arg	Lys	His	Glu	Ala	Lys	Lys	Ala	485	490	495	
Ala	Lys	Gln	Glu	Ala	Arg	Ser	Asp	Lys	Ser	Pro	Asp	Leu	Ala	Pro	Thr	500	505	510	
Pro	Ala	Pro	Gln	Ser	Thr	Pro	Arg	Asn	Thr	Val	Ser	Gln	Ser	Ile	Ser	515	520	525	
Gly	Asp	Pro	Glu	Ile	Asp	Lys	Lys	Ile	Lys	Asn	Leu	Lys	Lys	Lys	Leu	530	535	540	
Lys	Ala	Ile	Glu	Gln	Leu	Lys	Glu	Gln	Ala	Ala	Thr	Gly	Lys	Gln	Leu	545	550	555	560
Glu	Lys	Asn	Gln	Leu	Glu	Lys	Ile	Gln	Lys	Glu	Thr	Ala	Leu	Leu	Gln	565	570	575	
Glu	Leu	Glu	Asp	Leu	Glu	Leu	Gly	Ile								580	585		

<210> 7

<211> 1110

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (78)..._(642)

<400> 7

ragtggggccc cagtgttgcg ctctctggcc gtttccttaca ctttgettca ggctccagtg	60
cagggggcgtg gtgggat atg gcc aac tcg ggc tgc aag gac gtc acg ggt Met Ala Asn Ser Gly Cys Lys Asp Val Thr Gly 1 5 10	110
cca gat gag gag agt ttt ctg tac ttt gcc tac ggc agc aac ctg ctg Pro Asp Glu Glu Ser Phe Leu Tyr Phe Ala Tyr Gly Ser Asn Leu Leu 15 20 25	158
aca gag agg atc cac ctg cga aac ccc tcg gcg gcg ttc ttc tgt gtg Thr Glu Arg Ile His Leu Arg Asn Pro Ser Ala Ala Phe Phe Cys Val 30 35 40	206
gcc cgc ctg cag gat ttt aag ctt gac ttt ggc aat tcc caa ggc aaa Ala Arg Leu Gln Asp Phe Lys Leu Asp Phe Gly Asn Ser Gln Gly Lys 45 50 55	254
aca agt caa act tgg cat gga ggg ata gcc acc att ttt cag agt cct Thr Ser Gln Thr Trp His Gly Gly Ile Ala Thr Ile Phe Gln Ser Pro 60 65 70 75	302
ggc gat gaa gtg tgg gga gta gta tgg aaa atg aac aaa agc aat tta Gly Asp Glu Val Trp Gly Val Val Trp Lys Met Asn Lys Ser Asn Leu 80 85 90	350
aat tct ctg gat gag caa gaa ggg gtt aaa agt gga atg tat gtt gta Asn Ser Leu Asp Glu Gln Glu Gly Val Lys Ser Gly Met Tyr Val Val 95 100 105	398
ata gaa gtt aaa gtt gca act caa gaa gga aaa gaa ata acc tgt cga Ile Glu Val Lys Val Ala Thr Gln Glu Gly Lys Glu Ile Thr Cys Arg 110 115 120	446
agt tat ctg atg aca aat tac gaa agt gct ccc cca tcc cca cag tat Ser Tyr Leu Met Thr Asn Tyr Glu Ser Ala Pro Pro Ser Pro Gln Tyr 125 130 135	494
aaa aag att att tgc atg ggt gca aaa gaa aat ggt ttg ccg ctg gag Lys Lys Ile Ile Cys Met Gly Ala Lys Glu Asn Gly Leu Pro Leu Glu 140 145 150 155	542
tat caa gag aag tta aaa gca ata gaa cca aat gac tat aca gga aag Tyr Gln Glu Lys Leu Lys Ala Ile Glu Pro Asn Asp Tyr Thr Gly Lys 160 165 170	590
gtc tca gaa gaa att gaa gac atc atc aaa aag ggg gaa aca caa act Val Ser Glu Glu Ile Glu Asp Ile Ile Lys Lys Gly Glu Thr Gln Thr 175 180 185	638
ctt t agaacataac agaatatatc taagggtatt ctatgtgcta atataaata Leu	692
tttttaaacac ttgagaacag ggatctgggg gatctccacg tttgatccat tttcagcagt gctctgaagg agtatcttac ttgggtgatt ccttggtttt agactataaa aagaaactgg gataggagtt agacaattta aaaggggtgt atgagggcct gaaatatgtg acaaatgaat gtgagtaccc ctctctgtgaa caactgaaagc tattctcttg aattgatctt aagtgctctc ttgctctggt aaaagataga tttgtagctc acttgatgat ggtgctgggt aattgctctg ctctgtctga gatttttaaa aatcagctta atgagagtaa tctgcagaca attgataata acatttttgaa aattggaaag atggtatact gtttttagag gaataaacgt atttgtgg	752 812 872 932 992 1052 1110

<210> 8
 <211> 188
 <212> PRT
 <213> Homo sapiens
 <400> 8

Met	Ala	Asn	Ser	Gly	Cys	Lys	Asp	Val	Thr	Gly	Pro	Asp	Glu	Glu	Ser
1				5					10					15	
Phe	Leu	Tyr	Phe	Ala	Tyr	Gly	Ser	Asn	Leu	Leu	Thr	Glu	Arg	Ile	His
			20					25					30		
Leu	Arg	Asn	Pro	Ser	Ala	Ala	Phe	Phe	Cys	Val	Ala	Arg	Leu	Gln	Asp
		35					40					45			
Phe	Lys	Leu	Asp	Phe	Gly	Asn	Ser	Gln	Gly	Lys	Thr	Ser	Gln	Thr	Trp
	50					55					60				
His	Gly	Gly	Ile	Ala	Thr	Ile	Phe	Gln	Ser	Pro	Gly	Asp	Glu	Val	Trp
65					70					75					80
Gly	Val	Val	Trp	Lys	Met	Asn	Lys	Ser	Asn	Leu	Asn	Ser	Leu	Asp	Glu
				85					90					95	
Gln	Glu	Gly	Val	Lys	Ser	Gly	Met	Tyr	Val	Val	Ile	Glu	Val	Lys	Val
			100					105					110		
Ala	Thr	Gln	Glu	Gly	Lys	Glu	Ile	Thr	Cys	Arg	Ser	Tyr	Leu	Met	Thr
		115					120					125			
Asn	Tyr	Glu	Ser	Ala	Pro	Pro	Ser	Pro	Gln	Tyr	Lys	Lys	Ile	Ile	Cys
	130					135					140				
Met	Gly	Ala	Lys	Glu	Asn	Gly	Leu	Pro	Leu	Glu	Tyr	Gln	Glu	Lys	Leu
145					150					155					160
Lys	Ala	Ile	Glu	Pro	Asn	Asp	Tyr	Thr	Gly	Lys	Val	Ser	Glu	Glu	Ile
				165					170					175	
Glu	Asp	Ile	Ile	Lys	Lys	Gly	Glu	Thr	Gln	Thr	Leu				
			180					185							

<210> 9
 <211> 965
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> CDS
 <222> (79)..._(232)
 <400> 9

gtagtgggccc	ccagtgttgc	gctctctggc	cgttccttac	actttgcttc	aggctccagt	60
gcagggggcgt	agtgggat	atg gcc aac tcg ggc tgc aag gac gtc acg ggt	111			
	Met Ala Asn Ser Gly Cys Lys Asp Val Thr Gly					
	1 5 10					
cca gat gag gag agt ttt ctg tac ttt gcc tac ggc agc aac ctg ctg	159					
Pro Asp Glu Glu Ser Phe Leu Tyr Phe Ala Tyr Gly Ser Asn Leu Leu						
	15 20 25					
aca gag agy atc cac ctc cga aac ccc tcg gcg gcg ttc ttc tgt gtg	207					
Thr Glu Arg Ile His Leu Arg Asn Pro Ser Ala Ala Phe Phe Cys Val						
	30 35 40					
gcc cgc ctg cag gca aga agg ggt t aaaagtggaa tgtatgttgt	252					
Ala Arg Leu Gln Ala Arg Arg Gly						
	45 50					
aatagaagtt aaagttgcaa ctcaagaagg aaaagaaata acctgtcgaa gttatctgat	312					
gacaaattac gaaagtgcct ccccatcccc acagtataaa aagattatct gcattgggtgc	372					
aaaagaaaaat ggtttgcccc tggagtatca agagaagtta aaagcaatag aaccacaaatga	432					
ctatacagga aaggtctcag aagaaattga agacatcatc aaaaaggggg aaacacaaac	492					
tctttagaac ataacagaat atatctaagg gtattctatg tgctaataata aaatattttt	552					
aacacttgag aacagggatc tgggggatct ccacgtttga tccattttca gcagtgetct	612					
gaaggagtat cttacttggg tgattccttg tttttagact ataaaaagaa actgggatag	672					
gagttagaca atttaaaagg ggtgtatgag ggctgaaat atgtgacaaa tgaatgtgag	732					
taccccttct gtgaacactg aaagctattc tcttgaattg atcttaagtg tctccttgct	792					
ctggtaaaag atagatttgt agctcacttg atgatgggtc tgggtgaattg ctctgctctg	852					
tctgagattt ttaaaaatca gcttaatgag agtaatctgc agacaattga taataacatt	912					

ttgaaaattg gaaagatggt atactgtttt tagaggaata aacgtatttg tgg	965
--	-----

<210> 10

<211> 51

<212> PRT

<213> Homo sapiens

<400> 10

Met Ala Asn Ser Gly Cys Lys Asp Val Thr Gly Pro Asp Glu Glu Ser	
1 5 10 15	
Phe Leu Tyr Phe Ala Tyr Gly Ser Asn Leu Leu Thr Glu Arg Ile His	
20 25 30	
Leu Arg Asn Pro Ser Ala Ala Phe Phe Cys Val Ala Arg Leu Gln Ala	
35 40 45	
Arg Arg Gly	
50	

<210> 11
 <211> 658
 <212> DNA
 <213> Homo sapiens
 <400> 11

tgcgagaggag	ctgacagcct	tcctgcaccc	cgaggagtga	ggtggggccac	tgggtgctgc	60
cccctgoccca	ggaccagccc	ctggttggaa	ccaaccacct	gctgcacgag	agcgacacgg	120
acaaggacgg	gcggctgagc	aaagcggaaa	tcctgggtaa	ttggaacatg	tttgtgggca	180
gtcaggccac	caactatggt	gaggacctga	cccgccacca	cgatgagctg	tgagccccgc	240
gcacctgcca	cagcctcaga	ggccccgcaca	atgaccggag	gaggggcgcg	tgtggtctgg	300
ccccctccct	gtccaggccc	cgcaggaggc	agatgcagtc	ccaggcatcc	tcctgccccct	360
gggctctcag	ggaccccctg	ggtcggtctc	tgtccctgtc	acacccccaa	ccccagggag	420
gggctgtcat	agtcccagag	gataagcaat	acctatctct	gactgagtct	cccagcccag	480
acccaggggac	cctggcccca	agctcagctc	taagaaccgc	caccaccccc	tccagctcca	540
aatctgagcc	tccaccacat	agactgaaac	tcccttgccc	ccagccctct	cctgcctggc	600
ctggcctggg	acacctctct	tctgccaggga	ggcaataaaa	gccagcgccg	ggaccttg	658

<210> 12
 <211> 1507
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> misc feature
 <222> 1047, 1301
 <223> n = A, _T, _C or G
 <400> 12

ggaacgcaga	gcgagcgtg	gagagcggag	cgaagctgga	taacagggga	ccgatgatgt	60
ggcgaccatc	agttctgctg	cttctgttgc	tactgaggca	cggggcccag	gggaagccat	120
ccccagacgc	aggccctcat	ggccagggga	gggtgcacca	ggcgcccccc	ctgagcgacg	180
ctcccatga	tgacgccac	gggaacttcc	agtacgacca	tgaggctttc	ctgggacggg	240
aagtggccaa	ggaattcgac	caactcacc	cagaggaaag	ccaggcccg	ctggggcgga	300
tccgtggaccg	catggaccgc	gcgggggacg	gcgacggctg	ggtgtcgtg	gccgagcttc	360
gcgcgtggat	cgcgcacacg	cagcagcggc	acatacggga	ctcgggtgagc	gcggcctggg	420
acacgtacga	cacggaccgc	gacgggcgtg	tgggttggga	ggagctgcgc	aacgccacct	480
atggccacta	cgcgcccgg	gaagaatttc	atgacgtgga	ggatgcagag	acctacaaa	540
agatgctggc	tccggacgag	cggcgtttcc	gggtggccga	ccaggatggg	gactcgatgg	600
ccactcgaga	ggagctgaca	gccttcctgc	accccgagga	gttccctcac	atgcgggaca	660
togtgattgc	tgaaacctg	gaggacctgg	acagaaacaa	agatggctat	gtccagggtg	720
aggagtagat	cgcggatctg	tactcagccg	agcctgggga	ggaggagccg	gcgtgggtgc	780
agacggagag	gcagcagttc	cgggaacttc	gggatctgaa	caaggatggg	cacctggatg	840
ggagttaggt	gggccactgg	gtgctgcccc	ctgcccagga	ccagcccctg	gtggaagcca	900
accacctgct	gcacgaragc	gacacggaca	aggaygggcg	gctgagcaaa	gcgsaaatcc	960
tgggtaatgt	gaacatgttt	gtgggcagtc	aggccaccaa	ctatggygag	gacctgaccc	1020
ggcaccacga	tgagctgtga	gcmccngca	cctgccacag	cctcagaggc	ccgcacaatg	1080

accggaggag	gggcccgtgt	ggtctggccc	cctccctgtc	caggccccgc	aggaggcaga	1140
tgcagtccca	ggcatccctcc	tkcccctggg	ctctcaggga	ccccctgggt	cggtttctgt	1200
ccctgtcaca	cccccaaccc	cagggagggg	ctgtcatagt	cccagaggat	aagcaatacc	1260
tattttctgac	tgagtctccc	agcccagacc	cagggaccct	nggcccccaag	ctcagctcta	1320
agaaccgccc	caacccctcc	agctccaaat	ctgagcctcc	accacataga	ctgaaactcc	1380
cctggcccca	gcccctctct	gcctggcctg	gcctgggaca	cctcctctct	gccaggagggc	1440
aataaaaagcc	agcgccggga	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1500
aaaaaan						1507

<210> 13

<211> 661

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (79)..._(376)

<400> 13

ggcgcggtgc	agggctcttta	agaacgaacg	gcttggggcg	ggactgggtat	ccggggactg	60
tgacttgcag	ggtccgcc	atg gag cca gag cag atg ctg gag gga caa acg				111
		Met Glu Pro Glu Gln Met Leu Glu Gly Gln Thr				
		1 5 10				
cag gtt gca gaa aat cct cac tct gag tac ggt ctc aca gac aac gtt						159
Gln Val Ala Glu Asn Pro His Ser Glu Tyr Gly Leu Thr Asp Asn Val						
		15 20 25				
gag aga ata gta gaa aat gag aag att aat gca gaa aag tca tca aag						207
Glu Arg Ile Val Glu Asn Glu Lys Ile Asn Ala Glu Lys Ser Ser Lys						
		30 35 40				
cag aag gta gat ctc cag tct ttg cca act cgt gcc tac ctg gat cag						255
Gln Lys Val Asp Leu Gln Ser Leu Pro Thr Arg Ala Tyr Leu Asp Gln						
		45 50 55				
aca gtt gtg cct atc tta tta cag gga ctt gct gtg ctt gca aag gaa						303
Thr Val Val Pro Ile Leu Leu Gln Gly Leu Ala Val Leu Ala Lys Glu						
		60 65 70 75				
aga cca cca aat ccc att gaa ttt cta gca tct tat ctt tta aaa aac						351
Arg Pro Pro Asn Pro Ile Glu Phe Leu Ala Ser Tyr Leu Leu Lys Asn						
		80 85 90				
aag gca cag ttt gaa gat cga aac t gacttaatgg gaagaacaga						396
Lys Ala Gln Phe Glu Asp Arg Asn						
		95				
aaaatttagt	tgctactgta	gattttacatg	attaagaggc	agctttaatt	gccatgatca	456
ttccctcttt	ttggatgtat	aagaacottc	cggacaacag	aacctatttc	tggaaattgca	516
gaagataaca	tatttccott	attttgattt	aatcaccata	aaccatacct	atttaattgag	576
tgtattctgt	gcaatttttt	tctcagattg	tctttaactt	tgtttttaaa	atgaccttca	636
aaataaactg	tcaaaacacc	atttat				661

<210> 14
 <211> 99
 <212> PRT
 <213> Homo sapiens
 <400> 14

Met	Glu	Pro	Glu	Gln	Met	Leu	Glu	Gly	Gln	Thr	Gln	Val	Ala	Glu	Asn	
1				5					10					15		
Pro	His	Ser	Glu	Tyr	Gly	Leu	Thr	Asp	Asn	Val	Glu	Arg	Ile	Val	Glu	
			20					25					30			
Asn	Glu	Lys	Ile	Asn	Ala	Glu	Lys	Ser	Ser	Lys	Gln	Lys	Val	Asp	Leu	
		35					40					45				
Gln	Ser	Leu	Pro	Thr	Arg	Ala	Tyr	Leu	Asp	Gln	Thr	Val	Val	Pro	Ile	
	50					55					60					
Leu	Leu	Gln	Gly	Leu	Ala	Val	Leu	Ala	Lys	Glu	Arg	Pro	Pro	Asn	Pro	
65					70					75					80	
Ile	Glu	Phe	Leu	Ala	Ser	Tyr	Leu	Leu	Lys	Asn	Lys	Ala	Gln	Phe	Glu	
				85					90					95		
Asp	Arg	Asn														

<210> 15
 <211> 1507
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> misc__feature
 <222> 1047, 1301
 <223> n = A,_T,_C or G
 <400> 15

ggaacgcaga	gcggagcgtg	gagagcggag	cgaagctgga	taacagggga	ccgatgatgt	60
ggcgaccatc	agttctgtctg	cttctgtttgc	tactgaggca	cggggcccag	gggaagccat	120
ccccagacgc	aggccctcat	ggccagggga	gggtgcacca	ggcggccccc	ctgagcgacg	180
ctcccatga	tgaagccac	gggaacttcc	agtaagacca	tgaggctttc	ctgggacggg	240
aagtggccaa	ggaattogac	caactcaccc	cagaggaaag	ccaggcccg	ctggggcgga	300
tcgtggaccg	catggaccgc	gcgggggacg	gcgacggctg	ggtgtcgtg	gccgagcttc	360
gcgcgtggat	cgcgcacacg	cagcagcggc	acatacggga	ctcggtgagc	gcggcctggg	420
acacgtacga	cacggaccgc	gacgggcgtg	tgggttggga	ggagctgcgc	aacgccacct	480
atggccacta	cgcgcccgg	gaagaatttc	atgacgtgga	ggatgcagag	acctacaaaa	540
agatgctggc	tcgggacgag	cggcgtttcc	gggtggccga	ccaggatggg	gactcgatgg	600
ccactcgaga	ggagctgaca	gccttcctgc	accccgagga	gttcctcac	atgcgggaca	660
tcgtgattgc	tgaaccctg	gaggacctgg	acagaaacaa	agatggctat	gtccaggtgg	720
aggagtacat	cgcggatctg	tactcagccg	agcctgggga	ggaggagccg	gcgtgggtgc	780
agacggagag	gcagcagttc	cgggacttcc	gggatctgaa	caaggatggg	cacctggatg	840
ggagtggagt	gggccactgg	gtgctgcccc	ctgccagga	ccagccctg	gtggaagcca	900
accacctgct	gcacgaragc	gacacggaca	aggayggcg	gctgagcaaa	gcgsaaatcc	960
tgggtaattg	gaacatgttt	gtgggcagtc	aggccaccaa	ctatggygag	gacctgacct	1020
ggcaccacga	tgagctgtga	gcmccngca	cctgccacag	cctcagaggc	ccgcacaatg	1080
accggaggag	gggcccgtgt	ggtctggccc	cctccctgtc	caggccccgc	aggaggcaga	1140
tgcagtccca	ggcatcctcc	tkcccctggg	ctctcagggg	ccccctgggt	cggcttctgt	1200
ccctgtcaca	cccccaaccc	cagggagggg	ctgtcatagt	cccagaggat	aagcaatacc	1260
tatttctgac	tgagtctccc	agcccagacc	cagggacctc	nggccccaa	ctcagctcta	1320
agaaccgccc	caacccctcc	agctccaaat	ctgagcctcc	accacataga	ctgaaactcc	1380
cctggcccca	gccctctcct	gcctggcctg	gcctgggaca	cctcctctct	gccaggaggc	1440
aataaaagcc	agcgccggga	aaaaaaaaa	aaaaaaaaa	aaaaaaaaa	aaaaaaaaa	1500
aaaaaan						1507

<210> 16

<211> 716

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (16)..._(538)

<400> 16

ctcgatcgaa	gcgag	atg	gcg	gac	gtg	cta	gat	ctt	cac	gag	gct	ggg	ggc	51
		Met	Ala	Asp	Val	Leu	Asp	Leu	His	Glu	Ala	Gly	Gly	
		1				5						10		

gaa gat ttc gcc atg gat gag gat ggg gac gag agc att cac aaa ctg	99
Glu Asp Phe Ala Met Asp Glu Asp Gly Asp Glu Ser Ile His Lys Leu	
15 20 25	
aaa gaa aaa gcg aag aaa cgg aag ggt cgc ggc ttt ggc tcc gaa gag	147
Lys Glu Lys Ala Lys Lys Arg Lys Gly Arg Gly Phe Gly Ser Glu Glu	
30 35 40	
ggg tcc cga gcg cgg atg cgt gag gat tat gac agc gtg gag cag gat	195
Gly Ser Arg Ala Arg Met Arg Glu Asp Tyr Asp Ser Val Glu Gln Asp	
45 50 55 60	
ggc gat gaa ccc gga cca caa cgc tct gtt gaa ggc tgg att ctc ttt	243
Gly Asp Glu Pro Gly Pro Gln Arg Ser Val Glu Gly Trp Ile Leu Phe	
65 70 75	
gta act gga gtc cat gag gaa gcc acc gaa gaa gac ata cac gac aaa	291
Val Thr Gly Val His Glu Glu Ala Thr Glu Glu Asp Ile His Asp Lys	
80 85 90	
ttc gca gaa tat ggg gaa att aaa aac att cat ctc aac ctc gac agg	339
Phe Ala Glu Tyr Gly Glu Ile Lys Asn Ile His Leu Asn Leu Asp Arg	
95 100 105	
cga aca gga tat ctg aag ggg tat act cta gtt gaa tat gaa aca tac	387
Arg Thr Gly Tyr Leu Lys Gly Tyr Thr Leu Val Glu Tyr Glu Thr Tyr	
110 115 120	
aag gaa gcc cag gct gct atg gag gga ctc aat ggc cag gat ttg atg	435
Lys Glu Ala Gln Ala Ala Met Glu Gly Leu Asn Gly Gln Asp Leu Met	
125 130 135 140	
gga cag ccc atc agc gtt gac tgg tgt ttt gtt cgg ggt cca cca aaa	483
Gly Gln Pro Ile Ser Val Asp Trp Cys Phe Val Arg Gly Pro Pro Lys	
145 150 155	
ggc aag agg aga ggt ggc cga aga cgc agc aga agt cca gac cgg aga	531
Gly Lys Arg Arg Gly Gly Arg Arg Arg Ser Arg Ser Pro Asp Arg Arg	
160 165 170	
cgt cgc t gacaggctcct ctgttgtcca ggtgttctct tcaagattcc atttgaccat	588
Arg Arg	
gcagccttgg acaaatagga ctgggggtgga acttgctgtg tttatattta atctcttacc	648
gtatatgcgt agtatattgag ttgcgaataa atgttccatt ttgttttcta caaaaaaaaa	708
aaaaaaaa	716

<210> 17

<211> 174

<212> PRT

<213> Homo sapiens

<400> 17

Met	Ala	Asp	Val	Leu	Asp	Leu	His	Glu	Ala	Gly	Gly	Glu	Asp	Phe	Ala
1				5					10					15	
Met	Asp	Glu	Asp	Gly	Asp	Glu	Ser	Ile	His	Lys	Leu	Lys	Glu	Lys	Ala
			20					25					30		
Lys	Lys	Arg	Lys	Gly	Arg	Gly	Phe	Gly	Ser	Glu	Glu	Gly	Ser	Arg	Ala
		35					40					45			
Arg	Met	Arg	Glu	Asp	Tyr	Asp	Ser	Val	Glu	Gln	Asp	Gly	Asp	Glu	Pro
	50					55					60				

Gly	Pro	Gln	Arg	Sex	Val	Glu	Gly	Trp	Ile	Leu	Phe	Val	Thr	Gly	Val
65					70					75					80
His	Glu	Glu	Ala	Thr	Glu	Glu	Asp	Ile	His	Asp	Lys	Phe	Ala	Glu	Tyr
				85					90					95	
Gly	Glu	Ile	Lys	Asn	Ile	His	Leu	Asn	Leu	Asp	Arg	Arg	Thr	Gly	Tyr
			100					105					110		
Leu	Lys	Gly	Tyr	Thr	Leu	Val	Glu	Tyr	Glu	Thr	Tyr	Lys	Glu	Ala	Gln
		115					120					125			
Ala	Ala	Met	Glu	Gly	Leu	Asn	Gly	Gln	Asp	Leu	Met	Gly	Gln	Pro	Ile
	130					135					140				
Ser	Val	Asp	Trp	Cys	Phe	Val	Arg	Gly	Pro	Pro	Lys	Gly	Lys	Arg	Arg
145				150					155					160	
Gly	Gly	Arg	Arg	Arg	Ser	Arg	Ser	Pro	Asp	Arg	Arg	Arg	Arg		
				165					170						

<210> 18

<211> 763

<212> DNA

<213> Homo sapiens

 $\langle 220 \rangle$

<221> CDS

 $\langle 222 \rangle \quad (2) \dots (551)$

<400> 18

c atg gcc aag cag tgt ggg gtg cgc ctg agc ggg gaa gcc cgc aaa cag	49
Met Ala Lys Pro Cys Gly Val Arg Leu Ser Gly Glu Ala Arg Lys Gln	
1 5 10 15	
gtg gag gtc ttc aga cag aat ctt ttc cag gag gct gag gaa ttc ctc	97
Val Glu Val Phe Arg Gln Asn Leu Phe Gln Glu Ala Glu Glu Phe Leu	
20 25 30	
tac aga ttc ttg cca cag aaa atc ata tac ctg aat cag ctc ttg caa	145
Tyr Arg Phe Leu Pro Gln Lys Ile Ile Tyr Leu Asn Gln Leu Leu Gln	
35 40 45	
gag gac tcc ctc aat gtg gct gac ttg act tcc ctc cgg gcc cca ctg	193
Glu Asp Ser Leu Asn Val Ala Asp Leu Thr Ser Leu Arg Ala Pro Leu	
50 55 60	
gac atc ccc atc cca gac cct cca ccc aag gat gat gag atg gaa aca	241
Asp Ile Pro Ile Pro Asp Pro Pro Pro Lys Asp Asp Glu Met Glu Thr	
65 70 75 80	
gat aag cag gag aag aaa gaa gtc cct aag tgt gga ttt ctc cct ggg	289
Asp Lys Gln Glu Lys Lys Glu Val Pro Lys Cys Gly Phe Leu Pro Gly	
85 90 95	
aat gag aaa gtc ctg tcc ctg ctt gcc ctg gtt aag cca gaa gtc tgg	337
Asn Glu Lys Val Leu Ser Leu Leu Ala Leu Val Lys Pro Glu Val Trp	
100 105 110	
act ctc aaa gag aaa tgc att ctg gtg att aca tgg atc caa cac ctg	385
Thr Leu Lys Glu Lys Cys Ile Leu Val Ile Thr Trp Ile Gln His Leu	
115 120 125	
atc ccc aag att gaa gat gga aat gat ttt ggg gta gca atc cag gag	433
Ile Pro Lys Ile Glu Asp Gly Asn Asp Phe Gly Val Ala Ile Gln Glu	
130 135 140	
aag gtg ctg gag agg gtg aat gcc gtc aag acc aaa gtg aag ctt tcc	481
Lys Val Leu Glu Arg Val Asn Ala Val Lys Thr Lys Val Lys Leu Ser	
145 150 155 160	
aga caa cca ttt cca agt act tct cag aac gtg ggg atg ctg tgg cca	529
Arg Gln Pro Phe Pro Ser Thr Ser Gln Asn Val Gly Met Leu Trp Pro	
165 170 175	
agg cct cca agg aga ctc atg t aatggattac cgggccttgg tgcattgagcg	581
Arg Pro Pro Arg Arg Leu Met	
180	
agatgaggca gcctatgggg agctcagggc catggtgctg gacctgaggg ccttctatgc	641
tgagctttat catatcatca gcagcaacct ggagaaaatt gtcaacccaa aggggtgaaga	701
aaagccatct atgtactgaa cccgggacta gaaggaaaat aatgatcta tatgtttgt	761
gg	763

<210> 19

<211> 183

<212> PRT

<213> Homo sapiens

<400> 19

Met	Ala	Lys	Pro	Cys	Gly	Val	Arg	Leu	Ser	Gly	Glu	Ala	Arg	Lys	Gln	
1				5					10					15		
Val	Glu	Val	Phe	Arg	Gln	Asn	Leu	Phe	Gln	Glu	Ala	Glu	Glu	Phe	Leu	
			20					25					30			
Tyr	Arg	Phe	Leu	Pro	Gln	Lys	Ile	Ile	Tyr	Leu	Asn	Gln	Leu	Leu	Gln	
		35					40					45				
Glu	Asp	Ser	Leu	Asn	Val	Ala	Asp	Leu	Thr	Ser	Leu	Arg	Ala	Pro	Leu	
	50					55					60					
Asp	Ile	Pro	Ile	Pro	Asp	Pro	Pro	Pro	Lys	Asp	Asp	Glu	Met	Glu	Thr	
65					70					75				80		
Asp	Lys	Gln	Glu	Lys	Lys	Glu	Val	Pro	Lys	Cys	Gly	Phe	Leu	Pro	Gly	
				85					90					95		
Asn	Glu	Lys	Val	Leu	Ser	Leu	Leu	Ala	Leu	Val	Lys	Pro	Glu	Val	Trp	
			100					105					110			
Thr	Leu	Lys	Glu	Lys	Cys	Ile	Leu	Val	Ile	Thr	Trp	Ile	Gln	His	Leu	
	115					120						125				
Ile	Pro	Lys	Ile	Glu	Asp	Gly	Asn	Asp	Phe	Gly	Val	Ala	Ile	Gln	Glu	
	130					135					140					
Lys	Val	Leu	Glu	Arg	Val	Asn	Ala	Val	Lys	Thr	Lys	Val	Lys	Leu	Ser	
145				150						155				160		
Arg	Gln	Pro	Phe	Pro	Ser	Thr	Ser	Gln	Asn	Val	Gly	Met	Leu	Trp	Pro	
			165					170						175		
Arg	Pro	Pro	Arg	Arg	Leu	Met										
			180													

<210> 20

<211> 790

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (240)..._(585)

<400> 20

acgtttttaca	gtcttttaatt	aagcacataa	aactgtacta	tttaatatat	ttctccatga	60
aactttttgtg	aaattcagat	cgcagtgtgt	catttacaaa	tcttttgtct	ttctcttggt	120
catctacacc	ttttgcacag	ttcttgaaga	caaogtcato	atcccacctt	cttttaactt	180
tgaagttggc	ctgaggctgg	gatgggccag	tgagattaag	gagaggggtt	cogctcaga	239
atg ttt tcc	ata cga atc	ctc tct tct	tca gct ttt	tgt tct tgt	tcc	287
Met	Phe	Ser	Ile	Arg	Ile	Leu
1			5			10
						15

ttc ctg gcc tgc tct tca gct ctt tct ttt tta att ttt tcc agt tct	335
Phe Leu Ala Cys Ser Ser Ala Leu Ser Phe Leu Ile Phe Ser Ser Ser	
20 25 30	
gca aga aga gct gca gta tca tca tca tca ctt tct tct tca aaa tct	383
Ala Arg Arg Ala Ala Val Ser Ser Ser Ser Leu Ser Ser Ser Lys Ser	
35 40 45	
tca tct tcc tca tct gtt aga ggg tca tct gca tca agg ttg gcg gca	431
Ser Ser Ser Ser Ser Val Arg Gly Ser Ser Ala Ser Arg Leu Ala Ala	
50 55 60	
gga atc tgg tct aac cgt ggc ttt ttt gac act gaa gag gag gtt gta	479
Gly Ile Trp Ser Asn Arg Gly Phe Phe Asp Thr Glu Glu Glu Val Val	
65 70 75 80	
tgt tct cgg gtt gga cga tcc cta ttt ttc tct ctt gca gca gct ctc	527
Cys Ser Arg Val Gly Arg Ser Leu Phe Phe Ser Leu Ala Ala Ala Leu	
85 90 95	
tct ctt tct tcc aac tct ctc ctg aag tca cgg tta cga acc tct tca	575
Ser Leu Ser Ser Asn Ser Leu Leu Lys Ser Arg Leu Arg Thr Ser Ser	
100 105 110	
ggg gca tcc t gagtagctcg tctgtatttt atctttgtat gagagggtag	625
Gly Ala Ser	
115	
gtctctgctt gaatactgct ttgaaagttg gtcacaaatca ccttctcatt ttcccttcc	685
acctctggca gggtcaaagg ttggcctggc tgctgttgtc atcttttatg actggccgag	745
gtccgatgca gcaggctccg aagatcatatc agaagccatt accac	790

<210> 21

<211> 115

<212> PRT

<213> Homo sapiens

<400> 21

```

Met Phe Ser Ile Arg Ile Leu Ser Ser Ser Ala Phe Cys Ser Cys Ser
 1           5           10           15
Phe Leu Ala Cys Ser Ser Ala Leu Ser Phe Leu Ile Phe Ser Ser Ser
      20           25           30
Ala Arg Arg Ala Ala Val Ser Ser Ser Ser Leu Ser Ser Ser Lys Ser
      35           40           45
Ser Ser Ser Ser Ser Val Arg Gly Ser Ser Ala Ser Arg Leu Ala Ala
      50           55           60
Gly Ile Trp Ser Asn Arg Gly Phe Phe Asp Thr Glu Glu Val Val
      65           70           75           80
Cys Ser Arg Val Gly Arg Ser Leu Phe Phe Ser Leu Ala Ala Ala Leu
      85           90           95
Ser Leu Ser Ser Asn Ser Leu Leu Lys Ser Arg Leu Arg Thr Ser Ser
      100          105          110
Gly Ala Ser
      115

```

<210> 22
<211> 1939
<212> DNA
<213> Homo sapiens
<220>
<221> CDS
<222> (53)..._(1700)
<400> 22

gtggcgccag cggaggcagg ttgctgtgtt tgtgcttctt tctacagcca at atg aaa	58
Met Lys	
1	
agg cct aag tta aag aaa gca agt aaa cgc atg acc tgc cat aag cgg	106
Arg Pro Lys Leu Lys Lys Ala Ser Lys Arg Met Thr Cys His Lys Arg	
5 10 15	
tat aaa atc caa aaa aag gtt cga gaa cat cat cga aaa tta aga aag	154
Tyr Lys Ile Gln Lys Lys Val Arg Glu His His Arg Lys Leu Arg Lys	
20 25 30	
gag gct aaa aag cag ggt cac aag aag cct agg aaa gac cca gga gtt	202
Glu Ala Lys Lys Gln Gly His Lys Lys Pro Arg Lys Asp Pro Gly Val	
35 40 45 50	
cca aac agt gct ccc ttt aag gag gct ctt ctt agg gaa gct gag cta	250
Pro Asn Ser Ala Pro Phe Lys Glu Ala Leu Leu Arg Glu Ala Glu Leu	
55 60 65	
agg aaa cag agg ctt gaa gaa cta aaa cag cag cag aaa ctt gac agg	298
Arg Lys Gln Arg Leu Glu Glu Leu Lys Gln Gln Gln Lys Leu Asp Arg	
70 75 80	
cag aag gaa cta gaa aag aaa aga aaa ctt gaa act aat cct gat att	346
Gln Lys Glu Leu Glu Lys Lys Arg Lys Leu Glu Thr Asn Pro Asp Ile	
85 90 95	
aag cca tca aat gtg gaa cct atg gaa aag gag ttt ggg ctt tgc aaa	394
Lys Pro Ser Asn Val Glu Pro Met Glu Lys Glu Phe Gly Leu Cys Lys	
100 105 110	
act gag aac aaa gcc aag tgc ggc aaa cag aat tca aag aag ctg tac	442
Thr Glu Asn Lys Ala Lys Ser Gly Lys Gln Asn Ser Lys Lys Leu Tyr	
115 120 125 130	
tgc caa gaa ctt aaa aag gtg att gaa gcc tcc gat gtt gtc cta gag	490
Cys Gln Glu Leu Lys Lys Val Ile Glu Ala Ser Asp Val Val Leu Glu	
135 140 145	
gtg ttg gat gcc aga gat cct ctt ggt tgc aga tgt cct cag gta gaa	538
Val Leu Asp Ala Arg Asp Pro Leu Gly Cys Arg Cys Pro Gln Val Glu	
150 155 160	
gag gcc att gtc cag agt gga cag aaa aag ctg gta ctt ata tta aat	586
Glu Ala Ile Val Gln Ser Gly Gln Lys Lys Leu Val Leu Ile Leu Asn	
165 170 175	
aaa tca gat ctg gta cca aag gag aat ttg gag agc tgg cta aat tat	634
Lys Ser Asp Leu Val Pro Lys Glu Asn Leu Glu Ser Trp Leu Asn Tyr	
180 185 190	
ttg aag aaa gaa ttg cca aca gtg gtg ttc aga gcc tca aca aaa cca	682
Leu Lys Lys Glu Leu Pro Thr Val Val Phe Arg Ala Ser Thr Lys Pro	
195 200 205 210	
aag gat aaa ggg aag ata acc aag cgt gtg aag gca aag aag aat gct	730
Lys Asp Lys Gly Lys Ile Thr Lys Arg Val Lys Ala Lys Lys Asn Ala	
215 220 225	

gct cca ttc aga agt gaa gtc tgc ttt ggg aaa gag ggc ctt tgg aaa	778
Ala Pro Phe Arg Ser Glu Val Cys Phe Gly Lys Glu Gly Leu Trp Lys	
230 235 240	
ctt ctt gga ggt ttt cag gaa act tgc agc aaa gcc att cgg gtt gga	826
Leu Leu Gly Gly Phe Gln Glu Thr Cys Ser Lys Ala Ile Arg Val Gly	
245 250 255	
gta att ggt ttc cca aat gtg ggg aaa agc agc att atc aat agc tta	874
Val Ile Gly Phe Pro Asn Val Gly Lys Ser Ser Ile Ile Asn Ser Leu	
260 265 270	
aaa caa gaa cag atg tgt aat gtt ggt gta tcc atg ggg ctt aca agg	922
Lys Gln Glu Gln Met Cys Asn Val Gly Val Ser Met Gly Leu Thr Arg	
275 280 285 290	
agc atg caa gtt gtc ccc ttg gac aaa cag atc aca atc ata gat agt	970
Ser Met Gln Val Val Pro Leu Asp Lys Gln Ile Thr Ile Ile Asp Ser	
295 300 305	
ccg agc ttc atc gta tct cca ctt aat tcc tcc tct gcg ctt gct ctg	1018
Pro Ser Phe Ile Val Ser Pro Leu Asn Ser Ser Ser Ala Leu Ala Leu	
310 315 320	
cga agt cca gca agt att gaa gta gta aaa ccg atg gag gct gcc agt	1066
Arg Ser Pro Ala Ser Ile Glu Val Val Lys Pro Met Glu Ala Ala Ser	
325 330 335	
gcc atc ctt tcc cag gct gat gct cga cag gta gta ctg aaa tat act	1114
Ala Ile Leu Ser Gln Ala Asp Ala Arg Gln Val Val Leu Lys Tyr Thr	
340 345 350	
gtc cca ggc tac agg aat tct ctg gaa ttt ttt act atg ctt gct cag	1162
Val Pro Gly Tyr Arg Asn Ser Leu Glu Phe Phe Thr Met Leu Ala Gln	
355 360 365 370	
aga aga ggt atg cac caa aaa ggt gga atc cca aat gtt gaa ggt gct	1210
Arg Arg Gly Met His Gln Lys Gly Gly Ile Pro Asn Val Glu Gly Ala	
375 380 385	
gcc aaa ctg ctg tgg tct gag tgg aca ggt gcc tca tta gct tac tat	1258
Ala Lys Leu Leu Trp Ser Glu Trp Thr Gly Ala Ser Leu Ala Tyr Tyr	
390 395 400	
tgc cat ccc oct aca tct tgg act cct cct cca tat ttt aat gag agt	1306
Cys His Pro Pro Thr Ser Trp Thr Pro Pro Pro Tyr Phe Asn Glu Ser	
405 410 415	
att gtg gta gac atg aaa agc ggc ttc aat ctg gaa gaa ctg gaa aag	1354
Ile Val Val Asp Met Lys Ser Gly Phe Asn Leu Glu Glu Leu Glu Lys	
420 425 430	
aac aat gca cag agc ata aga gcc atc aag ggc cct cat ttg gcc aat	1402
Asn Asn Ala Gln Ser Ile Arg Ala Ile Lys Gly Pro His Leu Ala Asn	
435 440 445 450	
agc atc ctt ttc cag tct tcc ggt ctg aca aat gga ata ata gaa gaa	1450
Ser Ile Leu Phe Gln Ser Ser Gly Leu Thr Asn Gly Ile Ile Glu Glu	
455 460 465	
aag gac ata cat gaa gaa ttg cca aaa cgg aaa gaa agg aag cag gag	1498
Lys Asp Ile His Glu Glu Leu Pro Lys Arg Lys Glu Arg Lys Gln Glu	
470 475 480	

gag agg gag gat gac aaa gac agt gac cag gas act gtt gat gaa gaa	1546
Glu Arg Glu Asp Asp Lys Asp Ser Asp Gln Glu Thr Val Asp Glu Glu	
485 490 495	
gtt gat gaa aac agc tca ggc atg ttt gct gca gaa gag aca ggg gag	1594
Val Asp Glu Asn Ser Ser Gly Met Phe Ala Ala Glu Glu Thr Gly Glu	
500 505 510	
gca ctg tct gag gag act aca gca ggt gaa cag tct aca agg tct ttt	1642
Ala Leu Ser Glu Glu Thr Thr Ala Gly Glu Gln Ser Thr Arg Ser Phe	
515 520 525 530	
atc ttg gat aaa atc att gaa gag gat gat gct tat gac ttc agt aca	1690
Ile Leu Asp Lys Ile Ile Glu Glu Asp Asp Ala Tyr Asp Phe Ser Thr	
535 540 545	
gat tat gtg t aacagaacaa tggcttttta tgattttttt ttttaacatt	1740
Asp Tyr Val	
tttagcagac tgctaaactg ttctctgtat aagttatggt atgcatgagc tgtgtaaatt	1800
ttgtgaatat gtattatatt aaaccaggc aacttggaat ccctaaattc tgtaaaaaga	1860
caattcatct cattgtgagt ggaagtagtt atctggaata aaaaaagaag atacctattg	1920
aaaaaaaaa aaaaaaaaaa	1939

<210> 23

<211> 549

<212> PRT

<213> Homo sapiens

<400> 23

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

Val	Gly	Val	Ile	Gly	Phe	Pro	Asn	Val	Gly	Lys	Ser	Ser	Ile	Ile	Asn
			260					265					270		
Ser	Leu	Lys	Gln	Glu	Gln	Met	Cys	Asn	Val	Gly	Val	Ser	Met	Gly	Leu
		275					280					285			
Thr	Arg	Ser	Met	Gln	Val	Val	Pro	Leu	Asp	Lys	Gln	Ile	Thr	Ile	Ile
	290					295					300				
Asp	Ser	Pro	Ser	Phe	Ile	Val	Ser	Pro	Leu	Asn	Ser	Ser	Ser	Ala	Leu
305				310						315					320
Ala	Leu	Arg	Ser	Pro	Ala	Ser	Ile	Glu	Val	Val	Lys	Pro	Met	Glu	Ala
			325					330						335	
Ala	Ser	Ala	Ile	Leu	Ser	Gln	Ala	Asp	Ala	Arg	Gln	Val	Val	Leu	Lys
		340						345				350			
Tyr	Thr	Val	Pro	Gly	Tyr	Arg	Asn	Ser	Leu	Glu	Phe	Phe	Thr	Met	Leu
	355					360						365			
Ala	Gln	Arg	Arg	Gly	Met	His	Gln	Lys	Gly	Gly	Ile	Pro	Asn	Val	Glu
	370					375					380				
Gly	Ala	Ala	Lys	Leu	Leu	Trp	Ser	Glu	Trp	Thr	Gly	Ala	Ser	Leu	Ala
385				390						395					400
Tyr	Tyr	Cys	His	Pro	Pro	Thr	Ser	Trp	Thr	Pro	Pro	Pro	Tyr	Phe	Asn
			405					410						415	
Glu	Ser	Ile	Val	Val	Asp	Met	Lys	Ser	Gly	Phe	Asn	Leu	Glu	Glu	Leu
		420						425				430			
Glu	Lys	Asn	Asn	Ala	Gln	Ser	Ile	Arg	Ala	Ile	Lys	Gly	Pro	His	Leu
		435					440					445			
Ala	Asn	Ser	Ile	Leu	Phe	Gln	Ser	Ser	Gly	Leu	Thr	Asn	Gly	Ile	Ile
	450					455					460				
Glu	Glu	Lys	Asp	Ile	His	Glu	Glu	Leu	Pro	Lys	Arg	Lys	Glu	Arg	Lys
465				470					475						480
Gln	Glu	Glu	Arg	Glu	Asp	Asp	Lys	Asp	Ser	Asp	Gln	Glu	Thr	Val	Asp
			485					490					495		
Glu	Glu	Val	Asp	Glu	Asn	Ser	Ser	Gly	Met	Phe	Ala	Ala	Glu	Glu	Thr
		500						505					510		
Gly	Glu	Ala	Leu	Ser	Glu	Glu	Thr	Thr	Ala	Gly	Glu	Gln	Ser	Thr	Arg
		515					520					525			
Ser	Phe	Ile	Leu	Asp	Lys	Ile	Ile	Glu	Glu	Asp	Asp	Ala	Tyr	Asp	Phe
	530					535					540				
Ser	Thr	Asp	Tyr	Val											
545															

<210> 24

<211> 503

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (10)..._(400)

<400> 24

ggccacgta atg tcc gta gtt cgc tca tcc gtc cat gcc aga tgg att gtg	51
Met Ser Val Val Arg Ser Ser Val His Ala Arg Trp Ile Val	
1 5 10	
ggg aag gtg att ggg aca aaa atg caa aag act gct aaa gtg aga gtg	99
Gly Lys Val Ile Gly Thr Lys Met Gln Lys Thr Ala Lys Val Arg Val	
15 20 25 30	
acc agg ctt gtt ctg gat ccc tat tta tta aag tat ttt aat aag cgg	147
Thr Arg Leu Val Leu Asp Pro Tyr Leu Lys Tyr Phe Asn Lys Arg	
35 40 45	
aaa acc tac ttt gct cac gat gcc ctt cag cag tgc aca gtt ggg gat	195
Lys Thr Tyr Phe Ala His Asp Ala Leu Gln Gln Cys Thr Val Gly Asp	
50 55 60	
att gtg ctt ctc aga gct tta cct gtt cca cga gca aag cat gtg aaa	243
Ile Val Leu Leu Arg Ala Leu Pro Val Pro Arg Ala Lys His Val Lys	
65 70 75	
cat gaa ctg gct gag atc gtt ttc aaa gtt gga aaa gtc ata gat cca	291
His Glu Leu Ala Glu Ile Val Phe Lys Val Gly Lys Val Ile Asp Pro	
80 85 90	
gtg aca gga aag ccc tgt gct gga act acc tac ctg gag agt ccg ttg	339
Val Thr Gly Lys Pro Cys Ala Gly Thr Thr Tyr Leu Glu Ser Pro Leu	
95 100 105 110	
agt tcg gaa acc acc cag cta agc aaa aat ctg gaa gaa ctc aat atc	387
Ser Ser Glu Thr Thr Gln Leu Ser Lys Asn Leu Glu Glu Leu Asn Ile	
115 120 125	
tct tca gca cag t gaagcgggag tggaagaagg atctgaaggg aaaaactgac	440
Ser Ser Ala Gln	
130	
atgtttatgt tatggaaaaa gaaatttttc taagtttcat cacaaaaaaa aaaaaaaaaa	500
aaa	503

<210> 25

<211> 130

<212> PRT

<213> Homo sapiens

<400> 25

Met	Ser	Val	Val	Arg	Ser	Ser	Val	His	Ala	Arg	Trp	Ile	Val	Gly	Lys	
1				5					10					15		
Val	Ile	Gly	Thr	Lys	Met	Gln	Lys	Thr	Ala	Lys	Val	Arg	Val	Thr	Arg	
			20					25					30			
Leu	Val	Leu	Asp	Pro	Tyr	Leu	Leu	Lys	Tyr	Phe	Asn	Lys	Arg	Lys	Thr	
		35					40					45				
Tyr	Phe	Ala	His	Asp	Ala	Leu	Gln	Gln	Cys	Thr	Val	Gly	Asp	Ile	Val	
	50					55					60					
Leu	Leu	Arg	Ala	Leu	Pro	Val	Pro	Arg	Ala	Lys	His	Val	Lys	His	Glu	
65					70					75				80		
Leu	Ala	Glu	Ile	Val	Phe	Lys	Val	Gly	Lys	Val	Ile	Asp	Pro	Val	Thr	
			85						90					95		
Gly	Lys	Pro	Cys	Ala	Gly	Thr	Thr	Tyr	Leu	Glu	Ser	Pro	Leu	Ser	Ser	
		100						105					110			
Glu	Thr	Thr	Gln	Leu	Ser	Lys	Asn	Leu	Glu	Glu	Leu	Asn	Ile	Ser	Ser	
		115					120					125				
Ala	Gln															
	130															

<210> 26

<211> 651

<212> DNA

<213> Homo sapiens

<400> 26

ggaattcggc	acgaggtcga	ctcctgtgag	gtatgggtgct	gggtgcagat	gcagtgtggc	60
tctggatagc	accttatgga	cagttgtgtc	cccaagggaag	gatgagaata	gctactgaag	120
tctaaagag	caagcctaac	tcaagccatt	ggcacacagg	cattagacag	aaagctggaa	180
gttgaaatgg	tggagtccaa	cttgccctgga	ccagcttaat	ggttctgctc	ctggtaacgt	240
ttttatccat	ggatgacttg	cttgggttaag	gacatgaaga	cagttcctgt	catacctttt	300
aaagggtatgg	agagtcggct	tgactacact	gtgtggagca	agttttaaaag	aagcaaagga	360
ctcagaattc	atgattgaag	aaatgcaggc	agacctgtta	tcctaaacta	gggtttttta	420
tgaccacaac	aagcaagcat	gcagcttact	gcttgaaagg	gtcttgcttc	acccaagcta	480
gagtgcaagt	gcctttgaag	cttactacag	cctcaaaactt	ctgggctcaa	gtgatcctca	540
gcctcccagt	ggtcttttga	gactgcctga	tggagtctca	tggcacaaga	agattaaaac	600
agtgtctcca	attttaataa	atttttgcaa	tccaaaaaaa	aaaaaaaaaa	a	651

<210> 27

<211> 559

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (324)..._(519)

<400> 27

ggctctggat agcaccttat ggacagttgt gtccccaagg aaggatgaga atagctactg	60
aagtcctaaa gagcaagcct aactcaagcc attggcacac aggcattaga cagaaagctg	120
gaagttgaaa tggtaggagtc caacttgccct ggaccagctt aatggttctg ctcttggtaa	180
cgtttttata catggatgac ttgcttgggt atggagagtc ggcttgacta cactgtgtgg	240
agcaagtttt aaagaagcaa aggactcaga attcatgatt gaagaaatgc aggcagacct	300
gttatcctaa actagggttt tta atg acc aca aca agc aag cat gca gct tac	353
Met Thr Thr Thr Ser Lys His Ala Ala Tyr	
1 5 10	
tgc ttg aaa ggg tct tgc ctc acc caa gct aga gtg cag tgg cct ttg	401
Cys Leu Lys Gly Ser Cys Leu Thr Gln Ala Arg Val Gln Trp Pro Leu	
15 20 25	
aag ctt act aca gcc tca aac ttc tgg gct caa gtg atc ctc agc ctc	449
Lys Leu Thr Thr Ala Ser Asn Phe Trp Ala Gln Val Ile Leu Ser Leu	
30 35 40	
cca gtg gtc ttt gta gac tgc ctg atg gag tct cat ggc aca aga aga	497
Pro Val Val Phe Val Asp Cys Leu Met Glu Ser His Gly Thr Arg Arg	
45 50 55	
tta aaa cag tgt ctc caa ttt t aataaat ttt tgcaatccaa aaaaaaaaaa	549
Leu Lys Gln Cys Leu Gln Phe	
60 65	
aaaaaaaaa	559

<210> 28

<211> 65

<212> PRT

<213> Homo sapiens

<400> 28

Met Thr Thr Thr Ser Lys His Ala Ala Tyr Cys Leu Lys Gly Ser Cys	
1 5 10 15	
Leu Thr Gln Ala Arg Val Gln Trp Pro Leu Lys Leu Thr Thr Ala Ser	
20 25 30	
Asn Phe Trp Ala Gln Val Ile Leu Ser Leu Pro Val Val Phe Val Asp	
35 40 45	
Cys Leu Met Glu Ser His Gly Thr Arg Arg Leu Lys Gln Cys Leu Gln	
50 55 60	
Phe	
65	

<210> 29

<211> 623

<212> DNA

<213> Homo sapiens

<220>

<223> sequência *primer*

<400> 29

tgcagcggcc	gcccgggcag	gtgtcgactc	ctgtgaggtg	tggtgctggg	tcagatgca	60
gtgtggctct	ggatagcacc	ttatggacag	ttgtgtcccc	aaggaaggat	gagaatagct	120
actgaagtcc	taaagagcaa	gcctaaactca	agccattggc	acacaggcat	tagacagaaa	180
gctggaagtt	gaaatggtgg	agtccaaett	gcctggacca	gcttaatggt	tctgctcctg	240
gtaacgtttt	tatccatgga	tgacttgctt	gggtatggag	agtcggcttg	actacactgt	300
gtggagcaag	ttttaagaa	gcaaaggact	cagaattcat	gattgaagaa	atgcaggcag	360
acctgttatc	ctaaactagg	gtttttaatg	accacaacaa	gcaagcatgc	agcttactgc	420
ttgaaaggg	cttgcctcac	ccaagctaga	gtgcagtggc	ctttgaagct	tactacagcc	480
tcaaactttc	gggctcaagt	gacccctcagc	ctcccagtg	tctttgtaga	ctgcctgatg	540
gagtctcatg	gcacaagaag	attaaaaacag	tgtctccaat	tttaataaat	ttttgcaatc	600
caaaaaaaaa	aaaaaaaaaa	aaa				623

<210> 30

<211> 23

<212> DNA

<213> Sequência Artificial

<220>

<223> Primer

<400> 30

aggagtttct gaggaccatg cac 23

<210> 31

<211> 22

<212> DNA

<213> Sequência Artificial

<220>

<223> *Primer*

<400> 31

tcaagggttg gggatacaca cg 22

<210> 32

<211> 22

<212> DNA

<213> Sequência Artificial

<220>

<223> sequência *primer*

<400> 32

cttgcttget ttcttctctg gc 22

<210> 33
 <211> 24
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> sequência *primer*
 <400> 33
 agtctggaaa tccacatgac caag 24
 <210> 34
 <211> 23
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> sequência *primer*
 <400> 34
 cccaatgagg aacctaaagt tgc 2
 <210> 35
 <211> 23
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> sequência *primer*
 <400> 35
 ggtgccaaat ctggactctt gtc 23
 <210> 36
 <211> 23
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> *Primer*

<400> 36
gatccatttt cagcagtgct ctg 23
<210> 37
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> *Primer*
<400> 37
cagtgttcac agaaggggta ctcac 25
<210> 38
<211> 20
<212> DNA
<213> Sequência Artificial
<220>
<223> *Primer*
<400> 38
acgagagcga cacggacaag 20
<210> 39
<211> 20
<212> DNA
<213> Sequência Artificial
<220>
<223> *Primer*
<400> 39
tctgaggctg tggcaggtgc 20
<210> 40
<211> 21
<212> DNA
<213> Sequência Artificial

<220>
<223> *Primer*
<400> 40
ccagtctttg ccaactcgtg c 21
<210> 41
<211> 23
<212> DNA
<213> Sequência Artificial
<220>
<223> sequência *primer*
<400> 41
ttcgatcttc aaactgtgcc ttg 23
<210> 42
<211> 20
<212> DNA
<213> Sequência Artificial
<220>
<223> *Primer*
<400> 42
ttggcaacca gaccagcatc 20
<210> 43
<211> 22
<212> DNA
<213> Sequência Artificial
<220>
<223> *Primer*
<400> 43
tttcccatag gtgtgagtg cg 22
<210> 44
<211> 22

<212> DNA
<213> Sequência Artificial
<220>
<223> *Primer*
<400> 44
gactggtggtt ttgttcgggg tc 22
<210> 45
<211> 21
<212> DNA
<213> Sequência Artificial
<220>
<223> *Primer*
<400> 45
tttgtccaag gctgcatggt c 21
<210> 46
<211> 22
<212> DNA
<213> Sequência Artificial
<220>
<223> *Primer*
<400> 46
tgccctggtt aagccagaag tc 22
<210> 47
<211> 22
<212> DNA
<213> Sequência Artificial
<220>
<223> *Primer*
<400> 47
agcttcactt tggtcttgac gg 22

<210> 48
 <211> 22
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> *Primer*
 <400> 48
 ggtcatctgc atcaagggtg gc 22
 <210> 49
 <211> 22
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> *Primer*
 <400> 49
 ggttcgtaac cgtgacttca gg 22
 <210> 50
 <211> 21
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> *Primer*
 <400> 50
 gcatcctttt ccagtcttcc g 21
 <210> 51
 <211> 20
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> *Primer*

<400> 51
tgcagcaaac atgcctgagc 20
<210> 52
<211> 22
<212> DNA
<213> Sequência Artificial
<220>
<223> *Primer*
<400> 52
tgttccacga gcaaagcatg tg 22
<210> 53
<211> 22
<212> DNA
<213> Sequência Artificial
<220>
<223> *Primer*
<400> 53
atccttcttc cactcccgct tc 22
<210> 54
<211> 22
<212> DNA
<213> Sequência Artificial
<220>
<223> *Primer*
<400> 54
tcggcttgac tacactgtgt gg 22
<210> 55
<211> 22
<212> DNA
<213> Sequência Artificial

<220>
<223> *Primer*
<400> 55
tacaaagacc actgggaggc tg 22
<210> 56
<211> 24
<212> DNA
<213> Sequência Artificial
<220>
<223> *Primer*
<400> 56
cgggaaatcg tgcgtgacat taag 24
<210> 57
<211> 24
<212> DNA
<213> Sequência Artificial
<220>
<223> *Primer*
<400> 57
tgatctcctt ctgcatacctg tcgg 24
<210> 58
<211> 21
<212> DNA
<213> Sequência Artificial
<220>
<223> *Primer*
<400> 58
tttggctaca gcaacagggt g 21
<210> 59
<211> 22

<212> DNA
<213> Sequência Artificial
<220>
<223> *Primer*
<400> 59
tgtgaggagg ggagattcag tg 22
<210> 60
<211> 16
<212> DNA
<213> Sequência Artificial
<220>
<223> *Primer*
<400> 60
cgctgacctc aaccag 16
<210> 61
<211> 21
<212> DNA
<213> Sequência Artificial
<220>
<223> *Primer*
<400> 61
ctgtttgccc gttcttatta c 21
<210> 62
<211> 24
<212> DNA
<213> Sequência Artificial
<220>
<223> *Primer*
<400> 62
cgggaaatcg tgcgtgacat taag 24

<210> 63
 <211> 24
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> *Primer*
 <400> 63
 tgatctcctt ctgcatcctg tcgg 24
 <210> 64
 <211> 25
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo *Antisense*
 <400> 64
 atttgggcat cactggctac aagca 25
 <210> 65
 <211> 25
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo de controlo reverso
 <400> 65
 acgaacatcg gtcactacgg gttta 25
 <210> 66
 <211> 24
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo *Antisense*

<400> 66
cagagaggtg agacactcgc cgca 24
<210> 67
<211> 24
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo de controlo reverso
<400> 67
acgccgctca cagagtggag agac 24
<210> 68
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 68
ttggtgtcat tgggtcaagg gttgg 25
<210> 69
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 69
ggttgggaac tgggttactg tggtt 25
<210> 70
<211> 24
<212> DNA
<213> Sequência Artificial

<220>
<223> Oligonucleotídeo Sintético
<400> 70
acagggcaga tacggacctc ggtg 24
<210> 71
<211> 24
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 71
gtggctccag gcatagacgg gaca 24
<210> 72
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 72
ttgtgggtaa gcagtttcat gtcgc 25
<210> 73
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 73
cgctgtactt tgacgaatgg gtggt 25
<210> 74
<211> 25

<212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo Sintético
 <400> 74
 cctggatcag acgcaagtta tcggc 25
 <210> 75
 <211> 25
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo Sintético
 <400> 75
 cggctattga acgcagacta ggtcc 25
 <210> 76
 <211> 25
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo Sintético
 <400> 76
 ctactcccca cacttcatcg ccagg 25
 <210> 77
 <211> 25
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo Sintético
 <400> 77
 ggaccgctac ttcacacccc tcatc 25

<210> 78
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 78
ctcttgatac tccagcggca aacca 25
<210> 79
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 79
accaaacggc gacctcatag ttctc 25
<210> 80
<211> 24
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 80
gcgccaagc cggtcggttct taag 24
<210> 81
<211> 24
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético

<400> 81
gaattcttgc ttgccgaacc cgcg 24
<210> 82
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 82
ccaggtaggc acgagttggc aaaga 25
<210> 83
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 83
agaaacggtt gagcacggat ggacc 25
<210> 84
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 84
gccattgaag atgcccagat cccac 25
<210> 85
<211> 25
<212> DNA
<213> Sequência Artificial

<220>
<223> Oligonucleotídeo Sintético
<400> 85
caccctagac ccgtagaagt taccg 25
<210> 86
<211> 24
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 86
cctgcgtttg tccctccagc atct 24
<210> 87
<211> 24
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 87
tctacgacct ccctgtttgc gtcc 24
<210> 88
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 88
aagtcacagt ccccgatac cagtc 25
<210> 89
<211> 25

<212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo Sintético
 <400> 89
 ctgaccatag gccctgaca ctgaa 25
 <210> 90
 <211> 25
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo Sintético
 <400> 90
 ttgtcgcttt ggcaggcata aaacc 25
 <210> 91
 <211> 25
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo Sintético
 <400> 91
 tctggtcatt aacttgcttt ccgtg 25
 <210> 92
 <211> 25
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo Sintético
 <400> 92
 cagtgtttcg tgggtgtgctc tgttg 25

<210> 93
 <211> 23
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo Sintético
 <400> 93
 gctcaccatc cgggcaccaa gca 23
 <210> 94
 <211> 25
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo Sintético
 <400> 94
 tgagagacag tgtttcgtgg tgtgc 25
 <210> 95
 <211> 25
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo Sintético
 <400> 95
 tgccttcaca cgcttggtta tcttc 25
 <210> 96
 <211> 25
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo Sintético

<400> 96
gacaacatcg gaggcttcaa tcacc 25
<210> 97
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 97
gttgaggctc tgaacaccac tgttg 25
<210> 98
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 98
gtttggcagc accttcaaca tttgg 25
<210> 99
<211> 24
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 99
agcagtttgg cagcaccttc aaca 24
<210> 100
<211> 25
<212> DNA
<213> Sequência Artificial

<220>
<223> Oligonucleotídeo Sintético
<400> 100
cttctattgg ttcgcacact tccgt 25
<210> 101
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 101
ccactaactt cggaggctac aacag 25
<210> 102
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 102
gttgtcacca caagtctcgg agttg 25
<210> 103
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 103
ggttttacaac ttccacgacg gtttg 25
<210> 104
<211> 24

<212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo Sintético
 <400> 104
 acaacttcca cgacggtttg acga 24
 <210> 105
 <211> 24
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo Sintético
 <400> 105
 atctggcatg gacggatgag cgaa 24
 <210> 106
 <211> 24
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo Sintético
 <400> 106
 gctgggtggt ttccgaactc aacg 24
 <210> 107
 <211> 25
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo Sintético
 <400> 107
 gtcccaatca ccttccccac aatcc 25

<210> 108
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 108
tcagatcctt cttccactcc cgctt 25
<210> 109
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 109
tgctcgtgga acaggtaaag ctctg 25
<210> 110
<211> 24
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 110
aagcgagtag gcaggtagg tcta 24
<210> 111
<211> 24
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético

<400> 111
gcaactcaag ccttttggtgg gtcg 24
<210> 112
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 112
cctaacaccc cttccactaa ccctg 25
<210> 113
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 113
ttcgccctca ctttcttcct agact 25
<210> 114
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 114
gtctcgaaat ggacaaggtg ctcgt 25
<210> 115
<211> 25
<212> DNA
<213> Sequência Artificial

<220>
<223> Oligonucleotídeo Sintético
<400> 115
agcttcactt tggctcttgac ggcat 25
<210> 116
<211> 24
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 116
cggagggaag tcaagtcagc caca 24
<210> 117
<211> 24
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 117
cggcattcac cctctccagc acct 24
<210> 118
<211> 23
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 118
cctccacctg tttgcgggct tcc 23
<210> 119
<211> 25

<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 119
ccacattgag ggagtcctct tgcaa 25
<210> 120
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 120
tacggcagtt ctggtttcac ttcga 25
<210> 121
<211> 24
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 121
acaccgactg aactgaaggg aggc 24
<210> 122
<211> 24
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleotídeo Sintético
<400> 122
tccacgacct ctcccactta cggc 24

<210> 123
 <211> 23
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo Sintético
 <400> 123
 ccttcgggcg tttgtccacc tcc 23
 <210> 124
 <211> 24
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> Oligonucleotídeo sintético
 <400> 124
 ccccgaaaca aacaccagtc aacg 24
 <210> 125
 <211> 24
 <212 > DNA
 <213> Sequência Artificial
 <220>
 <223> oligonucleotídeo sintético
 <400> 125
 gcaactgacc acaaaacaag cccc 24
 <210> 126
 <211> 25
 <212> DNA
 <213> Sequência Artificial
 <220>
 <223> oligonucleotídeo sintético

<400> 126
ggccattgag tccctccata gcagc 25
<210> 127
<211> 25
<212> DNA
<213> Sequência Artificial
<220>
<223> oligonucleotídeo sintético
<400> 127
cgacgatacc tccctgagtt accgg 25

Lisboa, 28 de Agosto de 2008

REIVINDICAÇÕES

1. Um método de detecção de células cancerosas do cólon compreendendo:

- O contacto da amostra obtida a partir de uma célula de teste do cólon, com uma sonda para detecção de um produto genético de um gene expresso diferencialmente no cancro do cólon, sendo que o gene contém a sequência de SEQ ID NO: 1, sendo o dito contacto por um tempo suficiente para ligação da sonda ao produto genético; e
- comparação do nível de ligação da sonda à amostra com o nível de ligação da sonda a uma amostra de controlo obtida a partir de uma célula de controlo do cólon, sendo que a célula de controlo do cólon é de estado canceroso reconhecido;

caracterizado por um nível de ligação da sonda na amostra de células de teste do cólon, semelhante ao nível de ligação numa amostra de controlo, ser indicadora do estado canceroso da célula de teste do cólon.

2. O método, de acordo com a reivindicação N.º.1, caracterizado por a sonda ser uma sonda polinucleotídica e o produto genético ser um ácido nucleico.

3. O método, de acordo com a reivindicação N.º.1, caracterizado por o produto genético ser um polipeptídeo.

4. O método, de acordo com a reivindicação N.º.1, caracterizado por o produto genético se encontrar imobilizado numa matriz.

5. O método, de acordo com a reivindicação N.º.1, caracterizado por a sonda se encontrar imobilizada numa matriz.

6. Um método de identificação de células do cólon cancerosas, compreendendo o método os passos de:

detecção de, pelo menos, um produto genético expresso diferencialmente, sendo o produto genético codificado por um gene contendo a sequência de SEQ ID NO: 1 numa amostra de teste, sendo a amostra de teste obtida a partir de uma célula de teste suspeita de ser uma célula do cólon cancerosa; e

comparação do nível de expressão do produto genético expresso diferencialmente detectado, com o nível de expressão do produto genético expresso diferencialmente numa amostra de controlo, sendo a amostra de controlo obtida a partir de uma célula do cólon cancerosa;

caracterizado por a detecção do nível de expressão do produto genético expresso diferencialmente na amostra de teste, semelhante ao nível de expressão do produto genético na amostra de controlo, indicar que a célula de teste é uma célula do cólon cancerosa.

7. O método, de acordo com a reivindicação N°.6, caracterizado por a dita detecção ser feita por hibridação da amostra de teste com uma matriz de referência, a matriz de referência compreendendo um polinucleotídeo contendo pelo menos 12 nucleotídeos consecutivos da sequência de SEQ ID NO: 1.

8. O método, de acordo com a reivindicação N°.6, caracterizado por o produto genético detectado ser um polipeptídeo.

9. Um método de identificação de uma célula do cólon cancerosa, compreendendo o método os passos de:

detecção de, pelo menos, um produto genético expresso diferencialmente, sendo o produto genético codificado por um

gene compreendendo a sequência de SEQ ID NO:1 numa amostra de teste, sendo a amostra de teste obtida a partir de uma célula de teste suspeita de ser uma célula do cólon cancerosa; e
comparação do nível de expressão do produto genético detectado, expresso diferencialmente, com o nível de expressão do produto genético expresso diferencialmente numa amostra de controlo, sendo a amostra de controlo obtida a partir de uma célula do cólon normal;
caracterizado por a detecção de um aumento no nível de expressão do produto genético expresso diferencialmente na amostra de teste, relativamente ao nível de expressão do produto genético na amostra de controlo, indicar que a célula de teste é uma célula do cólon cancerosa.

10. O método, de acordo com a reivindicação N°.9, caracterizado por a detecção de um nível de expressão do produto genético expresso diferencialmente na amostra de teste, maior do que o nível de expressão do produto genético na amostra de controlo, indicar que a célula de teste é uma célula cancerosa do cólon.

11. O método, de acordo com a reivindicação N°.9, caracterizado por a detecção de um nível de expressão do produto genético, expresso diferencialmente, na amostra de teste, que é maior do que o nível de expressão do produto genético na amostra de controlo, indicar que a célula de teste é uma célula metastática de tumor do cólon.

12. Um método de identificação de uma célula do cólon cancerosa, caracterizado por compreender os passos de:

- detecção de, pelo menos, um produto genético expresso diferencialmente, sendo a detecção feita por detecção da

hibridação de um polinucleotídeo contendo a sequência de SEQ ID NO: 1 numa amostra de teste, sendo a amostra de teste obtida a partir de uma célula de teste suspeita de ser uma célula do cólon cancerosa; e

- comparação do nível de hibridação do produto genético, detectado, expresso diferencialmente, com o nível de hibridação do produto genético expresso diferencialmente numa amostra de controlo, sendo a amostra de controlo proveniente de uma célula do cólon cancerosa;

caracterizado por a detecção de um aumento no nível de hibridação do produto genético expresso diferencialmente, na amostra de teste, relativamente ao nível de hibridação do produto genético na amostra de controlo indicar que a célula de teste é uma célula do cólon cancerosa.

13. Um método de identificação de uma célula do cólon cancerosa, compreendendo os passos de:

- detecção de, pelo menos, um produto genético, expresso diferencialmente, em que a detecção é feita por detecção da hibridação de um polinucleotídeo contendo a sequência de SEQ ID NO: 1 numa amostra de teste, sendo a amostra de teste obtida a partir de uma célula de teste suspeita de ser uma célula do cólon cancerosa; e

- comparação do nível de hibridação do produto genético detectado, expresso diferencialmente, com o nível de hibridação do produto genético expresso diferencialmente numa amostra de controlo, sendo a amostra de controlo obtida a partir de uma célula do cólon normal;

caracterizado por a detecção de um aumento no nível de hibridação do produto genético expresso diferencialmente na amostra de teste, relativamente ao nível de hibridação do produto genético

na amostra de controlo, indicar que a célula de teste é uma célula do cólon cancerosa.

14. Utilização de um polinucleotídeo *antisense*, para inibição da expressão de um gene contendo a sequência de SEQ ID NO: 1, caracterizada por ser utilizada no fabrico de um medicamento para suprimir ou inibir um fenótipo canceroso de uma célula do cólon cancerosa.

15. A utilização, de acordo com a reivindicação N°.14, caracterizada por o fenótipo canceroso ser uma metástase.

16. A utilização, de acordo com a reivindicação N°.14, caracterizada por o fenótipo canceroso ser uma proliferação celular aberrante relativamente a uma célula normal.

17. A utilização, de acordo com a reivindicação N°.14, caracterizada por o fenótipo canceroso ser a perda da inibição por contacto do crescimento celular.

18. Utilização de um polinucleotídeo *antisense*, para inibição da produção de um produto genético codificado por um polinucleotídeo contendo a sequência de SEQ ID NO: 1, ou de um anticorpo que se liga especificamente a um polipeptídeo codificado por um polinucleotídeo contendo a sequência de SEQ ID NO: 1, caracterizada por ser utilizado no fabrico de um medicamento para inibir o crescimento do tumor do cólon, num sujeito com um tumor do cólon, que expressa um gene contendo a sequência de SEQ ID NO: 1.

19. Um método de identificação de um produto genético como alvo para uma terapêutica do cancro, compreendendo o método:

- contacto entre uma célula do cólon cancerosa que expressa um produto genético candidato com um agente anti-cancerígeno, sendo que o produto genético candidato corresponde à sequência de SEQ ID NO: 1; e

- análise do efeito do agente anti-cancerígeno sobre a expressão do produto genético candidato e sobre um fenótipo canceroso da célula cancerosa;

caracterizado por a modulação da actividade biológica do produto genético candidato, e a modulação do fenótipo canceroso da célula cancerosa, indicar que o produto genético candidato é um alvo para uma terapêutica do cancro.

20. O método, de acordo com a reivindicação N°.19, caracterizado por o agente anti-cancerígeno ser um oligonucleotídeo *antisense*.

21. O método, de acordo com a reivindicação N°.19, caracterizado por o fenótipo canceroso ser proliferação celular aberrante relativamente a uma célula normal.

22. O método, de acordo com a reivindicação N°.19, caracterizado por o fenótipo canceroso ser a formação de colónias devido à perda da inibição por contacto do crescimento.

23. Um método de identificação de agentes que diminuem a actividade biológica de um produto genético expresso diferencialmente numa célula cancerosa, compreendendo o método:

- contacto entre um agente candidato com um produto genético expresso diferencialmente, sendo que o produto genético expresso diferencialmente é um produto genético mRNA codificado pela sequência de SEQ ID NO: 1, um produto genético cDNA preparado a partir do produto genético mRNA, ou um produto genético polipeptídico expresso pelo produto genético mRNA ou cDNA; e
- detecção de uma diminuição na actividade biológica do produto genético, relativamente a um nível de actividade biológica do produto genético, na ausência do agente candidato.

24. O método, de acordo com a reivindicação N°.23, caracterizado por a dita detecção ser feita por detecção de uma diminuição na expressão de um produto genético expresso diferencialmente.

25. Um polinucleotídeo isolado, caracterizado por conter a sequência nucleotídica de SEQ ID NO: 1.

26. Uma matriz, caracterizada por conter o polinucleotídeo, de acordo com a reivindicação N°.25.

27. Uma matriz contendo pelo menos dois polinucleotídeos diferentes, caracterizada por os polinucleotídeos compreenderem uma sequência contendo a sequência de SEQ ID NO: 1.

28. Uma célula hospedeira recombinante, caracterizada por conter o polinucleotídeo, de acordo com a reivindicação N°.25.

29. Um polipeptídeo isolado, caracterizado por ser codificado pelo polinucleotídeo, de acordo com a reivindicação N°.25.

30. Um anticorpo, caracterizado por se ligar especificamente a um polipeptídeo, de acordo com a reivindicação N°.29.

31. Um polinucleotídeo, caracterizado por conter a sequência nucleotídica de uma inserção contida no clone SK-1, depositado na ATCC como Adesão No. PTA-1360.

32. Um polinucleotídeo isolado contendo a sequência codificadora do polipeptídeo de SEQ ID NO: 2.

33. A composição farmacêutica, caracterizada por conter um polinucleotídeo *antisense* para inibição da produção de um produto genético codificado por um polinucleotídeo contendo uma sequência de SEQ ID NO: 1.

34. A composição farmacêutica, de acordo com a reivindicação N°.33, caracterizada por o polinucleotídeo *antisense* compreender

uma sequência seleccionada do grupo consistindo das SEQ ID NOS:
68, 70, 72, e 74.

Lisboa, 28 de Agosto de 2008

Fig. 1 Níveis de Mensagem do Gene Correspondente a c9083

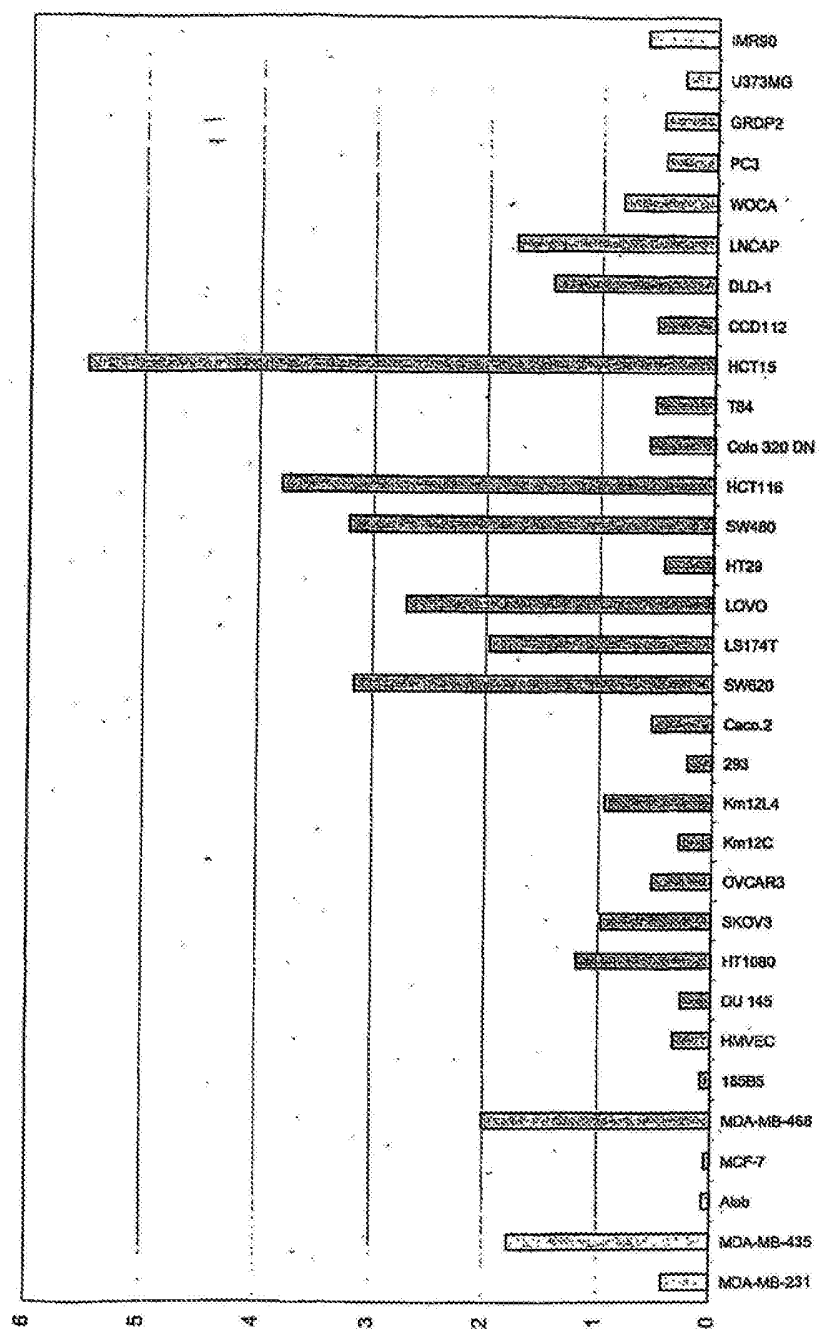


Fig. 2 Efeito dos oligonucleotídeos 9083 AS

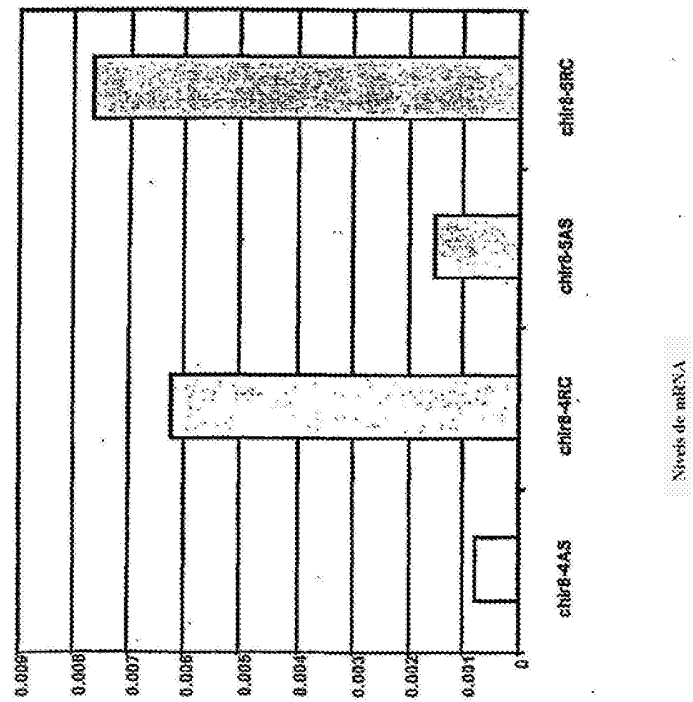
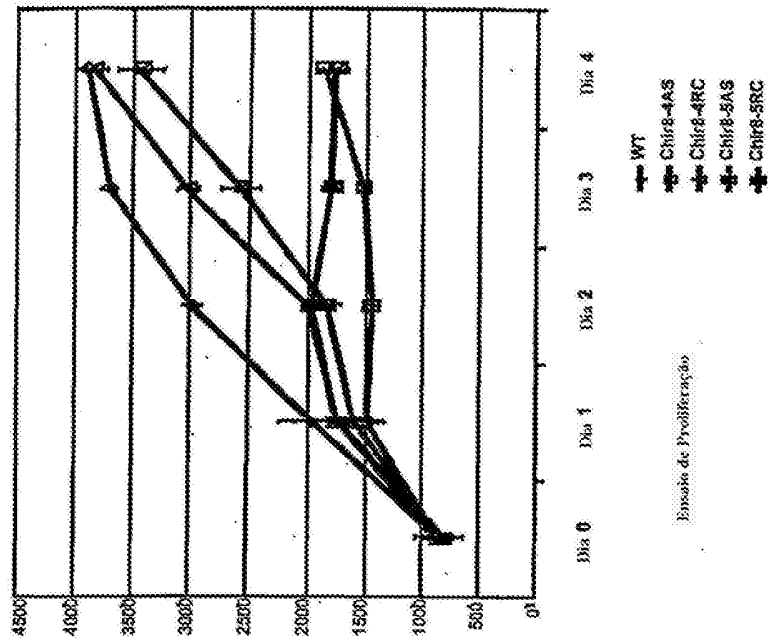


Fig. 3 Efeito de AS-chrb3 no crescimento de SV(C20)



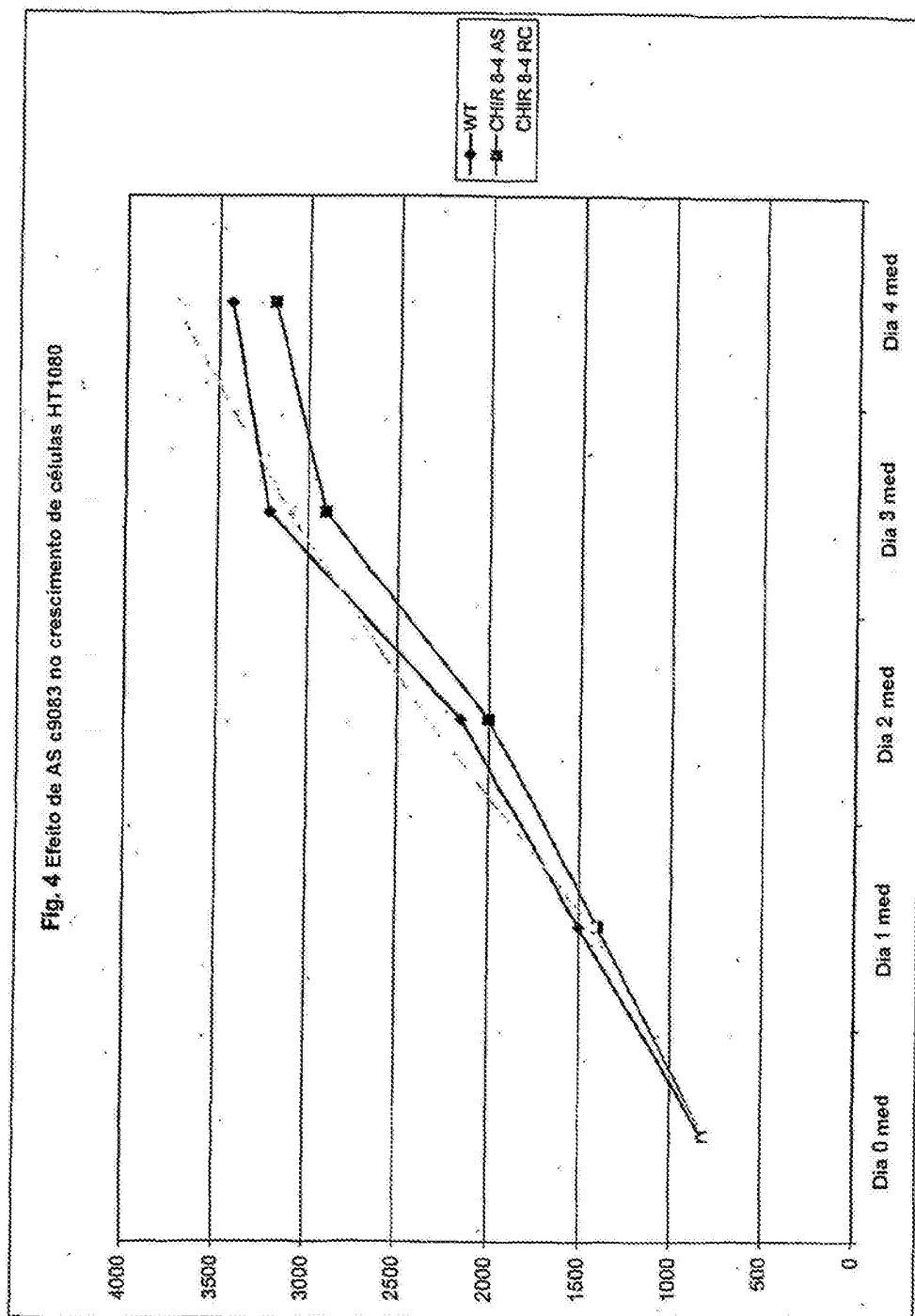


Fig. 5 Efeito do Gene Antisense contra o Gene correspondente ao aglomerado
378805 no Crescimento de Células SW620

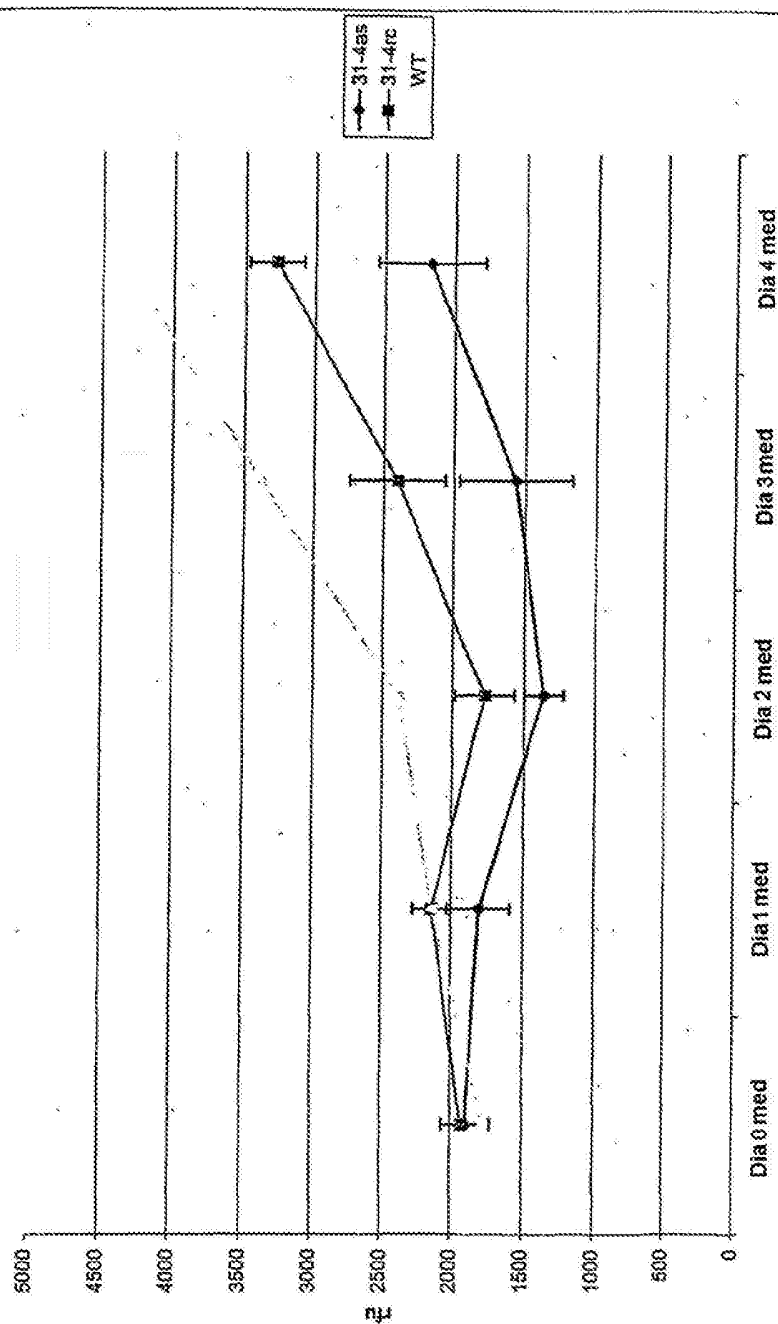


Fig. 6

Prolif SW620 c/oligo K-Ras 2576 (control)

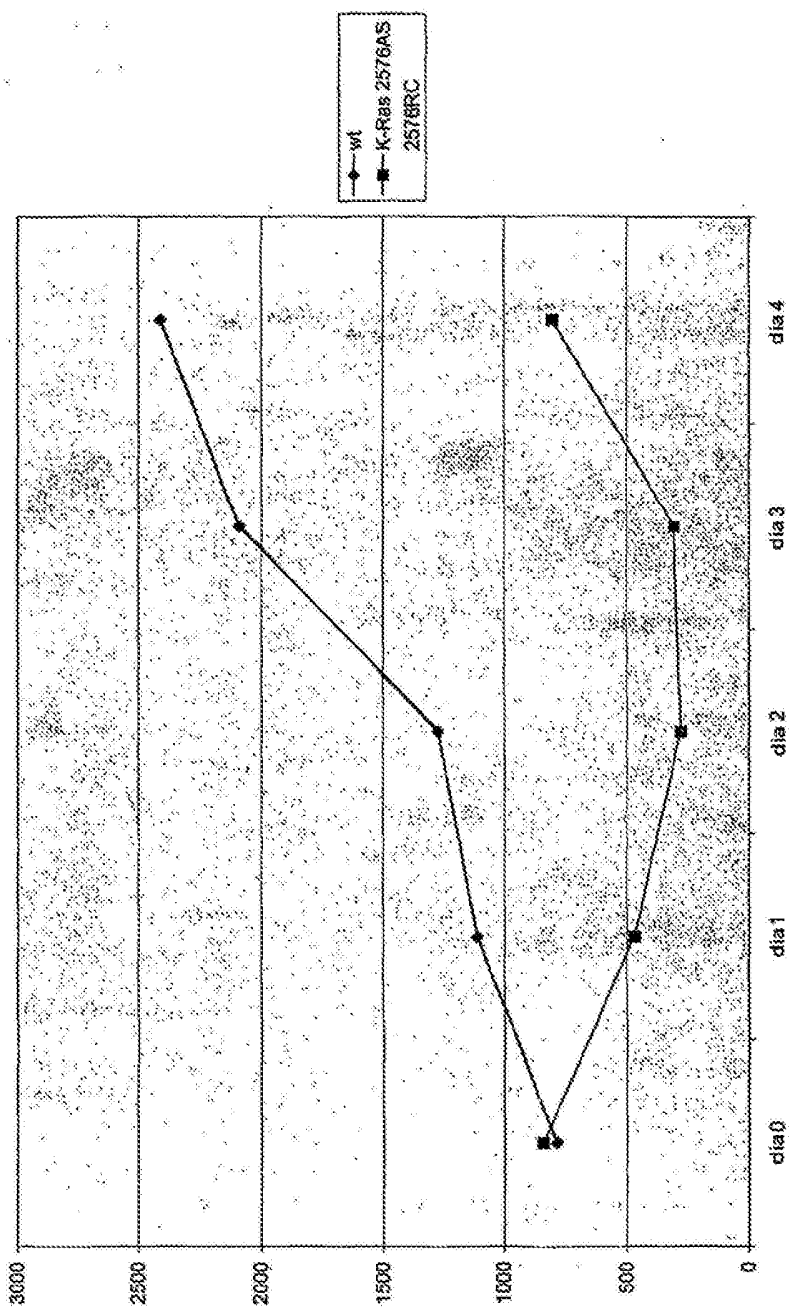


Fig. 7

SW620 transfetada com chir11-4 AS/RC 300nM

(c3376:4 AS/RL)

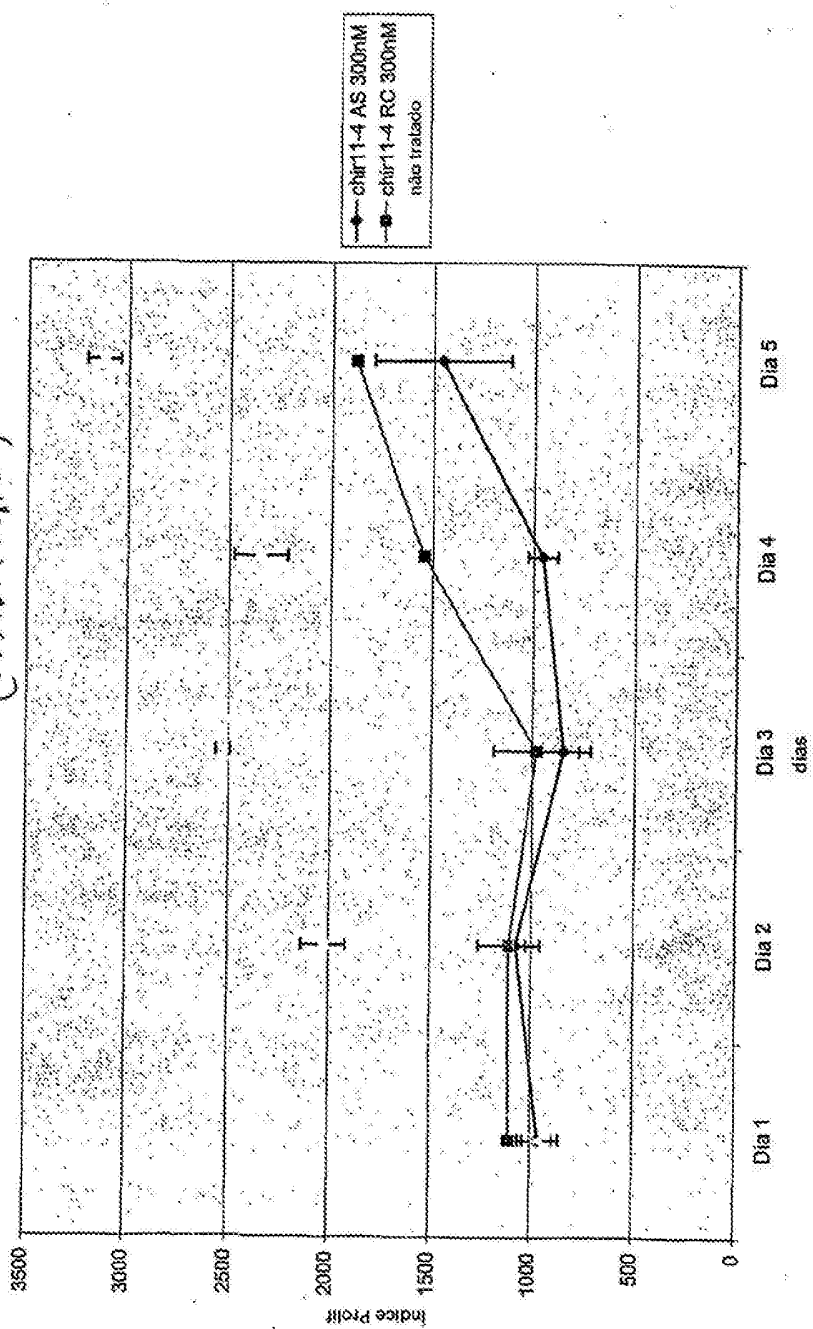


Fig.8

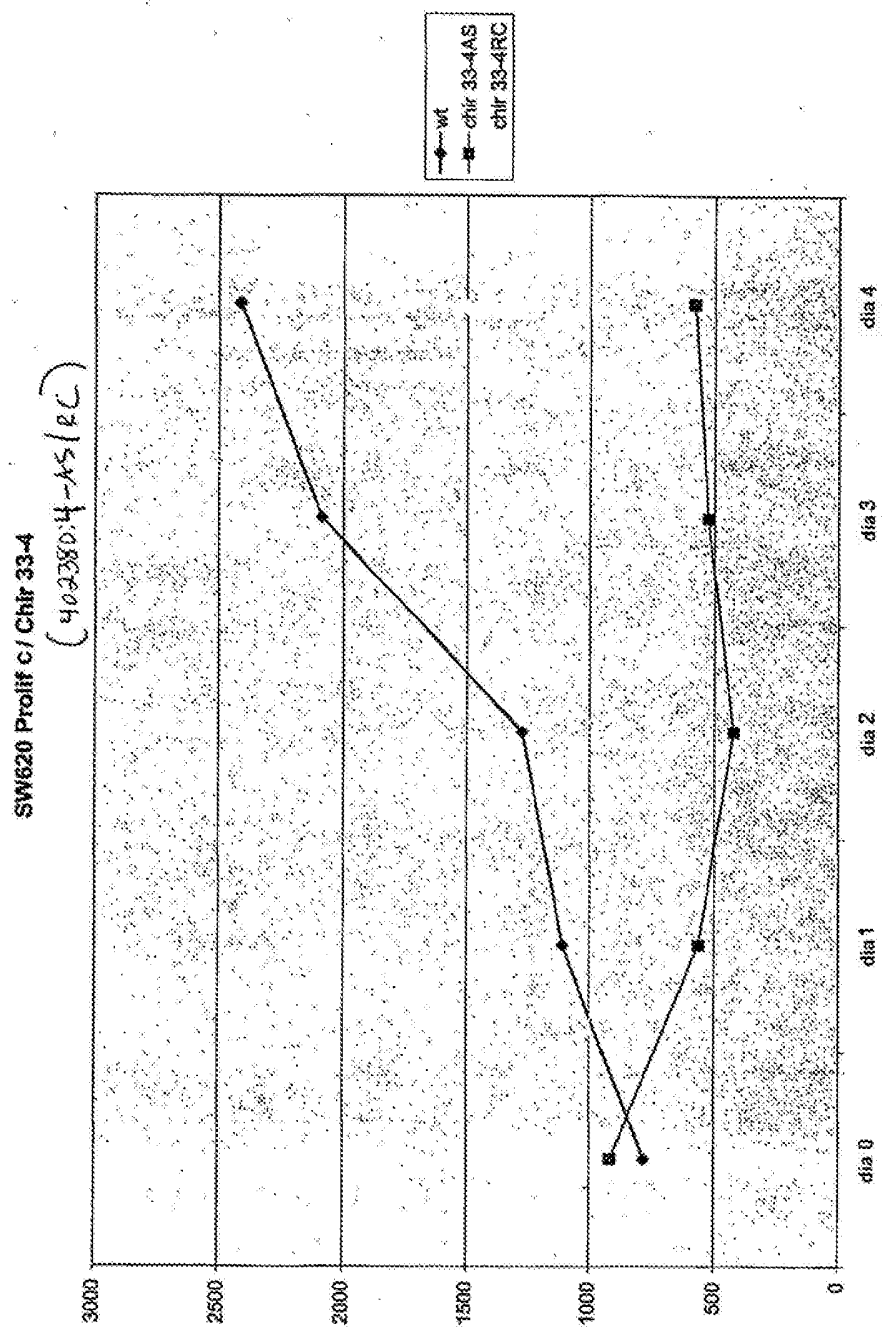


Fig. 9

Colônias de SW620 em Agar Mole normalizadas a WST1

