

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 399 246**

51 Int. Cl.:

C12Q 1/68

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **17.09.2004 E 04788855 (7)**

97 Fecha y número de publicación de la concesión europea: **07.11.2012 EP 1670946**

54 Título: **Predicción del resultado del tratamiento del cáncer de mama**

30 Prioridad:

19.09.2003 US 504087 P

02.12.2003 US 727100

06.02.2004 US 773761

23.02.2004 US 547199 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
27.03.2013

73 Titular/es:

**BIOETHERANOSTICS, INC. (100.0%)
9640 Towne Centre Drive, Suite 200
San Diego, CA 92121, US**

72 Inventor/es:

**MA, XIAO-JUN;
ERLANDER, MARK G.;
SGROI, DENNIS C. y
ENRIGHT, EDWARD**

74 Agente/Representante:

UNGRÍA LÓPEZ, Javier

ES 2 399 246 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Predicción del resultado del tratamiento del cáncer de mama

5 Campo de la invención

La invención se refiere a la identificación y uso de perfiles de expresión génica, o patrones, con importancia clínica para el tratamiento de cáncer de mama usando tamoxifeno (novaldex) y otros agentes "antiestrógeno" contra cáncer de mama, incluyendo otros "moduladores selectivos de los receptores de estrógenos" ("SERM"), "reguladores negativos selectivos de los receptores de estrógenos" ("SERD") e inhibidores de aromatasa ("AI"). En particular, la invención proporciona las identidades de secuencias génicas cuya expresión se correlaciona con la supervivencia de los pacientes y la protección de cáncer de mama en mujeres tratadas con tamoxifeno u otros agentes "antiestrógeno" contra cáncer de mama. Los perfiles de expresión génica, plasmados en expresión de ácido nucleico, expresión proteica u otros formatos de la expresión, pueden usarse para seleccionar sujetos aquejados de cáncer de mama que probablemente respondan de forma positiva al tratamiento con tamoxifeno u otro agente "antiestrógeno" contra cáncer de mama, así como los que probablemente no sean sensibles y por lo tanto candidatos para otros tratamientos. La invención también proporciona las identidades de conjuntos de secuencias de múltiples genes con patrones de expresión que son fuertemente predictivos de la sensibilidad a tamoxifeno y otros agentes "antiestrógeno" contra cáncer de mama.

20 Antecedentes de la invención

El cáncer de mama es con diferencia el cáncer más común entre mujeres. Cada año, se diagnostica cáncer de mama a más de 180.000 y 1 millón de mujeres en los Estados Unidos y en todo el mundo, respectivamente. El cáncer de mama es la causa principal de muerte para mujeres entre 50-55 años de edad, y es el tumor maligno no prevenible más habitual en mujeres en el Hemisferio Occidental. Actualmente viven con la enfermedad aproximadamente 2.167.000 mujeres en los Estados Unidos (Instituto del Cáncer Nacional, Programa de Vigilancia, Epidemiología y Resultados Sinales (NCI SEER), Cancer Statistics Review (CSR), www-seer.ims.nci.nih.gov/Publications/CSR1973 (1998)). Basándose en tasas de cáncer de 1995 a 1997, un informe del Instituto Nacional del Cáncer (NCI) estima que aproximadamente 1 de cada 8 mujeres en los Estados Unidos (aproximadamente 12,8 por ciento) desarrollarán cáncer de mama durante su vida (Programa de Vigilancia, Epidemiología y Resultados Finales del NCT (SEER) publicación *SEER Cancer Statistics Review 1973-1997*). El cáncer de mama es la segunda forma más común de cáncer, después de cáncer de piel, entre mujeres en los Estados Unidos. Se espera que se diagnostique una estimación de 250.100 nuevos casos de cáncer de mama en los Estados Unidos en 2001. De éstos, se espera que se produzcan 192.200 nuevos casos de cáncer de mama más avanzado (invasivo) entre mujeres (un aumento del 5% respecto al año pasado), se espera que se produzcan 46.400 nuevos casos de cáncer de mama de estadio temprano (*in situ*) entre mujeres (aumento del 9% respecto al año pasado), y se espera que se diagnostiquen aproximadamente 1.500 nuevos casos de cáncer de mama en hombres (Cancer Facts & Figures 2001 American Cancer Society). Se espera una estimación de 40.600 muertes (40.300 mujeres, 400 hombres) de cáncer mama en 2001. El cáncer de mama se clasifica en segunda posición solamente superado por el cáncer de pulmón entre las causas de muertes por cáncer en mujeres. Aproximadamente el 86% de mujeres a las que se les diagnostica cáncer de mama probablemente estén vivas cinco años después, aunque el 24% morirán de cáncer de mama después de 10 años y casi la mitad (47%) morirá de cáncer de mama después de 20 años.

Todas las mujeres están en riesgo de padecer cáncer de mama. Más del 70% de los cánceres de mama se producen en mujeres que no tienen factores de riesgo identificables aparte de la edad (U.S. General Accounting Office. Breast Cancer, 1971-1991: Prevention, Treatment and Research. GAO/PEMD-92-12; 1991). Solamente del 5 al 10% de los cánceres de mama están ligados a un historial familiar de cáncer de mama (Henderson IC, Breast Cancer. In: Murphy GP, Lawrence WL, Lenhard RE (eds). Clinical Oncology. Atlanta, GA: American Cancer Society; 1995:198-219).

Cada mama tiene de 15 a 20 secciones llamadas lóbulos. Dentro de cada lóbulo hay muchos lobulillos más pequeños. Los lobulillos terminan en docenas de bulbos minúsculos que pueden producir leche. Los lóbulos, lobulillos y bulbos están todos ligados por tubos estrechos llamados conductos. Estos conductos conducen al pezón en el centro de un área oscura de piel llamada la aureola. Los lóbulos y los conductos están rodeados de grasa. No hay músculos en la mama, pero hay músculos debajo de cada mama y cubren las costillas. Cada mama contiene también vasos sanguíneos y vasos linfáticos. Los vasos linfáticos portan un fluido incoloro llamado linfa, y conducen a los ganglios linfáticos. Los grupos de ganglios linfáticos se encuentran cerca de la mama en la axila (bajo el brazo), por encima de la clavícula, y en el pecho.

Los tumores de mama pueden ser benignos o malignos. Los tumores benignos no son cancerosos, no se expanden a otras partes del cuerpo y no son una amenaza para la vida. Habitualmente pueden retirarse y en la mayoría de los casos no regresan. Los tumores malignos son cancerosos y pueden invadir y dañar tejidos y órganos cercanos. Las células tumorales malignas pueden metastatizar, entrar en el torrente sanguíneo o en el sistema linfático. Cuando las células de cáncer de mama metastatizan fuera de la mama, con frecuencia se encuentran en los ganglios linfáticos bajo

el brazo (ganglios linfáticos axilares). Si el cáncer ha alcanzado estos ganglios, significa que las células cancerosas pueden haberse propagado a otros ganglios linfáticos u otros órganos, tales como huesos, hígado o pulmones.

Se ha centrado una investigación intensiva e importante en la detección temprana, tratamiento y prevención. Esto ha incluido un énfasis en la determinación de la presencia de células epiteliales ductales cancerosas o precancerosas. Estas células se analizan, por ejemplo, con respecto a morfología celular, con respecto a marcadores proteicos, con respecto a marcadores de ácido nucleico, con respecto a anomalías cromosómicas, con respecto a marcadores bioquímicos y con respecto a otros cambios característicos que señalarían la presencia de células cancerosas o precancerosas. Esto ha conducido a diversas alteraciones moleculares que se han indicado en cáncer de mama, pocas de las cuales se han caracterizado bien en muestras de ensayo de mama clínicas humanas. Las alteraciones moleculares incluyen presencia/ausencia de receptores esteroides de progesterona y estrógeno, expresión/amplificación de HER-2 (Mark HF, *et al.* HER-2/neu gene amplification in stages I-IV breast cancer detected by fluorescent *in situ* hybridization. *Genet Med*; 1(3): 98-103 1999), Ki-67 (un antígeno que está presente en todos los estadios del ciclo celular excepto G0 y se usa como un marcador para proliferación de células tumorales) y marcadores de pronóstico (incluyendo oncogenes, genes supresores de tumores y marcadores de angiogénesis) como p53, p27, Cathepsina D, pS2, gen de resistencia a multifármaco (MDR) y CD31. El documento WO 02/103320 se refiere a marcadores genéticos, cuya expresión se correlaciona con cáncer de mama, y en particular que pueden usarse para diferenciar afecciones clínicas asociadas con cáncer de mama tales como la presencia o ausencia del receptor de estrógenos ESR1 y BRCA1 y tumores esporádicos.

El tamoxifeno es el agente antiestrógeno más frecuentemente prescrito en mujeres con cáncer de mama positivo para receptor de hormonas tanto de estadio temprano como metastásico (para revisiones, véase Clarke, R *et al.* "Antiestrogen resistance in breast cancer and the role of estrogen receptor signalling". *Oncogene* 22, 7316-39 (2003) y Jordan, C. "Historical perspective on hormonal therapy of advanced breast Cancer". *Clin. Ther.* 24 Sup. A, A3-16 (2002)). En el entorno con adyuvante, la terapia con tamoxifeno da como resultado una reducción del 40-50% en el riesgo anual de reaparición, lo que conduce a una mejora del 5,6% en la supervivencia a 10 años en pacientes negativos para ganglios linfáticos y una mejora del 10,9% correspondiente en pacientes positivos para ganglios (Grupo, E.B.C.T.C. Tamoxifen for early breast cancer. *Cochrane Database Syst Rev*, CD000486 (2001)). Se cree que el tamoxifeno actúa principalmente como un inhibidor competitivo de la unión de estrógenos con el receptor de estrógenos (ER). Los niveles absolutos de expresión de ER, así como los del receptor de progesterona (PR, un indicador de una ruta de ER funcional), son en la actualidad los mejores predictores de respuesta a tamoxifeno en situaciones clínicas (Group, (2001) y Bardou, V.J. *et al.* "Progesterone receptor status significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases". *J Clin Oncol* 21, 1973-9 (2003)).

Sin embargo, el 25% de los tumores ER+/PR+, 66% de los casos de ER+/PR- y 55% de los casos de ER-/PR+ no responden, o desarrollan resistencia temprana a tamoxifeno, a través de mecanismos que en gran medida aún no se han aclarado (véase Clarke *et al.*; Nicholson, R.I. *et al.* "The biology of antihormone failure in breast cancer". *Breast Cancer Res Treat* 80 Supl. 1, S29-34; discussion S35 (2003) y Osborne, C.K. *et al.* "Growth factor receptor cross-talk with estrogen receptor as a mechanism for tamoxifen resistance in breast cancer". *Breast* 12, 362-7 (2003)). Daidone *et al.*, *British Journal of Cancer* 2000 Vol. 82 N° 2 pág. 270-277 buscan identificar biomarcadores y el resultado después del tratamiento con tamoxifeno en cáncer de mama positivo para ganglios en mujeres mayores. Ellis *et al.*, *J. Ster. Biochem. & Mol. Bio.* 86(2003) 301-307 evalúan perfiles de expresión en pacientes tratados con inhibidores de aromatasa y tamoxifeno. Van der Flier *et al.*, *J. Nat. Cane. Inst.* Vol. 92 N° 2 (2000) 120-127 revisan la expresión de Bcar1/p130Cas en la resistencia a terapia de tamoxifeno. Luo *et al.*, *Brit. J. Cancer* 86 (2002) 1790-1796 supervisan los niveles proteicos de calicreína 6 humana y calicreína 10 humana en pacientes con cáncer de mama y sensibilidad a tamoxifeno. Hilsenback *et al.*, *J. Nat. Cane. Inst.* Vol. 91 N° 5 (1999) pág. 453-459 analizan datos de expresión en matrices de resistencia a tamoxifeno adquirida.

En la actualidad, no existe ningún medio fiable para permitir la identificación de estos pacientes no sensibles. En estos pacientes, el uso de terapias hormonales alternativas, tales como los inhibidores de aromatasa, letrozol y anastrozol (Ellis, M.J. *et al.* "Letrozole is more effective neoadjuvant endocrine therapy than tamoxifen for ErbB-1- and/or ErbB-2-positive, estrogen receptor-positive primary breast cancer: evidence from a phase III randomized trial." *J Clin Oncol* 19, 3808-16 (2001); Buzdar, A.U. "Anastrozole: a new addition to the armamentarium against advanced breast cancer." *Am J Clin Oncol* 21, 161-6 (1998); y Goss, P.E. *et al.* "A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer." *N Engl J Med* 349, 1793-802 (2003)); agentes quimioterapéuticos o inhibidores de otras rutas de señalización, tales como trastuzumab y gefitinib podrían ofrecer la posibilidad de mejorar el resultado clínico. Por lo tanto, la capacidad para predecir de forma precisa el resultado del tratamiento con tamoxifeno debería avanzar significativamente el tratamiento de cáncer de mama de estadio temprano identificando pacientes que probablemente no se benefician de TAM de modo que puedan buscarse terapias adicionales o alternativas.

Sumario de la invención

La presente invención se refiere a la identificación y uso de patrones de expresión génica (o perfiles o "identificaciones") y los niveles de expresión de secuencias génicas individuales que son únicamente relevantes para

el cáncer de mama. En particular, se proporcionan las identidades de genes que se correlacionan con supervivencia de los pacientes y la reaparición del cáncer de mama (por ejemplo metástasis del cáncer de mama). Los perfiles de expresión génica, bien expresados en la expresión de ácido nucleico, expresión de proteínas o bien otros formatos de expresión, pueden usarse para predecir la supervivencia de los sujetos aquejados de cáncer de mama y la probabilidad de la reaparición de cáncer de mama, incluyendo metástasis de cáncer.

La invención posibilita de este modo la identificación y uso de patrones de expresión génica (o perfiles o "identificaciones") y los niveles de expresión de secuencias génicas individuales que se correlacionan con (y por lo tanto son capaces de diferenciar entre) pacientes con buenos o malos resultados de supervivencia. En una realización, la invención proporciona patrones que pueden distinguir en pacientes con tumores de mama positivos para receptor de estrógenos (una isoforma) (ER+) entre los que son sensibles, o que probablemente sean sensibles, al tratamiento con tamoxifeno (TAM), u otro agente "antiestrógeno" contra cáncer de mama (tal como un "modulador selectivo de los receptores de estrógenos" ("SERM"), "regulador negativo selectivo de los receptores de estrógenos" ("SERD") o inhibidores de aromatasas ("AI")) y los que no son sensibles, o probablemente no sean sensibles, a dicho tratamiento. La sensibilidad puede verse con respecto a mejora de resultados de supervivencia a lo largo del tiempo. Estos patrones pueden por lo tanto distinguir pacientes con tumores de mama ER+ en al menos dos subtipos.

En un primer aspecto, la presente invención proporciona un medio no subjetivo para la identificación de pacientes con cáncer de mama (ER+ o ER-) como probablemente con un buen o mal resultado de supervivencia después del tratamiento con TAM u otro agente "antiestrógeno" contra cáncer de mama ensayando con respecto a los patrones de expresión desvelados en este documento. Por lo tanto cuando podría haberse usado previamente la interpretación subjetiva para determinar el pronóstico y/o el tratamiento de pacientes con cáncer de mama, la presente invención proporciona patrones de expresión génica objetivos, que pueden usarse solos o en combinación con criterios subjetivos para proporcionar una evaluación más precisa de los resultados de paciente con cáncer de mama ER+ o ER- o resultados esperados, incluyendo supervivencia y la reaparición del cáncer, después del tratamiento con TAM u otro agente "antiestrógeno" contra cáncer de mama. Los patrones de expresión de la invención proporcionan de este modo un medio para determinar el pronóstico del cáncer de mama ER+ o ER-. Además, los patrones de expresión también pueden usarse como un medio para ensayar tumores pequeños, negativos para ganglios que no se ensayan fácilmente por otros medios.

Los patrones de expresión génica comprenden genes capaces de diferenciar entre resultados de cáncer de mama con precisión significativa. Las secuencias génicas se identifican como correlacionadas con resultados de cáncer de mama ER+ de modo que los niveles de su expresión son relevantes para una determinación de los protocolos de tratamiento preferidos para un paciente, bien ER+ o bien ER-.

Por lo tanto, en una realización, la invención proporciona un método para determinar el riesgo de reaparición de cáncer en un sujeto aquejado de cáncer de mama ER+ si se trata con tamoxifeno, comprendiendo dicho método ensayar una muestra de células de cáncer de mama ER+ de dicho sujeto con respecto a la expresión de ARNm de IL 17BR y HOXB13 humano para determinar una relación de los niveles de expresión de HOXB13 e IL17BR, siendo dicha relación indicativa de ausencia de reaparición de cáncer en dicho sujeto si se trata con tamoxifeno.

La capacidad para correlacionar la expresión génica con resultado de cáncer de mama y sensibilidad a TAM es particularmente ventajosa a la luz de la posibilidad de que hasta el 40% de los sujetos ER+ que se someten a tratamiento con TAM sean no sensibles. Por lo tanto, la capacidad para identificar, con confianza, estos pacientes no sensibles en un momento temprano permite la consideración y/o aplicación de terapias alternativas (tales como un agente "antiestrógeno" diferente contra cáncer de mama u otros tratamientos anticáncer de mama) a los pacientes no sensibles. Dicho de otro modo, la capacidad para identificar sujetos no sensibles a TAM permite que el personal médico considere y/o utilice terapias alternativas para el tratamiento de los sujetos antes de que se gaste tiempo en terapia de TAM ineficaz. El tiempo gastado en una terapia ineficaz con frecuencia permite el crecimiento adicional del cáncer, y la probabilidad de éxito con terapias alternativas se reduce a lo largo del tiempo dado dicho crecimiento. Por lo tanto, la invención también proporciona métodos para mejorar el resultado de supervivencia de pacientes no sensibles mediante el uso de los métodos desvelados en este documento para identificar pacientes no sensibles para el tratamiento con terapias alternativas.

Se identifican patrones de expresión génica de la invención como se describe posteriormente. En general, se obtiene una gran muestra del perfil de expresión génica de una muestra a través de la cuantificación de los niveles de expresión de ARNm correspondientes a muchos genes. Este perfil se analiza después para identificar genes, cuya expresión se correlaciona de forma positiva, o negativa, con el resultado de cáncer de mama ER+ tras el tratamiento con TAM u otro agente "antiestrógeno" contra cáncer de mama. Después puede identificarse un perfil de expresión de un subconjunto de genes humanos por los métodos de la presente descripción correlacionados con un resultado particular. El uso de múltiples muestras aumenta la confianza con la que puede creerse que un gen se correlaciona con un resultado de supervivencia particular. Sin suficiente confianza, sigue siendo impredecible si la expresión de un gen particular se correlaciona de hecho con un resultado y también es impredecible si la expresión de un gen particular puede usarse de forma exitosa para identificar el resultado para un paciente con cáncer de mama. La invención se practica basándose en las identidades de las secuencias génicas descritas en este documento o las secuencias reales usadas independientemente de la identificación.

Puede usarse un perfil de genes que se correlacionan en gran medida con un resultado en relación con otro para ensayar una muestra de un sujeto aquejado de cáncer de mama para predecir la sensibilidad probable (o falta de la misma) a TAM u otro agente "antiestrógeno" contra cáncer de mama en el sujeto del que se obtuvo la muestra. Dicho ensayo puede usarse como parte de un método para determinar el tratamiento terapéutico para dicho sujeto basándose en el resultado de cáncer de mama identificado.

Como se analiza posteriormente, los genes correlacionados pueden usarse en combinación para aumentar la capacidad para correlacionar de forma precisa un fenotipo de expresión molecular con un resultado de cáncer de mama. Esta correlación es un modo de posibilitar de forma molecular la determinación de los resultados de supervivencia como se describen en este documento. Hay usos adicionales de los genes correlacionados en la clasificación de células y tejidos; determinación del diagnóstico y/o pronóstico; y determinación y/o alteración de la terapia.

La capacidad para diferenciar se confiere por la identificación de la expresión de los genes individuales como relevantes y no por la forma del ensayo usada para determinar el nivel real de expresión. Un ensayo puede utilizar cualquier característica identificadora de un gen individual identificado como se describe en este documento siempre que el ensayo refleje, de forma cuantitativa o cualitativa, la expresión del gen en el "transcriptoma" (la fracción transcrita de genes en un genoma) o el "proteoma" (la fracción traducida de genes expresados en un genoma). Los ensayos adicionales incluyen los basados en la detección de fragmentos polipeptídicos del miembro o los miembros relevantes del proteoma. Las características identificadoras incluyen, pero sin limitación, secuencias de ácido nucleico únicas usadas para codificar (ADN) o expresar (ARN), dichos genes o epítomos específicos de, o actividades de una proteína codificada por dicho gen. Todo lo que se requiere es la secuencia o las secuencias génicas necesarias para diferenciar entre resultados de cáncer de mama y una muestra que contenga células apropiadas para su uso en un ensayo de expresión.

Los inventores también describen la identificación de los patrones de expresión génica analizando expresión génica global, o casi global, de células únicas o poblaciones celulares homogéneas que se han separado por disección de, o se han aislado o purificado de otro modo de, células contaminantes más allá de lo posible por una mera biopsia. Debido a que la expresión de numerosos genes fluctúa entre células de diferentes pacientes así como entre células de la muestra del mismo paciente, se usan múltiples datos de la expresión de genes individuales y patrones de expresión génica como datos de referencia para generar modelos que a su vez permiten la identificación de un gen o genes individuales, cuya expresión se correlaciona en mayor medida con resultados de cáncer de mama particulares.

Los inventores también describen medios físicos y metodológicos para detectar la expresión del gen o genes identificados por los modelos generados por patrones de expresión individuales. Estos medios pueden dirigirse a ensayar uno o más aspectos del molde o los moldes de ADN que subyacen a la expresión del gen o los genes, del ARN usado como un intermedio para expresar el gen o los genes, o del producto proteico expresado por el gen o los genes.

Los inventores también describen que el gen o los genes identificados por un modelo como capaces de diferenciar entre resultados de cáncer de mama pueden usarse para identificar el estado celular de una muestra desconocida de célula o células de la mama. Preferentemente, la muestra se aísla mediante medios no invasivos. La expresión de dicho gen o genes en dicha muestra desconocida puede determinarse y compararse con la expresión de dicho gen o genes en datos de referencia de patrones de expresión génica correlacionados con resultados de cáncer de mama. Opcionalmente, la comparación con muestras de referencia puede ser por comparación con el modelo o los modelos construidos basándose en las muestras de referencia.

Una ventaja proporcionada por la presente invención es que no están presentes células no de mama contaminantes (tales como linfocitos que se infiltran u otras células del sistema inmune) para afectar posiblemente a los genes identificados o el análisis posterior de la expresión génica para identificar los resultados de supervivencia de pacientes con cáncer de mama. Dicha contaminación está presente cuando se usa una biopsia para generar perfiles de expresión génica. Sin embargo, y como se observa en este documento, la invención incluye la identidad de genes que pueden usarse con precisión significativa, incluso con presencia de células contaminantes.

Los inventores también describen un medio no subjetivo basado en la expresión de múltiples genes, o combinaciones de los mismos, para la identificación de pacientes con cáncer de mama como probablemente con buen o mal resultado de supervivencia después del tratamiento con TAM u otro agente "antiestrógeno" contra cáncer de mama. Estos genes son miembros de los patrones de expresión descritos en este documento que se ha descubierto que son fuertemente predictivos del resultado clínico después del tratamiento con TAM de cáncer de mama ER+.

La presente invención proporciona de este modo secuencias génicas identificadas como diferencialmente expresadas en cáncer de mama ER+ en correlación con sensibilidad a TAM. Las secuencias de dos genes presentan expresión aumentada en células de mama ER+ que responden a tratamiento con TAM (y por lo tanto falta

de expresión aumentada en casos no sensibles). Las secuencias de otros dos genes presentan expresión reducida en células de mama ER+ que responden a tratamiento con TAM (y por lo tanto falta de expresión reducida en casos no sensibles).

5 El primer conjunto de secuencias que se ha descubierto que se expresan más en células de mama ER+ sensibles a TAM son las del receptor B de interleucina 17 (IL17RB), que se han mapeado en el cromosoma 3 humano en 3p21.1. IL17RB también se denomina receptor de interleucina 17B (IL17BR) y las secuencias que les corresponden, y por lo tanto pueden usarse en la práctica de la presente invención, se identifican por UniGene Cluster Hs.5470.

10 El segundo conjunto de secuencias que se ha descubierto que se expresan más en células de mama ER+ sensibles a TAM son las de una región transcrita recién identificada de colina deshidrogenasa (CHDH), que se han mapeado en el cromosoma 3 humano en 3p21.1. Esta está cerca de la localización mapeada para la subunidad 1D alfa, de tipo L, dependiente de voltaje del canal de calcio (CACNA1D) en 3p14.3. La invención se basa en parte en el descubrimiento inesperado de un error en bases de datos públicas que identificaron la secuencia de AI24093 como correspondiente a una parte transcrita de CACNA1D (en Hs.399966). Como se detalla posteriormente, las regiones transcritas de CHDH y CACNA1D se orientan de forma convergente de modo que la transcripción avance desde las regiones reguladoras de cada una hacia la región reguladora de la otra. Dicho de otro modo, se transcriben de forma convergente a partir de hebras complementarias en la misma región del cromosoma 3.

20 Por lo tanto, los inventores también describen la identificación de AI240933 que está en la orientación errónea con respecto a transcripción de CACNA1D pero en la orientación correcta como transcripción de CHDH y como se localiza en el extremo 3' de la transcripción de CHDH. Sin quedar ligado a la teoría, y aportado para mejorar el entendimiento de la descripción, se cree que la secuencia de AI240933 es una parte del extremo 3' del transcrito de CHDH. Es posiblemente parte de la región no traducida 3' (UTR) de CHDH. La descripción puede practicarse con secuencias correspondientes a CHDH así como las identificadas por Hs.126688.

25 El primer conjunto de secuencias que se ha descubierto que se expresa a niveles menores en células de mama ER+ sensibles a TAM son las de la caja homeótica B13 (HOXB13), que se ha mapeado en el cromosoma 17 humano en 17q21.2. Las secuencias correspondientes a HOXB13, y que pueden usarse por lo tanto en la práctica de la presente invención, se identifican por UniGene Cluster Hs.66731.

30 El segundo conjunto de secuencias que se ha descubierto que se expresa a niveles menores en células de mama ER+ sensibles a TAM son las de quinolinato fosforribosiltransferasa (QPRT, también conocida como nicotinato-nucleótido pirofosforilasa, carboxilante), que se ha mapeado en el cromosoma 16 humano en 16p12.1. Las secuencias correspondientes a QPRT que pueden usarse en la práctica de la presente invención, se identifican por UniGene Cluster Hs.335116.

35 La invención puede practicarse basándose en las identidades de estas secuencias génicas o las secuencias reales usadas independientemente de la identidad asignada.

40 Las secuencias identificadas pueden por lo tanto usarse en métodos para determinar la sensibilidad, o ausencia de sensibilidad, del cáncer de mama ER+ o ER- de un sujeto a tratamiento con TAM, o tratamiento con otro agente "antiestrógeno" contra cáncer de mama, mediante análisis de células de mama en una muestra que contenga tejidos o células de un sujeto. Como ejemplos no limitantes, puede usarse la falta de expresión aumentada de secuencias de IL17BR y/o CHDH y/o la falta de expresión reducida de secuencias de HOXB 13 y/o QPRT como un indicador de casos no sensibles. La presente invención proporciona un medio no empírico para determinar la sensibilidad a TAM u otro SERM en pacientes ER+. Esto proporciona ventajas frente al uso de un enfoque de "esperar y ver" después del tratamiento con TAM u otro agente "antiestrógeno" contra cáncer de mama. Los niveles de expresión de estas secuencias también pueden usarse como un medio para ensayar tumores pequeños, negativos para ganglios que no se evalúan fácilmente por medios convencionales.

45 Los niveles de expresión de las secuencias identificadas pueden usarse solos o en combinación con otras secuencias capaces de determinar la sensibilidad a tratamiento con TAM u otro agente "antiestrógeno" contra cáncer de mama. Preferentemente, las secuencias de la invención se usan solas o en combinación entre sí, tal como en el formato de una relación de niveles de expresión que pueda tener capacidad predictiva mejorada frente a análisis basado en la expresión de secuencias correspondientes a genes individuales. La invención proporciona relaciones del nivel de expresión de una secuencia que se infraexpresa y el nivel de expresión de una secuencia que se sobreexpresa como un indicador de sensibilidad o ausencia de sensibilidad.

50 La presente invención proporciona medios para correlacionar un fenotipo de expresión molecular con una respuesta fisiológica en un sujeto con cáncer de mama ER+. Esta correlación proporciona un medio para diagnosticar y/o determinar molecularmente tratamiento para un sujeto aquejado de cáncer de mama. Hay usos adicionales de las secuencias en la clasificación de células y tejidos; y determinación de diagnóstico y/o pronóstico. Puede usarse el uso de las secuencias para identificar células de una muestra como sensibles, o no, a tratamiento con TAM y otro agente "antiestrógeno" contra cáncer de mama para determinar la selección, o alteración, de la terapia usada para tratar tales células en el sujeto, así como el sujeto en sí mismo, del que se originó la muestra.

Tales métodos de la invención pueden usarse para ayudar en la determinación de proporcionar tamoxifeno u otro agente "antiestrógeno" contra cáncer de mama como un agente quimiopreventivo o quimioprotector a un sujeto en alto riesgo de desarrollo de cáncer de mama. Estos métodos de la invención son un avance frente a los estudios de Fabian *et al.* (J Natl Cancer Inst. 92(15): 1217-27, 2000), que propusieron una combinación de citomorfología y el modelo de riesgo de Gail para identificar pacientes con alto riesgo. Los métodos pueden usarse en combinación con evaluaciones de riesgo relativo de cáncer de mama tales como la analizada por Tan-Chiu *et al.* (J Natl Cancer Inst. 95(4): 302-307, 2003). Los ejemplos no limitantes incluyen ensayar muestras mínimamente invasivas, tales como aspirados de aguja fina aleatorios (perialveolares) o muestras de lavado de conductos (tales como las descritas por Fabian *et al.* y opcionalmente en combinación con o como una adición a un mamograma positivo para cáncer de mama benigno o maligno), de células de mama con respecto a los niveles de expresión de secuencias génicas como se describen en este documento para ayudar en la determinación de la administración de terapia con un agente "antiestrógeno" contra cáncer de mama, tal como la que puede producirse en casos de sujetos de alto riesgo (tales como los descritos por Tan-Chiu *et al.*). Los ensayos conducirían de este modo a la identificación de sujetos para los que la aplicación de un agente "antiestrógeno" contra cáncer de mama probablemente sería beneficiosa como un agente quimiopreventivo o quimioprotector. Se contempla que dicha aplicación como se permite por la presente invención podría conducir a efectos beneficiosos como los vistos con la administración de tamoxifeno (véase por ejemplo, Wickerham D. L., Breast Cancer Res. and Treatment 75 Supl 1:S7-12, Discussion S33-5, 2000). Otras aplicaciones de la invención incluyen ensayar cáncer de mama avanzado, incluyendo cáncer metastásico, para determinar la sensibilidad, o ausencia de sensibilidad, del mismo a tratamiento con un agente "antiestrógeno" contra cáncer de mama.

Un ensayo de la invención puede utilizar un medio relacionado con el nivel de expresión de las secuencias descritas en este documento siempre que el ensayo refleje, de forma cuantitativa o cualitativa, la expresión de la secuencia. Preferentemente, sin embargo, se prefiere un medio de ensayo cuantitativo. Se proporciona la capacidad para determinar sensibilidad a TAM u otro agente "antiestrógeno" contra cáncer de mama y por lo tanto el resultado del tratamiento con el mismo por el reconocimiento de la relevancia del nivel de expresión de las secuencias identificadas y no por la forma del ensayo usada para determinar el nivel real de expresión. Las características identificadoras de la secuencia incluyen, pero sin limitación, secuencias de ácido nucleico únicas usadas para codificar (ADN), o expresar (ARN), las secuencias descritas o epítomos específicos de, o actividades de, proteínas codificadas por las secuencias. Los medios alternativos incluyen detección de amplificación de ácido nucleico como indicativa de aumento de los niveles de expresión e inactivación, delección o metilación del ácido nucleico, como indicativa de la reducción de los niveles de expresión. Dicho de otro modo, la invención puede practicarse ensayando uno o más aspectos del molde o los moldes de ADN que subyacen a la expresión de la secuencia o las secuencias descritas, del ARN usado como un intermedio para expresar la secuencia o las secuencias, o del producto proteico expresado por la secuencia o las secuencias, así como fragmentos proteolíticos de tales productos. Como tal, la detección de la presencia de, cantidad de, estabilidad de, o degradación (incluyendo velocidad) de, dicho ADN, ARN y moléculas proteicas puede usarse en la práctica de la descripción.

La práctica de la presente invención no se ve afectada por la presencia de emparejamientos erróneos menores entre las secuencias descritas y las expresadas por células de la muestra de un sujeto. Un ejemplo no limitante de la existencia de dichos emparejamientos erróneos se ve en casos de polimorfismos de secuencia entre individuos de una especie, tales como pacientes humanos individuales dentro de *Homo sapiens*. El conocimiento de que la expresión de las secuencias descritas (y secuencias que varían debido a emparejamientos erróneos menores) se correlaciona con la presencia de células de mama no normales o anómalas y cáncer de mama es suficiente para la práctica de la invención con una muestra que contenga células apropiada mediante un ensayo con respecto a expresión.

En una realización, la invención posibilita la identificación de los niveles de expresión de las secuencias descritas por análisis de su expresión en una muestra que contenga células de mama ER+ o ER-. En una realización preferida, la muestra contiene células únicas o poblaciones celulares homogéneas que se han separado por disección de, o aislado o purificado de otro modo de, células contaminantes más allá de lo posible por una mera biopsia. Como alternativa, pueden usarse células no diseccionadas dentro de una "sección" de tejido. Están disponibles múltiples medios para dicho análisis, incluyendo detección o expresión dentro de un ensayo con respecto a expresión génica global, o casi global, en una muestra (por ejemplo como parte de un análisis de perfil de expresión génica tal como en una micromatriz) o por detección específica, tal como PCR cuantitativa (Q-PCR), o PCR cuantitativa en tiempo real.

Preferentemente, la muestra se aísla mediante medios no invasivos o mínimamente invasivos. La expresión de la secuencia o secuencias descritas en la muestra puede determinarse y compararse con la expresión de dicha secuencia o secuencias en datos de referencia de células de mama cancerosas o no normales. Como alternativa, el nivel de expresión puede compararse con los niveles de expresión en células normales o no cancerosas, preferentemente de la misma muestra o sujeto. En realizaciones de la invención que utilizan Q-PCR, el nivel de expresión puede compararse con los niveles de expresión de genes de referencia en la misma muestra o puede usarse una relación de niveles de expresión.

Cuando se aíslan células de mama individuales en la práctica de la invención, un beneficio es que no están presentes células no de mama contaminantes (tales como linfocitos que se infiltran u otras células del sistema inmune) para afectar posiblemente a la detección de la expresión de la secuencia o las secuencias descritas. Dicha contaminación está presente cuando se usa una biopsia para generar perfiles de expresión génica. Sin embargo, el análisis de la expresión génica diferencial y correlación con resultados de cáncer de mama ER+ con muestras tanto aisladas como no aisladas, como se describe en este documento, aumenta el nivel de confianza de las secuencias descritas como capaces de tener potencia predictiva significativa con uno de los tipos de muestra.

Aunque la presente invención se describe principalmente en el contexto de cáncer de mama humano, puede practicarse en el contexto de cáncer de mama de cualquier animal que se sepa que está aquejado potencialmente de cáncer de mama. Los animales preferidos para la aplicación de la presente invención son mamíferos, particularmente los importantes para aplicaciones agrícolas (tales como, pero sin limitación, vacas, ovejas, caballo y otros "animales de granja"), animales modelo de cáncer de mama y animales para compañía humana (tales como, pero sin limitación, perros y gatos).

Los anteriores aspectos y realizaciones de la invención pueden aplicarse igualmente con respecto al uso de más de un agente "antiestrógeno" contra cáncer de mama. En el caso de una combinación de agentes, puede usarse cualquier combinación de más de un SERM, SERD o AI en lugar de TAM u otro agente "antiestrógeno" contra cáncer de mama. La aromatasa es una enzima que proporciona una fuente principal de estrógeno en tejidos corporales incluyendo la mama, hígado, músculo y grasa. Sin quedar ligado a la teoría, y proporcionado solamente para ayudar en un mejor entendimiento de la invención, se entiende que los AI actúan de una manera comparable a TAM y otros agentes "antiestrógeno" contra cáncer de mama, que se cree que actúan como antagonistas de receptor de estrógenos en tejidos mamarios y por lo tanto contra cáncer de mama. Los AI pueden ser agentes no esteroideos o esteroideos. Los ejemplos de los primeros, que inhiben la aromatasa mediante el grupo protético hemo incluyen, pero sin limitación, anastrozol (arimidex), letrozol (femara) y vorozol (rivisor), que se han usado o contemplado como tratamientos para cáncer de mama metastásico. Los ejemplos de AI esteroideos, que inactivan aromatasa, incluyen, pero sin limitación, exemestano (aromasina), androstenediona y formestano (lentaron).

Otras formas de terapia para reducir los niveles de estrógenos incluyen ablación ovárica quirúrgica o química. La primera es la retirada física de los ovarios mientras que la segunda es el uso de agentes para bloquear la producción ovárica de estrógeno. Un ejemplo no limitante de la segunda son agonistas de la hormona liberadora de gonadotropina (GnRH), tales como goserelina (zoladex).

La invención descrita en este documento se basa en parte en el rendimiento de un análisis de micromatriz de todo el genoma de tumores de mama invasivos positivos para el receptor de hormonas de 60 pacientes tratados con adyuvante tamoxifeno solamente, que conduce a la identificación de una relación de expresión de dos genes que es altamente predictiva del resultado clínico. Esta relación de expresión, que se adapta fácilmente a análisis basado en PCR de muestras de ensayo clínicas incluidas en parafina convencionales, se validó en un conjunto independiente de 20 pacientes como se describe posteriormente.

Breve descripción de las figuras

La Figura 1 muestra análisis de las características operativas del receptor (ROC) de los niveles de expresión de IL17BR, HOXB13 y CACNA1D como predictores de los resultados del cáncer de mama en secciones de tejido completo (los 3 gráficos de la parte superior) y células microdisseccionadas con láser (los 3 gráficos de la parte inferior). AUC se refiere al área bajo la curva.

La Figura 2 contiene seis partes relacionadas con la validación de una proporción entre la expresión de HOXB 13 y la expresión de IL17BR como indicador de la capacidad de respuesta, o la falta de la misma, a TAM. Las partes a y b muestran los resultados del análisis de la expresión génica de las secuencias de HOXB13 y IL17BR mediante Q-PCR en muestras de respondedores y de no respondedores. Los gráficos de los conjuntos de datos de formación y validación de los respondedores y los no respondedores se muestran en las partes c y d, en las que "0" indica los puntos de datos de Respondedores en ambas y "1" indica los puntos de datos de los no respondedores en ambas. Las partes e y f muestran gráficos de los conjuntos de datos de formación y validación de los respondedores y los no respondedores en función de la supervivencia, en los que la línea superior en cada parte representa los respondedores y la línea inferior representa los no respondedores.

La Figura 3 muestra una representación esquemática de la región 3' conocida de la secuencia del gen CHDH en combinación con las secuencias en 3' no traducidas de CHDH identificadas mediante la presente divulgación.

La Figura 4 muestra los resultados de una reacción de amplificación mediante PCR en la que se produce un amplicón consistente con la prevista a partir del esquema de la Figura 1. Los cebadores para PCR usados fueron los siguientes: cebador de CHDH directo: 5'-AAAGTCTT-GGGAAATGAGACAAGT-3'; cebadores inversos 83R: 5'-AGCTGTCATTTGCCAGTGAGA-3' y 81R: 5'-CTGT-CATTTGCCAGTGAGAGC-3'.

La Figura 5 muestra la alineación de 28 secuencias para identificar un cóntigo que comprende la región del extremo 3' de CHDH. La alineación incluye la secuencia de AI240933, que incluye el extremo 3' de la secuencia consenso ensamblada.

La Figura 6 muestra la secuencia de un cóntigo ensamblado que contiene el nuevo extremo 3' de CHDH.

La Figura 7 muestra una representación de una región del cromosoma 3 humano en la que se identifica la localización de CACNA1D mediante "Hs.399966" y la localización de CHDH se identifica mediante "Hs.126688".

Figura 8, la Parte A contiene seis partes relacionadas con la validación de una proporción entre la expresión de QPRT y la expresión de CHDH como indicador de la capacidad de respuesta, o la falta de la misma, a TAM. Las tres porciones identificadas por "QPRT:CHDH AI240933" reflejan la proporción usando una sonda para la expresión de la secuencia GenBank AI240933. Las tres porciones identificadas por "QPRT:CHDH AJ272267" reflejan la proporción usando una sonda para la expresión de la secuencia GenBank AJ272267, identificada como la de un ARNm parcial para CHDH. La parte B contiene el uso análogo de una proporción de la expresión de HOXB13 y la expresión de IL17BR como indicador de la capacidad de respuesta a TAM. Se muestran los gráficos de los conjuntos de datos de respondedores ("R") y no respondedores ("NR"). Los valores P son la prueba t de dos muestras.

Modos de llevar a cabo la invención

Definiciones de los términos y expresiones como se usan en el presente documento:

Un "patrón" o "perfil" o "firma" de expresión génica hace referencia a la expresión relativa de genes correlacionada con la capacidad de respuesta al tratamiento del cáncer de mama ER+ con TAM u otro agente "antiestrógenos" contra el cáncer de mama. La capacidad de respuesta, o la falta de la misma, se puede expresar como los resultados de la supervivencia que se correlacionan con un "patrón" o "perfil" o "firma" de la expresión que es capaz de diferenciar dichos resultados y predecirlos.

Un "modulador selectivo del receptor de estrógenos" o SERM es un agente "antiestrógenos" que en algunos tejidos actúa como estrógenos (agonistas) pero bloquea la acción de los estrógenos en otros tejidos (antagonista). Un "regulador por disminución selectivo del receptor de estrógenos" o ("SERD") o antiestrógenos "puros" incluye agentes que bloquean la actividad de los estrógenos en todos los tejidos. Véase Howell y col. (Best Practice & Res. Clin. Endocrinol. Metab. 18(1):47-66, 2004). Los SERM preferidos de la invención son aquéllos que son antagonistas de los estrógenos en tejidos y células de mama, incluidos los de cáncer de mama. Ejemplos no limitantes de estos incluyen TAM, raloxifeno, GW5638 e ICI 182,780. Se han revisado los posibles mecanismos de acción por varios SERM (véase, por ejemplo, Jordan y col., 2003, Breast Cancer Res. 5:281-283; Hall y col., 2001, J. Biol. Chem. 276(40): 36869-36872; Dutertre y col., 2000, J. Pharmacol. Exp. Therap. 295(2):431-437; y Wijayaratne y col., 1999, Endocrinology 140(12):5828-5840). Otros ejemplos no limitantes de SERM en el contexto de la invención incluyen trifeniletlenos, tales como tamoxifeno, GW5638, TAT-59, clomifeno, toremifeno, droloxifeno e idoxifeno; benzotiofenos, tales como arzoxifeno (LY353381 o LY353381-HCI); benzopiranos, tal como EM-800; naftalenos, tales como CP-336,156; y ERA-923.

Ejemplos no limitantes de SERD o antiestrógenos "puros" incluyen agentes tales como ICI 182,780 (fulvestrant o faslodex) o el análogo oral SR16243 y ZK 191703, así como inhibidores de la aromatasa y agentes químicos de ablación ovárica como se describe en el presente documento.

Otros agentes abarcados por los SERM como se usan en el presente documento incluyen inhibidores del receptor de la progesterona y fármacos relacionados, tales como progestomiméticos como acetato de medroxiprogesterona, megace y RU-486; e inhibidores basados en péptidos de la acción de los ER, tales como análogos de la LH-RH (leuporelina, zoladex, [D-Trp6]LH-RH), análogos de la somatostatina y miméticos de motivo LXXLL de los ER así como tibolona y resveratrol. Como se ha indicado anteriormente, los SERM preferidos de la invención son aquéllos que son antagonistas de los estrógenos en tejidos y células de mama, incluidos los de cáncer de mama. Ejemplos no limitantes de SERM preferidos incluyen los metabolitos reales o contemplados (in vivo) de cualquier SERM, tal como, entre otros, 4-hidroxitamoxifeno (metabolito del tamoxifeno), EM652 (o SCH 57068 en el que EM-800 es un profármaco de EM-652) y GW7604 (metabolito de GW5638). Véase en Willson y col., (1997, Endocrinology 138(9):3901-3911) y Dauvois y col., (1992, Proc. Nat'l. Acad. Sci., USA 89:4037-4041) debates de algunos SERM específicos.

Otros SERM preferidos son los que producen el mismo perfil de expresión génica relevante que el tamoxifeno o el 4-hidroxitamoxifeno. Un ejemplo de medios para identificar dichos SERM lo proporcionan Levenson y col., (2002, Cancer Res. 62:4419-4426).

Un "gen" es un polinucleótido que codifica un producto pequeño, de naturaleza ARN o proteínico. Se ha apreciado que un polinucleótido puede ser capaz de codificar un producto pequeño. El término incluye alelos y polimorfismos

de un gen que codifica el mismo producto, o un análogo del mismo funcionalmente asociado (incluidos ganancia, pérdida o modulación de la función) en base a la localización cromosómica y a la capacidad para recombinarse durante la mitosis normal.

5 Una "secuencia" o "secuencia génica" como se usa en el presente documento es una molécula de ácido nucleico o polinucleótido compuesta por un orden pequeño de bases nucleotídicas. El término incluye el orden de las bases que codifica un producto pequeño (es decir, "región de codificación"), de naturaleza ARN o proteínica, así como las bases en orden que preceden o siguen a una "región de codificación"). Ejemplos no limitantes de esto último incluyen las regiones no traducidas en 5' y 3' de un gen. Se ha apreciado que más de un polinucleótido puede ser
10 capaz de codificar un producto pequeño. También se aprecia que pueden existir alelos y polimorfismos de las secuencias divulgadas y se pueden usar en la práctica de la invención para identificar el o los niveles de expresión de las secuencias divulgadas o el alelo o polimorfismo. La identificación de un alelo o polimorfismo depende en parte de la localización cromosómica y la capacidad para recombinarse durante la mitosis.

15 Los términos "correlacionan" o "correlación" o equivalentes de los mismos hacen referencia a una asociación entre la expresión de uno o más genes y una respuesta fisiológica de una célula de cáncer de mama y/o un paciente de cáncer de mama en comparación con la falta de respuesta. Un gen se puede expresar a niveles más altos o más bajos y seguir correlacionado con la capacidad de respuesta, la falta de capacidad de respuesta o la supervivencia o el resultado del cáncer de mama. Los inventores describen, por ejemplo, la correlación entre incrementos de la
20 expresión de las secuencias de IL17BR y/o CHDH y la capacidad de respuesta de las células de mama ER+ a TAM o a otro agente "antiestrógenos" contra el cáncer de mama. Por tanto, los incrementos son indicativos de la capacidad de respuesta. Por el contrario, la falta de incrementos, incluidos los niveles de expresión no modificados, son indicadores de falta de capacidad de respuesta. De un modo similar, los inventores describen, por ejemplo, la correlación entre disminuciones de la expresión de las secuencias de HOXB13 y/o QPRT y la capacidad de
25 respuesta de las células de mama ER+ a TAM o a otro SERM. Por tanto, las disminuciones son indicativas de la capacidad de respuesta, mientras que la falta de disminuciones, incluidos los niveles de expresión no modificados, son indicadores de la falta de capacidad de respuesta. Los incrementos y disminuciones se pueden expresar fácilmente en forma de una proporción entre la expresión en una célula no normal y una célula normal, de modo que una proporción de uno (1) indica ausencia de diferencia entre las proporciones de dos (2) y un medio indica el doble y la mitad de la expresión en células no normales frente a las células normales, respectivamente. Los niveles de
30 expresión se pueden determinar fácilmente mediante procedimientos cuantitativos como se describe más adelante.

Por ejemplo, los incrementos en la expresión génica pueden indicarse mediante proporciones de o
35 aproximadamente 1,1, de o aproximadamente 1,2, de o aproximadamente 1,3, de o aproximadamente 1,4, de o aproximadamente 1,5, de o aproximadamente 1,6, de o aproximadamente 1,7, de o aproximadamente 1,8, de o aproximadamente 1,9, de o aproximadamente 2, de o aproximadamente 2,5, de o aproximadamente 3, de o aproximadamente 3,5, de o aproximadamente 4, de o aproximadamente 4,5, de o aproximadamente 5, de o aproximadamente 5,5, de o aproximadamente 6, de o aproximadamente 6,5, de o aproximadamente 7, de o aproximadamente 7,5, de o aproximadamente 8, de o aproximadamente 8,5, de o aproximadamente 9, de o
40 aproximadamente 9,5, de o aproximadamente 10, de o aproximadamente 15, de o aproximadamente 20, de o aproximadamente 30, de o aproximadamente 40, de o aproximadamente 50, de o aproximadamente 60, de o aproximadamente 70, de o aproximadamente 80, de o aproximadamente 90, de o aproximadamente 100, de o aproximadamente 150, de o aproximadamente 200, de o aproximadamente 300, de o aproximadamente 400, de o aproximadamente 500, de o aproximadamente 600, de o aproximadamente 700, de o aproximadamente 800, de o
45 aproximadamente 900, de o aproximadamente 1000. Una proporción de 2 es un incremento del 100 % (o del doble) en la expresión. Las disminuciones en la expresión génica pueden estar indicadas por proporciones de o aproximadamente 0,9, de o aproximadamente 0,8, de o aproximadamente 0,7, de o aproximadamente 0,6, de o aproximadamente 0,5, de o aproximadamente 0,4, de o aproximadamente 0,3, de o aproximadamente 0,2, de o aproximadamente 0,1, de o aproximadamente 0,05, de o aproximadamente 0,01, de o aproximadamente 0,005, de o
50 aproximadamente 0,001, de o aproximadamente 0,0005, de o aproximadamente 0,0001, de o aproximadamente 0,00005, de o aproximadamente 0,00001.

Para un fenotipo dado, una proporción de la expresión de una secuencia génica expresada a niveles incrementados en la correlación con el fenotipo y la expresión de una secuencia génica expresada a niveles disminuidos en
55 correlación con el fenotipo también se puede usar como indicador del fenotipo. Como ejemplo no limitante, el fenotipo de la falta de capacidad de respuesta al tratamiento con tamoxifeno del cáncer de mama se correlaciona con un incremento de la expresión de HOXB13 y/o QPRT así como una expresión disminuida de IL17BR y/o CHDH. Por tanto, una proporción de los niveles de expresión de HOXB13 o QPRT a IL17BR o CHDH se puede usar como
60 indicador de la falta de capacidad de respuesta.

Un "polinucleótido" es una forma polimérica de nucleótidos de cualquier longitud, ya sea ribonucleótidos o desoxirribonucleótidos. Este término hace referencia únicamente a la estructura primaria de la molécula. Por tanto, este término incluye formas monocatenarias y bicatenarias de ADN y ARN. También incluye tipos conocidos de modificaciones, incluidos marcadores conocidos en la técnica, metilación, "capuchones", sustitución de uno o más
65 de los nucleótidos naturales con un análogo y modificaciones internucleotídicas tales como enlaces sin carga (p. ej., fosforotioatos, fosforditioatos etc.), así como formas no modificadas del polinucleótido.

El término “amplificar” se usa en su sentido más amplio para querer decir que la creación de un producto de amplificación se puede efectuar enzimáticamente con ADN o ARN polimerasas. “Amplificación”, como se usa en el presente documento, se refiere, en general, al proceso de producir múltiples copias de una secuencia deseada, en particular las de una muestra. “Múltiples copias” quiere decir al menos 2 copias. Una “copia” no necesariamente significa una perfecta complementariedad de secuencia o identidad con la secuencia molde. En general, los procedimientos para amplificar el ARNm se conocen en la técnica e incluyen PCR con transcripción inversa (RT-PCR) y los descritos en la solicitud de patente de EE.UU. 10/062,857 (presentada el 25 de octubre de 2001), así como las solicitudes de patente provisionales de EE.UU. 60/298,847 (presentada el 15 de junio de 2001) y 60/257,801 (presentada el 22 de diciembre de 2000). Otro procedimiento que se puede usar es la PCR cuantitativa (o Q-PCR). Como alternativa, el ARN puede marcarse directamente como el ADNc correspondiente mediante procedimientos conocidos en la técnica.

Con “correspondiente” se quiere decir que una molécula de ácido nucleico comparte una cantidad sustancial de identidad de secuencia con otra molécula de ácido nucleico. Cantidad sustancial significa al menos 95 %, normalmente al menos 98 % y más normalmente al menos 99 % y la identidad de secuencia se determina usando el algoritmo BLAST, como se describe en Altschul y col., (1990), J. Mol. Biol. 215:403-410 (usando los parámetros por defecto publicados, es decir los parámetros $w=4$, $t=17$).

Una “micromatriz” es una matriz lineal o bidimensional o tridimensional (y de fase sólida) de regiones preferentemente pequeñas, teniendo cada una un área definida, formadas sobre la superficie de un soporte sólido tales como, entre otros, cristal, plástico o membrana sintética. La densidad de las regiones pequeñas en una micromatriz se determina con el número total de polinucleótidos inmovilizados que se han de detectar sobre la superficie de un único soporte de fase sólida, preferentemente de al menos aproximadamente $50/\text{cm}^2$, más preferentemente de al menos aproximadamente $100/\text{cm}^2$, incluso más preferentemente de al menos aproximadamente $500/\text{cm}^2$, pero preferentemente inferior a aproximadamente $1.000/\text{cm}^2$. Preferentemente, las matrices contienen menos de aproximadamente 500, aproximadamente 1.000, aproximadamente 1.500, aproximadamente 2.000, aproximadamente 2.500 o aproximadamente 3.000 polinucleótidos inmovilizados en total. Como se usa en el presente documento, una micromatriz de ADN es una matriz de oligonucleótidos o polinucleótidos colocados sobre un circuito u otras superficies usadas para hibridar con polinucleótidos amplificados o clonados de una muestra. Dado que la posición en la matriz de cada grupo concreto de cebadores se conoce, las identidades de los polinucleótidos de una muestra se pueden determinar en base a su unión a una posición concreta en la micromatriz. Como alternativa al uso de una micromatriz, en la práctica de la invención se puede usar una matriz de cualquier tamaño, incluida una disposición de una o más posiciones de una disposición bidimensional o tridimensional en una fase sólida para detectar expresión de una única secuencia génica.

Dado que la invención depende de la identificación de los genes que están sobre o subexpresados, una realización de la invención implica determinar la expresión mediante hibridación de ARNm o una versión clonada o amplificada del mismo, de una célula de muestra con un polinucleótido que es único de una secuencia génica concreta. Los polinucleótidos de este tipo preferidos contienen al menos aproximadamente 16, al menos aproximadamente 18, al menos aproximadamente 20, al menos aproximadamente 22, al menos aproximadamente 24, al menos aproximadamente 26, al menos aproximadamente 28, al menos aproximadamente 30 o al menos aproximadamente 32 pares de bases consecutivos de una secuencia génica que no se encuentra en otras secuencias génicas. El término “aproximadamente”, como se usa en la frase anterior, hace referencia a un incremento o disminución de 1 con respecto al valor numérico indicado. Incluso más preferidos son los polinucleótidos de al menos o aproximadamente 50, al menos o aproximadamente 100, al menos o aproximadamente 150, al menos o aproximadamente 200, al menos o aproximadamente 250, al menos o aproximadamente 300, al menos o aproximadamente 350, al menos o aproximadamente 400, al menos o aproximadamente 450 o al menos o aproximadamente 500 pares de bases consecutivos de una secuencia que no se encuentra en otras secuencias génicas. El término “aproximadamente”, como se usa en la frase anterior, hace referencia a un incremento o disminución del 10 % con respecto al valor numérico indicado. Por supuesto, polinucleótidos más largos pueden contener errores de apareamiento minoritarios (p. ej., mediante la presencia de mutaciones) que no afectan a la hibridación con los ácidos nucleicos de una muestra. Dichos polinucleótidos también se pueden denominar sondas polinucleotídicas que son capaces de hibridar con las secuencias de los genes, o con porciones únicas de las mismas, descritas en el presente documento. Dichos polinucleótidos pueden estar marcados para ayudar a su detección. Preferentemente, las secuencias son las de ARNm codificadas por los genes, el correspondiente ADNc de dichos ARNm y/o versiones amplificadas de dichas secuencias. En realizaciones preferidas de la invención, las sondas polinucleotídicas están inmovilizadas sobre una matriz, otros dispositivos de soporte sólido o manchas individuales que localizan las sondas.

En otra realización de la invención, todo o parte de una secuencia divulgada se puede amplificar y detectar mediante procedimientos tales como la reacción en cadena de la polimerasa (PCR) y variaciones de la misma, tales como, entre otras, PCR cuantitativa (Q-PCR), PCR con transcripción inversa (RT-PCR) y PCR en tiempo real (incluida como medio de medición de las cantidades iniciales de las copias de ARNm para cada secuencia en una muestra), opcionalmente RT-PCR en tiempo real o Q-PCR en tiempo real. Dichos procedimientos usarían uno o dos cebadores que son complementarios a porciones de una secuencia divulgada, en los que los cebadores se usan

para cebar la síntesis de ácidos nucleicos. Los ácidos nucleicos recién sintetizados están opcionalmente marcados y se pueden detectar directamente o mediante hibridación con un polinucleótido de la invención. Los ácidos nucleicos recién sintetizados pueden ponerse en contacto con polinucleótidos (que contienen secuencias) de la invención en condiciones que permitan su hibridación. Procedimientos adicionales para detectar la expresión de ácidos nucleicos expresados incluyen ensayos de protección con RNasa, incluidas hibridaciones en fase líquida e hibridación *in situ* de células.

Como alternativa, y en otra realización más de la divulgación, la expresión génica se puede determinar mediante análisis de la proteína expresada en una muestra celular de interés mediante el uso de uno o más anticuerpos específicos de uno o más epítomos de productos génicos individuales (proteínas) o fragmentos proteolíticos de los mismos, en dicha muestra celular o en un fluido corporal de un sujeto. La muestra celular puede ser una de células epiteliales de cáncer de mama enriquecida de la sangre de un sujeto, tal como mediante el uso de anticuerpos marcados contra marcadores de superficie celular, seguido de clasificación celular activada por fluorescencia (FACS). Dichos anticuerpos están marcados, preferentemente, para permitir su fácil detección tras la unión al producto génico. Las metodologías de detección adecuadas para usar en la práctica de la divulgación incluyen, entre otras, inmunohistoquímica de muestras o tejido que contienen células, ensayos de inmunosorción ligado a enzimas (ELISA) incluyendo los ensayos con anticuerpos de tipo sándwich de tejidos que contienen células o muestras de sangre, espectroscopia de masas e inmuno-PCR.

El término “marcador” se refiere a una composición capaz de producir una señal detectable indicativa de la presencia de la molécula marcada. Marcadores adecuados incluyen radioisótopos, cromóforos nucleotídicos, enzimas, sustratos, moléculas fluorescentes, restos quimioluminiscentes, partículas magnéticas, restos bioluminiscentes y similares. Como tal, un marcador es cualquier composición detectable por medios espectroscópicos, fotoquímicos, bioquímicos, inmunoquímicos, eléctricos, ópticos o químicos.

El término “soporte” se refiere a soportes convencionales tales como soportes de perlas, partículas, tiras reactivas, fibras, filtros, membranas y silano o silicato, tales como portaobjetos de cristal.

Como se usa en el presente documento, una “muestra de tejido de mama” o “muestra de célula de mama” hace referencia a una muestra de tejido o fluido de mama aislada de un individuo que se sospecha que está afectado con cáncer de mama o en riesgo de desarrollarlo. Dichas muestras son aislados primarios (en contraste con las células cultivadas) y se pueden obtener mediante cualquier medio no invasivo o mínimamente invasivo, incluidos, entre otros, lavado ductal, aspiración con aguja fina, biopsia con aguja, los dispositivos y procedimientos descritos en la patente de EE.UU. 6,328,709, o cualquier otro medio adecuado reconocido en la técnica. Como alternativa, la “muestra” se puede obtener mediante un procedimiento invasivo, incluida, entre otras, la biopsia quirúrgica.

“Expresión” y “expresión génica” incluyen transcripción y/o traducción de material de ácido nucleico.

Como se usa en el presente documento, el término “que comprende” y sus afines se usan en su sentido de incluido; es decir, equivalente al término “que incluye” y sus correspondientes afines.

Condiciones que “permiten” que se produzca un acontecimiento o condiciones que son “adecuadas” para que se produzca un acontecimiento, tal como hibridación, extensión de hebra y similares, o condiciones “adecuadas” son condiciones que no impiden que se produzcan dichos acontecimientos. Por tanto, estas condiciones permiten, potencian, facilitan y/o conducen al acontecimiento. Dichas condiciones, conocidas en la técnica y descritas en el presente documento, dependen de, por ejemplo, la naturaleza de la secuencia nucleotídica, la temperatura y las condiciones tampón. Estas condiciones también dependen de qué acontecimiento se desee, tal como hibridación, escisión, extensión de hebra o transcripción.

“Mutación” de secuencia, como se usa en el presente documento, hace referencia a cualquier alteración de secuencia en la secuencia de un gen de interés divulgado en el presente documento en comparación con una secuencia de referencia. Una mutación de secuencia incluye cambios en un solo nucleótido o alteraciones de más de un nucleótido en una secuencia debido a mecanismos tales como sustitución, delección o inserción. Un polimorfismo de un solo nucleótido (SNP) es también una mutación de secuencia como se usa en el presente documento. Dado que la presente invención se basa en el nivel relativo de la expresión génica, las mutaciones en las regiones no codificantes de genes tal como se divulgan en el presente documento también pueden analizarse en la práctica de la invención.

“Detección” incluye cualquier medio de detección, incluida la detección directa e indirecta de la expresión génica y cambios en la misma. Por ejemplo, los productos “menos detectables” se pueden observar directa o indirectamente, y el término indica cualquier reducción (incluida la ausencia de señal detectable). De un modo similar, producto “más detectable” significa cualquier incremento observado tanto directa como indirectamente.

Los incrementos y disminuciones de la expresión de las secuencias divulgadas se definen en los términos siguientes en base al porcentaje de cambios o al número de cambios sobre la expresión en células normales. Los incrementos pueden ser 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180 o 200 % respecto a los niveles de expresión en

células normales. Como alternativa, el número de incrementos puede ser 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5, o 10 veces sobre los niveles de expresión en células normales. Las disminuciones pueden ser 10, 20, 30, 40, 50, 55, 60, 65, 70, 75, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 99 o 100 % respecto a los niveles de expresión en células normales.

A menos que se defina otra cosa, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado que un experto en la técnica a la que la presente invención pertenece entiende habitualmente.

Realizaciones de la invención

En un primer aspecto, la invención divulgada se refiere a la identificación y uso de patrones (o perfiles o "firmas") de expresión génica que discriminan entre (o se correlacionan con) supervivencia del cáncer de mama en un sujeto tratado con tamoxifeno (TAM) u otro agente "antiestrógenos" contra el cáncer de mama. Dichos patrones se pueden determinar mediante los procedimientos de la invención mediante el uso de una serie de muestras de células o tejido de referencia, tales como las revisadas por un patólogo experto en la anatomía patológica del cáncer de mama, que reflejan células de cáncer de mama frente a células normales u otras células no cancerosas. Los resultados experimentados por los sujetos de los que las muestras se pueden correlacionar con los datos de expresión para identificar patrones que se correlacionen con los resultados tras el tratamiento con TAM u otro agente "antiestrógenos" contra el cáncer de mama. Dado que el perfil de expresión génica global difiere de una persona a otra, de un cáncer a otro y de una célula cancerosa a otra, las correlaciones entre ciertas células y genes expresados o subexpresados se pueden realizar como se divulga en el presente documento para identificar genes que son capaces de discriminar entre resultados de cáncer de mama.

La presente divulgación se puede poner en práctica con un número cualquiera de los genes que se cree que se expresan de forma diferencial, o que probablemente lo hagan, con respecto a los resultados de cáncer de mama, en particular en los casos de cáncer de mama ER+. La identificación se puede realizar usando los perfiles de expresión de varias poblaciones de células de cáncer de mama homogéneas que se aislaron mediante microdissección, tales como, entre otras, microdissección por captura por láser (LCM) de 100 - 1.000 células. El nivel de expresión de cada gen del perfil de expresión se puede correlacionar con un resultado concreto. Como alternativa, los niveles de expresión de múltiples genes se pueden agrupar para identificar correlaciones con resultados concretos.

Los genes con correlaciones significativas con la supervivencia del cáncer de mama cuando el sujeto es tratado con tamoxifeno se pueden usar para generar modelos de expresiones génicas que discriminarían de forma máxima entre resultados cuando un sujeto responde al tratamiento con tamoxifeno u otro agente "antiestrógenos" contra el cáncer de mama y resultados en los que el tratamiento no ha tenido éxito. Como alternativa se pueden usar genes con correlaciones significativas en combinación con genes con correlaciones menores sin una pérdida significativa de capacidad para discriminar entre resultados. Dichos modelos se pueden generar por cualquier medio adecuado reconocido en la técnica, incluyendo, entre otros, análisis por grupos, máquinas de vector con soporte, redes neurales u otro algoritmo conocido en la técnica. Los modelos son capaces de predecir la clasificación de una muestra desconocida en base a la expresión de los genes usados para discriminación en los modelos. Se puede usar la validación cruzada de "dejar uno fuera" para analizar el funcionamiento de varios modelos y ayudar a identificar pesos (genes) que no proporcionan información o son perjudiciales para la capacidad de predicción de los modelos. La validación cruzada también se puede usar para identificar genes que potencian la capacidad predictiva de los modelos.

El o los genes identificados que se correlacionan con resultados concretos del cáncer de mama en relación con el tratamiento con tamoxifeno mediante los modelos anteriores proporcionan la capacidad de centrar el análisis de la expresión génica en solo los genes que contribuyen a la capacidad de identificar un sujeto como que probablemente tenga un resultado concreto respecto a otro. La expresión de otros genes en una célula de cáncer de mama sería relativamente incapaz de proporcionar información pertinente y, por tanto, ayudar a la discriminación de un resultado de cáncer de mama.

Como apreciarán los expertos en la técnica, los modelos son muy útiles con incluso un conjunto pequeño de datos de expresión génica de referencia y pueden ser cada vez más precisos con la inclusión de más datos de referencia, aunque el incremento incremental en la precisión probablemente disminuya con cada dato adicional. La preparación de datos de expresión génica de referencia adicionales usando genes identificados y divulgados en el presente documento para discriminar entre diferentes resultados en el cáncer de mama tras el tratamiento con tamoxifeno u otro agente "antiestrógenos" contra el cáncer de mama es rutinario y un experto en la técnica la puede realizar fácilmente para permitir la generación de modelos como se ha descrito en lo que antecede para predecir el estado de una muestra desconocida en base a los niveles de expresión de dichos genes.

Para determinar los niveles de expresión (aumentados o disminuidos) de los genes en la práctica de la presente invención se puede usar cualquier procedimiento conocido en la técnica. En una realización preferida de la invención, se usa la expresión basada en la detección de ARN que hibrida con los genes identificados y divulgados en el presente documento. Esto se realiza fácilmente mediante cualquier procedimiento de detección de ARN o amplificación + detección o reconocido como equivalente en la técnica, tal como, entre otros, PCR-transcripción

inversa, los procedimientos divulgados en la patente de EE.UU. 6,794,141 y los procedimientos para detectar la presencia o ausencia de secuencias estabilizantes o desestabilizantes de ARN.

Como alternativa se puede usar la expresión basada en la detección del estado de ADN. La detección del ADN de un gen identificado como metilado o delecionado se puede usar para genes que tienen una expresión disminuida en correlación con un resultado concreto del cáncer de mama. Esto se puede realizar fácilmente mediante procedimientos basados en PCR conocidos en la técnica, incluidos, entre otros, Q-PCR. Por el contrario, la detección del ADN de un gen identificado como amplificado se puede usar para genes que tienen una expresión aumentada en correlación con un resultado concreto del cáncer de mama. Esto se puede realizar fácilmente mediante procedimientos basados en PCR, hibridación fluorescente *in situ* (FISH) e hibridación cromosómica *in situ* (CISH) conocidos en la técnica.

También se puede usar la expresión basada en la detección de una presencia, incremento o disminución de los niveles o actividad de proteínas. La detección se puede realizar mediante cualquier procedimiento basado en inmunohistoquímica (IHC), basado en sangre (especialmente para proteínas secretadas), basado en anticuerpos (incluidos autoanticuerpos contra la proteína), basado en exfoliados celulares (de cáncer), basado en espectroscopia de masas y basado en imágenes (incluido el uso de ligando marcado) conocido en la técnica y reconocido como adecuado para la detección de la proteína. Los procedimientos basados en anticuerpos e imágenes son útiles adicionalmente para la localización de tumores tras la determinación del cáncer mediante el uso de células obtenidas mediante un procedimiento no invasivo (tal como lavado ductal o aspiración con aguja fina), en los que la fuente de las células cancerosas no se conoce. Un anticuerpo o ligando marcado se puede usar para localizar el o los carcinomas en un paciente o para ayudar en el enriquecimiento de células de cáncer exfoliadas de un fluido corporal.

Una realización preferida que usa un ensayo basado en ácido nucleico para determinar la expresión es mediante inmovilización de una o más secuencias de los genes identificadas en el presente documento en un soporte sólido, incluidos, entre otros, un sustrato sólido como matriz o en perlas o en tecnología a base de perlas tal como se conoce en la técnica. Como alternativa, también se pueden usar ensayos de expresión basados en solución conocidos en la técnica. El o los genes inmovilizados pueden estar en forma de polinucleótidos que son únicos o, de otro modo, específicos del o los genes de forma que los polinucleótidos serían capaces de hibridar con un ADN o ARN correspondiente al o los genes. Estos polinucleótidos pueden ser la longitud total del o los genes o ser secuencias cortas de los genes (hasta de un nucleótido más corto que la secuencia de longitud total conocida en la técnica mediante deleción del extremo 5' o 3' de la secuencia) que están opcionalmente mínimamente interrumpidos (tal como mediante errores de apareamiento o pares de bases no complementarias insertados) de modo que no afecta a la hibridación con un ADN o ARN correspondiente al o los genes. Preferentemente, los polinucleótidos usados son del extremo 3' del gen, tal como en aproximadamente 350, aproximadamente 300, aproximadamente 250, aproximadamente 200, aproximadamente 150, aproximadamente 100 o aproximadamente 50 nucleótidos de la señal de poliadenilación o sitio de poliadenilación de un gen o secuencia expresada. Los polinucleótidos que contienen mutaciones respecto a las secuencias de los genes divulgados se pueden usar siempre que la presencia de las mutaciones siga permitiendo hibridación para producir una señal detectable.

El o los genes inmovilizados, o secuencias complementarias de los mismos, se pueden usar para determinar el estado de las muestras de ácido nucleico preparadas de célula(s) de mama de muestra para las que no se conoce el resultado del sujeto de la muestra (p. ej., paciente del que se obtiene la muestra) o para la confirmación de un resultado que ya se ha asignado al sujeto de la muestra. Sin limitar la divulgación, dicha célula puede proceder de un paciente con cáncer de mama ER+ o ER-. El o los polinucleótidos inmovilizados solo tienen que ser suficientes para hibridar específicamente con las correspondientes moléculas de ácido nucleico derivadas de la muestra en condiciones adecuadas. Aunque incluso una única secuencia génica correlacionada puede proporcionar la precisión adecuada en la discriminación entre dos resultados de cáncer de mama, se pueden usar dos o más, tres o más, cuatro o más, cinco o más, seis o más, siete o más, ocho o más, nueve o más, diez o más, once o más, o cualquier número entero de los genes identificados en el presente documento, como subpoblación capaz de discriminar en combinación con un incremento de la precisión del procedimiento. La divulgación contempla específicamente la selección de más de uno, dos o más, tres o más, cuatro o más, cinco o más, seis o más, siete o más, ocho o más, nueve o más, diez o más, once o más o cualquier número entero de los genes divulgados en las tablas y figuras en el presente documento para usar como una subpoblación en la identificación del resultado de supervivencia del cáncer de mama.

Por supuesto, se puede usar 1 o más, 2 o más, 3 o más, 4 o más, 5 o más, 6 o más, 7 o más, 8 o más, 9 o más, cualquier número entero o todos los genes proporcionados en las Tablas 2, 3 y/o XXX más adelante. "Acceso" como se usa en el contexto de las Tablas del presente documento hace referencia al número de acceso en GenBank de una secuencia de cada gen. El valor P hace referencia a los valores asignados como se describe en los ejemplos siguientes. Las indicaciones de "E-xx", en el que "xx" es un número de dos dígitos se refiere a una notación alternativa para cifras exponenciales en las que "E-xx" es " 10^{-xx} ". Por tanto, en combinación con los números a la izquierda de "E-xx", el valor que se está representando es el número de la izquierda por 10^{-xx} . "Descripción", como se usa en las Tablas, proporciona un breve identificador de lo que codifica la secuencia/gen.

Los genes con una correlación identificada por un valor p inferior o de aproximadamente 0,02, inferior o de aproximadamente 0,01, inferior o de aproximadamente 0,05 o inferior o de aproximadamente 0,001 se prefieren para usar en la práctica de la divulgación. La presente divulgación incluye el uso de gen(es) cuya expresión identifica diferentes resultados del cáncer de mama tras el tratamiento con TAM u otro agente "antiestrógenos" contra el

En un segundo aspecto, que también sirve como realizaciones del uso de una subpoblación de los genes divulgados en el presente documento, la presente divulgación se refiere a la identificación y uso de múltiples conjuntos de secuencias para la determinación de la capacidad de respuesta del cáncer de mama ER+ al tratamiento con TAM u otro agente "antiestrógenos" contra el cáncer de mama. La expresión diferencial de estas secuencias en el cáncer d

Para identificar los patrones de expresión génica en cánceres de mama invasivos, en fase temprana, ER positivos, que podrían predecir la respuesta a terapia hormonal, se realizó el análisis de expresión génica en microserie sobre tumores de 60 mujeres tratadas de forma uniforme trató con adyuvante tamoxifeno solo. Estas pacientes se identificaron de un total de 103 casos ER+ en fase temprana que se presentaron en el Hospital General de Massachusetts entre 1987 y 1997, de los cuales se congelaron instantáneamente muestras tumorales y de los cuales se disponía de un seguimiento mínimo de 5 años (véase la Tabla 1 para los detalles). Dentro de esta cohorte, 28 (46%) mujeres desarrollaron metástasis distante con un tiempo medio hasta la recurrencia de 4 años ("no respondedores a tamoxifeno") y 32 (54%) mujeres permanecieron sin enfermedad con un seguimiento medio de 10 años ("respondedores a tamoxifeno"). Los respondedores se hicieron coincidir con los casos de no respondedores con respecto a la estadificación TNM (véase Singletary, S.E. et al. "Revision of the American Joint Committee on Cancer staging system for breast cancer." J Clin Oncol 20, 3628-36 (2002)) y el grado del tumor (véase Dalton, L.W. et al. "Histologic grading of breast cancer: linkage of patient outcome with level of pathologist agreement." Mod Pathol 13, 730-5. (2000)).

Estudios previos que vinculan los perfiles de expresión génica con el resultado clínico en cáncer de mama han demostrado que el potencial de metástasis distante y la probabilidad de supervivencia global puede ser predecible a través de características biológicas del tumor primario en el momento del diagnóstico (véase Huang, E. et al. "Gene expression predictors of breast cancer outcomes." Lancet 361,1590-6 (2003); Sorlie, T. et al. "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications." Proc Natl Acad Sci U S A 98:10869-74 (2001); Sorlie, T. et al. "Repeated observation of breast tumor subtypes in independent gene expression data sets." Proc Natl Acad Sci U S A 100, 8418-23 (2003); Sotiriou, C. et al. "Breast cancer classification and prognosis based on gene expression profiles from a population-based study." Proc Natl Acad Sci U S A 100, 10393-8 (2003); van't Veer, L.J. et al. "Gene expression profiling predicts clinical outcome of breast cancer." Nature 415, 530-6 (2002); y van de Vijver, M.J. et al. "A gene-expression signature as a predictor of survival in breast cancer." N Engl J Med 347, 1999-2009 (2002)). En particular, una firma de expresión de 70 genes ha demostrado ser un potente factor pronóstico, superando todos los parámetros clínico-patológicos conocidos. Sin embargo, en esos estudios las pacientes recibieron terapia sin adyuvante (van 't Veer, L.J. et al. Nature 2002) o se trataron de forma no uniforme con regímenes hormonales y quimioterapéuticos (Huang, E. et al.; Sorlie, T. et al.; Sorlie, T. et al.; Sotiriou, C. et al.; y van de Vijver, M.J. et al. N Engl J Med 2002). Las pacientes con cáncer de mama ER+ en fase temprana tratadas con tamoxifeno solo, tal como la cohorte estudiada aquí, representan solamente un conjunto de la población ensayada con la firma de 70 genes. Es de observar que 61 de los genes en la firma de 70 genes estuvieron presentes en la microserie usada como se describe a continuación, pero no se observó asociación significativa con el resultado clínico en el subconjunto definido de pacientes.

En comparación con biomarcadores existentes, incluyendo ER1, PGR, ERBB2 y EGFR, las series de secuencias génicas descritas en este documento son significativamente más predictivas de sensibilidad al tratamiento con TAM. Un análisis multivariable indicó que estos tres genes eran predictores significativos de resultado clínico independiente del tamaño del tumor, el estado nodal y el grado del tumor. La expresión de ER y el receptor de progesterona (PR) han sido los predictores clínico-patológicos principales de de respuesta a TAM. Sin embargo, hasta el 40% de los tumores ER+ no lograron responder o desarrollaron resistencia a TAM. La invención por tanto proporciona el uso de los biomarcadores identificados para permitir un mejor tratamiento del paciente identificando pacientes que tienen mayor probabilidad de beneficiarse de TAM u otra terapia endocrina y aquellos que tienen probabilidad de desarrollar resistencia y recurrencia tumoral.

Como se indica en este documento, la secuencia o secuencias identificadas por la presente invención se expresan en correlación con células cancerosas de mama ER+. Por ejemplo, se ha descubierto que IL17BR, identificado por los grupos I.M.A.G.E. Consortium NM_018725 y NM_172234 ("The I.M.A.G.E. Consortium: An Integrated Molecular Analysis of Genomes and their Expression," Lennon et al., 1996, Genomics 33:151-152; véase también image.llnl.gov) es útil para predecir la sensibilidad al tratamiento con TAM.

En realizaciones preferidas de la invención, puede usarse cualquier secuencia, o parte única de la misma, de las secuencias IL17BR del grupo, así como el grupo UniGene *Homo sapiens* Hs.5470. Asimismo, puede usarse

cualquier secuencia que codifique todo o una parte de la proteína codificada por cualquier secuencia de IL17BR descrita en este documento. Las secuencias consenso de los grupos I.M.A.G.E. Consortium son las siguientes, con la región codificada asignada (finalizando con un codón de terminación) subrayada y precedida por la región no traducida y/o no codificante 5' y seguida de la región no traducida y/o no codificante 3':

agcgcagcgt gcggtggcc tggatccgc gcagtggccc ggcgatgtcg ctcggtgtgc
taagcctggc cgcgctgtgc aggagcgccg taccgccaga gccgaccgtt caatgtggct
ctgaaactgg gccatctcca gagtggatgc tacaacatga tctaattccc ggagacttga
gggacctccg agtagaacct gttacaacta gtgttgcaac aggggactat tcaattttga
tgaatgtaag ctgggtactc cgggcagatg ccagcatccg cttgttgaag gccaccaaga
tttgtgtgac gggcaaaagc aacttccagt cctacagctg tgtgaggtgc aattacacag
aggccttcca gactcagacc agacctctg gtggtaaag gacattttcc tacatcggt
tcctgtaga gctgaacaca gtctatttca ttggggccca taatattcct atgcaaata
tgaatgaaga tggcccttcc atgtctgtga atttcacctc accaggtgc ctagaccaca
taatgaata taaaaaaag tgtgtcaagg ccggaagcct gtgggatccg aacatcactg
cttgaagaa gaatgaggag acagtagaag tgaacttcac aacctccc ctgggaaca
gatacatggc tcttatccaa cacagcacta tcatcggtt ttctcaggtg ttgagccac
accagaagaa acaaacgcga gcttcagtgg tgattccagt gactggggat agtgaaggtg
ctacggtgca gctgactcca tattttccta cttgtggcag cgactgcac cgacataaag
gaacagttgt gctctgcccc caaacaggcg tccctttccc tctggataac aacaaaagca
agccgggagg ctggctgctt ctctctctgc tgtctctgct ggtggccaca tgggtgctgg
tggcagggat ctatctaag tggaggcagc aaaggatcaa gaagacttcc tttctacca
ccacactact gccccccatt aagggtcttg tggtttacc atctgaaata tgtttccatc
acacaatttg ttacttctact gaatttcttc aaaaccattg cagaagtga gtcctcttg
aaaagtggca gaaaaagaaa atagcagaga tgggtccagt gcagtggctt gccactcaa
agaaggcagc agacaaagtc gtcttcttcc tttccaatga cgtcaacagt gtgtgcgatg
gtacctgtgg caagagcgag ggcagtccca gtgagaactc tcaagacctc tcccccttg
cctttaacct tttctgcagt gatctaagaa gccagattca tctgcacaaa tacgtgggtg
tctactttag agagattgat acaaaagacg attacaatgc tctcagtgtc tgcccaagt
accacctcat gaaggatgcc actgctttct gtgcagaact tctccatgtc aagcagcagg

tgtcagcagg aaaaagatca caagcctgcc acgatggctg ctgctccttg taqcccacc
atgagaagca agagacctta aaggcttctt atcccaccaa ttacaggga aaaaactgtg
atgatcctga agcttactat gcagcctaca aacagcctta gtaattaaaa cattttatac
caataaaatt ttcaaatatt gctaactaat gtagcattaa ctaacgattg gaaactacat
ttacaacttc aaagctgttt tatacataga aatcaattac agttttaatt gaaaactata
accattttga taatgcaaca ataaagcatc ttcagccaaa catctagtct tccatagacc
atgcattgoa gtgtaccag aactgtttag ctaatattct atgtttaatt aatgaatact
aactctaaga acccttact gattcactca atagcatctt aagtgaaaaa cttctatta
catgcaaaaa atcattgttt ttaagataac aaaagtaggg aataaacaag ctgaaccac
ttttaaaaa aaaaaaaaa aaaaaaaaa aaaaaaa

5

SEC ID N° 2 (secuencia consenso para IL17BR, variante de transcrito 2, identificada como NIVM72234 o NM_172234.1):

agcgcagcgt gcggggtggcc tggatccgc gcagtggccc ggcgatgtcg ctcggtctgc
 taagcctggc cgcgctgtgc aggagcgccg taccocgaga gccgaccgtt caatgtggct
 ctgaaactgg gccatctcca gagtggatgc tacaacatga tctaattccc ggagacttga
 gggacctccg agtagaacct gttacaacta gtgttgcaac aggggactat tcaattttga
 tgaatgtaag ctgggtactc cgggcagatg ccagcatccg cttgttgaaag gccaccaaga
 tttgtgtgac gggcaaaagc aacttccagt cctacagctg tgtgaggtgc aattacacag
 aggccttcca gactcagacc agaccctctg gtggtaaagt gacattttcc tacatcggct
 tccctgtaga gctgaacaca gtctatttca ttggggccca taatattcct aatgcaaata
 tgaatgaaga tggcccttcc atgtctgtga atttcacctc accaggctgc ctgaccaca
 taatgaaata taaaaaaaaag tgtgtcaagg ccggaagcct gtgggatccg aacatcactg
 cttgtgaaga gaatgaggag acagtagaag tgaacttcac aaccactccc ctgggaaaca
 gatacatggc tcttatccaa cacagcaacta tcatcgggtt ttctcaggtg tttgagccac
 accagaagaa acaaacgcga gcttcagtgg tgattccagt gaotggggat agtgaaggtg
 ctacggtgca ggtaaaqtcc agtgagctgc tctggggagg gaagggacat agaagactgt
 tccatcattc attgctttta aggatgagtt ctctcttctc aaatgcactt ctgccagcag
 acaccagtta agtggcgttc atgggggctc ttctcgtgca gcctccaccg tgctgaggtc
 aggaggccga cgtggcagtt gtggctccct ttgcttgtat taatggctgc tgaccttcca
 aagcactttt tattttcatt ttctgtcaca gacactcagg gatagcagta ccattttact
 tccgaagcc ttttaactgca agatgaagct gcaaagggtt tgaaatggga aggtttgagt
 tccaggcagc gtatgaactc tggagagggg ctgccagtcc tctctgggcc gcageggacc
 cagctggaac acaggaagtt ggagcagtag gtgctccttc acctctcagt atgtctcttt
 caactctagt ttttgaggtg gggacacagg aggtccagtg ggacacagcc actcccaaaa
 gagtaaggag cttccatgct tcatccctg gcataaaaag tgcataaaca caccagaggg
 ggcaggcacc agccagggtg tgatggctac tacccttttc tggagaacca tagacttccc
 ttactacagg gacttgcatt tctaaagca ctggctgaag gaagccaaga ggatcactgc
 tgctcctttt ttctagagga aatgtttgtc taogtggtaa gatatgacct agccctttta
 ggtaagcgaa ctggtatgtt agtaacgtgt acaaagttta ggttcagacc ccgggagtct
 tgggacgtg ggtctcgggt cactggtttt gactttaggg ctttgttaca gatgtgtgac
 caaggggaaa atgtgcatga caacactaga ggtatgggcg aagccagaaa gaagggaggt
 tttggctgaa gtaggagctc tggtagatt ttgctctgat gcattgggtg aactttctga
 gcctcttgtt ttctctcagc tgactccata ttttctact tgtggcagcg actgcatccg
 acataaagga acagtgtgtc tctgccaca aacaggcgct cctttccctc tggataacaa
 caaaagcaag ccgggaggct ggctgcctct cctcctgctg tctctgctgg tggccacatg
 ggtgctggtg gcagggatct atctaattgt gaggcacgaa aggatcaaga agacttctt
 ttctaccacc aactactgc cccccattaa ggttcttgtg gtttaccat ctgaaatatg
 tttccatcac acaatttgtt acttccactga atttcttcaa aaccattgca gaagtgaggt
 catccttgaa aagtggcaga aaaagaaaat agcagagatg ggtccagtgc agtggcttgc
 cactcaaaag aaggcagcag acaaagtctt ctctcttctt tccaatgaog tcaacagtgt

gtgcgatggt acctgtggca agagcgaggg cagtcccagt gagaactctc aagacctott
 ccccttggc ttttaacctt tctgcagtga tctaagaagc cagattcate tgcacaaata
 cgtgggtggt tacttttagag agattgatac aaaagacgat tacaatgctc tcaagtgtctg
 ccccaagtac cacctcatga aggatgccac tgctttctgt gcagaacttc tccatgtcaa
 gcagcaggtg tcagcaggaa aaagatcaca agcctgccac gatggctgct gctccttcta
 gccacccat gagaagcaag agaccttaaa ggcttctat cccaccaatt acagggaaaa
 aacgtgtgat gatcctgaag cttactatgc agcctacaaa cagccttagt aattaaaaca
 ttttatacca ataaaaat tcaaatattgc taactaatgt agcattaaact aacgattgga
 aactacattt acaacttcaa agctgtttta tacatagaaa tcaattacag ttttaattga
 aaactataac cattttgata atgcaacaat aaagcatctt cagccaaaca tctagtcttc
 catagaccat gcattgcagt gtaccagaa ctgttttagct aatattctat gtttaattaa
 tgaatactaa ctctaagaac cctcactga ttactcaat agcatcttaa gtgaaaaacc
 ttctattaca tgcaaaaaat cattgttttt aagataacaa aagtagggaa taaacaagct
 gaaccactt ttaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaa

ES 2 399 246 T3

- 5 A continuación se enumeran los números de ID de clon de I.M.A.G.E. Consortium y los correspondientes números de acceso a GenBank de las secuencias identificadas como pertenecientes a los grupos I.M.A.G.E. Consortium y UniGene. También se incluyen secuencias que no están identificadas con un número de ID de clon pero aún así son de IL17BR. Las secuencias incluyen aquellas de las secuencias de la hebra "con sentido" y complementaria correspondientes a IL17BR. La secuencia de cada número de acceso a GenBank se presenta en el Apéndice adjunto.

Número de ID de clon	Números de acceso a GenBank
2985728	AW675096, AW673932, BC000980
5286745	BI602183
5278067	B1458542
5182255	BI823321
924000	AA514396
3566736	IBF110326
3195409	BE466508
3576775	BF740045
2772915	AW299271
1368826	AA836217
1744837	AI203628
2285564	AI627783
2217709	AI744263
2103651	AI401622
2419487	AI826949
3125592	BE047752
2284721	AI911549
3643302	BF194822
1646910	AI034244
1647001	AI033911
3323709	BF064177
1419779	AA847767
2205190	AI538624
2295838	AI913613
2461335	AI942234
12130362	JAI580483
2385555	AI831909
2283817	AI672344
2525596	AW025192
454687	AA677205
1285273	AA721647
3134106	BF115018
342259	W61238, W61239

ES 2 399 246 T3

Número de ID de clon	Números de acceso a GenBank
1651991	AI032064
2687714	AW236941
3302808	BG057174
2544461	AW058532
122014	T98360, T98361
2139250	AI470845
2133899	AI497731
121300 T96629, T96740	T96629, T96740
162274	H25975, He941
3446667	BE539514, BX282554
156864	R74038, R74129
4611491	BG433769
4697316	BG530489
429376	AA007528, AA007529
5112415	BI260259
701357	AA287951, AA287911
121909	T97852, T97745
268037	N40294
1307489	AA809841
1357543	AA832389
48442	H14692
1302619	AA732635
1562857	AA928257
1731938	AI184427
1896025	AI298577
2336350	AI692717
1520997	AA910922
240506	H90761
2258560	AI620122
1569921	AI793318, AA962325, AI733290
6064627	BQ226353
299018	W04990
5500181	BM455231
2484011	BI492426
4746376	BG674622
233783	BX111256
1569921	BX1117618
450450	AA682806

ES 2 399 246 T3

Número de ID de clon	Números de acceso a GenBank
1943085	AI202376
2250390	AI658949
4526156	BG403405
3249181	BE673417
2484395	AW021469
30515867	CF455736
285815	AW339874
4556884	BG399724
3254505	BF475787
3650593	BF437145
233783	H64601
Ninguno (secuencias de ARNm)	AF212365, AKp95091 AF208110, AF208111, AF250309, AK095091
Ninguno	BM983744, CB305764, BM715988, BM670929, BI792416, BI715216, N56060, CB241389, AV660618, BX088671, CB154426, CA434589, CA412162, CA314073, BF921554, BF920093, AV685699, AV650175, BX483104, CD675121, BE081436, AW970151, AW837146, AW368264, D25960, AV709899, BX431018, AL535617, AL525465, BX453536, BX453537, AV728945, AV728939, AV727345

En una realización preferida, puede usarse cualquier secuencia, o parte única de la misma, de la siguiente secuencia de IL17BR, identificada por AF208111 o AF208111.1, en la práctica de la invención.

- 5 SEC ID N° 3 (secuencia para IL17BR):

CGGCGATGTCGCTCGTGCTGATAAGCCTGGCCGCGCTGTGCAGGAGCGCCGTACCCCGAG
 AGCCGACCGTTCAATGTGGCTCTGAACTGGGCCATCTCCAGAGTGGATGCTACAACATG
 ATCTAATCCCGGAGACTTGAGGGACCTCCGAGTAGAACCTGTTACAACACTAGTGTGCAA
 CAGGGGACTATTCAATTTTGATGAATGTAAGCTGGGTACTCCGGGCAGATGCCAGCATCC
 GCTTGTGTGAAGGCCACCAAGATTTGTGTGACGGGCAAAAGCAACTTCCAGTCTACAGCT
 GTGTGAGGTGCAATTACACAGAGGCCTTCCAGACTCAGACCAGACCCTCTGGTGGTAAAT
 GGACATTTTCTATATCGGCTTCCCTGTAGAGCTGAACACAGTCTATTTTCATTGGGGCCC
 ATAATATTCTAATGCAAATATGAATGAAGATGGCCCTTCCATGTCTGTGAATTTACCT
 CACCAGGCTGCCTAGACCACATAATGAAATATAAAAAAAGTGTGTCAAGGCCGGAAGCC
 TGTGGGATCCGAACATCACTGCTTGTAGAAGAATGAGGAGACAGTAGAAGTGAACTTCA
 CAACCACTCCCCTGGGAAACAGATACATGGCTCTTATCCAACACAGCACTATCATCGGGT
 TTTCTCAGGTGTTTGAGCCACACCAGAAGAAACAAACCGGAGCTTCAGTGGTGAATCCAG
 TGAAGGGGATAGTGAAGGTGCTACGGTGCAGGTAAAGTTCAGTGAGCTGCTCTGGGGAG
 GGAAGGGACATAGAAGACTGTTCCATCATTGCTTTTAAGGATGAGTTCTCTCTGT
 CAAATGCACCTCTGCCAGCAGACACCAGTTAAGTGGCGTTTCATGGGGGTTCTTTGCTGC
 AGCCTCCACCGTCTGAGGTGAGGAGGCGGACGTGGCAGTTGTGGTCCCTTTTGTGTGTA
 TTAATGGCTGCTGACCTTCCAAAGCACTTTTTATTTTCATTTTCTGTACAGACACTCAG
 GGATAGCAGTACCATTTTACTTCCGCAAGCCTTTAACTGCAAGATGAAGCTGCAAAGGT
 TTGAATGGGAAGGTTTGAGTTCCAGGCAGCGTATGAACTCTGGAGAGGGGCTGCCAGTC
 CTCTCTGGGCCGACGCGGACCCAGCTGGAACACAGGAAGTTGGAGCAGTAGGTGCTCCTT
 CACCTCTCAGTATGTCTCTTTCAACTCTAGTTTTTGAAGTGGGGACACAGGAAGTCCAGT
 GGGGACACAGCCACTCCCCAAAGAATAAGGAAGTTCCATGCTTCATTCCCTGGCATAAAA
 AGTGNTCAAACACACCAGAGGGGGCAGGCACCAGCCAGGGTATGATGGGTACTACCTTTT
 TCTGGAGAACCATAGACTTCCCTTACTACAGGGACTTGCATGTCTTAAAGCACTGGCTGA
 AGGAAGCCAAGAGGATCACTGCTGCTCCTTTTTTGTAGAGGAAATGTTTGTGTACGTGGT
 AAGATATGACCTAGCCCTTTTAGGTAAGCGAACTGGTATGTTAGTAACGTGTACAAAGTT
 TAGGTTTCAAGCCCCGGGAGTCTTGGGCATGTGGGTCTCGGGTCACTGGTTTTGACTTTAG
 GGCTTTGTTACAGATGTGTGACCAAGGGGAAATGTGCATGACAACACTAGAGGTAGGGG
 CGAAGCCAGAAAGAAGGGAAGTTTTGGCTGAAGTAGGAGTCTTGGTGAGATTTTGTGTG
 ATGCATGGTGTGAACTTTCTGAGCCTCTTGTTTTTCTCAGCTGACTCCATATTTTCTTA
 CTTGTGGCAGCGACTGCATCCGACATAAAGGAACAGTTGTGCTCTGCCACAAACAGGCG
 TCCCTTTTCCCTCTGGATAACAACAAAGCAAGCCGGGAGGCTGGCTGCCTCTCTCCTGCT
 TGTCTCTGCTGGTGGCCACATGGGTGCTGGTGGCAGGGATCTATCTAATGTGGAGGCAG
 AAAGGATCAAGAAGACTTCCCTTTTCTACCACCACACTACTGCCCCCATTAAGGTTCTTG
 TGGTTTACCCTCTGAAATATGTTTCCATCACACAATTTGTTACTTCACTGAATTTCTTC
 AAAACCATTGCAGAAGTGAGGTCACTCTTGAAGAGTGGCAGAAAAGAAAATAGCAGAGA
 TGGGTCCAGTGCAGTGGCTTGCCACTCAAAGAAGGCAGCAGACAAAGTCTCTCTCTTC
 TTTCCAATGACGTCAACAGTGTGTGCGATGGTACCTGTGGCAAGAGCGAGGGCAGTCCCA
 GTGAGAACTCTCAAGACCTCTTCCCCCTTGCCCTTAACTTTTCTGCAAGTGTCTAAGAA
 GCCAGATTCTCTGCACAAATACGTGGTGGTCTACTTTAGAGAGATTGATACAAAAGACG
 ATTACAATGCTCTCAGTGTCTGCCCCAAGTACCACTTCATGAAGGATGCCACTGCTTTCT
 GTGCAGAACTTCTCCATGTCAAGCAGCAGGTGTGAGCAGGAAAAAGATCACAAGCCTGCC
 ACGATGGCTGCTGCTCCTTGTAGCCACCCATGAGAAGCAAGAGACCTTAAAGGCTTCCT
 ATCCCACCAATTACAGGGAAAAACGTGTGATGATCCTGAAGCTTACTATGCAGCCTACA

AACAGCCTTAGTAATTAAACATTTTATACCAATAAAATTTTCAAATATTACTAACTAAT
 GTAGCATTAACTAACGATTGGAACTACATTTTCAACTTCAAAGCTGTTTTATACATAGA
 AATCAATTACAGCTTTAATTGAAAAGTGAACCATTTTGATAATGCAACAATAAAGCATC
 TTCCAAAAA

- 5 En realizaciones preferidas de la descripción, puede usarse cualquier secuencia, o parte única de la misma, de las secuencias CHDH del grupo, así como el grupo UniGene Homo sapiens Hs. 126688. Asimismo, puede usarse cualquier secuencia que codifique todo o una parte de la proteína codificada por cualquier secuencia de CHDH descrita en este documento, incluyendo secuencias del nuevo contig ensamblado. Las secuencias consenso de los grupos I.M.A.G.E. Consortium son los siguientes, con la región codificante asignada (finalizando con un codón de

terminación) subrayada y precedida por la región no traducida y/o no codificante 5' y seguida de la región previamente identificada no traducida y/o no codificante 3':

SEC ID Nº 4 (secuencia consenso para CHDH, identificada como NM_018397 o NM_018397.1):

5

```

agcggggcgc gccacccgc tcctcccgt ccggtcccga ctgtcgggct ctcgcccgag
tcgccccgga caatcacaaa gagtgtgtag gccagccccg gtcacagagt gcaccgtatc
ctgtcacttc tggatgtgag ggagaagtga gtcattctcat tcccctccgt ggatcagagg
acttggaacta gatagaagca tgtggtgtct cctacgaggc ctggggccggc ctggagccct
ggcacgggga gccctggggc agcagcaatc cctgggtgcc cgggccctgg ccagcgcagg
ctctgagagc cgggacgagt acagctatgt ggtggtgggc gcgggctcgg cgggctgcgt
gctggctggg aggcctcacgg aggacccgc cgagcgcgtg ctgctgctgg aggcggggcc
caaggacgtg cgcgcgggga gcaagcggct ctggtggaag atccacatgc ccgcggccct
ggtggccaac ctgtgcgacg acaggtacaa ctggtgctac cacacagagg tgcagcgggg
cctggacggc cgcgtgctgt actggccacg cggccgcgtc tggggtggct cctcatccct
caatgccatg gtctacgtcc gtgggcacgc cgaggactac gagcgtggc agcgcaggg
cgcccgcggc tgggactacg cgcactgct gccctacttc cgcaaggcgc agggccacga
gctgggcgcc agccggtacc ggggcgcoga tggcccgctg cgggtgtcc ggggcaagac
caaccaccgc ctgcaactgcg cattcctgga ggccacgcag caggccggct acccgctcac
cgaggacatg aatggcttcc agcaggaggg cttcggctgg atggacatga ccatccatga
aggcaaacgg tggagcgcag cctgtgccta cctgcaccca gcactgagcc gcaccaacct
caaggccgag gccgagacgc ttgtgagcag ggtgctatct gagggcaacc gtgcagtggg
cgtggagtat gtcaagaatg gccagagcca cagggttat gccagcaagg aggtgattct
gagtggaggt gccatcaact ctccacagct gctcatgctc tctggcatcg ggaatgctga
tgacctcaag aaactgggca tcctgtggt gtgccaccta cctgggggttg gccagaacct
gcaagaccac ctggagatct acattcagca ggcattgcacc cgccctatca cctccattc
agcacagaag cccctgcgga aggtctgcat tggctcggag tggctctgga aattcacagg

```

```

ggagggagcc actgcccac tggaaacagg tgggttcac cgcagccagc ctgggggtccc
ccacccggac atccagttcc atttcctgcc atcccaagtg attgaccacg ggcgggtccc
caccagcag gaggttacc aggtacatgt ggggcccacg cggggcacga gtgtgggctg
gctcaaactg agaagtgcc atccccaaga ccaccctgtg atccagccca actacttgtc
aacagaaact gatattgagg atttcctgtc gtgtgtgaag ctcaccagag aaatttttgc
acaggaagcc ctggctccgt tccgagggaa agagctccag ccaggaagcc acattcagtc
agataaagag atagatgcct ttgtgcgggc aaaagccgac agcgcctacc acccctcgtg
cacctgtaag atgggcccag cctccgatcc cactgccgtg gtggatccgc agacaagggc
cctcgggggtg gaaaacctca gggcgtcga tgcctccac atgcctagca tggtcagcgg
caacctgaac gccccacaa tcatgatgc agagaaggca gctgacatta tcaaggggca
gcctgcactc tgggacaaag atgtccctgt ctacaagccc aggacgctgg ccaccagcg
ctaagacagt tgctgtgga ggatgaccag ggaagcccc tgataagcca agagggccag
cacagccctt gctcccaggc tcctgcctga aactatctag cacactagga cccaggtggt
accctactca gtggctgaga attggataaa gtcttgggaa atgagacaaa aaaaaaaaaa
aa

```

- 5 A continuación se enumeran los números de ID de clon I.M.A.G.E. Consortium y los correspondientes números de acceso a GenBank de secuencias identificadas como pertenecientes a los grupos I.M.A.G.E. Consortium y UniGene. También se incluyen secuencias que no están identificadas con un número de ID de clon pero aún así son de CHDH. Las secuencias incluyen aquellas de las secuencias de hebra "con sentido" y complementaria correspondientes a CHDH. Secuencias adicionales para su uso en la práctica de la descripción son aquellas alineadas en la Figura 5, que también se proporcionan en el Apéndice adjunto.

Número de ID de clon / número(s) de acceso a GenBank
4824572/BC034502 y BG720228
5191415/BI765156
5311690/B1667529
5267676/BI460380
11031605/AA609488
3842653/BE732217
4543273/BG336766
3504516/BE279319
3140587BE279968
6297066/BQ648069
2714263/AW449121
2735859/AW450678
2720363/AW139168
3642981/BF195860
5931105/BQ066460
3574335/BF430927
3268842/BF435866
3267752/BF435185

ES 2 399 246 T3

Número de ID de clon / número(s) de acceso a GenBank
1868020/AI264647 y BX116752 y AI733810 y AI792632
2365837/AI741739
3085519BF510364
1647746/AI034449
2695349/AW194S22
2285283/AI628996
2694067/AW235087
2285315/AI629023
2463061/AI928186
2462306/AI927042
2381448/AI768443
2298488/AI650346
3134601/BF197300
2300327/AI631941
2697626/AW167538
3034918/AW779820
2525301/AW024823
2300291/AI631914
2137091/AI473735
4147169/BG060119
2772286/AW299654
2172535/AI564145
2690214/AW241612
1868068/AI241086
1608918/AA991365
3134810/BF197431
1869723/AI245204
2691133/AW242403
6109050/BU500214
2384051/AI796286
12055388/AI308167
3032446/AW771262
2907815/AW340332
1636795//AI792354 y AI017355
2299592/AI640195
2054920/AI334627
2690173/AW237735
1869819/AI245373
3195030BE464406

Número de ID de clon / número(s) de acceso a GenBank
1646613/AI025866
2773291/AW299629
2461358/AI942245
5678397/BM142449 y BM142311
5672209/BM052814 y BM053126
2137904/AI800207
511224/A088689 y M088826
2734357/AW449405
381379/AA052926 y AA052927
2337545/AI914219
2528186/AW337722
2028284/AI262965
3436048/BF940636
2344677/AI695649
123940/R00867 y R01524/
240988/H90906 y H91018
240077/H82409 y H82667
Ninguno (secuencias de ARNm)/NM_018397.1 y AJ272267.1 y AK055402.1
Ninguno/AA772473.1 y BM682615.1 y BM713059.1 y BM716959.1 y BU738538.1 y AA324019.1 y AA302740.1 y C20981.1 y BF930030.1 y BQ303877.1 y BM769931.1 y AW900269.1 y F26419.1 y CB147231.1 y BE765491.1 AV656671.1

En una realización preferida, puede usarse cualquier secuencia, o parte única de la misma, de las secuencias CHDH de las Figuras 5 ó 6 en la práctica de la descripción.

- 5 En otro conjunto de realizaciones preferidas de la descripción, puede usarse cualquier secuencia, o parte única de la misma, de las secuencias QPRT del grupo I.M.A.G.E. Consortium NM_014298, así como los grupos UniGene Homo sapiens Hs. 126688. Asimismo, puede usarse cualquier secuencia que codifique todo o parte de la proteína codificada por cualquier secuencia de QPRT descrita en este documento. La secuencia consenso del grupo I.M.A.G.E. Consortium es la siguiente, con la región codificada asignada (finalizando con un codón de terminación) subrayada y precedida por la región no traducida y/o no codificada 5' y seguida de la región no traducida y/o no codificada 3':
- 10

SEC ID Nº 5 (secuencia consenso para QPRT, identificada como NM_014298 o NM_014298.2):

```

gtcctgagca gccaacacac cagcccagac agctgcaagt caccatggac gctgaaggcc
tggcgctgct gctgccgccc gtcaccctgg cagccctggg ggacagctgg ctccgagagg
actgcccagg gctcaactac gcagccttgg tcagcggggc aggccccctc caggcggcgc
tgtgggcoaa atcccctggg gtactggcag ggcagccttt ctccgatgcc atatttacct
aactcaactg ccaagtctcc tggttcctcc ccgagggatc gaagctggtg ccggtggcca
gagtggccga ggtccggggc cctgcccact gcctgctgct gggggaacgg gtggccctca
acacgctggc ccgctgcagt ggcattgcca gtgctgccgc cgtgcagtg gaggccgcca
ggggggccgg ctggactggg cacgtggcag gcacgaggaa gaccacgcca ggcttccggc
tgggtggagaa gtatgggctc ctgggtggcg gggccgcctc gcaccgctac gacctgggag
ggctgggtgat gttgaaggat aaccatgtgg tgcctcccg ggcgtggag aaggcgggctc
gggcggccag acaggcggct gacttcgctc tgaagggtga agtggaatgc agcagcctgc
aggaggctgt ccaggcagct gaggtggcg ccgacctgt cctgctggac aacttcaagc
cagaggagct gcacccacg gccaccgcgc tgaaggccca gttcccgagt gtggctgtgg
aagccagtgg gggcatcacc ctggacaacc tccccagtt ctgcggggcc caccatagacg
tcatctccat ggggatgctg acccaggcgg tccagacct tgatttctcc ctcaagctgt
ttgccaaaga ggtggctcca gtgccccaaa tccactagtc ctaaaccgga agaggatgac
accggccatg ggttaacgtg gctcctcagg accctctggg tcacacatct ttagggctcag
tgaacaatgg ggcacatttg gcaactagctt gagcccaact ctggctctgc cactgctgc
tcctgtgacc tgtcagggt gacttcacct ctgctcatct cagtttccta atctgtaaaa
tgggtctaataaagatcaa ccaaaaaaaaaa aaaaaaaaaa aa

```

- 5 A continuación se enumeran los números de ID de clon I.M.A.G.E. Consortium y los correspondientes números de acceso a GenBank de secuencias identificadas como pertenecientes a los grupos I.M.A.G.E. Consortium y UniGene. También se incluyen secuencias que no están identificadas con un número de ID de clon pero aún así son de QPRT. Las secuencias incluyen aquellas de las secuencias de hebra "con sentido" y complementaria correspondientes a QPRT. Secuencias representativas de los siguientes números de acceso a GenBank se presentan en el Apéndice adjunto.

Número de ID de clon / número(s) de acceso a GenBank
2960170/BC005060 y BE299670 y BE299712 □
3506460/BE273102 y BC010033 □
3959973/BC018910 y BE902622 □
267692/N23182 y N32648 □
3843834/BE735342 □
4872092/BX283118 y BG769505 □
4845859/13G750434 □
4868806/BG766440 □
4594651/BG441877 □
4553618/BG337811 □
4554044/BG338063 □

ES 2 399 246 T3

Número de ID de clon / número(s) de acceso a GenBank
44731611/BG251163 ☐
14581127/BG396079 ☐
4136221/BF316915 ☐
4127089/BF313098 ☐
4508387/BG257831 ☐
4125826/BF312975 ☐
4411920/BG115486
4842556/BG748194
4395232/BF980859
4122808/BF304964
6305325/BQ643384
4107138/BF204965
4875437/BG753310/
6337913/BU501237
4136491/BF317004
4131857/13F307788
4302204/BF684687
5092370/B1195027
3353576/BE257622
4473768/BG252578
3912491/BE887856
160124003/BU186666
4873695/BG751315
4873694/BG751234
4080072/BF237708
6300166/BQ876922
896716/B1198351
4877853/BG770209
4896715/BI198375
5087154/BI252426 y BI252874
6085289/BU174626
5741237/BM558378
4995462/BI088884
6720160/CA488404
5764841/BM926410
6208509/BQ879962
4581968/BG396587
5554997/BM477735

Número de ID de clon / número(s) de acceso a GenBank
3451668/BE538581
5803440/BQ069150
6250974/BQ688755
6251079/BQ685759
5535758/BM468306
6146330/BU165540
1740729/AI191477
2729947/AW293885
6082577BU174653
2753118/AW275889
2437568/AI884372
2507497/AI961218
5207705/BI771713
2067750/AI383718
263894/N28522 y H99843
1148416/AA627205
138014/R63144
70610/T49073 y T49074
5531252/BM800219
3629874/BE409186
5001398/BI093643
4361336/BF971224
4451023/BG121013
3844815/BE730924
4361451/BF970190
4154033/BF346117
4915206/BG818225
4444686/BG118070
6086243/BU149745
4899066/BG829478
6086128/BU180123
4366207/BG108477
3140355/BE280221
5459527/BM012505
3627907/BE382922
5418599/BM016313
4862852/BG765156
14877780/BG769917

Número de ID de clon / número(s) de acceso a GenBank
3162024/BE262076
5182393/BI518189 y BI517759
5417445/BM015407
16015713/BU175170
417111/W87557 y W87461
4580196/BG395022
16271908/BQ648651
6298174BQ652789
6271910/BQ653475
6271630/BQ647246
798664/BM928534
6652195/BU860925
6299767/BQ651366
5225259/BI838658
4895399/BI198873
740128/A477534 y AA479051
617256/BU178924
4562784/BG326197
3957711/BE902093
6293406/BQ650920
Ninguno (secuencias de ARNm)/BT007231.1 y NM_014298.2 y AK090801.1 y D78177.1
Ninguno/CB156177.1 y BM711970.1 y BM675916.1 y BM675420.1 y BM714918.1 y AA337770.1 y AA305670.1 y AA305611.1 y BU622082.1 y AV705250.1 y AL528086.2 y AL531128.2 y AL543783.2 y AL548817.2 y AL554386.2 y AL563056.2 y AL563955.2 y AL570131.2 y AL573234.2 y BF956608.1 y AV648116.1 y AV645766.1 y BF742969.1 y AL577191.2 y CD050133.1 y CD049103.1 y BX417895.1 y CB529044.1 y CD250136.1 y AA054830.1 y BX508036.1 y BX454610.1

En otro conjunto de realizaciones preferidas de la invención, puede usarse cualquier secuencia, o parte única de la misma, de las secuencias HOXB13 del grupo I.M.A.G.E. Consortium NM_006361, así como los grupos UniGene Homo sapiens Hs.66731. Asimismo, puede usarse cualquier secuencia que codifique todo o parte de la proteína codificada por cualquier secuencia de HOXB13 descrita en este documento. La secuencia consenso del grupo I.M.A.G.E. Consortium es la siguiente, con la región codificada asignada (finalizando con un codón de terminación) subrayada y precedida por la región no traducida y/o no codificada 5' y seguida de la región no traducida y/o no codificada 3':

5

10 SEC ID N° 6 (secuencia consenso para HOXB13, identificada como NM_006361 o NM_006361.2):

cgaatgcagg cgacttgcca gctgggagcg atttaaaacg ctttggatcc ccccggcctg
 ggtggggaga gcgagctggg tgccccctag attccccgcc ccgcacctc atgagccgac
 cctcggtccc atggagcccc gcaattatgc caccttggat ggagccaagg atatcgaagg
 cttgctggga gcgggagggg ggcggaatct ggtcgccac tccccctga ccagccccc
 agcggcgcc acgctgatgc ctgctgtcaa ctatgcccc ttggatctgc caggctcggc
 ggagccgcca aagcaatgcc acccatgccc tggggtgccc caggggacgt cccagctcc
 cgtgccttat ggttactttg gaggcgggta ctactcctgc cgagtgtccc ggagctcgt
 gaaacccgtg gccaggcag ccaccctggc cgcgtacccc ggggagactc ccacggcggg
 ggaagagtac cccagtcgcc ccactgagtt tgccttctat cgggatatc cgggaacctc
 ccacgctatg gccagttacc tggacgtgtc tgtggtgag actctgggtg ctctggaga
 accgcgacat gactccctgt tgctgtgga cagttaccag tcttgggtc tcgctgggtg
 ctggaacagc cagatgtgtt gccagggaga acagaacca ccaggtccct tttggaaggc
 agcatttgca gactccagcg ggcagcacc tctgacgcc tgcgccttcc gtcgcggccg
 caagaaacgc attccgtaca gcaaggggca gttgcgggag ctggagcggg agtatgcggc
 taacaagtcc atcaccaagg acaagaggcg caagatctcg gcagccacca gcctctcgga
 ggcgcagatt accatctggt ttcagaaccg ccgggtcaaa gagaagaagg ttctcgccaa
 ggtgaagaac agcgtacccc ctttaagagat ctcttgcct ggggtgggagg agcgaaagtg
 ggggtgtcct ggggagacca gaaacctgcc aagcccaggc tggggccaag gactctgtg
 agaggccct agagacaaca ccttcccag gccactggt gctggactgt tctcaggag
 cggcctgggt acccagtatg tgcagggaga cggaacccca tgtgacaggc ccactccacc
 agggttccca aagaacctgg cccagtcata atcattcacc ctacagtgg caataatcac
 gataaccagt

- 5 A continuación se enumeran los números de ID de clon I.M.A.G.E. Consortium y los correspondientes números de acceso a GenBank de secuencias identificadas como pertenecientes a los grupos I.M.A.G.E. Consortium y UniGene. También se incluyen secuencias que no están identificadas con un número de ID de clon pero aún así son de HOXB13. Las secuencias incluyen aquellas de las secuencias de hebra "con sentido" y complementaria correspondientes a HOXB13. La secuencia de cada número de acceso a GenBank se presenta en el Apéndice adjunto.

Número de ID de clon	Números de acceso a GenBank
4250486	BF676461, BC007092
5518335	BM462617
4874541	BG752489
4806039	BG778198
3272315	CB050884, CB050885
4356740	BF96519
6668163	BU930208
1218366	AA807966
2437746	AI884491
11187697	AA652388
3647557	BF446158
1207949	AA657924
1047774	AA644637
3649397	BF222357
971664	AA527613
996191	AA533227
813481	AA456069, AA455572, BX117624

ES 2 399 246 T3

Número de ID de clon	Números de acceso a GenBank
6256333	BQ673782
2408470	AI814453
2114743	AI417272
998548	AA535663
2116027	AI400493
3040843	AW779219
1101311	AA594847
1752062	AI150430
898712	M494387
1218874	AA662643
2460189	AI935940
986283	AA532530
1435135	AA857572
1871750	AI261980
3915135	BE888751
2069668	AI378797
667188	AA234220, AA236353
1101561	AA588193
1170268	AI821103, AI821851, AA635855
2095067	AI420753
4432770	BG180547
783296	AA468306, AA468232
3271646	CB050115, CB050116
1219276	AA661819
30570598	CF146837
30570517	CF146763
30568921	CF144902
3099071	CF141511
3096992	CF139563
3096870	CF139372
3096623	CF139319
3096798	CF139275
30572408	CF122893
2490082	AI972423
2251055	AI918975
2419308	AI826991
2249105	AI686312
2243362	AI655923

Número de ID de clon	Números de acceso a GenBank
30570697	CF146922
3255712	BF476369
3478356	BF057410
3287977	BE645544
3287746	BE645408
3621499	BE388501
30571128	CF147366
30570954	CF147143
Ninguno (secuencias de ARNm)	BT007410, BC007092, U57052, U81599
Ninguno	CB120119, CB125764, AU098628, CB126130, BI023924, BM767063, BM794275, BQ363211, BM932052, AA357646, AW609525, CB126919, AW609336, AW609244, BF855145, AU126914, CB126449, AW582404, BX641644

En una realización preferida, puede usarse cualquier secuencia, o parte única de la misma, de la siguiente secuencia de HOXB13, identificada por BC007092 o BC007092.1, en la práctica de la invención.

5 SEC ID N° 7 (secuencia para HOXB 13):

GGATTCCCCCGGCTGGGTGGGGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCG
CACCTCATGAGCCGACCTCGGCTCCATGGAGCCCGCAATTATGCCACCTTGGATGGAG
CCAAGGATATCGAAGGCTTGCTGGGAGCGGGAGGGGGCGGAATCTGGTCGCCCCACTCCC
CTCTGACCAGCCACCCAGCGGCGCTACGCTGATGCCTGCTGTCAACTATGCCCCCTTGG
ATCTGCCAGGCTCGGCGGAGCCGCCAAAGCAATGCCACCCATGCCCTGGGGTGGCCCGAGG
GGACGTCGCCAGCTCCCGTGCCCTTATGGTTACTTTGGAGGCGGGTACTACTCCTGCCGAG
TGTCCCGGAGCTCGCTGAAACCCTGTGCCAGGCAGCCACCCTGGCCGCGTACCCCGCGG

AGACTCCACGGCCGGGAAGAGTACCCAGCCGCCCCACTGAGTTTGCCTTCTATCCGG
GATATCCGGGAACCTACCAGCCTATGGCCAGTTACCTGGACGTGTCTGTGGTGACAGCTC
TGGGTGCTCCTGGAGAACCGGACATGACTCCCTGTTGCCTGTGGACAGTTACCACTCTT
GGGCTCTCGCTGGTGGCTGGAACAGCCAGATGTGTTGCCAGGGAGAACAGAACCCACCAG
GTCCCTTTTGAAGGCAGCATTTCAGACTCCAGCGGGCAGCACCCCTCCTGACGCTTGCG
CCTTTTCGTGCGGCGCCGCAAGAAACGCATTCCGTACAGCAAGGGGCAGTTGCGGGAGCTGG
AGCGGGAGTATGCGGCTAACAAGTTTCATCACCAGGACAAGAGGCGCAAGATCTCGGCAG
CCACCAGCCTCTCGGAGCGCCAGATTACCATCTGGTTTCAGAACCGCCGGGTCAAAGAGA
AGAAGGTTCTCGCCAAGGTGAAGAACAGCGCTACCCCTTAAGAGATCTCCTTGCCCTGGGT
GGGAGGAGCGAAAAGTGGGGGTGTCTTGGGGAGACCAGGAACCTGCCAAGCCAGGCTGGG
GCCAAGGACTCTGCTGAGAGGCCCTAGAGACAACACCCTTCCCAGGCCACTGGCTGCTG
GACTGTTCTCAGGAGCGGCCTGGGTACCCAGTATGTGCAGGGAGACGGAACCCCATGTG
ACAGCCCACTCCACCAGGGTTCCCAAAGAACCTGGCCAGTCATAATCATTCATCCTGAC
AGTGGCAATAATCACGATAACCAGTACTAGCTGCCATGATCGTTAGCCTCATATTTTCTA
TCTAGAGCTCTGTAGAGCACTTTAGAAACCGCTTTCATGAATTGAGCTAATTATGAATAA
ATTTGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

- 10 Las secuencias identificadas por SEC ID N° se proporcionan usando representaciones convencionales de una hebra de ADN empezando a partir del extremo ligado a 5' fosfato hasta el extremo ligado a 3' hidroxilo. La asignación de las regiones codificantes es generalmente por comparación con la secuencia o secuencias consenso y por lo tanto pueden contener inconsistencias relativas a otras secuencias asignadas al mismo grupo. Éstas no tienen efecto sobre la práctica de la invención porque la invención puede ponerse en práctica por el uso de segmentos más cortos (o combinaciones de los mismos) de secuencias únicas para cada una de las tres series descritas anteriormente y afectadas por inconsistencias. Como ejemplos no limitantes, puede usarse un segmento de la secuencia de ácido nucleico IL17BR, CHDH, QPRT, o HOXB13 compuesta por una secuencia de la región no traducida 3' y/o una
- 15 secuencia desde el extremo 3' de la región codificante como sonda para la detección de la expresión de IL17BR,

CHDH, QPRT, o HOXB13, respectivamente, sin verse afectada por la presencia de ninguna inconsistencia en las regiones codificantes debidas a diferencias entre las secuencias. Asimismo, el uso de un anticuerpo que reconozca específicamente una proteína, o fragmento de la misma, codificada por las secuencias IL17BR, CHDH, QPRT, o HOXB13 descritas en este documento, para detectar su expresión no se vería afectada por la presencia de ninguna inconsistencia en la representación de las regiones codificantes proporcionadas anteriormente.

Como apreciarán los especialistas en la técnica, algunas de las secuencias anteriores incluyen tramos 3' poli A (o poli T en la hebra complementaria) que no contribuyen a la singularidad de las secuencias descritas. La invención por tanto puede ponerse en práctica con secuencias que carecen de los tramos 3' poli A (o poli T). La singularidad de las secuencias descritas se refiere a las partes o las totalidades de las secuencias que se encuentran solamente en los ácidos nucleicos IL17BR, CHDH, QPRT, o HOXB13, incluyendo las secuencias únicas encontradas en la parte no traducida 3' de los genes. Secuencias únicas preferidas para la práctica de la descripción son aquellas que contribuyen a las secuencias consenso para cada una de las tres series de modo que las secuencias únicas serán útiles para detectar la expresión en una diversidad de individuos en lugar de ser específicas para un polimorfismo presente en algunos individuos. Como alternativa, pueden usarse secuencias únicas para un individuo o una subpoblación. Las secuencias únicas preferidas son preferiblemente de las longitudes de los polinucleótidos de la invención como se analiza en este documento.

Para determinar los niveles de expresión (aumentados o disminuidos) de las secuencias descritas anteriormente en la práctica de la presente invención, puede utilizarse cualquier método conocido en la técnica. En una realización preferida de la invención, se usa la expresión basada en la detección de ARN que hibrida con polinucleótidos que contienen las secuencias descritas anteriormente. Esto se realiza fácilmente por cualquier método de detección o amplificación+detección de ARN conocido o reconocido como equivalente en la técnica tal como, aunque sin limitación, PCR de transcripción inversa (opcionalmente PCR a tiempo real), los métodos descritos en la patente de Estados Unidos 6.794.141, los métodos descritos en la patente de Estados Unidos 6.291.170, y PCR cuantitativa. También pueden usarse métodos para identificar la estabilidad aumentada del ARN (que provoca la observación de una expresión aumentada) o estabilidad disminuida del ARN (que provoca una observación de expresión disminuida). Estos métodos incluyen la detección de secuencias que aumentan o disminuyen la estabilidad de los ARNm que contienen las secuencias IL17BR, CHDH, QPRT, o HOXB13 descritas en este documento. Estos métodos también incluyen la detección de degradación aumentada de ARNm.

En realizaciones particularmente preferidas de la descripción, se usan polinucleótidos que tienen secuencias presentes en las regiones no traducidas y/o no codificantes 3' de las secuencias descritas anteriormente para detectar la expresión o ausencia de expresión de las secuencias IL17BR, CHDH, QPRT, o HOXB13 en células mama en la práctica de la descripción. Dichos polinucleótidos pueden contener opcionalmente secuencias en las partes 3' de las regiones codificantes de las secuencias descritas anteriormente. Los polinucleótidos que contienen una combinación de secuencias de las regiones codificantes y no codificantes 3' preferiblemente tienen las secuencias dispuestas de forma contigua, sin secuencias o secuencias heterólogas intermedias.

Como alternativa, la descripción puede ponerse en práctica con polinucleótidos que tienen secuencias presentes en las regiones no traducidas y/o no codificantes 5' de las secuencias IL17BR, CHDH, QPRT, o HOXB13 en células de mama para detectar sus niveles de expresión. Dichos polinucleótidos pueden contener opcionalmente secuencias encontradas en las partes 5' de las regiones codificantes. Los polinucleótidos que contienen una combinación de secuencias de las regiones codificantes y no codificantes 5' preferiblemente tienen las secuencias dispuestas de forma contigua, son secuencia o secuencias intermedias. La descripción también puede ponerse en práctica con secuencias presentes en las regiones codificantes de IL17BR, CHDH, QPRT, o HOXB13.

Los polinucleótidos preferidos contienen secuencias de regiones no traducidas y/o no codificantes 3' ó 5' de al menos aproximadamente 16, al menos aproximadamente 18, al menos aproximadamente 20, al menos aproximadamente 22, al menos aproximadamente 24, al menos aproximadamente 26, al menos aproximadamente 28, al menos aproximadamente 30, al menos aproximadamente 32, al menos aproximadamente 34, al menos aproximadamente 36, al menos aproximadamente 38, al menos aproximadamente 40, al menos aproximadamente 42, al menos aproximadamente 44, o al menos aproximadamente 46 nucleótidos consecutivos. El término "aproximadamente" como se usa en la fresa previa se refiere a un aumento o disminución de 1 desde el valor numérico establecido. Son incluso más preferidos polinucleótidos que contienen secuencias de al menos o aproximadamente 50, al menos o aproximadamente 100, al menos o aproximadamente 150, al menos o aproximadamente 200, al menos o aproximadamente 250, al menos o aproximadamente 300, al menos o aproximadamente 350, o al menos o aproximadamente 400 nucleótidos consecutivos. El término "aproximadamente" como se usa en la frase precedente se refiere a un aumento o disminución del 10% a partir del valor numérico establecido.

Las secuencias del extremo 3' o 5' de las regiones codificantes descritas anteriormente encontradas en polinucleótidos de la invención son de las mismas longitudes que los descritos anteriormente, excepto que estarían limitadas de forma natural por la longitud de la región codificante. El extremo 3' de una región codificante puede incluir secuencias hasta la mitad 3' de la región codificante. A la inversa, el extremo 5' de una región codificante puede incluir secuencias hasta la mitad 5' de la región codificante. Por supuesto, las secuencias descritas

anteriormente, o las regiones codificantes y polinucleótidos que contienen partes de las mismas, pueden usarse en sus totalidades.

Los polinucleótidos que combinan las secuencias de una región no traducida y/o no codificante 3' y el extremo 3' asociado de la región codificante son preferiblemente de al menos o aproximadamente 100, al menos o aproximadamente 150, al menos o aproximadamente 200, al menos o aproximadamente 250, al menos o aproximadamente 300, al menos o aproximadamente 350, o al menos o aproximadamente 400 nucleótidos consecutivos. Preferiblemente, los polinucleótidos usados son del extremo 3' del gen, tal como en aproximadamente 350, aproximadamente 300, aproximadamente 250, aproximadamente 200, aproximadamente 150, aproximadamente 100, o aproximadamente 50 nucleótidos desde la señal de poliadenilación o sitio de poliadenilación de un o gen o secuencia expresada. También pueden usarse polinucleótidos que contienen mutaciones relativas a las secuencias de los genes descritos también siempre que la presencia de las mutaciones aún permita la hibridación para producir una señal detectable.

En otra realización de la invención, pueden usarse polinucleótidos que contienen delecciones de nucleótidos desde el extremo 5' y/o 3' de las secuencias descritas anteriormente. Las delecciones son preferiblemente de 1-5, 5-10, 10-15, 15-20, 20-25, 25-30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-125, 125-150, 150-175, ó 175-200 nucleótidos desde el extremo 5' y/o 3', aunque la extensión de las delecciones estaría limitada de forma natural por la longitud de las secuencias descritas y la necesidad de ser capaz de usar los polinucleótidos para la detección de los niveles de expresión.

Otros polinucleótidos de la invención desde el extremo 3' de las secuencias descritas anteriormente incluyen aquellos de cebadores y sondas opcionales para PCR cuantitativa. Preferiblemente, los cebadores y sondas son aquellos que amplifican una región de menos de aproximadamente 350, menos de aproximadamente 300, menos de aproximadamente 250, menos de aproximadamente 200, menos de aproximadamente 150, menos de aproximadamente 100, o menos de aproximadamente 50 nucleótidos desde la señal de poliadenilación o sitio de poliadenilación de un gen o secuencia expresada.

En otra realización de la invención, pueden usarse polinucleótidos que contienen partes de las secuencias descritas anteriormente que incluyen el extremo 3' en la práctica de la invención. Dichos polinucleótidos contendrían al menos o aproximadamente 50, al menos o aproximadamente 100, al menos o aproximadamente 150, al menos o aproximadamente 200, al menos o aproximadamente 250, al menos o aproximadamente 300, al menos o aproximadamente 350, o al menos o aproximadamente 400 nucleótidos consecutivos desde el extremo 3' de las secuencias descritas.

La descripción, por tanto, también incluye polinucleótidos usados para detectar la expresión de IL17BR, CHDH, QPRT, o HOXB13 en células de mama. Los polinucleótidos pueden comprender un polinucleótido más corto que consiste en secuencias halladas en las SEC ID N° proporcionadas anteriormente en combinación con secuencias heterólogas no halladas de forma natural en combinación con las secuencias IL17BR, CHDH, QPRT, o HOXB13.

Como ejemplos no limitantes, puede usarse un polinucleótido que comprende una de las siguientes secuencias en la práctica de la invención.

SEC ID N° 8:

GCTCTCACTGGCAAATGACAGCTCTGTGCAAGGAGCACTCCCAAGTATAAAAAATTATTAC

SEC ID N° 9:

TGCCTAATTTCACTCTCAGAGTGAGGCAGGTAAGTGGGGCTCCACTGGGTCCTCTGAGA

SEC ID N° 10:

GATCGTTAGCCTCATATTTTCTATCTAGAGCTCTGTAGAGCACTTTAGAAACCGCTTTCA

La SEC ID N° 8 es una parte de la secuencia AI240933 mientras que la SEC ID N° 9 es una parte de la secuencia AJ272267 (ARNm de CHDH). Corresponden a las dos posiciones "60meras" indicadas en la Figura 3. La SEC ID N° 10 es un polinucleótido capaz de hibridar con algunas secuencias de HOXB13 como se describe en este documento.

Por tanto, la descripción puede ponerse en práctica con un polinucleótido que consiste en la secuencia de las SEC ID N° 8, 9 ó 10 en combinación con una o más secuencias heterólogas que no se hallan normalmente con las SEC ID N° 8, 9 ó 10. Como alternativa, la descripción también puede ponerse en práctica con un polinucleótido que consiste en la secuencia de las SEC ID N° 8, 9 ó 10 en combinación con una o más secuencias de origen natural que se hallan normalmente con las SEC ID N° 8, 9 ó 10.

Pueden usarse polinucleótidos con secuencias que comprenden las SEC ID N° 8 ó 9, de origen natural o sintético, para detectar ácidos nucleicos que se sobre-expresan en células de cáncer de mama que son sensibles, y aquellos que no se sobre-expresan en células de cáncer de mama que son no sensibles, al tratamiento con TAM u otro agente "antiestrógenos" contra el cáncer de mama. Pueden usarse polinucleótidos con secuencias que comprenden

la SEC ID Nº 10, de origen natural o sintético, para detectar ácidos nucleicos que se infra-expresan en células de cáncer de mama que son sensibles, y aquellos que no se infra-expresan en células de cáncer de mama que son no sensibles, al tratamiento con TAM u otro agente "antiestrógenos" contra el cáncer de mama.

- 5 Secuencias adicionales que pueden usarse en polinucleótidos como se ha descrito anteriormente para la SEC ID Nº 8 y 9 es la siguiente, que es complementaria a una parte de las secuencias de IL17BR descritas en este documento:

SEC ID Nº 11: TCCAATCGTTAGTTAATGCTACATTAGTT

- 10 Secuencias adicionales que pueden usarse en polinucleótidos como se ha descrito anteriormente para la SEC ID Nº 10 son la siguiente, que es complementaria a una parte de las secuencias de IL17BR descritas en este documento:

SEC ID Nº 12: CAATTCATGAAAGCGGTTTCTAAAG

- 15 Además, pueden usarse cebadores de secuencias definidas para amplificar por PCR partes de las secuencias de CHDH para determinar su nivel de expresión. Por ejemplo, pueden usarse cebadores que comprenden las siguientes secuencias para amplificar una parte de la secuencia AI240933.

Cebador directo (SEC ID Nº 13): TGAAGTGTTTTGCCTGGATCA

- 20 Cebador inverso (SEC ID Nº 14): CACCACTTTGTTATGAAGACCTTACAA

- 25 En algunas realizaciones de la descripción, los cebadores pueden usarse en métodos de RT-PCR cuantitativa conocidos en la técnica, opcionalmente en presencia de una sonda marcada o detectable que se une a ácidos nucleicos bicatenarios (tal como Sybr Green™) o una sonda específica tal como una sonda "TaqMan". En una realización, dicha sonda puede comprender la secuencia AGTAAGAATGTCTTAAGAAGAGG (SEC ID Nº 15) para la detección de la expresión de AI240933.

- 30 Además, pueden usarse polinucleótidos que contienen otras secuencias, particularmente secuencias únicas, presentes en moléculas de ácido nucleico de origen natural que comprenden las SEC ID Nº 8-15 en la práctica de la descripción.

- 35 Otros polinucleótidos para su uso en la práctica de la descripción incluyen aquellos que tienen suficiente homología con aquellos descritos anteriormente para detectar la expresión por el uso de técnicas de hibridación. Dichos polinucleótidos preferiblemente tienen aproximadamente o el 95%, aproximadamente o el 96%, aproximadamente o el 97%, aproximadamente o el 98%, o aproximadamente o el 99% de identidad con secuencias de IL17BR, CHDH, QPRT, o HOXB13 descritas en este documento. La identidad se determina usando el algoritmo BLAST, como se ha descrito anteriormente. Los otros polinucleótidos para su uso en la práctica de la descripción también pueden describirse en base a la capacidad de hibridar con polinucleótidos de la invención en condiciones rigurosas de aproximadamente el 30% v/v a aproximadamente el 50% de formamida y de aproximadamente 0,01 M a aproximadamente 0,15 M de sal para hibridación y de aproximadamente 0,01 M a aproximadamente 0,15 M de sal para las condiciones de lavado a aproximadamente 55 a aproximadamente 65°C o superior, o condiciones equivalentes a las mismas.

- 45 En una realización adicional de la descripción, se proporciona una población de moléculas de ácido nucleico monocatenario que comprenden una o ambas hebras de una secuencia humana de IL17BR, CHDH, QPRT, o HOXB13 como sonda, de modo que pueda hibridarse al menos una parte de dicha población con una o ambas hebras de una molécula de ácido nucleico amplificada cuantitativamente a partir del ARN de una célula de cáncer de mama. La población puede ser solamente la hebra antisentido de una secuencia humana de IL17BR, CHDH, QPRT, o HOXB13 de modo que pueda hibridarse una hebra con sentido de una molécula a partir de, o amplificarse a partir de, una célula de cáncer de mama con una parte de dicha población. En el caso de IL17BR o CHDH, la población preferiblemente comprende una cantidad suficientemente en exceso de dicha una o ambas hebras de una secuencia humana de IL17BR o CHDH en comparación con la cantidad de moléculas de ácido nucleico expresadas (o amplificadas) que contienen una secuencia de IL17BR o CHDH complementaria a partir de una célula de mama normal. Esta condición de exceso permite que la cantidad aumentada de expresión de ácido nucleico en una célula de cáncer de mama sea fácilmente detectable como un aumento.

- 60 Como alternativa, la población de moléculas monocatenarias es igual a o en exceso de todo de una o ambas hebras de las moléculas de ácido nucleico amplificadas a partir de una célula de cáncer de mama de modo que la población es suficiente para hibridar con todo de una o ambas hebras. Las células preferidas son aquellas de una paciente con cáncer de mama que es ER+ o para la cual se contempla el tratamiento con tamoxifeno o uno o más agentes "antiestrógenos" diferentes contra el cáncer de mama. Las moléculas monocatenarias pueden, por supuesto, ser la forma desnaturalizada de cualquier secuencia de IL17BR, CHDH, QPRT, o HOXB13 que contenga la molécula de ácido nucleico bicatenaria o polinucleótido descrito en este documento.

- 65 La población también puede describirse como hibridada con una secuencia de IL17BR o CHDH que contiene moléculas de ácido nucleico a un nivel al menos dos veces tan elevado como por moléculas de ácido nucleico de

una célula de mama. Como en las realizaciones descritas anteriormente, las moléculas de ácido nucleico pueden ser aquellas amplificadas cuantitativamente a partir de una célula de cáncer de mama de modo que reflejen la cantidad de expresión en dicha célula.

La población preferiblemente se inmoviliza en un soporte sólido, opcionalmente en forma de una localización en una microserie. Una parte de la población se hibrida preferiblemente con moléculas de ácido nucleico amplificadas cuantitativamente a partir de una célula de mama no normal o anormal por amplificación de ARN. El ARN amplificado puede ser el obtenido de una célula de cáncer de mama, siempre que la amplificación usada fuera cuantitativa con respecto a secuencias que contienen IL17BR, CHDH, QPRT, o HOXB13.

En otra realización de la descripción, puede usarse expresión basada en la detección del estado de ADN. La detección del ADN de QPRT o HOXB13 metilado, deleciónado o inactivado de otro modo, puede usarse como indicación de expresión disminuida hallada en células de mama no normales. Esto puede realizarse fácilmente por métodos basados en PCR conocidos en la técnica. El estado de las regiones promotoras de QPRT o HOXB13 también pueden ensayarse como una indicación de expresión disminuida de secuencias de QPRT o HOXB13. Un ejemplo no limitante es el estado de metilación de secuencias halladas en la región promotora.

A la inversa, puede usarse la detección del ADN de una secuencia amplificada como indicación de expresión aumentada hallada en células de mama no normales. Esto puede realizarse fácilmente por métodos basados en PCR, de hibridación fluorescente *in situ* (FISH) e hibridación de cromosomas *in situ* (CISH) conocidos en la técnica.

Una realización preferida que usa un ensayo basado en ácido nucleico para determinar la expresión es por inmovilización de una o más de las secuencias identificadas en este documento en un soporte sólido incluyendo, aunque sin limitación, un sustrato sólido como una serie o perlas o tecnología basada en perlas como se sabe en la técnica. Como alternativa, también pueden usarse ensayos de expresión basados en solución conocidos en la técnica. La secuencia o secuencias inmovilizadas pueden estar en forma de polinucleótidos como se describe en este documento, de modo que el polinucleótido sería capaz de hibridar con un ADN o ARN correspondiente a la secuencia o secuencias.

El polinucleótido o polinucleótidos inmovilizados pueden usarse para determinar el estado de muestras de ácido nucleico preparadas a partir de una célula o células de cáncer de mama de muestra, opcionalmente como parte de un método para detectar el estado ER en dicha célula o células. Sin limitar la descripción, dicha célula puede ser de un paciente sospechoso de estar afectado con, o en riesgo de desarrollar, cáncer de mama. El polinucleótido o polinucleótidos inmovilizados tiene que ser solamente suficientes para hibridar específicamente con las moléculas de ácido nucleico correspondientes obtenidas de la muestra (y para la exclusión de hibridación detectable o significativa con otras moléculas de ácido nucleico).

En otra realización de la descripción, puede usarse una proporción de los niveles de expresión de dos de los genes descritos para predecir la respuesta al tratamiento con TAM u otro SERM. Preferiblemente, la proporción es la de dos genes con patrones opuestos de expresión, tal como un gen infra-expresado a un gen sobre-expresado, en correlación con el mismo fenotipo. Ejemplos no limitantes incluyen la proporción de HOXB13 sobre IL17BR o la proporción de QPRT sobre CHDH. Este aspecto de la invención se basa en parte en la observación de que dicha proporción tiene una correlación más fuerte con el resultado del tratamiento con TAM que el nivel de expresión de cualquier gen solo. Por ejemplo, la proporción de HOXB13 sobre IL17BR tiene una precisión de clasificación observada del 77%.

Como ejemplo no limitante, pueden usarse los valores Ct de detección basada en Q-PCR de los niveles de expresión génica para obtener una proporción para predecir la respuesta al tratamiento con uno o más agentes "antiestrógenos" contra el cáncer de mama.

Realizaciones adicionales de la invención

En realizaciones en las que tiene que analizarse solamente uno o unos pocos genes, el ácido nucleico obtenido de la célula o células de cáncer de mama de muestra puede amplificarse preferentemente por el uso de cebadores apropiados de modo que se amplifiquen solamente los genes a analizar para reducir las señales de fondo contaminantes de otros genes expresados en la célula de mama. Como alternativa, y cuando tienen que analizarse múltiples genes o cuando se usan muy pocas células (o una célula), el ácido nucleico de la muestra puede amplificarse de forma global antes de la hibridación de los polinucleótidos inmovilizados. Por supuesto, puede marcarse directamente y usarse el ARN o el equivalente de ADNc del mismo, sin amplificación, por métodos conocidos en la técnica.

También puede usarse la expresión de secuencias basada en la detección de una presencia, aumento o disminución en los niveles de proteínas o actividad. La detección puede realizarse por cualquier procedimiento basado en inmunohistoquímica (IHC), basado en fluido corporal (en el que un polipéptido IL17BR, CHDH, QPRT o HOXB13, o fragmento del mismo, se encuentra en un fluido corporal tal como, pero no se limita, a sangre), basado en anticuerpos (incluyendo autoanticuerpos contra la proteína, si están presente), basado en células exfoliadas (del

cáncer), basado en espectroscopía de masas y basado en imágenes (incluyendo el uso de ligando marcado, cuando esté disponible) conocido en la técnica y reconocido según convenga para la detección de la proteína. Los procedimientos basados en anticuerpos y en imágenes son adicionalmente útiles para la localización de tumores después de la determinación de cáncer usando células obtenidas por un procedimiento no invasivo (tal como lavado ductal o aspiración con aguja fina), en el que la fuente de las células cancerosas es desconocida. Un anticuerpo marcado o ligando puede usarse para localizar el (los) carcinoma(s) dentro de una paciente.

Anticuerpos para su uso en tales procedimientos de detección incluyen anticuerpos policlonales, opcionalmente aislados de fuentes que se producen naturalmente cuando están disponibles, y anticuerpos monoclonales, que incluyen aquellos preparados por el uso de polipéptidos IL17BR, CHDH, QPRT o HOXB13 (o fragmentos de los mismos) como antígenos. Tales anticuerpos, además de fragmentos de los mismos (que incluyen, pero no se limitan a, fragmentos Fab) sirven para detectar o diagnosticar células de mama no normales o cancerosas en virtud de su capacidad para unirse específicamente a polipéptidos IL17BR, CHDH, QPRT o HOXB13 para la exclusión de otros polipéptidos que producen una señal detectable. También pueden usarse anticuerpos recombinantes, sintéticos e híbridos con la misma capacidad en la práctica de la divulgación. Los anticuerpos pueden generarse fácilmente por inmunización con un polipéptido IL17BR, CHDH, QPRT o HOXB13 (o fragmento del mismo), y también pueden usarse sueros policlonales en la práctica de la divulgación.

Los procedimientos de detección basados en anticuerpos son muy conocidos en la técnica e incluyen ensayo tipo sándwich y de ELISA, además de ensayos basados en transferencia Western y citometría de flujo como ejemplos no limitantes. Muestras para análisis en tales procedimientos incluyen cualquiera que contenga polipéptidos IL17BR, CHDH, QPRT o HOXB13 o fragmentos de los mismos. Ejemplos no limitantes incluyen aquellos que contienen células de mama y contenidos de células, además de fluidos corporales (incluyendo sangre, suero, saliva, líquido linfático, además de secreciones de la mucosa y otras celulares como ejemplos no limitantes) que contienen los polipéptidos.

Las realizaciones de ensayo anteriores pueden usarse de varias formas diferentes para identificar o detectar la respuesta al tratamiento con TAM u otro agente "antiestrógeno" contra cáncer de mama basándose en la expresión génica en una muestra de células de cáncer de mama de una paciente. En algunos casos, esto reflejaría un cribado secundario para la paciente, que puede ya haberse sometido a una mamografía o examen físico como cribado primario. Si el cribado primario es positivo, la posterior biopsia con aguja, lavado ductal, aspiración con aguja fina u otro procedimiento mínimamente invasivo análogo puede proporcionar la muestra para su uso en las realizaciones de ensayo antes, simultáneamente con o después de ensayar el estado de RE. La presente invención es particularmente útil en combinación con protocolos no invasivos tales como lavado ductal o aspiración con aguja fina, para preparar una muestra de células de mama.

La presente invención proporciona un conjunto más objetivo de criterios en forma de perfiles de expresión génica de un conjunto discreto de genes para discriminar (o delinear) un conjunto discreto de genes, para discriminar (o delinear) entre desenlaces de cáncer de mama. En realizaciones particularmente preferidas de la invención, los ensayos se usan para discriminar entre desenlaces buenos y malos después del tratamiento con tamoxifeno u otro agente "antiestrógeno" contra cáncer de mama. Pueden realizarse comparaciones que discriminan entre desenlaces después de aproximadamente 10, aproximadamente 20, aproximadamente 30, aproximadamente 40, aproximadamente 50, aproximadamente 60, aproximadamente 70, aproximadamente 80, aproximadamente 90, aproximadamente 100 o aproximadamente 150 meses.

Aunque los desenlaces de supervivencia buenos y malos pueden definirse relativamente en comparación entre sí, un desenlace "bueno" puede visualizarse como una tasa de supervivencia mejor del 50% después de aproximadamente 60 meses de la intervención posquirúrgica para extirpar tumor(es) de cáncer de mama. Un desenlace "bueno" también puede ser una tasa de supervivencia mejor de aproximadamente el 60%, aproximadamente el 70%, aproximadamente el 80% o aproximadamente el 90% después de aproximadamente 60 meses de la intervención posquirúrgica. Un desenlace "malo" puede visualizarse como una tasa de supervivencia del 50% o menos después de aproximadamente 60 meses de la intervención posquirúrgica para extirpar tumor(es) de cáncer de mama. Un desenlace "malo" también puede ser una tasa de supervivencia de aproximadamente el 70% o menos después de aproximadamente 40 meses, o aproximadamente una tasa de supervivencia de aproximadamente el 80% o menos después de aproximadamente 20 meses de la intervención posquirúrgica.

En otra realización de la divulgación basada en la expresión de algunos genes, el aislamiento y el análisis de una muestra de células de cáncer de mama puede realizarse del siguiente modo:

- (1) Se realiza lavado ductal u otro procedimiento no invasivo en una paciente para obtener una muestra.
- (2) La muestra se prepara y se recubre sobre un portaobjetos de microscopio. Obsérvese que el lavado ductal produce agrupaciones de células que se examinan citológicamente como se ha establecido anteriormente.
- (3) El anatomopatólogo o software de análisis de imágenes barre la muestra para la presencia de células atípicas.

(4) Si se observan células atípicas, aquellas células se recogen (por ejemplo, por microdissección tal como LCM).

(5) Se extrae ARN de las células recogidas.

(6) Se ensaya el ARN, directamente o después de la conversión en ADNc o amplificación del mismo, para la expresión de secuencias de IL17BR, CHDH, QPRT y/o HOXB13.

Con el uso de la presente invención, los médicos expertos pueden prescribir o detener el tratamiento con TAM u otro agente "antiestrógeno" contra cáncer de mama basándose en el pronóstico determinado por la práctica de la presente invención.

La discusión anterior también es aplicable cuando se detecta una lesión palpable, seguido de aspiración con aguja fina o biopsia con aguja de células de la mama. Las células se siembran en placa y son revisadas por un anatomopatólogo o sistema de obtención de imágenes automatizado que selecciona células para el análisis como se ha descrito anteriormente.

Sin embargo, la presente invención también puede usarse con biopsias de tejido sólido, que incluyen aquellas guardadas como un espécimen FFPE. Por ejemplo, una biopsia sólida puede recogerse y prepararse para visualización, seguido de determinación de la expresión de uno o más genes identificados en el presente documento para determinar el desenlace del cáncer de mama. Como otro ejemplo no limitante, una biopsia sólida puede recogerse y prepararse para visualización, seguido de determinación de la expresión de IL17BR, CHDH, QPRT y/o HOXB13. Un medio preferido es mediante el uso de hibridación *in situ* con sonda(s) que identifican polinucleótidos o proteínas para ensayar la expresión de dicho(s) gen(es).

En un procedimiento alternativo, la biopsia de tejido sólido puede usarse para extraer moléculas, seguido de análisis de la expresión de uno o más genes. Esto proporciona la posibilidad de excluir la necesidad de visualización y recogida de sólo células cancerosas o células que se sospecha que son cancerosas. Este procedimiento puede por supuesto modificarse de forma que sólo se recojan células que han sido seleccionadas positivamente y se usan para extraer moléculas para el análisis. Esto requeriría la visualización y selección como un requisito previo al análisis de la expresión génica. En el caso de una muestra FFPE, las células pueden obtenerse seguido de extracción, amplificación y detección de ARN como se describe en el presente documento.

En una alternativa a lo anterior, la(s) secuencia(s) identificada(s) en el presente documento puede(n) usarse como parte de una PCR simple o ensayo basado en matriz simplemente para determinar la respuesta al tratamiento con TAM u otro agente "antiestrógeno" contra cáncer de mama mediante el uso de una muestra de un procedimiento de muestreo no invasivo o mínimamente invasivo. La detección de la expresión de secuencias de muestras puede ser mediante el uso de una única micromatriz capaz de ensayar la expresión de las secuencias desveladas, además de otras secuencias, que incluyen secuencias conocidas por no variar en los niveles de expresión entre células de mama normales y no normales, por comodidad y precisión mejorada.

Otros usos de la presente invención incluyen proporcionar la capacidad para identificar muestras de células de cáncer de mama que tienen diferentes respuestas a tratamiento con TAM u otro agente "antiestrógeno" contra cáncer de mama para investigación o estudio adicional. Esto proporciona un avance basado en criterios genéticos/moleculares objetivos.

En otra realización más de la divulgación basada en la expresión de múltiples genes en un patrón o perfil de expresión, el aislamiento y análisis de una muestra de células de cáncer de mama puede realizarse del siguiente modo:

(1) Se realiza lavado ductal u otro procedimiento no invasivo en un paciente para obtener una muestra.

(2) La muestra se prepara y se recubre sobre un portaobjetos de microscopio. Obsérvese que el lavado ductal produce agrupaciones de células que se examinan citológicamente como se ha establecido anteriormente.

(3) El anatomopatólogo o software de análisis de imágenes barre la muestra para la presencia de células no normales y/o de cáncer de mama atípicas.

(4) Si se observan células atípicas, aquellas células se recogen (por ejemplo, por microdissección tal como LCM).

(5) Se extrae ARN de las células recogidas.

(6) El ARN se purifica, se amplifica y se marca.

(7) El ácido nucleico marcado se pone en contacto con una micromatriz que contiene polinucleótidos complementarios a todo o parte de uno o más de los genes identificados en el presente documento como que guardan relación con discriminaciones entre desenlaces de cáncer de mama bajo condiciones de hibridación adecuadas, luego se procesan y se barren para obtener un patrón de intensidades de cada sitio (con respecto a un control para la expresión génica general en células) que determinan el nivel de expresión del (de los) gen(es) en las células.

(8) El patrón de intensidades se analiza por comparación con los patrones de expresión de los genes en muestras de células de cáncer de mama conocidas que guardan relación con desenlaces (con respecto al mismo control).

Un ejemplo específico del procedimiento anterior sería realizar lavado ductal tras un cribado primario, observar y recoger células no normales y/o atípicas para el análisis. La comparación con patrones de expresión conocidos tales como aquellos hechos posibles por un modelo generado por un algoritmo (tal como, pero no se limitan a, análisis tipo vecino más próximo, SVM o redes neurales) con datos de expresión génica de referencia para los diferentes desenlaces de supervivencia de cáncer de mama identifica las células que guardan relación con sujetos con desenlaces buenos o malos. Otro ejemplo sería tomar un tumor de mama extirpado de un sujeto después de intervención quirúrgica, opcionalmente convirtiendo todo o parte de él en una muestra FFPE antes del posterior aislamiento y preparación de células de cáncer de mama del tumor para la determinación/identificación de células atípicas, no normales o cancerosas, y aislamiento de dichas células seguido de las etapas 5 a 8 anteriores.

Alternativamente, la muestra puede permitir la recogida de células tanto normales, además de cancerosas, para análisis. Los patrones de expresión génica para cada una de estas dos muestras se compararán entre sí, además del modelo, y dichas comparaciones normales frente a individuales basadas en el conjunto de datos de referencia. Este enfoque puede ser significativamente más poderoso que el que las células cancerosas sólo se aproximan debido a que utiliza significativamente más información de las células normales y las diferencias entre células normales y cancerosas (en tanto los conjuntos de datos de la muestra como de referencia) para determinar el desenlace de cáncer de mama de la paciente basándose en expresión génica en las células cancerosas de la muestra.

Los genes identificados en el presente documento también pueden usarse para generar un modelo que puede predecir los desenlaces de supervivencia y reaparición de cáncer de mama de una muestra de células de mama RE+ basándose en la expresión de los genes identificados en la muestra. Un modelo tal puede generarse por cualquiera de los algoritmos descritos en el presente documento o de otro modo conocidos en la técnica, además de aquellos reconocidos como equivalentes en la materia usando gen(es) (y subconjuntos de los mismos) desvelados en el presente documento para la identificación de desenlaces de cáncer de mama. El modelo proporciona un medio para comparar los perfiles de expresión de gen(es) del subconjunto de la muestra con los perfiles de datos de referencia usados para construir el modelo. El modelo puede comparar el perfil de la muestra con cada uno de los perfiles de referencia o con un modelo que define delineaciones hechas basándose en los perfiles de referencia. Adicionalmente puede usarse valores relativos del perfil de la muestra en comparación con el modelo o perfiles de referencia.

En una realización preferida de la invención, muestras de células de mama identificadas como normales y cancerosas del mismo sujeto pueden analizarse, opcionalmente mediante el uso de una única micromatriz, para sus perfiles de expresión de los genes usados para generar el modelo. Esto proporciona un medio ventajoso de identificar desenlaces de supervivencia y reaparición basados en diferencias relativas de los perfiles de expresión de la muestra normal. Esas diferencias pueden luego usarse en la comparación con diferencias entre datos de referencia normales y cancerosos individuales que también se usaron para generar el modelo.

Artículos de fabricación

Los materiales y procedimientos de la presente invención son idealmente aptos para la preparación de kits producidos según procedimientos muy conocidos. Los presentes inventores describen kits que comprenden agentes (como los polinucleótidos y/o anticuerpos descritos en el presente documento como ejemplos no limitantes) para la detección de la expresión de las secuencias desveladas. Se proporcionan tales kits, que opcionalmente comprenden el agente con una descripción identificadora o etiqueta o instrucciones referentes a su uso en los procedimientos de la presente invención. Un kit tal puede comprender recipientes, cada uno con uno o más de los diversos reactivos (normalmente en forma concentrada) utilizados en los procedimientos que incluyen, por ejemplo, micromatrices prefabricadas, tampones, los trifosfatos de nucleótidos apropiados (por ejemplo, dATP, dCTP, dGTP y dTTP; o rATP, rCTP, rGTP y UTP), transcriptasa inversa, ADN polimerasa, ARN polimerasa, y uno o más complejos de cebadores de la presente invención (por ejemplo, poli(T) de longitud apropiada o cebadores al azar ligados a un promotor reactivo con la ARN polimerasa). Normalmente también se incluirá un conjunto de instrucciones.

Los procedimientos proporcionados por la presente invención también pueden automatizarse por completo o en parte. Todos los aspectos de la presente invención también pueden ponerse en práctica de forma que consistan esencialmente en un subconjunto de los genes desvelados para la exclusión de material irrelevante para la identificación de desenlaces de supervivencia de cáncer de mama mediante una muestra que contiene células.

Habiendo descrito ahora generalmente la invención, la misma se entenderá más fácilmente mediante referencia a los siguientes ejemplos que se proporcionan a modo de ilustración, y que no pretenden ser limitantes de la presente invención, a menos que se especifique.

Ejemplos

Ejemplo 1

5 Procedimientos generales

Criterios de selección de pacientes y tumores y diseño del estudio

10 Los criterios de inclusión de pacientes para este estudio fueron: mujeres diagnosticadas en el Hospital general de Massachusetts (MGH) entre 1987 y 2000 con cáncer de mama positivo para RE, tratamiento con cirugía de mama convencional (mastectomía radical modificada o lumpectomía) y radiación, seguido de cinco años del adyuvante sistémico tamoxifeno; ninguna paciente recibió quimioterapia antes de la reaparición. Los datos clínicos y de seguimiento se derivaron del registro de tumores del MGH. No faltaron datos del registro y todos los registros médicos disponibles se registraron como un segundo escalón de la confirmación de datos.

15 Todos los especímenes de tumor recogidos en el momento del diagnóstico inicial se obtuvieron de repositorios de tejidos congelados e incorporados en parafina fijada a formalina (FFPE) en el Hospital general de Massachusetts. Las muestras tumorales con más del 20% de células tumorales se seleccionaron con una mediana superior al 75% para todas las muestras. Cada muestra se evaluó para las siguientes características: tipo de tumor (ductal frente a lobular), tamaño del tumor y grado histológico combinado de Nottingham. La expresión del receptor de estrógeno y progesterona se determinó por análisis de unión a hormona bioquímicos y/o por tinción inmunohistoquímica como se ha descrito (Long, A.A. y col. "High-specificity in-situ hybridization. Methods and application." *Diagn Mol Pathol* 1, 45-57 (1992)); la positividad del receptor se definió como superior a 3 fmol/mg de tejido tumoral (Long y col.) y superior al 1% de tinción nuclear para los ensayos bioquímicos e inmunohistoquímicos, respectivamente.

25 El diseño del estudio es del siguiente modo: se seleccionó un conjunto de entrenamiento de 60 especímenes de cáncer de mama congelados para identificar firmas de expresión génica predictivas de desenlace o respuesta, en el entorno de terapia de tamoxifeno adyuvante. Los tumores de pacientes que respondieron al tratamiento se hicieron corresponder con los de pacientes que no respondieron al tratamiento con respecto a estadificación de TNM y grado del tumor. La expresión génica diferencial identificada en el conjunto de entrenamiento se validó en un grupo independiente de 20 tumores de mama invasivos con muestras de tejido incorporado en parafina fijado a formalina (FFPE).

LCM, aislamiento y amplificación de ARN

35 Con cada muestra tumoral congelada dentro de la cohorte de 60 casos, el ARN se aisló de tanto una sección de tejido completo de 8 µm de espesor como de una población altamente enriquecida de 4.000-5.000 células epiteliales malignas adquiridas por una microdissección de captura láser usando un sistema de LCM PixCell Ile (Arcturus, Mountain View, CA). De cada muestra tumoral dentro del conjunto de prueba de 20 casos, el ARN se aisló de cuatro secciones de tejido FFPE de 8 µm de espesor. El ARN aislado se sometió a una ronda de transcripción *in vitro* con T7 polimerasa usando el kit RiboAmp™ (muestras congeladas) u otro sistema para muestras FFPE según instrucciones del fabricante (Arcturus Bioscience, Inc., Mountain View, CA para RiboAmp™). Se generó ARNc marcado por una segunda ronda de transcripción *in vitro* de ARN basada en T7 en presencia de 5-[3-aminoalil]uridin-5'-trifosfato (Sigma-Aldrich, St. Louis, MO). El ARN de referencia humano universal (Stratagene, San Diego, CA) se amplificó del mismo modo. El ARN purificado se conjugó después con colorante Cy5 (muestras experimentales) o Cy3 (muestra de referencia) (Amersham Biosciences).

Análisis de micromatrices

50 Se fabricó una micromatriz de oligonucleótidos (60-meros) de 22.000 genes diseñada por encargo usando tecnología de síntesis *in situ* de chorro de tinta (Agilent Technologies, Palo Alto, CA). Se co-hibridaron ARN de muestra marcado con Cy5 y ARN de referencia marcado con Cy3 a 65°C, 1X tampón de hibridación (Agilent Technologies). Los portaobjetos se lavaron a 37°C con 0,1X SSC/ 0,005% de Triton X-102. El análisis de imágenes se realizó usando el software de análisis de imágenes Agilent. Las relaciones brutas de Cy5/Cy3 se normalizaron usando regresión no lineal dependiente de la intensidad.

55 Una matriz de datos matriz que consistía en relaciones de Cy5/Cy3 normalizadas de todas las muestras se centró en la mediana para cada gen. Se calculó la varianza de cada gen con respecto a todas las muestras y se seleccionó el primer 25% de genes de alta varianza (5.475) para análisis adicionales. Las pruebas de identificación y permutación para la significancia de la expresión génica diferencial se realizaron usando BRB ArrayTools, desarrollado por Dr. Richard Simon y Amy Peng (véase el sitio <http://linus.nci.nih.gov/BRB-Array-Tools.html>). El análisis de agrupaciones jerárquicas se realizó con el software GeneMaths (Applied-Maths, Bélgica) usando correlación de cosenos y enlace completo. Todos los otros procedimientos estadísticos (prueba de la t de dos muestras, análisis de la característica de operación del receptor, regresión logística multivariable y análisis de supervivencia) se realizaron en el entorno estadístico R de fuente abierta (véase el sitio <http://www.r-project.org>). La prueba estadística de significancia de curvas ROC fue mediante el procedimiento de DeLong ("Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach". *Biometrics* 44, 837-45 (1988)). La

supervivencia sin enfermedad se calculó a partir de la fecha del diagnóstico. Los acontecimientos se puntuaron como la primera metástasis distante, y se censuraron los pacientes que siguieron sin enfermedad en el último seguimiento. Las curvas de supervivencia se calcularon por las estimaciones de Kaplan-Meier y se compararon por pruebas del orden logarítmico.

5 Análisis por PCR cuantitativa en tiempo real

La PCR en tiempo real se realizó en muestras de entrenamiento de 59 de los 60 casos (se excluyó un caso debido a materiales insuficientes) y las muestras de validación de 20 casos. Brevemente, 2 µg de ARN amplificado se convirtieron en ADNc bicatenario. Para cada caso se usaron 12 ng de ADNc por triplicado para PCR en tiempo real con un ABI 7900HT (Applied Biosystems) como se ha descrito (Gelmini, S. y col. "Quantitative polymerase chain reaction-based homogeneous assay with fluorogenic probes to measure c-erbB-2 oncogene amplification". Clin Chem 43, 752-8 (1997)). Las secuencias de los pares de cebadores de PCR y la sonda MGB fluorogénica (5' a 3'), respectivamente, que se usaron para cada gen son del siguiente modo:

HoxB13
TTCATCCTGACAGTGGCAATAATC,
CTAGATAGAAAATATGAGGCTAACGATCAT,
VIC- CGATAACCACTAGCTG;
IL17BR
GCATTAACCTAACGATTGGAACTACATT,
GGAAGATGCTTTATTGTTGCATTATC,
VIC-ACAACCTCAAAGCTGTTTAA.

Los niveles de expresión relativa de HOXB13 en muestras normales, de DCIS y de IDC se calcularon del siguiente modo. Primero, todos los valores de CT se ajustan restando la mayor CT (40) de entre todas las muestras, luego la expresión relativa = $1 / 2^{\Delta CT}$.

30 Hibridación in situ

Se prepararon sondas de ARN marcadas con Dig usando el kit de marcado de ADN con DIG (SP6/T7) de Roche Applied Science siguiendo el protocolo proporcionado con el kit. La hibridación *in situ* se realizó en secciones de tejido congeladas como se ha descrito (Long y col.).

35 Tabla 1. Pacientes y características de tumor del conjunto de entrenamiento.

ID de muestra	Tipo de tumor	Tamaño	Grado	Nodos	RE	RP	Edad	SSE	Estado
1389	D	1,7	2	0/1	Pos	Pos	80	94	0
648	D	1,1	2	0/15	Pos	ND	62	160	0
289	D	3	2	0/15	Pos	ND	75	63	1
749	D	1,8	2	2/9	Pos	Pos	61	137	0
420	D/L	2	3	ND	Pos	Pos	72	58	
633	D	2,7	3	0/11	Pos	ND	61	20	
662	D	1	3	6/11	Pos	Pos	79	27	
849	D	2	1	0/26	Pos	Neg	75	23	
356	D	1	2	2/20	Pos	ND	58	24	
1304	D	2	3	0/14	Pos	Pos	57	20	
1419	D	2,5	2	1/8	Pos	Pos	59	86	0
1093	D	1	3	1/14	Pos	Pos	66	85	0
1047	D/L	2,6	2	0/18	Pos	Neg	70	128	0
1037	D/L	1,5	2	0/4	Pos	Pos	85	83	0
319	D	4	2	1/13	Pos	ND	67	44	1
25	D	3,5	2	0/9	Neg	Pos	62	75	1
180	D	1,6	2	2/19	Pos	Pos	69	169	0
687	D	3,5	3	3/16	Pos	ND	73	142	0
856	D	1,6	2	0/16	Pos	Pos	73	88	0
1045	D	2,5	3	1/12	Pos	Neg	73	121	0
1205	D	2,7	2	1/19	Pos	Pos	71	88	0
1437	D	1,7	2	2/22	Pos	Pos	67	89	0
1507	D	3,7	3	0/40	Pos	Pos	70	70	0

469	D	1	1	0/19	Pos	ND	66	161	0
829	D	1,2	2	0/9	Pos	ND	69	136	0
868	D	3	3	0/13	Pos	Pos	65	130	0
1206	D	4,1	3	0/15	Pos	Neg	84	56	
843	D	3,4	2	11/20	Pos	Neg	76	122	
342	D	3	2	9/21	Pos	ND	62	102	
1218	D	4,5	1	3/16	Pos	Pos	62	10	
547	D/L	1,5	2	ND	Pos	ND	74	129	
1125	D	2,6	2	0/18	Pos	Pos	54	123	0
1368	D	2,6	2	ND	Pos	Pos	82	63	0
605	D	2,2	2	6/18	Pos	ND	70	110	0
59	L	3	2	33/38	Pos	ND	70	21	
68	D	3	2	0/17	Pos	ND	53	38	
317	D	1,2	3	1/10	Pos	Pos	71	5	
374	D	1	3	0/15	Pos	Neg	57	47	
823	D	2	2	0/6	Pos	Pos	51	69	
280	D	2,2	3	0/12	Pos	ND	66	44	
651	D	4,7	3	10/13	Pos	ND	48	137	
763	D	1,8	2	0/14	Pos	Pos	63	118	0
1085	D	4,7	2	0/8	Pos	Pos	48	101	
1363	D	2,1	2	0/15	Pos	Pos	56	114	0
295	D	3,5	2	3/21	Pos	Pos	52	118	
871	D	4	3	0/16	Pos	Neg	61	6	
1343	D	2,5	3	ND	Pos	Pos	79	21	
140	L	>2,0	2	18/28	Pos	ND	63	43	
260	D/L	0,9	2	1/13	Pos	ND	73	42	
297	D	0,8	2	1/16	Pos	Pos	66	169	0
1260	D	3,5	2	0/14	Pos	Pos	58	79	0
1405	D	1	3	ND	Pos	Pos	81	95	0
518	L	5,5	2	3/20	Pos	ND	68	156	0
607	D	1,2	2	5/14	Pos	Pos	76	114	0
638	D	2	2	1/24	Pos	Pos	67	148	0
655	D	2	3	ND	Pos	Pos	73	143	0
772	D	2,5	2	0/18	Pos	Pos	68	69	1
878	D/L	1,6	2	0/9	Pos	Neg	76	138	0
1279	D	2	2	0/12	Pos	Pos	68	102	0
1370	D	2	2	ND	Pos	Pos	73	61	0

Abreviaturas: D, ductal; L, lobular; D/L, características de ductal y lobular; pos, positivo; neg, negativo; ND, no determinado; RE, receptor de estrógeno; RP, receptor de progesterona; SSE, supervivencia sin enfermedad (número de meses); estado = 1, recurrido; restado = 0, sin enfermedad.

Ejemplo 2

Identificación de genes diferencialmente expresados

- 5 El perfilado de la expresión génica se realizó usando una micromatriz de oligonucleótidos de 22.000 genes como se ha descrito anteriormente. En el análisis inicial se usó ARN aislado de secciones de tejido de tumor congeladas tomadas de las biopsias primarias archivadas. El conjunto de datos de la expresión resultante se filtró primero basándose en la varianza global de cada gen con los primeros 5.475 genes de alta varianza (centil 75)
- 10 seleccionados para análisis adicional. Usando este conjunto de datos reducido se realizó la prueba de la t en cada gen comparando los pacientes que respondieron y los pacientes que no respondieron a tamoxifeno, conduciendo a la identificación de 19 genes diferencialmente expresados en el corte de valor p de 0,001 (Tabla 2). Se estimó que la probabilidad de seleccionar muchos genes o genes más diferencialmente expresados por casualidad era 0,04 permutando al azar la clase de paciente con respecto al desenlace del tratamiento y repitiendo el procedimiento de
- 15 la prueba de la t 2.000 veces. Por tanto, este análisis demostró la existencia de diferencias estadísticamente significativas en la expresión génica entre los cánceres de mama primarios de pacientes que respondieron y pacientes que no respondieron a tamoxifeno.

Tabla 2. Firma de 19 genes identificados por la prueba de la t en el conjunto de datos de secciones

	Valor p paramétrico	Media en pacientes que responden al tratamiento	Media en pacientes que no responden al tratamiento	Veces de diferencia de las medias	Acc de GB	Descripción
1	1,96E-05	0,759	1,317	0,576	AW006861	SCYA4 citocina A4 inducible pequeña
2	2,43E-05	1,31	0,704	1,861	AI240933	EST
3	8,08E-05	0,768	1,424	0,539	X59770	IL1R2 receptor de interleucina 1, tipo II
4	9,57E-05	0,883	1,425	0,62	AB000520	Proteína adaptadora APS con homología con pleckstrina y dominios de homología 2 con src
5	9,91 E-05	1,704	0,659	2,586	AF208111	IL17BR receptor de interleucina 17B
6	0,0001833	0,831	1,33	0,625	AI820604	EST
7	0,0001935	0,853	1,459	0,585	AI087057	DOK2 proteína de acoplamiento 2, 56 kD
8	0,0001959	1,29	0,641	2,012	AJ272267	CHDH colina deshidrogenasa
9	0,0002218	1,801	0,943	1,91	N30081	EST, débilmente similares a la proteína hipotética 138022 [H. sapiens]
10	0,0004234	1,055	2,443	0,432	AI700363	EST
11	0,0004357	0,451	1,57	0,287	AL117406	ABCC11 casete de unión a ATP, subfamilia C (CFTR/MRP), miembro 11
12	0,0004372	1,12	3,702	0,303	BC007092	Homeocaja de HOXB13 B13
13	0,0005436	0,754	1,613	0,467	M92432	guanilato ciclase 2D GUCY2D, membrana (específica para retina)
14	0,0005859	1,315	0,578	2,275	AL050227	ARNm de Homo sapiens; ADNc DKFZp586M0723 (del clon DKFZp586M0723)
15	0,000635	1,382	0,576	2,399	AW613732	ADNc de Homo sapiens FLJ31137 fis, clon IMR322001049
16	0,0008714	0,794	1,252	0,634	BC007783	SCYA3 citocina A3 inducible pequeña
17	0,0008912	2,572	1,033	2,49	X81896	C11 orf25 marco de lectura abierto 25 del cromosoma 11
18	0,0009108	0,939	1,913	0,491	BC004960	MGC10955 proteína hipotética MGC10955
19	0,0009924	1,145	0,719	1,592	AK027250	ADNc de Homo sapiens: FLJ23597 fis, clon LNG15281

- 5 Para perfeccionar nuestro análisis con respecto a las células tumorales y evitar una posible variabilidad atribuible a una contaminación por células estromales, se reanalizó la misma cohorte después de una microdissección por captura láser (LCM, *Lasser Capture Microdissection*) de las células tumorales dentro de cada sección de tejido. Usando varianza basada en exploración por filtrado génico y ensayo *t* idéntico al utilizado para el conjunto de datos de todas las secciones tisulares, se identificaron 9 secuencias génicas expresadas diferencialmente con un valor de
- 10 $P < 0,001$ (Tabla 3).

Tabla 3. Firma de 9 genes identificados por ensayo t en el conjunto de datos de LCM

	Valor p paramétrico	Media de respondedores	Media de no respondedores	Factor de diferencia de medias	GBacc	Descripción
1	2,67E-05	1,101	4,891	0,225	BC007092	HOXB13homeoboxB13
2	0,0003393	1,045	2,607	0,401	AI700363	MSE (marcador de secuencia expresada)
3	0,0003736	0,64	1,414	0,453	NM_ 014298	QPRT quinolinato fosforribosiltransferasa (nicotinato-nucleótido pirofosforilasa (carboxilación))
4	0,0003777	1,642	0,694	2,366	AF208111	IL17BR receptor de interleucina 17B
5	0,0003895	0,631	1,651	0,382	AF033199	ZNF204 proteína de dedos de cinc 204
6	0,0004524	1,97	0,576	3,42	AI688494	FLJ13189 proteína teórica FLJ 13189
7	0,0005329	1,178	0,694	1,697	AI240933	MSE
8	0,0007403	0,99	1,671	0,592	AL157459	ARNm de Homo sapiens; ADNc DKFZp434B0425 (del clon DKFZp434B0425)
9	0,0007739	0,723	1,228	0,589	BC002480	FLJ 13352 proteína teórica FLJ 13352

- 5 Solo se identificaron 3 genes como diferencialmente expresados tanto en el análisis de LCM como de sección de tejido completo: el gen homeobox HOXB13 (identificado dos veces como AI700363 y BC007092), el receptor 17B de interleucina 17B, IL17BR (AF208111) y el canal de calcio regulado por voltaje CACNA1D (AI240933). HOXB 13 se sobreexpresó diferencialmente en casos no sensibles a tamoxifeno mientras que IL17BR y CACNA1D se sobreexpresaron en casos sensibles a tamoxifeno. Cabe destacar, que la secuencia QPRT TUVO similitudes con la
- 10 secuencia HOXB13 en relación con los niveles de expresión en respondedores y no respondedores. Basándose en su identificación como marcadores derivados de tumores significativamente asociados con resultados clínicos en dos análisis independientes, la utilidad de cada uno de estos genes se evaluó por sí mismo y en combinación con el resto.
- 15 Para definir la sensibilidad y especificidad de la expresión de HOXB13, IL17BR y CACNA1D como marcadores de resultado clínico, se utilizó un análisis Característico Operativo del Receptor (ROC) (Pepe, M.S. "An interpretation for the ROC curve and inference using GLM procedures." Biometrics 56, 352-9 (2000)). Para datos derivados de secciones de tejido completo, los valores del Área Bajo la Curva (ABC) para IL17BR, HOXB 13 y CACNA1D fueron de 0,79, 0,67 y 0,81 para IL17BR, HOXB 13 y CACNA1D, respectivamente (véase la Tabla 4 y la Figura 1, parte superior). El análisis ROC de los datos generados a partir de las células tumorales microdisecionadas produjo
- 20 valores del ABC de 0,76, 0,8 y 0,76 para estos genes (véase la Tabla 4 y la Figura 1, parte inferior).

Tabla 4. Análisis ROC del uso de la expresión de IL17BR, CACNA1D y HOXB13 para predecir respuestas a tamoxifeno

	Secciones de tejido			
	ABC	Valor P	ABC	valor P LCM
IL17BR	0,79	1,58E-06	0,76	2,73E-05
CACNA1D	0,81	3,02E-08	0,76	1,59E-05
HOXB13	0,67	0,012	0,79	9,94E-07
ESR1	0,55	0,277	0,63	0,038
PGR	0,63	0,036	0,63	0,033
ERBB2	0,69	0,004	0,64	0,027
EGFR	0,56	0,200	0,61	0,068

ABC, área bajo la curva; los valores de *P* son $ABC > 0,5$

- 5 Un ensayo de significado estadístico indicó que todos estos valores del ABC son significativamente mayores de 0,5, el valor esperado del modelo nulo que predice al azar resultados clínicos. Por lo tanto, estos tres genes tienen posible utilidad para predecir resultados clínicos de terapia adyuvante con tamoxifeno. Como comparativa, se analizaron marcadores que son normalmente útiles en la evaluación de la probabilidad de respuesta a tamoxifeno en comparación. Los niveles del ER (símbolo del gen ESR1) y del receptor de progesterona (RP, símbolo del gen PGR) se sabe que se correlacionan positivamente con la respuesta a tamoxifeno (véase Fernández, M. D., et al. "Quantitative oestrogen and progesterone receptor values in primary breast cancer and predictability of response to endocrine therapy." Clin Oncol 9, 245-50 (1983); Ferno, M. et al. "Results of two or five years of adjuvant tamoxifen correlated to steroid receptor and S-phase levels." South Sweden Breast Cancer Group, and South-East Sweden Breast Cancer Group. Breast Cancer Res Treat 59, 69-76 (2000); Nardelli, G. B., et al. "Estrogen and progesterone receptors status in the prediction of response of breast cancer to endocrine therapy (preliminary report)." Eur J Gynaecol Oncol 7, 151-8 (1986); and Osborne, C. K., et al. "The value of estrogen and progesterone receptors in the treatment of breast cancer." U 46, 2884-8 (1980)).

Además, se piensa que las rutas de señalización del factor de crecimiento (EGFR, ERBB2), regulan negativamente la señalización dependiente de estrógeno y por tanto contribuye a pérdida de sensibilidad a tamoxifeno (véase Dowsett, M. "Overexpression of HER-2 as a resistance mechanism to hormonal therapy for breast cancer." Endocr Relat Cancer 8, 191-5 (2001)). El análisis ROC de estos genes confirmó su correlación con resultados clínicos, pero con valores de ABC que variaban solamente de 0,55 a 0,69 alcanzando un significado estadístico para PGR y ERBB2 (véase la Tabla 4).

El conjunto de datos de LCM es particularmente relevante, ya que EGFR, ERBB2, ESR1 y PGR se miden normalmente a nivel de células tumorales usando inmunohistoquímica o como hibridación *in situ* con fluorescencia. Como marcadores individuales de resultado clínico, HOXB13, IL17BR y CAC1D sobrepasaron el rendimiento de ER1, PGR, EGFR y ERBB2 (véase la Tabla 4).

Ejemplo 3

Identificación de la proporción de la expresión de HOXB13:IL17BR

- La proporción de la expresión de HOXB13:IL17BR se identificó de la siguiente manera como un fuerte indicador de resultado compuesto. Dado que HOXB13 e IL17BR tienen modelos de expresión opuestos, la proporción de la expresión de HOXB13 sobre IL17BR se examinó para determinar si esto proporciona un mejor indicador compuesto de respuesta a tamoxifeno. De hecho, tanto el ensayo *t* como el análisis ROC demostraron que la proporción de los dos genes tenía una correlación más fuerte con resultados de tratamiento que cualquier gen en solitario, tanto en los conjuntos de datos de secciones de tejido completo como de LCM (véase la Tabla 5). Los valores del ABC de HOXB13:IL17BR alcanzaron 0,81 para el conjunto de datos de las secciones de tejidos y 0,84 para el conjunto de datos de LCM. El emparejamiento de HOXB13 con CACNA1D o el análisis de los tres marcadores juntos no proporcionó un poder predictivo adicional.

Tabla 5. La proporción de HOXB13:IL17BR es un fuerte indicador de resultados de tratamiento

Sección de tejido LCM		Ensayo <i>t</i>		ROC
		t-estadístico	Valor <i>P</i>	Valor <i>P</i> ABC
	IL17BR	4,15	1,15E-04	0,79 1,58E-06
	HOXB13	-3,57	1,03E-03	0,67 0,01
	HOXB13:IL17B R	-4,91	1,48E-05	0,81 1,08E-07
	IL17BR	3,70	5,44E-04	0,76 2,73E-05
	HOXB13	-4,39	8,00E-05	0,79 9,94E-07
	HOXB13:IL17BR	-5,42	2,47E-06	0,84 4,40E-11

ABC, área bajo la curva; los valores de *P* son ABC > 0,5.

La proporción HOXB13/IL17BR se comparó con factores de pronóstico bien establecidos para cáncer de mama, tales como edad del paciente, tamaño del tumor, grado y estado del nódulo linfático (véase Fitzgibbons, P. L. et al. "Prognostic factors in breast cancer. College of American Pathologists Consensus Statement 1999." Arch Pathol Lab Med 124, 966-78 (2000)). El análisis univariado de regresión logística indicó que solo el tamaño del tumor fue marginalmente significativo en esta cohorte ($P=0,04$); esto no fue sorprendente dado que el grupo de respondedores era muy parecido al grupo de no respondedores con respecto al tamaño del tumor, grado del tumor y estado del nódulo linfático durante la selección de los pacientes. Entre los indicadores conocidos positivos (ESR1 y PGR) y negativos (ERBB2 y EGFR) de respuesta a tamoxifeno, el análisis ROC de los datos de las secciones de tejidos indicó que solo PGR y ERBB2 fueron significativos (véase la Tabla 4). Por lo tanto, se realizó una comparación de los modelos de regresión logística que contenía la proporción HOXB13: IL17BR en sí misma o en combinación con el tamaño del tumor, y los niveles de expresión de PGR y ERBB2 (véase la Tabla 6). La proporción HOXB13:IL17BR en solitario fue un indicador muy significativo ($P=0,0003$) y tuvo una razón de probabilidades de 10,2 (IC del 95% de 2,9-35,6). En el modelo multivariado, la proporción HOXB13:IL17BR es la única variable significativa ($P=0,002$) con una razón de probabilidades de 7,3 (IC del 95% de 2,1-26). Por tanto, la proporción de la expresión de HOXB13:IL17BR es un fuerte indicador independiente del resultado de tratamiento en el caso de terapia adyuvante con tamoxifeno.

Tabla 6. Análisis de Regresión Logística

Modelo Univariado			
Indicador	Razón de Probabilidades	IC del 95%	Valor <i>P</i>
HOXB13:IL17BR	10,17	2,9-35,6	0,0003
Modelo Multivariado			
Indicadores	Razón de Probabilidades	IC del 95%	Valor <i>P</i>
Tamaño del tumor	1,5	0,7-3,5	0,3289
PGR	0,8	0,3-1,8	0,5600
ERBB2	1,7	0,8-3,8	0,1620
HOXB13:IL17BR	7,3	2,1-26,3	0,0022

Todos los indicadores son variables continuas. Los valores de expresión de genes eran de mediciones de micromatrices. La razón de probabilidades es la proporción de probabilidad intercuartil, en base a la diferencia de un indicador de su cuartil inferior (0,25) con respecto a su cuartil superior (0,75); IC, intervalo de confianza.

Ejemplo 4

Validación independiente de la proporción de la expresión HOXB13:IL17BR

La reducción de una firma de micromatriz compleja a una proporción de expresión de dos genes permite el uso de estrategias de detección más sencillas, tales como análisis por PCR cuantitativa en tiempo real (RT-QPCR). Se analizó la proporción de la expresión HOXB13:IL17BR mediante RT-QPCR utilizando secciones de tejidos congeladas que se encontraban disponibles en 59 de los 60 casos de entrenamiento (Figura 2, parte a). Los datos de la RT-QPCR eran fuertemente concordantes con los datos de la micromatriz de los especímenes tumorales

congelados (coeficiente de correlación $r=0,83$ para HOXB13, $0,93$ para IL17BR). Además, las proporciones HOXB13:IL17BR derivadas de la PCR, representadas como las ΔCT , en el que CT son los ciclos de amplificación de la PCR para conseguir una cantidad umbral predeterminada (por ejemplo, Figura 2, partes a y b) y ΔCT es la diferencia de CT entre HOXB13 e IL17BR, se correlacionaron altamente con los datos derivados de micromatrices ($r=0,83$) y con resultados de tratamiento (ensayo t $P=0,0001$, Figura 2, parte c). Por tanto, parece que el análisis de RT-QPCR convencional para la proporción de la expresión de HOXB13 con respecto a IL17BR es equivalente al análisis basado en micromatrices de especímenes tumorales congelados.

Para validar la utilidad predictiva de la proporción de la expresión HOXB13:IL17BR en una cohorte de pacientes independiente, se identificaron 20 tumores de mama primarios de fase temprana ER-positivos adicionales, de mujeres tratadas con tamoxifeno adyuvante solo a MGH entre 1991 y 2000, y para los que se disponía de registros médicos y tejidos incluidos en parafina. De los 20 casos del archivo, 10 habían sido recurrentes con un tiempo medio de recurrencia de 5 años y 10 permanecieron libres de enfermedad con un seguimiento medio de hasta 9 años (véase la Tabla 7 para más detalles).

Tabla 7. Características de los pacientes y de los tumores del conjunto de validación.

Muestra	Tipo de Tumor	Tamaño	Grado	Nódulos	ER	PR	Edad	DFS	Estado
Ensayo 1	D	1,9	3	0/10	Pos	Pos	69	15	1
Ensayo 2	D	1,7	3	0/19	Pos	Pos	61	117	1
Ensayo 3	D	1,7	2	0/26	Pos	ND	65	18	1
Ensayo 4	D	1,2	2	0/19	Pos	Pos	63	69	1
Ensayo 5	D	1,7	2	2/2	Pos	Pos	60	52	1
Ensayo 6	D	1,1	1	0/10	Pos	Pos	54	59	1
Ensayo 7	D	>1,6	2	0/17	Pos	Neg	66	32	1
Ensayo 8	L	2,6	1-2	0/14	Pos	Pos	58	67	1
Ensayo 9	D	1,2	2	ND	Pos	Pos	93	58	1
Ensayo 10	D	4	3	0/20	Pos	Pos	66	27	1
Ensayo 11	D	1,1	2	0/19	Pos	Pos	64	97	0
Ensayo 12	D	2,7	2	0/10	Pos	Pos	66	120	0
Ensayo 13	D	0,9	1	0/22	Pos	Pos	66	123	0
Ensayo 14	D	2,1	2	0/16	Pos	Pos	57	83	0
Ensayo 15	D	0,8	1-2	0/8	Pos	Pos	74	80	0
Ensayo 16	D	1	2	0/13	Pos	Pos	74	93	0
Ensayo 17	D	1,6	2	0/29	Pos	Pos	66	121	0
Ensayo 18	L	1,5	1-2	0/8	Pos	Pos	65	25	0
Ensayo 19	D	1,5	3	0/19	Pos	Pos	60	108	0
Ensayo 20*	L	4	1-2	0/19	Pos	Pos	60	108	0

Abreviaturas: Las mismas que las de la Tabla 1 complementaria. * El paciente recibió tamoxifeno durante 2 años

El ARN se extrajo de secciones de tejido completo incluidas en parafina fijadas con formalina (FFPE, *Formalin-Fixed Paraffin-Embedded*), amplificadas linealmente y se usaron como molde para análisis RT-QPCR. Coherente con los resultados de la cohorte de entrenamiento, la proporción de la expresión HOXB13:IL17BR en esta cohorte de pacientes independiente estuvo muy correlacionada con el resultado clínico (ensayo t $P=0,035$) con una mayor expresión de HOXB13 (ΔCT inferior) correlacionada con un mal resultado (Figura 2, parte d). Para ensayar la

precisión predictiva de la proporción HOXB13:IL17BR, se usaron los datos de la RT-QPCR de las secciones tisulares congeladas (n=59) para construir un modelo de regresión logística. En este conjunto de entrenamiento, el modelo predijo un resultado de tratamiento con una precisión global del 76% (P=0,000065, intervalo de confianza del 95% del 63%-86%). Los valores predictivos positivos y negativos fueron 78% y 75%, respectivamente. Aplicando este modelo a los 20 pacientes independientes en la cohorte de validación, el resultado del tratamiento para 15 de los 20 pacientes se predijo correctamente (precisión global 75%, P=0,04, intervalo de confianza del 95% del 51%-91%), con valores predictivos positivos y negativos de 78% y 73%, respectivamente.

El análisis de Kaplan-Meier de los grupos de pacientes como se predice mediante el modelo dio como resultado curvas de supervivencia libre de enfermedad significativamente diferentes tanto en el conjunto de entrenamiento con en el conjunto de ensayo independiente (Figura 2, partes e y f).

Un ejemplo adicional representativo de la aplicación de la proporción a las muestras de sección y muestras LCM de la cohorte de 60 pacientes se muestra en la Figura 8, Parte B (indicado por "Secciones" y "LCM", respectivamente). En esta figura también se muestra una aplicación ejemplar de la proporción de 31 muestras FFPE (indicado por "FFPE").

Ejemplo 5

Identificación de secuencias adicionales como resultado de la expresión de CHDH

El hecho de que la secuencia de AI240933 fuese complementaria con la cadena codificante de CACNA1D conduce a la cuestión de si en el Ejemplo 2 anterior se detectó la expresión de una secuencia distinta de la de CACNA1D. Por lo tanto se realizó un ensamblaje de secuencias que condujo a la identificación de una región 3' más larga de las secuencias conocidas expresadas como parte de CHDH (véanse las Figuras 5 y 3). Esta secuencia más larga se confirmó que se expresaba por análisis PCR usando sondas que eran capaces de amplificar una región calculada de 4283 nucleótidos que abarcaba las partes de la secuencia de CHDH previamente identificada y las de la secuencia de AI240933 (véase la Figura 4).

El ensamblaje de las secuencias de la Figura 5 condujo a la identificación de un cóntigo de secuencias más largo mostrado en la Figura 6. Se proporcionó un refuerzo adicional para la probabilidad de un cóntigo más largo mediante un alineamiento del cóntigo con las secuencias CHDH de ratón.

En la Figura 7 se muestra la relación de probabilidad entre las secuencias de CHDH y CACNA1D.

Ejemplo 6

Identificación y uso de la proporción de expresión de QPRT:CHDH

La proporción de la expresión de AQPRT:CHDH se identificó como un identificador fuerte y compuesto de resultado de una manera similar a la descrita en el Ejemplo 3 anterior. Dado que QPRT y CHDH tienen patrones de expresión opuestos, la proporción de la expresión de QPRT sobre CHDH se examinó para determinar su capacidad de funcionar como un indicador compuesto de respuesta a tamoxifeno. Los resultados de la aplicación de la proporción con respecto a las muestras de sección y muestras de LCM de la cohorte de 60 pacientes se muestran en la Figura 8, Parte A (indicado como "Secciones" y "LCM" respectivamente). En la figura también se muestra una aplicación ejemplar de la proporción con respecto a 31 muestras FFPE (indicadas como "FFPE").

Referencias adicionales

- Ma, X. J. et al. Gene expression profiles of human breast cancer progression. Proc Natl Acad Sci U S A 100, 5974-9 (2003).
- Nicholson, R. I. et al. Epidermal growth factor receptor expression in breast cancer: association with response to endocrine therapy. Breast Cancer Res Treat 29, 117-25 (1994).

Anexo

Secuencias identificadas como las del grupo de IL17BR

AW675096

CCGGCGATGTCGCTCGTGCTGCTAAGCCTGGCCGCGCTGTGCAGGAGCGCCGTACCCCGA
GAGCCGACCGTTCAATGTGGCTCTGAAACTGGGCCATCTCCAGAGTGGATGCTACAACAT
GATCTAATCCCGGGAGACTTGAGGGACCTCCGAGTAGAACCTGTTACAAC TAGTGTTGCA
ACAGGGGACTATTCAATTTTGATGAATGTAAGCTGGGTACTCCGGGCAGATGCCAGCATC
CGCTTGTTGAAGGCCACCAAGATTTGTGTGACGGGCAAAAGCAACTTCCAGTCCTACAGC
TGTGTGAGGTGCAATTACACAGAGGCCTTCCAGACTCAGACCAGACCCTCTGGTGGTAAA
TGGACATTTTCCTACATCGGCTTCCCTGTAGAGCTGAACACAGTCTATTTTCATTGGGGCC
CATAATATTCCTAATGCAAATATGAATGAAGATGGCCCTTCCATGTCTGTGAATNTCACC
TCACCAGGCTGCCTAGACCACATAATGAAATATAAAAAAAGTGTGTCAAGGCCGGAAGC
CTGTGGGATCCGAACATCACT

AW673932

TTTTTTTTTTTTTTTTTTAAAAAGTGGGTTTACGCTTGTATTATCCCTACTTTTGTATCT
 TAAAAACAATGATTTTTTGCATGTAATAGAAGGTTTTTCACCTAAGATGCTATTGAGTGA
 ATCAGTGAGGGGTTCTTAGAGTTAGTATTCATTAATTAAACATAGAATATTAGCTAAACA
 GTTCGGGTACACTGCAATGCATGGTCTATGGAAGACTAGATGTTTGGCTGAAGATGCTT
 TATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAAACTGTAATTGATTTCTATGT
 ATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGTTAATGCTACATTAGTT
 AGCAATATTTGAAAATTTTATTGGTATAAAATGTTTAAATTACTAAGGCTGTTTGTAGGC
 TGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATTGG

BC000980

ggcgcggcga tgcgcgcgt getgctaagc ctggcgcgc tgtgcaggag cgccgtaccc
 cgagagccga ccgttcaatg tggtctgaa actgggccat ctccagagt gatgctacaa
 catgatctaa tcccgggaga cttgagggac ctccagtag aacctgttac aactagtgtt
 gcaacagggg actattcaat tttgatgaat gtaagctggg tactccgggc agatgccagc
 atccgcttgt tgaaggccac caagatttgt gtgacgggca aaagcaactt ccagtcctac
 agctgtgtga ggtgcaatta cacagaggcc ttccagactc agaccagacc ctctggtggt
 aaatggacat tttcctacat cggcttcct gttagagctga acacagtcta tttcattggg
 gcccataata ttcctaattgc aaatatgaat gaagatggcc cttccatgtc tgtgaatttc
 acctcaaccag gctgcctaga ccacataatg aaatataaaa aaaagtgtgt caaggccgga
 agcctgtggg atccgaacat cactgcttgt aagaagaatg aggagacagt agaagtgaac
 ttcacaacca ctcccctggg aaacagatac atggctotta tccaacacag cactatcatc
 ggggtttctc aggtgtttga gccacaccag aagaaacaaa cgcgagcttc agtggtgatt
 ccagtgactg gggatagtga aggtgctacg gtgcagctga ctccatattt tctacttgt
 ggcagcgact gcatccgaca taaaggaaca gttgtgctct gccacaaac aggcgtccct
 ttccctctgg ataacaacaa aagcaagccg ggaggtggc tgctctcct cctgctgtct
 ctgctggtgg ccacatgggt gctggtggca gggatctatc taatgtggag gcacgaaagg
 atcaagaaga cttccttttc taccaccaca ctactgccc ccattaagg tcttgtggtt
 taccatctg aaatatgttt ccatcacaca atttgttact tcaactgaatt tcttcaaac
 cattgcagaa gtgaggtcat ccttgaaaag tggcagaaaa agaaaatagc agagatgggt
 ccagtgcagt ggcttgccac tcaaaagaag gcagcagaca aagtcgtctt cttctttcc
 aatgacgtca acagtgtgtg cgatggtacc tgtggcaaga gcgagggcag tcccagtga

aactctcaag acctcttccc ccttgccttt aaccttttct gcagtgatct aagaagccag
 attcatctgc acaaatcgt ggtggtctac tttagagaga ttgatacaaa agacgattac
 aatgctctca gtgtctgcc caagtaacc ctcatgaagg atgccactgc tttctgtgca
 gaacttctcc atgtcaagca gcaggtgtca gcaggaaaaa gatcacaagc ctgccacgat
 ggtgctgct cctgttagcc caccatgag aagcaagaga ccttaaaggc ttcctatccc
 accaattaca gggaaaaaac gtgtgatgat cctgaagctt actatgcagc ctacaaacag
 ccttagtaat taaaacattt tataccaata aaattttcaa atattgctaa ctaatgtagc
 attaaactaac gattggaac tacatttaca acttcaaagc tgttttatac atagaaatca
 attacagttt taattgaaa ctataaccat tttgataatg caacaataaa gcactttcag
 ccaaacatct agtcttccat agaccatgca ttgcagtgt cccagaactg ttagctaat
 attctatggt taattaatga atactaactc taagaacccc tcactgattc actcaatagc
 atcttaagtg aaaaaccttc tattacatgc aaaaaatcat tgtttttaag ataacaaaag
 tagggaataa acaagctgaa cccactttta aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
 aa

BI602183

AGCGGAGCTGCGGGTGGCCTGGATCCCGCGCAGTGGCCCGGCGATGTCGCTCGTGCTGCT
AAGCCTGGCCACGCTGTGCAGGAGCGCCGTACCCCGAGAGCCGACCGTTCAATGTGGCTC
TGAAACTGTGGACATTTTCTATATCGGCTTCCCTGTAGAGCTGAAAACAGTCTATTTCA
TTGGGGCCCATATATTTCTAATGCAAATATGAATGAAGATGGCCCTTCCATGTCTGTGA
ATTTACCTCACCAGGCTGCCTAGACCACATAATGAAATATAAAAAAAGTGTGTCAAGGC
CGGAAGCCTGTGGGATCCGAACATCACTGCTTGTAAAGAAGATGAGGAGACAGTAGAAGT
GAACTTCACAACCACTCCCTGGGAAACAGATACATGGCTCATCCAACACAGCACTATCA
TCGGGTTTTCTCAGGTGTTTGAGCCACACCAGAAGAAACAAACGCGAGCTTCAGTGGTGA
TTCCAGTGACTGGGGATAGTGAAGGTGCTACGGTGCAGCTGACTCCATATTTTCTACTT
GTGGCAGCGACTGCATCCGACATAAAGGAACAGTTGTGCTCTGCCCACAAACAGGCGTCC
CTTTCCCTCTGGATAACAACAAAAGCAAGCCGGGAGGCTGGCTGCCTCTCCTCCTGCTG
TCTCTGCTGGTTGGCCACATTGGGTGCTGGTGGCAGGGATCTATCTAATGTGGAGGCACG
AAAGGATCCAGAAGACTTCCTTTTCTACCACAACTACTGCCCCCATTAAGGTCTGTGG
TTACCCATCTTGAAATATGTTCCCTCACACAATTTGTTACTTCACTGAATTTCTTCAAACC
TG

BI458542

AGCGGAGCGTGCGGGTGGCCTGGATCCCGCGCAGTGGCCCGGCGATGTCGCTCGTGCTGC
TAAGCCTGGCCACGCTGTGCAGGAGCGCCGTACCCCGAGAGCCGACCGTTCAATGTGGCT
CTGAAACTGTGGACATTTTCTATATCGGCTTCCCTGTAGAGCTGAAAACAGTCTATTTT
ATTGGGGCCCATATATTTCTAATGCAAATATGAATGAAGATGGCCCTTCCATGTCTGTG
AATTTACCTCACCAGGCTGCCTAGACCACATAATGAAATATAAAAAAAGTGTGTCAAG
GCCGGAAGCCTGTGGGATCCGAACATCACTGCTTGTAAAGAAGATGAGGAGACAGTAGAA
GTGAACTTCACAACCACTCCCTGGGAAACAGATACATGGCTCATCCAACACAGCACTAT
CATCGGGTTTTCTCAGGTGTTTGAGCCACACCAGAAGAAACAAACGCGAGCTTCAGTGGT
GATTCAGTGACTGGGGATAGTGAAGGTGCTACGGTGCAGCTGACTCCATATTTTCTACT
TTGTGGCAGCGACTGCATCCGACATAAAGGAACAGTTGTGCTCTGCCCACAAACAGGCGT
CCCTTTCCCTCTGGATAACAACAAAAGCAAGCCGGGAGGCTGGCTGCCTCTCCTCCTGCT
GTCTCTGCTGGTGGNCACATTGGGTGCTGGTGGCAGGGATCTATCTAATGTGGAGGCACG
AAAGGATCAGAAGACTTCCTTTTCTACCACACATACTGCCCCCATTAAGGTCTTGTG
GTTTACCC

BI823321

GGCGATGTCGCTCGTGCTGCTAAGCCTGGCCGCGCTGTGCAGGAGCGCCGTACCCCGAGA

GCCGACCGTTCAATGTGGCTCTGAACTGGGCCATCTCCAGAGTGGATGCTACAACATGA
TCTAATCCCGGAGACTTGAGGGACCTCCGAGTAGAACCTGTTACAACAGTGTGGCAAC
AGGGGACTATTCAATTTTGATGAATGTAAGCTGGGTACTCCGGGCAGATGCCAGCATCCG
CTTGTGAAGGCCACCAAGATTTGTGTGACGGGCAAAAGCAACTTCCAGTCCTACAGCTG
TGTGAGGTGCAATTACACAGAGGCCTTCCAGACTCAGACCAGACCCTCTGGTGGTAAATG
GACATTTTCTATATATCGGCTTCCCTGTAGAGCTGAACACAGTCTATTTTATTGGGGCCCA
TAATATTTCTAATGCAAATATGAATGAAGATGGCCCTTCCATGTCTGTGAATTTACCTC
ACCAGGAAGCCTGTGGGATCCGAACATCACTGCTTGTAAAGAAGATGAGGAGACAGTAG
AAGTGAACCTTCACAACCACTCCCTGGGAAACAGATACATGGCTCTTATCCAACACAGCA
CTATCATCGGGTTTTCTCAGGTGTTTGAGCCACACCAGAAGAAACAAACGCGAGCTTCAGT
GGTGATTCCAGTGACTGGGGATAGTGAAGGTGCTACGGTGCAGCTGACTCCATATTTTCC
TACTTGTGGCAGCGACTGCAATCCGACATAAAGGAACAGTTGTGCTCTGCCCACAAACAG
GCGTCCCTTTCCCTCTTGGATAGCAACAGAAGCAAGCCGGGAGGCTGGTGCCTCTTCTTC
TGGTGTCTCTGCTGGTGGCACATTGAGTGCTGGTGGCAGGATCCATCTAATGTGGAGGCC
CCAAAGGACCAGGAAGACTTCCTTTATTAGCACCAAGTATTGCC

AA514396

TGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAAACTGT
AATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGTT
AATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTA
AGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATT
GGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCA
GCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTACTTGACATGGAGAAG
TTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGC
ATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGATG
AATCTGGC

BF110326

TTTGTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAA
AACTGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCG
TTAGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAA
TTACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCT
GTAATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTTTCATGGGTGGGCTACAAGGA
GCAGCAGCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATG
GAGAAGTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACT
GAGAGCATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTG
CAGATGAATCTGGCTTCTTAGATCACTGC

BE466508

TGGCATGAGATGCTATATTGTTGCATTATCAAAATGGGTTTAGTCTTCAATTAACACTGT
AATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTGGTTTCCAATCGTCAGTT
AATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTA
AGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATT
GGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCA
GCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGGAGAAG
TTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGC
ATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGATG
AATCTGGCTTCTTAGATCACTG

BF740045

GTTTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAAAC
TGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTT
AGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATT
ACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGT
AATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGC
AGCAGCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGGA
GAAGTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGA
GAGCATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTA

AW299271

TGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAAACTGT
 AATTGATTTCTATGTATAAACAGCGTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGTT
 AATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTA
 AGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATT
 GGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCA
 GCCATCGTGGCAGGCTTGTGATCTTTTCCCTGCTGACACCTGCTGCTTGACATGGAGAAG
 TTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGC
 ATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGATG
 AATCTGGCTTCTTAGATCACTGCAGAAAAG

AA836217

TTTTTTTTTTTACAACCTTCAAAGCTGTTTTATACATAGAAATCAATTACAGTTTTAATTGA
 AAACATAACCATTTTGATAATGCAACAATAAAGCATCTTCAGCCAAACATCTAGTCTTC
 CATAGACCATGCATTGCAGTGTACCCAGAACTGTTTAGCTAATATCTATGTTTAATTAA
 TGAATACTAACTCTAAGAACCCCTCACTGATTCACCTCAATAGCATCTTAAGTGA AAAAC
 TTCTATTACATGCAAAAAATCATTGTTTTTAAGATAACAAAAGTAGGGAATAAACAAGCT
 GAACCCACTTTTACTGGACCAAATGATCTATTATATGTGTACCACTTGTATGATTTGGTA
 TTTGCATAAGACCTTCCCTCTACAACTAGATTTCATATCTTGATTCTTGTACAGGTGCCT
 TTTAACATGAACAACAAAATACCCACAACTTGTCTACTTTTGCC

AI203628

TAGTAATTA AACATTTTATACCAATAAAATTTTCAAATATTGCTAACTAATGTAGCATT
 AACTAACGATTGGAACTACATTTACAACCTCAAAGCTGTTTTATACATAGAAATCAATT
 ACAGTTTTAATTGAAAACTATAACCATTTTGATAATGCAACAATAAAGCATCTTCAGCCA
 AACATCTAGTCTTCCATAGACCATGCATTGCAGTGTACCCAGAACTGTTTAGCTAATATT
 CTATGTTTTAATTAATGAATACTAACTCTAAGAACCCCTCACTGATTCACCTCAATAGCATC
 TTAAGTGA AAAACCTTCTATTACATGCAAAAAATCATTGTTTTTAAGATAACAAAAGTAG
 GGAATAAACAAGCTGAACCCACTTTTACTGGACCAAATGATCTATTATATGTGTAAACCAC
 TTGTATGATTTGGTATTTGCATAAGACCTTCCCTCTACAACTAGATTTCATATCTTGATT
 CTTGTACAGGTGCCTTTTAACATGAA

AI627783

TTTTTTTTTTTTTTTTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACT
 AAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAAT
 TGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGC
 AGCCATCGTGGCAGGCTTGTGATCTTTTCCCTGCTGACACCTGCTACTTGACATGGAGAA
 GTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAG

CATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGAT
 GAATCTGGCTTCTTAGATCACTGCAGAAAAGGTTAAAGGCAAGGGGGAAGAGGTCTTGAG
 AGTTCTC

AI744263

TTAAAGTGGGTTTCAGCTTGTTTATTCCCTACTTTTGTATCTTAAAAACAATGATTTTTT
GCATGTAATAGAAGGTTTTCTACTTAAGATGCTATTGAGTGAATCAGTGAGGGGTTCTTA
GAGTTAGTATTCAATTAATAACATAGAATATTAGCTAAACAGTTCTGGGTACACTGCAA
TGCATGGTCTATGGAAGACTAGATGTTTGGCTGAAGATGCTTTATTGTTGCATTATCAAA
ATGGTTACAGTTTTCAATTAAAGCTGTAATTGATTTCTATGTATAAAACAGCTTTGAAGT
TGTAATGTAGTTTCCAATCGTTAGTTAATGCTACATTAGTTAGCAATATTTGAAAATTT
TATTGGTATAAAATGTTTAAATTACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGG
ATCATCACACGTTNTTTCCTGTAATTGGTGGGATAGGAAGCCTTA

AI401622

AGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAAGGCTGTTTGT
AGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATTGGTGGGATAGG
AAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAGCCATCGTGGC
AGGCTTGATCTTTTTCTGCTGACACCTGCTACTTGACATGGAGAAAGTTCTGCACAGA
AAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTGTAATCGT
CTTTTGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGATGAATCTGGCTTC
TTAGATCACTGCAGAAAAGGTTAAAGGCAAGGGGGAAGAGGTCTTGAGAGTTCTCACTGG

AI826949

TTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAACTG
TAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGT
TAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACT
AAGGCTGTTTGTAGGCTTGATAGAAGCTTCAGGATCATCACACGTTTTTCCCTGTAAT
TGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGC
AGCCATCGTGGCAGGCTTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGGAGAA
GTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAG
CATTGTAATCGTCT

BE047352

TTTTTTTTTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAAGGCT
GTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATTGGTGG
GATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAGCCAT
CGTGGCAGGCTTGATCTTTTTCTGCTGACACCTGCTACTTGACATGGAGAAAGTTCTG
CACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTGT
AATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGATGAATCT
GGCTTCTTAGATCACTGCAGAAAAGGTTAAAGGCAAGGGGGAAGAGGTCTTGAGAG

AI911549

TTTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTACAGTTTTCAATTAAAGCT
GTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAG
TTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTAC
TAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAA

TTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAG
CAGCCATCGTGGCAGGCTTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGGAGA
AGTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACA

BF194822

TTCTCTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTACAGTTTTCAATTAAA
GCTGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGT
TAGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTAAAT
TACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTG
TAATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAG
CAGCAGCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGG
AGAAGTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGG

AI034244

TTTTTTTTTTTTTTTTTACAACCTTGAAAGCTGTTTTATACATAGAAATCAATTACAGTTT
TAATTGAAAACCTATAACCATTTTGATAATGCAACAATAAAGCATCTTCAGCCAAACATCT
AGTCTTCCATAGACCATGCATTGCAGTGTACCCAGAAGCTTTAGCTAATATTCTATGTT
TAATTAATGAATACTAACTCTAAGAACCCCTCACTGATTCACTCAATAGCATCTTAAGTG
AAAAACCTTCTATTACATGCAAAAAATCATTGTTTTTAAGATAACAAAAGTAGGGAATAA
ACAAGCTGAACCCACTTTTACTGGACCAAATGATCTATTATATGTGTAACCACTTGTATG
ATTTGGATTGTCATAAGACCTTCCCTCTACAAACTAGATTCATATCTTGATTCT

AI033911

TTTTTTTTTTTTTTTTTACAACCTGCAAGCTGTTTTATACATAGAAATCAATTACAGTTT
AATTGAAAACCTATAACCATTTTGATAATGCAACAATAAAGCATCTTCAGCCAAACATCTA
GTCTTCCATAGACCATGCATTGCAGTGTACCCAGAAGCTTTAGCTAATATTCTATGTT
AATTAATGAATACTAACTCTAAGAACCCCTCACTGATTCACTCAATAGCATCTTAAGTG
AAAACCTTCTATTACATGCAAAAAATCATTGTTTTTAAGATAACAAAAGTAGGGAATAA
CAAGCTGAACCCACTTTTACTGGACCAAATGATCTATTATATGTGTAACCACTTGTATGA
TTTGGTATTGTCATAAGACCTTCCCTCTACAAACTAGATTCATATCTTGATTCT

BF064177

TTTTTTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAAGGCT
GTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATTGGTGG
GATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAGCCAT
CGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTACTTGACATGGAGAAGTTCTG
CACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTGT
AATCGTCTTTGTATCAATCTCTCTAAAGTAGACCACCGTATTTGTGCAGATGAATCT
GGCTTCTTAGATCACTGCAGAAAAGGTTAAAGGCAAGGGGAAGAGGTCTTGAGAGTTCT
CACTGGGACTGCCCTCGCTCTTGCCACAGGTACCATCGCACACACTGTTGACGTCATTGG
AAAG

AA847767

GGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAACTGTA
ATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGTTA
ATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAA
GGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATTG

GTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAG
CCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGGAGAAGT
TCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTA

AI538624

TTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTACAGTTTTCAATTAAAGCTG
TAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGT
TAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACT
AAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAAT
TGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGC
AGCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGGAGAA
GTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTAC

AI913613

TTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAACTG
TAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGT
TAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACT
AAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTNTTCCCTGTAAT
TGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGC
AGCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGGAGAA
GTTCTGCACAGAAAGCAGTGGCATCCTTCATG

AI942234

GTTTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAAC
TGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTA
GTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTA
CTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTA
ATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCA
GCAGCCATCGTGGCAGGCTTGGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGAAG
AAGTTCTGCACAGAAAGCAGTGGCAT

AI580483

GTTTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAAC
TGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTA
GTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTA
CTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTA
ATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCA
GCAGCCATCGTGGCAGGCTTGGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATTGGAA
AGTTCTGCACAGAAAGCAGTGGCATC

AI831909

TTTTGGCTGATGATGCTTTATTGTTGCATTATCAAAATGGTTACAGTTTTCAATTAAAGC
TGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTA
GTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTA
CTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTA
ATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCA
GCAGCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGGAG

AAGTTCTGCACAGAAAGCAGTGGCAT

AI672344

GGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTACAGTTTTCAATTAAAGCTGTA
ATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGTTA
ATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAA
GGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATTG
GTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAG
CCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGGAGAAGT
TCTGCACAGAAAG

AW025192

GATTGGCTGTTTTATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAAA
CTGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTT
AGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATT
ACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGT
TATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGC
AGCAGCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGGA
GAAGTTCTGCACAAAAGCAGTGGCATCCTTCATGAGGTGGTA

AA677205

GCAATATTTTAAAAATTTTATTGGTATAAAATGTTTTAATTACTAAGGCTGTTTGTAGGCT
GCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATTGGTGGCATAGGAAGCC
TTTAAGGTCTCTTGCTTCTCATGGTGTGGGCTACAAGGAGCAGCAGCCATCGTGGCAGGC
TTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGGAGAAGTTCTGCACAGAAAGC
AGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTGTAATCGTCTTT
TGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGATGAATCTGGCTTCTTAG
ATCACTGCAGAAAAGGTTAAAGGCAAGGGGAAGAGGTCTTGAGAGTTCTCACTGGGACT
GCCCTGCTCTTGCCACAGGTACCATCGCACACACTG

AA721647

TTTTTTTTTTTACAACTTCAAAGCTGTTTTATACATAGAAATCAATTACAGTTTTAATTGA
AAACTATAACCATTTTGATAATGCAACAATAAAGCATCTTCAGCCAAACATCTAGTCTTC
CATAGACCATGCATTGCAGTGTACCCAGAAGTCTTAGCTAATAATTCTATGTTTAATTAA
TGAATACTAAGTCTAAGAAGCCCTCACTGATTCATCAATAGCATCTTAAGTGAAAAACC
TTCTATTACATGCAAAAAATCATTGTTTTTAAGATAACAAAAGTAGGGAATAAACAAGCT
GAACCCACTTTTACTGGACCAATGATCTATTATATGTGTAACCACTTGTATGATTTGGT
ATTTG

BF115018

GTTTCGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAAAC
TGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTA
GTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATT
CTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTAAGGCCCATCACACGTTTTTCCCTGTA
ATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTNTCATGGGTGGGCTACAAGGAGCA
GCAGCCATCGTGGCAGGCTTGNATCTTTTTCTGCTGGCCCTGCTGCTTGACAT

W61238

NAAAGCACTGGCTGAAGGAAGCCAAGAGGATCACTGCTGCTCCTTTTTCTAGAGGAAAT
GTTTGTCTACGTGGTAAGATATGACCTAGCCCTTTAGGTAAGCGAACTGGTATGTTAGT
AACGTGTACAAAGTTTAGGTTTACAGACCCCGGGAGTCTTGGGCACGTGGGTCTCGGGTCAC
TGGTTTTGACTTTAGGGCTTTGTTACAGATGTGTGACCAAGGGGAAAATGTGCATGACAA
CACTAGAGGTATGGGCGACACGANAACGAACGGGAAGTTTGGCTGAAGTAGGAGTCTTG
GTGAGATTTTGTCTGTATGCATGGTGTGAACCTTCTGAGCCTCTTGTTTTCTCAAGCT
GACTCCATATTTCTACTTGTGGCAGCGACTGCATCCGACATAAAGGAACAG

W61239

TAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAAGGCTGTTTGTAGG
CTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATTGGTGGGATAGGAAG
CCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAGCCATCGTGGCAGG
CTTGATGATCTTTTTCTGCTGACACCTGCTGCTTGACATGGAGAAGTTCTGCACAGAAAG
CAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTGTAATCGTCTT
TTGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGATGAATCTGGCTTCTTA
GATCACTGCAGAAAAGGTTAAAGGCAAGGGGGGA

AI032064

AGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAAGGCTGTTTGTAGGC
TGATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATTGGTGGCATAGGAAGC
CTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAGCCATCGTGGCAGGC
TTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGGAGAAGTTCTGCACAGAAAGC
AGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTGTAATCGTCTT
TGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGATGAATCTGGCTTCTTAG
ATCACTGCAGAAAAGGTTAAAGGCAAGGGGGAAGAGGTCTTGAGAGTTCTCACTGGGACT
GCCCTCGCTCTTGCCAC

AW236941

TTTTTTTTTTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAAGGC
TGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATTGGTG
GGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAGCCA
TCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGGAGAAGTTCT
GCACAAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTG
TAATCGTCTTTTGTATCAATC

AW236941

TTTTTTTTTTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAAGGC
TGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATTGGTG
GGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAGCCA
TCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGGAGAAGTTCT
GCACAAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTG
TAATCGTCTTTTGTATCAATC

BG057174

TTTATACATAGAAATCAATTACAGCTTTAATTGAAAACATAACCATTTTGATAATGCA
ACAATAAAGCATCTTCAGCCAAACATCTAGTCTCCATAGACCATGCATTGCAGTGTACC

CAGAACTGTTTAGCTAATATTCTATGTTTAATTAATGAATACTAACTCTAAGAACCCCTC
ACTGATTCACTCAATAGCATCTTAAGTGAAAAACCTTCTATTACATGCAAAAAATCATTG
TTTTTAAGATAACAAAAGTAGGGAATAACAAGCTGAACCCACTTTTACTGGACCAATG
ATCTATTATATGTG

AW058532

GGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAAACCTGTA
ATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGTTA
ATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAA
GGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCCTGTATGG
GTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCT

T98360

TNAGGAANGAGAAGAAGCGAGATNNANNTNNAGAAATANGTGGTGGCNTANTTTAGAGAG
ATTGATNCAAAAGCNGATTNCAATNNNCTCAGTGNCTNCCCAAGTNCNCCTCATGAAGG
ATNCACTNCTTTCTGTGCAGACTNNNCATGTCAAGCAGCAGGTGTCAGCAGGAAAAAGAN
CACAAGCTCCNCGATGGCTGCTGCTCCTTGTAGCCCNCCATGAGAAGCAAGAGNCTTAAA
GGCTTCCTATCCCACCAATTACAGGGAAAAACGTGTGATGACCTGAGCTTACTATGCAGC
CTACAANCAGCCTTAGTAATTAAACNTTTTATT

T98361

NANNATGAAGATGCTTTATTGTTGCATTATCAAAATGGTTACAGTTTTCAATTAAAGCTG
TAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGT
TAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGNATAAAATGTTTTAATTACT
AAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTNCCCTGTAAT
TGGGTGGGGATAGGGAAGCCCTTTAAGGGTCTCTTGCTTCTCATGGGGTGGGGCCTACNA
AGGGAGCAGCCAGCCCATCGTGGCCAGGGCCTTGTGGANCCTTTTTCCCTGCCTGGACAC
CCTGCCTGCCCTTGGACCATGGGGAGGAAGGTTCTGGCACCAGGAAAGCCAGGTGGCCCAT
CCCTTCCATGAGGGTGGGGTACTTNGGGGGGCCAGGACCACTGAGGNGCCATTGGTAATC
CGTCTTTTNGTATCCAATCCCCTCCTAAGGTAGGNCCCCC

AI470845

TTTTGTGGGTTTCACTTGTTTATTCCTACTTTTGTATCTTAAAAACAATGATTTTTTG
CATGTAATAGAAGGTTTTTCACTTAAGATGCTATTGAGTGAATCAGTGAGGGGTCTTAG
AGTTAGTATTCAATTAATAACATAGAATATTAGCTAAACAGTTCTGGGTACACTGCAAT
GCATGGTCTATGGAAGACTAGATGTTTGGCTGAAGATGCTTTTATTGTTGCATTATCAAN
ATGGTTTATAGTTTTCAATTAAAACCTGTAATTGATT

AI497731

GGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAAACCTGTA
ATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGTTA
ATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAA
GGCTGTTTGTAGGCTGCATAGTAAGCTTAANGATCATACNCACGTTTTTCCCTGAATTTG
GTGGGATAANGAAGCCTTTAAAGGT

T96629

TTGAAAATTTTATTGGNATAAAATGTTTTAATTACTAAGGCTGTTTGTAGGCTGCATAGT
 AAGCTTCAGGANCATCACACGTTTTTCCCTGTAATTGGTGGCATAGGAAGCCTTTAAGG
 TCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAGCCATCGTGGCAGGCTTGTGATC
 TTTTTCCCTGCTGACACCTGCTGCTTGACATGGAGAAGTTCTGCACAGAAAGCAGTGGCAT
 CCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTGTAATCGTCTTTTGTATCAA
 TCTCTCTAAAGTAGACCACCACCGTNTTTGTGCAGATGGANTCTGGCTTC

T96740

AGGCACATATCATCGGGTTTTCTCAGGTGTTTGAGCCACACCAGAAGAAACAAACGCGAGC
 TTCAGTGGTGATTCCAGTGACTGGGGATAGTGAAGGTGCTACGGTGCAGCTGACTCCATA
 TTTTCCCTACTTGTGGCAGCGACTGCATCCGACATAAAGGAACAGTTGTGCTCTGCCCCA
 AACAGGCGTCCCTTTCCCTCTGGATAACAACAAAGCAAGCCGGGANGNCTGNCCTCTC
 CTCCTGCTGTCTCTGCTGGTGGCCACATGGGTGCTGGTGGCAGGGATCTATCTAATGTGG
 AGGCACGAAAGGATCAAGAAGACTTCCTTTTCTAACCACCACATTACTGCCCCCATTTA
 AGGTTCTTGTGGTTTTACCCATCTGGAAATATGTTTTCCCTTCACACATTTGTTTATTTT
 ATTGATTTNTTTCAAAACCTTGGCAGGAGTTT

H25975

GGGTCCAGTGCAGTGGCTTGCNTGCAGAAAGAAGGCAGCAGACAAAGTCGTCTTCCTTCT
 TTCCAATGACGTCAACAGTGTGTGCGATGGTACCTGTGGCAAGAGCGAGGGCAGTCCCAG
 TGAGAACTCTCAAGACCTCTTCCCCCTTGCCCTTTAACCTTTTCTGCAGTGATCTAAGAAG
 CCAGATTCTCTGCACAAATACGTGGTGGTCTACTTTAGAGAGATTGATACAAAAGACGA
 TTACAATGCTCTCAGTGTCTGCCCCAAGTACCACCTCATGAAGGATGCCACTGCTTTCTG
 TGCAGAACTTCTCCATGTCAAGCAGCAGGTGTCAGCAGGAAAAAGATTCAAGCCTGCC
 ACGATGGCTGCTTGCTTCCTTTGTAGCCACCCATGAGGAAGNCAAGAGACCTTNAAAGG
 GTTCCTTTTCCCATCANTTTACAGGGGANAAAACGTGTGATGATC

H25941

TTTTGTTTGGCTNATNTNNTTCTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATT
 AAACTGTAATTGATTNCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAAT
 CGTTAGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAANGTTTT
 AATTACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCC
 CTGTAATTGGTGGGATAGGAAGCCTTTAAGGTCTCTNGCTTCTCATGGGTGGGCTACAAG
 GAGCAGCAGCCATCGTGGCAGGCTTGTGANCTTTTNCCTGCTGACACCTGCTGCTTGACA
 TGGGAGAAGTTCTGCACAGAAAGGCAGTGGGCATCCTTCATGAGGTGGGTACTTGGGGN
 CAGACACTGAGGAGCATTGT

BE539514

ACTCAAAGAAGGCAGCAGACAAAGTCGTCTTCCTTCTTTCCAATGACGTCAACAGTGTG
 TGCGATGGTACCTGTGGCAAGAGCGAGGGCAGTCCCAGTGAGAACTCTCAAGACCTCTTC
 CCCCTTGCCTTTAACCTTTTCTGCAGTGATCTAAGAAGCCAGATTCTCTGCACAAATAC
 GTGGTGGTCTACTTTAGAGAGATTGATACAAAAGACGATTACAGTGCTCTCAGTGTCTGC
 CCCAAGTACCACCTCATGAAGGATGCCACTGCTTTCTGTCAGAACTTCTCCATGTCAAG
 CAGCAGGTGTGAGCAGGAAAAAGATCAAGCCTGCCACGATGGCCGCTGCTCCTTGTAG
 CCCACCCATGAGAAGCAAGAGACCTTAAAGGCTTCCTATCCCACCAATTACAGGGAAAAA
 ACGTGTGATGATCCTGAAGCTTACTATGCAGCCTACAAACAGCCTTAGTAATTAACAT
 TTTATACCAATAAAATTTTCAAATATGCTAACTAATGTAGCATTAACCTAACGATTGGAAA
 CTACATTTACAACCTTCAAAGCTGTTTTATACATAGAAATCAATTACAGCTTTAATTGAAA
 ACTGTAACCATTTTGATAATGCAACAATAAGCATCTTCAG

BX282554

GTCCAGTGCAGTGGCTTGCCACTCAAAGAAGGCAGCAGACAAAGTCGTCTTCCTTCTTT
CCAATGACGTCAACAGTGTGTGCGATGGTACCTGTGGCAAGAGCGAGGGCAGTCCCAGTG
AGAACTCTCAAGACCTCTTCCCCCTTGCCCTTTAACCTTTTCTGCAGTGATCTAAGAAGCC
AGATTTCATCTGCACAAATACGTGGTGGTCTACTTTAGAGAGATTGATACAAAAGACGATT
ACAGTGCTCTCAGTGTCTGCCCCAAGTACCACCTCATGAAGGATGCCACTGCTTTCTGTG
CAGAACTTCTCCATGTCAAGCAGCAGGTGTGAGCAGGAAAAAGATCACAAGCCTGCCACG
ATGGCCGCTGCTCCTTGTAGCCCAACCATGAGAAAGCAAGAGACCTTAAAGGCTTCCTATC
CCACCAATTACAGGGGAAAAAACGTGTGATGATCCTGAAGCTTACTAT

R74038

TATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAACTGTAATTGATTTCTATGT
ATAAAACAGCTTTGAAGTTGTAAATGTAGTTTTCAATCGTTAGTTAATGCTACATTAGTT
AGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAAGGCTGTTGTAGGC
TGCAATAGTAAGCTTCAGGATCATCACACGTTTTTNCCTGTAAATTGGGTGGGGATAGGGA
AGCCTTTAAGGTCTCTTGCTTCTCATGGGGTGGGGCTACAAGGGAGGCAGGCAGCCATCG
TGGGCAGGGCTTGTGATCTTTTCCCTGCTGACACCTGCTGCTTGACATGGGGGGAAGGT
TCTGGCACAGAAAGCAGTGGGCATCCTTCATGAGGGTGGTACTTGGGGGGCAGACACTGA
GGAGGCNTTGTAAATCGNCTTTTTNGTATCCAANCTCTNCTAAAGTAGGGNCCACCNCGT
TTTTNTTGCAGGTGGATNCGGGGCTN

R74129

GGGTCCAGTGCAGTGGCTTGCNTNCAAAGAAGGCAGCAGACAAAGTCGTCTTCCTTCTT
TCCAATGACGTCAACAGTGTGTGCGATGGTACCTGTGGCAAGAGCGAGGGCAGTCCCAGT
GAGAACTCTCAAGACCTCTTCCCCCTTGCCCTTTAACCTTTTCTGCAGTGATCTAAGAAGC
CAGATTTCATCTGCACAAATACGTGGTGGTCTACTTTAGAGAGATTGATACAAAAGACGAT
TACAATGCTCTCAGTGTCTGCCCCAAGTACCACCTCATGAAGGATGCCACTGCTTTCTGT
GCAGAACTTCTCCATGTCAAGCAGCAGGTGTGAGCAGGAAAAAGATCACAAGCCTGCCAC
GATNGCTGCTGCTCCTTGTAGNCCACCCATGAGAAAGCAAGTGACCTTTAAAGGNTTTCCT
ATTNCCACCNATTTACAGGG

BG433769

GACTAGATGTTTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCA
ATTAAAACTGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCC
AATCGTTAGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGT
TTTAATTACTAAGGCTGTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTT
TCCCTGTAATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTAC
AAGGAGCAGCAGCCATCGTGGCAGGCTTGTGATCTTTTCCCTGCTGACACCTGCTGCTTG
ACATGGAGAAGTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAG
ACACTGAGAGCATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACCACGTAT
TTGTGCAGATGAATCTGGCTTCTTAGATCACTGCAGAAAAGGTTAAAGGCAAGGGGGAAG
AGGTCTTGAGAGTTCTCACTGGGACTGCCCTCGCTCTTGCCACAGGTACCATCGCACACA
CTGTTGACGTCATTGGAAAAGAAGGAAGAC

BG530489

GAGTTCTCACTGGGACTGCCCTCGCTCTTGCCACAGGTACCATCGCACACACTGTTGACG

TCATTGGAAAGAAGGAAGACGACCTTGTCTGCTACCTTCTTTTGAGTGGCAAGCCACTGC
ACTGGACCCATCTCTGCTATTTTCTTTTCTGCCACTTTTCAAGGATGACCTCACTTCTG
CAATGGTTTTGAAGAAATTCAGTGAAGTAACAAATTGTGTGATGGAAACATATTTCAAGAT
GGGTAAACCACAAGAACCTTAATGGGGGGCAGTAGTGTGGTGGTAGAAAAGGAAGTCTTC
TTGATCCTTTCTGTGAGAGGAGAAAAGCATTGTATCTGTGAATAGCAAACAGCAGGCT
TTCACCTCTGTAAACCATCCCTGACAAATGATCCCTTGCTAGAGAATGTCAGCTGAGCACC
AAGGGCCTTGTTAGTGACAGCAAGGAAAAACATCCTGATGTTCTTTTGAACACATCACC
TGAAACACACTGATGCTTAAACCTTAACCTTTTTTTTTTTGGGGGACATAGTCTCACTCTG
TCGCCCAGGCTGGAGTGCGTGGGAGAGGACCTCGGAAAGACTGGCAAGCATCCGCATACA
AGGGAGTAACAGCACAACTACTCCGTGAACCTTCGGAGCCCTCCAAAGGAATACTCAAGGGC
GGGTAAAGGATGGCAAGGGTCGACGGAGAGCCCACGAGGAGAGCGGAAGGTAGAGAGGAG
ACAAGCATAAGACGCGAGAGGAACCTCAAGGCGGGGCCAAAGAGAGAAACCACGGTCACC
AACAGAAG

AA007528

AGAAGCCAGATTTCATCTGCACAAATACGTGGTGNTCTACTTTAGAGAGATTGATACAAAA
GACGATTACAATGCTCTCAGTGTCTGCCCCAAGTACCACCTCATGAAGGATGCCACTGCT
TTCTGTGCAGAACTTCTCCATGTCAAGCAGCAGGTGTGAGCAGGAAAAAGATCACAAGCC
TGCCACGATGGCTGCTGCTCCTTGTAGCCACCCATGAGAAGCAAGAGACCTTAAAGGCT
TCCTATCCACCAATTACAGGGNAAAAACNGTAGTGATNATCCCTGACAGCTTACTATGC
CAGCCNT

AA007529

TTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATCGGTTACAGTTTTCAATTAAAGCT
GTAATTNGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTTCAATCGTTA
GTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTA
CTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACAGTTTTTTCCCTGTA
ATTGGGTGGGATAGGAAGCCTTAAAGGTCTCTTGCTTCTCATTGGGTGGGCTACAAGGAG
CAGCAGCCATCCGTNGGCAAGGCTTTGTGGATNCT

BI260259

GGAAGAGAAAGATCGTCCAGAGGTTCCATCGCACACACTGTATGACGTCATTGGAAATGA
AGGAAGACGACTTTGTCTGCTGGCTTCTTGTGAGTGGCAAGCCACTGCAGTGGACCCATC
TCTGCTATTTTCTTTATTCTGCCACTTTTCAAGGATGACCTCACTTCTGCAATGGTTTTG
AAGAAAGTTTCAGTGAAGTAACAAATTGTGTGATGGAAACATATTTCAAGATGGGTAAACCA
CAAGAACCTTAATGGGGGGCAGTAGTGTGGTGGTAGAAAAGGAAGTCTTCTTGATCCTTT
CTGTGAGAGGAGAAAAGCATTAGTTATCTGTGAACAGCAAACAGCAGGCATTTACATCT
GTAAACCATCCCTGACAAATGATCCCTTGCTAGAGAATGTCAGCTGAGCACCAAGGGGCC
TTGTTAGTGACAGCAAGGACAAAACATCCTGATGTTCTTTTGAACACATCAGCTGAAAC
ACACTGATGCTCTAAACCGTTAACTATTTATTAATGGGGGAACATAGGTCTCAACTCATG
TACGACCAGGCTGGAGTGCAAGTGGGGTTGAACATCGACAGACATAGCAAACCACCGATCA
CTAGGGAAACAACGCACAGAACTCCAGACTTAAACACC

AA287951

ATTCCGCACCTGGGGGGCAGACACTGAGAGCATTGTAATCGTCTTTTGTATCAATCTCTC
TAAAGTAGACCACCACGTATTTGTGCAGATGAATCTGGCTTCTTAGATCACTGCAGAAAA
GGTTAAAGGCAAGGGGGAAGAGGTCTTGAGAGTTCTCACTGGGACTGCCCTCGCTCTTGC
CACAGGTACCATCGCACACACTGTTGACGTCAATTGGAAAGAAGGAAGACGACTTTGTCTG
CTGCCCTTCTTTTGAAGTGGCAAGCCACTGCAGTGGACCCATCTCTGCTATTTTCTTTTCT

GCCACTTTTCAAGGATGACCTCACTTCTGCAATGGTTTTGAAGAAATTCAGTGAAGTAAC
AAATNTGTGTGATGGAAACATATTTTCAGATGGGTAAACCACAAGAACCCTTAATGGGGGGC
AGTAGTGTGGTGGTAGAAAAGGAAGTCTTCTTGATCCTTTCTGTGAGAGGAGAAAAGC

AA287911

TTTTGATGGTCCACTTCCATTAAATGAATTAGTAAATATCTTTTCTCATGATTTTAATTA
CATTTTTTTCTCTAGCTTACTTTATTATAATACAGCACATAATACACCTAACATGCAAAA
TATGTGTTAATTGGCTGTTTATGTTATTGGTAAGACTTCCAGTCAACAGTAGGCTATTAG
AAGTTAAGTTGTGGGAAAATCAAAGGTTATAGGAGATTTCAACTGCATGCAGGGCCGGT
GCCCTCCCCACTGTGTTGTTCAAGGGTCAGCTGTACTCTCTAAGGGCTTTGCTAACTTCA
AAACATGGAGTATTTGAATACAGAAACCAGAGCATTACATACTCAGCTCAAGGCAGAGC
TATTAATAAACTCCTCTTCTCCATATGTAGGAAAGGAAATACAAATGCATCCTTTGAGT
CATTTGTGATGT

T97852

AACAGTTGTGCTCTGCCCACAAACAGGCGTCCCTTTCCCTCTGGATAACAACAAAAGCAA
GCCGGGANGNCTGNCGCTCTCCTCCTGCTGTCTCTGCTGGTGGCCACATGGGTGCTGGTN
GCAGGGATCTATCTAATGTNGAGGCACGAAAGGGATCAAGAGGACTTCCTTTTCTACCAC
CACACTACTGCCCCCATTAAGGTTCTTGTTNGGTTTACCCATCTGGAAATATGTTTCCAT
CACACAATTTGTTACTTCACTGGAATTTCTTCAAAACCATTGGCAGGANGTGAGGGTCAT
CCTTGGAAGAGTGGGC

T97745

CCTCACTTCTGCAATGGTTTTGAAGAAATTCAGTGAAGTAACAAATTGTGTGATGGAAAC
ATATTTTCAGATGGGTAAACCACAAGAACCCTTAATGGGGGGCAGTAGTGTGGTGGTAGAAA
AGGAAGTCTTCTTGATCCTTTTCGTGCCTCCACATTAGATAGATCCCTGCCACCAGCACCC
ATGTGGCCACCAGCAGAGACAGCAGGAGGAGAGGCAGCCAGCCTCCCGGCTTTGCTTTTG
TTGTTATCCAGAGGGGAAAGGGGACGCCTGTTTNTGGGGCAGAGCACAACGTTCCTC
GTGCCCGAATTCTTTGGGCCTTCGAGGGGCCAAATTTCCCTATTAGGTGAGGTGATTTT
TAAATTTCCGTAATTATGGTCATAGGCTTGTTTTTCCCCG

N40294

GTTTCAACACAATTTTGGATCAGCTGCCTGTTTGCAAAAACATAATATATTTCTGTAA
CAGTTCTTCACCTAACAGCATATTGCTCTTATAACTGGTAGAGCTGTTTCAAAGGAAGTT
GGTTTCTGGTCCAAGTTTGGACCTAAACCATGTCCATCTTCTATTACCAGCACTTACAAG
CACTGTGAAAACCTGATCATGACAAATAAGTAAAATTTGCTACATTAAACATATTGCCTCA
GCCATTACTAAGCGTCCACTTGTAAGCTGGACACAGTTTTTACTTTATGCTTCATTTTG
ATTTTTTATCCGTAAGACATAAATTAGAAGGCATGAGGTGGCCCTTTAAGGATAATCTGC
AAATATACACATTTTAAATAGTCATCCATCTGGAAATCGNTCCACCATTCCAGGGGAAGG
ATTCCAGGTATTGGTGCTGTGGTGGAAATAAAGCATTCCCCNGGGAAAAAACCATTTTA
TGNCTAAATAATTACCACCATTAACTCNTGGGGTT

AA809841

GAATACTAACTCTAAGAACCCCTCACTGATTCACTCAATAGCATCTTAAGTGAAAAACCT
TCTATTACATGCAAAAAATCATTTGTTTTTAAGATAACAAAGTAGGGAATAAACAAGCTG
AACCACCTTTTACTGGACCAATGATCTATTATATGTGTAACCACTTGTATGATTTGGGA
TTTGCAT

AA832389

TTTTTTTACAACCTTCAAAGCTGTTTTATACATAGAAATCAATTACAGTTTTAATTGAAAAC
TATAACCATTTTGATAATGCAACAATAAAGCATCTTCAGCCAAACATCTAGTCTTCCATA
GACCATGCATTGCAGTGTACCCAGAAGCTGTTTAGCT

H14692

CTGAGTGTGATGGTGTAAGCCTGTGGTCCCAGCTACTAGGGAGGCTGAGATGGGATTACA
GGTGTGAGCCACGGCGCCTGGCCTAAAAGCATCTTTTTCTTTAACGCAGAGGTTATGTTG
TATTATTAGCATAAATGTTTTTTCTGGGAATGCTTATTTACACAGCACAATACTGAAT
CTTCTCTGGAATGTGGATCGATTTAGATGGATGACTATTAAAATGTGTATATTTGCAGA
TTATCCTTAAAGGGCCACCTCATGCCCTTAATTTATGTCTTACGGATAAAAAATCAAAA
TGAAGCATAAAGTAAAACTGTGTCCAGCTTTACAAGTGGACGCTTAGTAATGGCTGAGG
CAATATGTTTTAATGTAGCCAAATTTTACTTATTTGTCCATGATCCAGTTTTTACAGTGC
TTGTTAAGTGCTGGTAATTAGGAAGGTGGGACATGGGTAGGTCAAACTTGGGACCNGA
AACCAACTTGN

AA732635

TTTTTTTTTTTACAACCTTCAAAGCTGTTTTATACATAGAAATCAATTACAGTTTTAATTGA
AAACTATAACCATTTTGATAATGCAACAATAAAGCATCTTCAGCCAAACATCTAGTCTTC
CATAGACCATGCATTGCATTGTACCCAGAAGCTTTAGCTAATATTCTATGTTTTAATTAA
TGAATACTAAGTCTAAGAACCCCTCACTGATTCACTCAATAGCATCTTAAGTGAAAAACC
TTCTATTACATGCAAAAAATCATGGTTTT

AA928257

TTTTCTGAGTAAGAACAGGCTTTATTTGTAAAACCACTCGTGACTCTTTACAAAGCAGGA
TACACAGAAGGGAAAAAATACACAGTGCAAAATGGATGTTCTGAGTGCCACAAGGATCT
GCTGAAAAAAGCCAAAGATGTAAGATGGCTGGGTATATATGAGAATGAATATTTCACTAT
ATTCTGATTCAATTACCAGTCTCAGTGGCCAGGATGAGCTTTTGGTGTGGTCACATGGC
CAACATTTGGATAACAAATGAGGAATAATGGTACCGCCTCACTAGTGCCTGAGAACAGCA
TGTTCTGGAATGTCTCTGGAGTTAGAGATGTGTAGCTTTTTTATTACAGATGGAGAA
ATACAATGTTTACACAACAGTCCAGGGGTGGGGTCAAAAGTTGGAAGGTGTCATTAGACG
CAGCCAAATAAAGTGAAGACAACCCAGGTGACTGGCAGCCCTGACTTGTGCGTGGGCG

AI184427

TTTCTGAGTAAGAACAGGCTTTATTTGTAAAACCACTCGTGACTCTTTACAAAGCAGGAT
ACACAGAAGGGAAAAAATACACAGTGCAAAATGGATGTTCTGAGTGCCACAAGGATCTG
CTGAAAAAAGCCAAAGATGTAAGATGGCTGGGTATATATGAGAATGAATATTTCACTATA
TTCTGATTCAATTACCAGTCTCAGTGGCCAGGATGAGCTTTGGTGGTGGTCACATGGCC
AACATTTGGATAACAAATGAGGA

AI298577

GAGATGGAGGTCTCGCTTTGTGACGTAGCCTGGTCTTGAGCGATCCTTTTGCCTTGGCCT
TGCCAAAGTGCTGGGATTGGAGGCATGAGCCACTGCACCCACCCCTGTTTTTTTTTTAAG
TAAACCATTATAATAACTCATTTATAAAAAGGTTACTTCAAGAGGGCTTTCAACTTAAGA
ATTATTTTTCATTTTGAACATGAAAAGTTAAATAGTAACCTAAGAACTGAGAAGCTTGACA

GTGACCTCTAATAGGTAACCTTTAGGCAAAAGTAGACAAGTTTGTGGGTATTTTGTGTTC
ATGTTAAAAGGCACCTGTACAAGAATCAAGATATGAATCTAGTTTGTAGAGGGAAGGTCT
TATGCAAATACCAAATCATACAAGTGGT

AI692717

AGAGATGTTGGTCTCGCTTTGTGACGTAGCCTGGGCTTGAGCGATCCTTTTGCCTTGGCC
TTGCCAAAGTGCTGGGATTGGAGGCATGAGCCACTGCACCCACCCCTGTTTTTTTTTTAA
GTAAACCATTATAATAACTCATTTATAAAAAGGTTACTTCAAGAGGGCTTTCAACTTAAG
AATTATTTTCATTTTGAACATGAAAAGTTAAATAGTAACTAAGAACTGAGAACTCTGAC
AGTGACCTCTAATAGGTAACCTTTAGGCAAAAGTAGACAAGTTTGTGGGTATTTTGTGTTC
CATGTTAAAAGGCACCTGTACAAGAATCAAGATATGAATCTAGTTTGTAGAGGGAAGGTC
TTATGCAAATACCAAATCATACAAGTGGTTACACATATAATAGATCATTTGGTCCAGTAA
AAGTGGGTTTCAGCTTGTATTATCCCTACTT

AA910922

GAGATGGAGGTCTCGCTTTGTGACGTAGCCTGGTCTTGAGCGATCCTTTTGCCTTGGCTT
GCAAAGTGCTGGGATTGGAGGCATGAGCACTGCACCCACCCCTGTTTTTTTTTTAAGTA
AACCATTATAATAACTCATTTATAAAAAGGTTACTTCAAGAG

H90761

TTCACTCAATAGCATCTTAAGTGAAAAACCTTCTATTACATGCAAAAAATCATTGTTTTT
AAGATAACAAAAGTAGGGAATAACAAGCTGAACCCACTTTTACTGGACCAAATGANCTA
TTATATGTATAACCACTTGTATGATTGGTATTTGCATAAGACCTTCCCTCTACAACTA
GATTCATATCTTGATTCTTGACAGGTGCCCTTTTAATATTCTGTGATGAAATCGTTCAC
AGTCAGAGTACATGTCTGCTGCATATGGGAAATAGGGACTGTTGTTCTGAGGGACAAGGC
ACTCAATTGAGCCGTAAAGGCTGACCCGGGCTACTTTTTTCCANGGGAATACAATTTTT
TTACCTTGGGAATAAAATNGGGCCCGACNGGAC

AI620122

TTTTTTTTTTTGTAGTAAGAACAGGCTTTATTTGTAAAACCACTCGTGACTCTTTACAAAG
CAGGATACACAGAAGGGAAAAAATACACAGTGCAAAATGGATGTTCTGAGTGCCACAAG
GATCTGCTGAAAAAAGCCAAAGATGTAAGATGGCTGGGTATATATGAGAATGAATATTT
CACTATATTCTGATTCAATTACCAGTCTCAGTGGCCCAGGATGAGCTTTTGGTGTGGTCA
CATGGCCAACATTTGGATAACAAATGAGGAATAATGGTACCGCCTCACTAGTGCCTGAGA
ACAGCATGTTCTGGAAAATGTCTCTGGAGTTAGAGATGTGTTAGCTTTTTTCATTACAGAT
GGAGAAATACAATGTTTACACAACAGTCCAGGGGTGGGGTCAAAAGTTGGAAGGTGTCAT
TAGACGCA

AI793318

AAATTTTAACTTTTAATAGTTAAAATAGTTAACTATTGGTATGGTAGGAAATGATAAAG
TAGACTAGTATCTGTATACATTTCTGCATTTATGACATACCTTTTTCTTCATTTTTTC
AATATTTTAAATTGAAAAGTTCATCCGAGTTTCATCTAAGTTTTTCAAAGTGATACAAAT
CTCCAAAAATTTTCCAATATATGTATTGAAAAATCCAGGTGTAAGTGGCTCTGCGCAG
TCCAAACCTGTGTTGTTCAAGGGTCACTGTGTATGAATCCAAGCGAAAGCTTTTCTTAA
CACCTCATAAGAACTATTTTTTAAAAAACAGGAACTAGCATAGAGTAACCATCACAGGTA
AAGTGTAATTTGTTATCAGCCATCTTTTGCCATTTCAGTACTGGTAGAAGGCTCAATGG
TAAAAATAAA

AA962325

TTTTTTTTTTTTTTTTTTTTTTTTCTGACTGTCCCGTTTTTATTTTTACCATTGAGCCTT
CTACCAGTACTGAAATGGGCAAAGATGGCTGATAACAAATTACACTTTACCTGTGATGG
TTACTCTATGCTAGTTCCTGTTTTTTAAAAAATAGTTCTTATGAGGTGTTAAGAAAAGCT
TTCGCTTGGATTATACACAGTTGACCCCTTGAACAACACAGGTTTGGACTGCGCAGACCA
CTTACACCTGGATTTTTTCAATACATATATTGGAAAAATTTTTTGGGGATTTGTATCACTT
TGAAAAAAGCTTAGATGAAACTCGGATGGACTTTTCCATTAAAAATATTGAAAAAATGAAG
AAAAAGGT

AI733290

TTTTTTTTTTTTTTTTTTTTTTTTCTGACTGGCCCGTTTTTATTTTTACCATTGAGCCTT
CTACCAGTACTGAAATGGGCAAAGATGGCTGATAACAAATTACACTTTACCTGGGATGG
TTACTCTATGCTAGTTCCTGTTTTTTAAAAAATAGTTCTTATGAGGGGTAAAAAAGCT
TTCGCTTGGATTATACACAGTTGACCCCTTGAACAACACAGGTTTGGACTGCGCAGAGCC
ACTTACACCTGGATTTTTTCAATACATATATTGGAAAAATTTTTTGGAGATTTGTATCACT
TTGAAAAAAGCTTAGATGAAACTCGGATGAACTTTTCAATTAAAAATATTGAAAAAATGAA
GAAAAAGGTATGTATAAATGCAGAAATGTATACAGATACTAGTCTACTTTATCATTTTC
CTACCATACCAATAG

BQ226353

TAAAGGAACAGTTGTGCTCTGCCCCACAAACAGGCGTCCCTTTCCCTCTGGATAACAGTAA
GTGCCCAGTAACTTCAACCAGATGATCAAAGTGGCTCACACACAGTCACTGCCCCCACT
CAGTATGTGGAAGGGTTGTGTGTATGTGGGCAGTGCAAGGGGTCGCTGCCTGTGTACACT
GAACTGGGGTGCAGAGAAAGCCAAACAGTGCTGTCCAGAGAACCTAGAATCTGAGTAAGA
ACAGGCTTTATTTGTAAAACCACTCGTGACTCTTTACAAAGCAGGATACACAGAAGGGAA
AAAAATACACAGTGCAAAATGGATGTTCTGAGTGCCACAAGGATCTGCTGAAAAAAGCCA
AAGATGTAAGATGGCTGGGTATATATGAGAATGAATATTTCACTATATTCTGATTCAATT
ACCAGTCTCAGTGGCCCAAGATGAGCTTTTGGTGTGGTCACATGGCCCAACATTTGGATAA
CAAATGAGGAATAATGGTACCGCCTCACTAGTGCCTGAGAACAGCATGTTCTGAAAAATG
TCTCTGGAGTTAGAGATGTGTTAGCTTTTTTATTACAGATGGAGAAATACAATGTTTACA
CAACAGTCCAGGGGTGGGGGTCAAAGTTGGAAGGTGTCATTAGACGCAGCCAAATAAAG
TGAAGACCACCCAGGTGACTGGCAGCCCTGACTTGTGCGTGGGCGAAACCTTACAGATTC
CTGGGGCACTCTGTGCCTGAACTTACCTGGATGGTCTTTGTGAGGCGGGTGGGCACTTAT
CCTCCATNAATGGTCAGTCTAACAAGACCGGCCTGTAAAAATGGCATCTAATAGGGGCTA
TGGAATGGAACAGTTGGTACCCAGAAATAACTTTAATT

W04890

GACAGTCTGGGAGCCCAGAGCTCTGGGAGGAGTNGGGAAAAATGCTGCTTCTGCTGCTTG
CTTCTAGGCACCTGCTTCCGCCATCTCACTTACCATGGCTAGAGATGGGGGTGAGACTGG
GGAAGGACAAAAGCAGGGAACAGATAAGGGATGGAAATCAGAAGGGAATATAGAAAGAAC
TCTGGATATGCNAGAAATGCCGCTACCTGAGCATTTTGTATCAATGGGAGTACCCTCTGT
AACTGCTCAGTAGGTTACAATGAAGAGTCCACCAGTATTAGAAACAATTTAAACTTGCC
AGTACCAACTGGGATGTGTGCCTCAATTTGAAAAATTGTATGTTTTATTTTTTAAATTTG
GTTAACAGCATTAAATTTATAGAGTATTTGATGTCATTTATGGTTCCCGAGGTGTTTCCAA
CACAATTTTTGGGATCA

BM455231

CTTTTAATAGTTAAATAGTTAACTATTGGTATGGTAGGAAATGATAAAGTAGACTAGTA
TCTGTATACATTTTCTGCATTTATGACATACCTTTTCTTCATTTTTTTCAATATTTTAA
TTGAAAAGTTTCATCCGAGTTTCATCTAAGTTTTTCAAAGTGATACAAATCTCCAAAAA
TTTTCCAATATATGTATTGAAAAATCCAGGTGTAAGTGGCTCTGCGCAGTCCAAACCTG
TGTTGTTCAAGGGTCAACTGTGTATGAATCCAAGCGAAAGCTTTTCTTAACACCTCATAA
GAACTATTTTTTAAAAAACAGGAAC TAGCATAGAGTAACCATCACAGGTAAAGTGTAATT
TGTTATCAGCCATCTTTTGCCCATTTTCAGTACTGGTAGAAGGCTCAATGGTAAAAATAAA
AACGGGACAGTCAGAAGATCTGGAAGTCCTGACCCTGCTTTCACCTGGCATGTGTAATCC
AGTCATGCTCGTATCAGTCTCTGTAGGAGCACTTGAAGGTATTACATAAATGCTATCTAA
CTCTGGGAAACGCCAACATGTGATTGCCTCCAGAGGAATCTTCTTTAAAAAAAATTCAA
AATGTTATTTCTTACTAGGATGTCTTTAAAGAATTATAACCTTACCGTGCCTCCACAT
TAGATAGATCCCTGCCACCAGCACCCTATGTGGCCACCAGCAGAGACAGCAGGAGGAGAGG
CAGCCAGCCTCCCGGCTTGCTTTTGTCTGGAAAAAACAAAGCTTATTACCTTTGGAAA
AAAATCCACACTTATCTCTAATTTAAAACTAAGACTTGGTATACTTTATAGAGGGTTA
TTTATTTTTTTATTATTTTTTAGTTTTGAGACAGAGTCTCGCTTTGTTGCCTANGCTGGAG
TGCAGTGGCGCAATCTCGTTCACTGCAGCCTCCGTTCTCCGGGGTTCAAGGCATGCTGG
CTCAGCCTCCTGTATAGCTGGGGATTAAAGGCATGTGTTACGCGGCCAGCCCCCTTTTG
TAAAAGATTTAGATCCCTTTTAAACCATCAGTCAGGAGGCTCCTTTAAAAAGTCTGGCC
ATCTAATCTTTTTTCCCCAAAAGGGG

BI492426

TTTTTTTTTTCTTTTTCTGAGTAAGAACAGGCTTTATTTGTAAAACCACTCGTGACTC
TTTACAAAGCAGGATACACAGAAGGGAAAAAATACACAGTGCAAAATGGATGTTCTGAG
TGCCACAAGGATCTGCTGAAAAAGCCAAAGATGTAAGATGGCTGGGTATATATGAGAAT
GAATATTTCACTATATTCTGATTCATTACCAGTCTCAGTGGCCCAGGATGAGCTTTTGG
TGTGGTCACATGGCCAACATTTGGATAACAAATGAGGAATAATCTCGTGC

BG674622

AAATTTATAGAGTATTGATGTCATTTATGTTTCTGAGGTGTTTCAACACAATTTTGGATCA
GCTGCCTGTTTGCAAAAACATAATATATTTCTGTAAACAGTTCTTCACCTAACAGCATA
TTGCTCTTATAACTGGTAGAGCTGTTTCAAAGGAAGTTGGTTTCTGGTCCAAGTTTTGAC
CTAAACCATGTCCATCTTCTATTACCAGCACTTACAAGCACTGTGAAAACCTGATCATGAC
AATAAGTAAATTTGCTACATTAAACATATTGCCTCAGCCATTACTAAGCGTCCACTTG
TAAAGCTGGACACAGTTTTTACTTTATGCTTCATTTTGATTTTTTATCCGTAAGACATAA
ATTAGAAGGCATGAGGTGGCCCTTTAAGGATAATCTGCAATATACACATTTTAATAGTC
ATCCATCTGAAATCGATCCACATTCCAGAGAAGATTCAAGTATTGTGCTGTGTGAAATAAG
CATTTCCAGAAAAAAACATTTATGCTAATAATACAACATAACCTCTGCATTAAAGAAAA
AGATGCTTTTAGGCCAGGCGCCGTGGCTCACGCCTGTAATCCCTGCACTTTGAGAGGCTG
AGGTGGGTGGATCATGAGGTCAGGAGATCAAGACCATCCTGGCTAACAGGGTGAAACCCC
GTCTCTACTGGGGATATAACAAAGTTAGCTGGGTGTGGTGGTGGGTGCTTGTGGTCCCAG
CTACTCAGGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGAAGGCAGAGGTTGTAGTGAC
GCGAGGTTACGCCACTGCATTCCAGTCTGGG

BX111256

CAGGAAGNTAAGAACAGTCCTAAAATCTCTTTGGCTTCTTTGTCCTGATATGCACCGGCA
TTTTACAGTAGGAACTAGGGTTTCTGTCCAGTTTTTTGGTTCTTTAAGGAATTAATGT
TATTCTGGGTACAACCTGCTTACATACATAGCACATATAGATGACATTTTTACAGGCCGTC
TTGTTAGACTGACATACATGGAGGATAGTGCCACCCGCTCACAAGAACATCAGGTAAGC
TCAGGCACAGAGTGCCAGGAATCTGTAAGGCTTCGCCACGCACAAGTCAGGGCTGCCA
GTCACCTGGGTGTCTTCACTTTATTGGCTGCGTCTAATGACACCTTCCAACTTTTGAC

CCCACCCCTGGACTGTTGTGTAAACATTGTATTTCTCCATCTGTAATGAAAAAGCTAACA
CATCTCTAACTCCAGAGACATTTTCCAGAACATGCTGTTCTCAGGCACTAGTGAGGCGGT
ACCATTATTCTTCATTTGTTATCCAAATGTTGGCCATGTGACCACACCAAAGCTCATCC
TGGGCCACTGAGACTGGTAATTGAATCAGAATATAGTGAAATATTCTTCTCATATATAC
CCAGCCATCTTACATCTTTGGCTTTTTTTCAGCAGATCCTTGTGGCACTCAGAACATCCAT
TTTGCACTGTGTATTTTTT

BX117618

AAATTTTTAACTTTTAAATAGTTAAAATAGTTAACTATTGGTATGGTAGGAAATGATAAAG
TAGACTAGTATCTGTATACATTTTCTGCATTTATGACATACCTTTTTCTTCATTTTTTTC
AATATTTTAAATTGAAAAGTTCATCCGAGTTTCATCTAAGTTTTTCAAAGTGATACAAAT
CTCCAAAAAATTTTCCAATATATGTATTGAAAAAATCCAGGTGTAAGTGGCTCTGCGCAG
TCCAAACCTGTGTTGTTCAAGGGTCAACTGTGTATGAATCCAAGCGAAAGCTTTTCTTAA
CACCTCATAAGAACTATTTTTTAAAAAACAGGAAGTAGCATAGAGTAACCATCACAGGTA
AAGTGTAATTTGTTATCAGCCATCTTTTGCCCATTTTCACTACTGGTAGAAGGCTCAATGG
TAAAAATAAAAAACGGGACAGTCAGAAAAA

AA682806

TCTGAGTAAGAACAGGCTTTATTTGTAAAACCACTCGTGACTCTTTACAAAGCAGGATAC
ACAGAAGGGAAAAAATACACAGTGCAAAATGGATGTTCTGAGTGCCACAAGGATCTGCT
GAAAAAAGCCAAAGATGTAAGATGGCTGGGTATATATGAGAAATGAATATTTCACTATATT
CTGATTCAATTACCAGTCTCAGTGGCCAGGATGAGCTTTTGGTGTGGTCACATGGCCAA
CATTTGGATAACAAATGAGGAATAATGGTACCGCTCACTAGTGCCTGAGAACAGCATGT
TCTGGAAAATGTCTCTGGAGTTAGAGATGTGTTAGCTTTTTCATTACAGATGGAGAAATA
CAATGTTTACACAACAGTCCAGGGGTGGGGTCAAAG

AI202376

CTGACTGTCCCGTTTTTATTTTTACCATTGAGCCTTCTACCAGTACTGAAATGGGCAAAA
GATGGCTGATAACAAATTACACTTTACCTGTGATGGTTACTCTATGCTAGTTCCTGTTTT
TTAAAAAATAGTTCTTATGAGGTGTTAAGAAAAGCTTTCGCTTGGATTATACACAGTTG
ACCCTTGAACAACACAGGTTTGGACTGCGCAGAGCCACCCTCGTGCCGAATT

AI658949

CTGACTGTCCCGTTTTTATTTTTACCATTGAGCCTTCTACCAGTACTGAAATGGGCAAAA
GATGGCTGATAACAAATTACACTTTACCTGTGATGGTTACTCTATGCTAGTTCCTGTTTT
TTAAAAAATAGTTCTTATGAGGTGTTAAGAAAAGCTTTCGCTTGGATTATACACAGTTG
ACCCT

BG403405

GGAAATGATAAAGTAGACTAGTATCTGTATACATTTTCTGCATTTATGACATACCTTTTT
CTTCATTTTTTTCAATATTTTAAATTGAAAAGTTCATCCGAGTTTCATCTAAGTTTTTCA
AAGTGATACAAATCTCCAAAAAATTTTCCAATATATGTATTGAAAAAATCCAGGTGTAAG
TGGCTCTGCGCAGTCCAAACCTGTGTTGTTCAAGGGTCAACTGTGTATGAATCCAAGCGA
AAGCTTTTCTTAACACCTCATAAGAACTATTTTTTAAAAAACAGGAAGTAGCATAGAGTA
ACCATCACAGGTAAAGTGTAATTTGTTATCAGCCATCTTTGCCCATTTTCACTACTGGTAG
AAGGCTCAATGGTAAAAATAAAAAACGGGACAGTCAGAAGATCTGGAAGTCTGACCCTGC
TTTACCTGGCATGTGTAATCCAGTCATGCTCGTATCAGTCTCTGTAGGAGCACTTGAAG

GTATTACATAAATGCTATCTAACTCTGGGAAACGCCAACATGTGATTGCCTCCAGAGGAA
TCTTCTTTAAAAAAAATTCAAAATGTTATTTCTTACTAGGATGTCTTTAAAGAATTAT
AACCCTTACCGTGCCTCCACATTAGATAGATCCCTGCAACAGACCCATGTGGCACCAGCA
GAGACAGCAGGAGGAGAGGCAGCAGCTCCCGTTGTTTGTCTGGAAAAACAAAGGTTATC
ACTTTG

BE673417

CTGACTGTCCCGTTTTTATTTTTACCATTGAGCCTTCTACCAGTACTGAAATGGGCAAAA
GATGGCTGATAACAAATTACACTTTACCTGTGATGGTTACTCTATGCTAGTTCTGTTTT
TTAAAAATAGTTCTTATGAGGTGTTAAGAAAAGCTTTCGCTTGGATTATACACAGTTG
ACCTT

AW021469

GCACGAGATTATTCCTCATTTGTTATCCAAATGTTGGCCATGTGACCACACCAAAAGCTC
ATCCTGGGCCACTGAGACTGGTAATTGAATCAGAATATAGTGAAATATTCATTCTCATAT
ATACCCAGCCATCTTACATCTTTGGCTTTTTTTCAGCAGATCCTTGTGGCACTCAGAACAT
CCATTTTGCAGTGTGTATTTTTTCCCTTCTGTGTATCCTGCTTGTAAAGAGTCACGAG
TGGTTTTACAAATAAAGCCTGTTCTTACTCAGAAAAAAAAAAAAAAAAAAAAA

CF455736

NNTTGAACAGGCGTGACGGTCCGGATTCCCGGGATGTTGTGCTCTGCCCCAACAAACAGGCG
TCCCTTTCCCTCTGGATAACACAAAAGCAAGCCGGGAGGCTGGCTGCCTCTCCTCCTGC
TGTCTCTGCTGGTGGCCACATGGGTGCTGGTGGCAGGGATCTATCTAATGTGGAGGCACG
AAAGGATCAAGAAGACTTCCTTTTCTACCACCACACTACTGCCCCCATTAAAGGTTCTTG
TGGTTTACCCATCTGAAATATGTTTCCATCACACAATTGTTACTTCACTGAATTTCTTC
AAAACCATTGCAGAAGTGAGGTCATCCTTGAAAAGTGGCAGAAAAAGAAAATAGCAGAGA
TGGGTCCAGTGCAGTGGCTTGCCACTCAAAGAAGGCAGCAGACAAAGTCGTCTTCCTTC
TTTCCAATGACGTCAACAGTGTGTGCGATGGTACCTGTGGCAAGAGCGAGGGCAGTCCCA
GTGAGAACTCTCAAGACCTCTTCCCCCTTGCCCTTTAACCTTTTCTGCAGTGATCTAAGAA
GCCAGATTCTATCTGCACAAATACGTGGTGGTCTACTTTAGAGAGATTGATACAAAAGACG
ATTACAATGCTCTCAGTGTCTGCCCCAAGTACCACCTCATGAAGGATGCCACTGCTTTCT
GTGCAGAACTTCTCCATGTCAAGCAGCAGGTGTGAGCAGGAAAAAGATCACAAGCCTGCC
ACGATGGCTGCTGCTCCTGTAGCCACCCATGAGAAGCAAGAGACCTTNAAGGCTTCCT
ATCCCACCATTACAG

AW339874

TTTTTTTTTTTTCTGAGTAAGAACAGGCTTTATTTGTAAAACCACTCGTGACTCTTTAC
AAAGCAGGATACACAGAAGGGAAAAAATACACAGGGCAAAATGGATGTTCTGAGTGCCA
CAAGGATCTGCTGAAAAAAGCCAAAGATGTAAGATGGCTGGGTATATATGAGAATGAATA
TTTCACTATATTCTGATTCAATTACCAGTCTCAGTGGCCCAGGATGAGCTTTTGGTGTGG
TCACATGGCCAACATTTGGATAACAAATGAGGAATAATGGTACCGCCTCACTAGTGCCTG
AGAACAGCATGTTCTGAAAAATGTCTCTGGAGTTAGAGATGTGTTAGCTTTTTTCATTACA
GATGGAGAAATACAATGTTTACACAAC

BG399724

CATGATGTTTCACTATGATCAGTTAACCTTAACCTCTGAGCATCCTGAAGCAAAATCTAAA
TAATGCAGCTATTACCACTGGTGGTCCAGGCTCTGGTGAAGCCCTCTGAGCCCAGGAGGA

AGAGAAAGCATTGTCCAGAGGTAGGAACACAGTCTGGGAGCCCAGAGCTCTGGGAGGAGT
GGGAAAATGCTGCTTCCTGCTGCTTGCTTCTAGGCACCTGCTTCCGCCATCTCACTTACC
ATGGCTAGAGATGGGGGTGAGACTGGGGAAGGACAAAAGCAGGGAACAGATAAGGGATGG
AAATCAGAAGGGAATATAGAAAAGAACTCTGGATGTGGAGAAATGCCGGTACCTGAGCATT
TTGTATCAATGGGAGTACCTCTGTAACTGCTCAGTAGGTTACAAATGAAGAGTCCACCA
GTATTAGAAACAATTTAACTTGCCAGTACCAACTGGGATGTGTGCCTTCAATTTGAAAA
TTGTATGTTTTATTTTTTAAATTTGTTAACAGCATTAAATTTATAGAGTATTGATGTCATT
TATGTTTCTGAGGTGTTTCAA

BF475787

TCTGAGTAAGAACAGGCTTTATTTGTAAAACCACTCGTGACTCTTTACAAAGCAGGATAC
ACAGAAGGGAAAAAATACACAGTGCAAAATGGATGTTCTGAGTGCCACAAGGATCTGCT
GAAAAAGCCAAAGATGTAAGATGGCTGGGTATATATGAGAATGAATATTTCACTATATT
CTGATTCAATTACCAGTCTCAGTGGCCAGGATGAGCTTTTGGTGTGGTCACATGGCCAA
CATTTGGATAACAAATGAGGAATAATGGTACCGCCTCACTAGTGCCTGAGAAAGAGCATGT
TCTGGAAAATGTCTCTGGAGTTAGAGATGTGTTAGCTTTTTTCATTACAGATGGAGAAATA
CAATGTTTACACAACAGTCCAGGGGTGGGGTCAAAAAGTTGGAAGGTGTCATTAGACGCAG
CCAAATAAAGTGAAGACAACCCAGGTGACTGGCAGCCCTGACTTGTGCGTGGGCGA

BF437145

CTGACTGTCCCGTTTTTTATTTTTACCATTGAGCCTTCTACCAGTACTGAAATGGGCAAAA
GATGGCTGATAACAAATTACACTTTACCTGTGATGGTTACTCTATGCTAGTATCCTGTTT
TTTAAAAAATAGTTCTTATGAGGTGTTAAGAAAAGCTTTCGCTTGGATTACATACAGTT
GACCTT

H64601

AGGAAGTTAAGAACAGTCCTAAAATCTCTTTGGCTTCTTTGTCCTGATATGCACCGGCAT
TTTCACAGTAGGAACTAGGGTTTCTGTCCAGTTTCTTTGGTTCTTTAAGGAATTAATGTT
ATTCTGGGTACAACCTGCTTACATACATAGCACATATAGATGACATTTTACAGGCCGTCT
TGTTAGACTGACATACATGGAGGATAGTGCCACCCGCTCACAAGAACATCAGGTAAGCT
CAGGCACAGAGTCCNAGGNNATCTGTAAGGGCTTCGCCCACGCACAAGTCAGGGCTGCCA
GTCACCNGGGTTGTCTTCACTTTATTTGGGCTGCGTCTAATGACACCTTNCCAACCTTTT
GACCCACCCCTGGGGCTTGTGTGTAAACCATGTTATTTCTCCNTCTGTAATGAAAA
AGGTTAACACNTTTTTAACCTCCGGNGACATTTTTC

AF212365

gcacgagcga	tgtcgctcgt	gctgctaagc	ctggcgcgcg	tgtgcaggag	cgccgtaccc
cgagagccga	ccgttcaatg	tggtcttgaa	actggggccat	ctccagagt	gatgctacaa
catgatctaa	tccccggaga	cttgagggac	ctccgagtag	aacctgttac	aactagtgtt
gcaacagggg	actattcaat	tttgatgaat	gtaagctggg	tactccgggc	agatgccagc
atccgcttgt	tgaaggccac	caagatttgt	gtgacgggca	aaagcaactt	ccagtcctac
agctgtgtga	ggtgcaatta	cacagaggcc	ttccagactc	agaccagacc	ctctgggtggt
aaatggacat	tttctacat	cggtctccct	gtagagctga	acacagtcta	tttcattggg
gcccataata	ttcctaattg	aaatatgaat	gaagatggcc	cttccatgtc	tgtgaatttc
acctcaccag	gctgcctaga	ccacataatg	aatataaaaa	aaaagtgtgt	caaggccgga
agcctgtggg	atccgaacat	cactgcttgt	aagaagaatg	aggagacagt	agaagtgaac
ttcacaacca	ctcccctggg	aaacagatac	atggctctta	tccaacacag	cactatcatc
gggtttttct	aggtgtttga	gccacaccag	aagaacaaaa	cgcgagcttc	agtggtgatt
ccagtgaactg	gggatagtga	aggtgctacg	gtgcagctga	ctccatattt	tcctacttgt

```

ggcagcgact gcateccgaca taaaggaaca gttgtgctct gccacaaaac aggcgtccct
ttccctctgg ataacaacaa aagcaagccg ggaggctggc tgccctctct cctgctgtct
ctgctggtgg ccacatgggt gctgggtggc gggatctatc taatgtggag gcacgaaagg
atcaagaaga cttccttttc taccaccaca ctactgcccc ccattaagggt tcttgtgggt
taccatctg aaatatgttt ccatcacaca atttgttact tcaactgaatt tcttcaaaaac
cattgcagaa gtgaggtcat ccttgaaaag tggcagaaaa agaaaatagc agagatgggt
ccagtgcagt ggcttgccac tcaaaagaag gcagcagaca aagtcgtctt ccttctttcc
aatgacgtca acagtgtgtg cgatggtacc tgtggcaaga gcgagggcag tcccagttag
aactotcaag actcttcccc ttgcttttaa ccttttctgc agtgatctaa gaagccagat
tcatctgcac aaatacgtgg tggctacttt tagagagatt gatacaaaaag acgattacaa
tgcctctcagt gtctgcccc agtaccacct catgaaggat gccactgctt tctgtgcaga
acttctccat gtcaagtagc aggtgtcagc aggaaaaaga tcacaagcct gccacgatgg
ctgctgctcc ttgtagccca cccatgagaa gcaagagacc ttaaaggctt cctatcccac
caattacagg gaaaaaacgt gtgatgatcc tgaagcttac tatgcagcct acaaacagcc
ttagtaatta aaacatttta taccaataaa attttcaaat attgctaact aatgtagcat
taactaacga ttggaaacta catttacaac ttcaaagctg ttttatacat agaaatcaat
tacagtttta attgaaaact ataaccattt tgataatgca acaataaagc atcttcagcc
aaaaaaaaa aaaaaa

```

AF208110

```

cggcgatgtc gctcgtgctg ataagcctgg ccgcgctgtg caggagcgcc gtaccccgag
agccgaccgt tcaatgtggc tctgaaaact ggccatctcc agagtggatg ctacaacatg
atctaattcc cggagacttg agggacctcc gagtagaacc tgttacaact agtgttgcaa
caggggacta ttcaattttg atgaatgtaa gctgggtact ccgggcagat gccagcatcc
gcttgttgaa ggccaccaag atttgtgtga cgggcaaaag caacttccag tctacagct
gtgtgaggtg caattacaca gaggccttcc agactcagac cagacctctt ggtggtaaat
ggacattttc ctatatcggc ttccctgtag agctgaacac agtctatttc attggggccc
ataatattcc taatgcaaat atgaatgaag atggccttcc catgtctgtg aatttcacct
caccaggctg cctagaccac ataatgaaat ataaaaaaa gtgtgtcaag gccggaagcc
tgtgggatcc gaacatcact gcttgtgaaga agaattgagga gacagtagaa gtgaacttca
caaccactcc cctgggaaac agatacatgg ctcttatcca acacagcact atcatcgggt
tttctcaggt gtttgagcca caccagaaga aacaaacggc agcttcagtg gtgattccag
tgactgggga tagtgaagggt gctacgggtg agctgactcc atattttcct acttgtggca
gcgactgcat ccgacataaa ggaacagttg tgcctctgcc acaaacagge gtccctttcc
ctctggataa caacaaaagc aagccgggag gctggctgcc tctcctcctg ctgtctctgc
tgggtggccac atgggtgctg gtggcagggg tctatctaag gtggaggcac gaaaggatca
agaagacttc cttttctacc accacactac tgccccccat taagggttctt gtggtttacc
catctgaaat atgtttccat cacacaattt gttacttcac tgaatttctt caaaaccatt
gcagaagtga ggtcatcctt gaaaagtggc agaaaaagaa aatagcagag atgggtccag
tgcaagtggc tgccactcaa aagaaggcag cagacaaagt cgtcttccct ctttccaatg
acgtcaacag tgtgtgcgat ggtacctgtg gcaagagcga gggcagtccc agtgagaact
ctcaagacct cttccccctt gcctttaacc ttttctgcag tgatctaaga agccagattc
atctgcacaa atacgtgggt gtctacttta gagagattga taaaaagac gattacaatg
ctctcagtggt ctgccccaa gaccacttca tgaaggatgc cactgcttcc tgtgcagaac
ttctccatgt caagcagcag gtgtcagcag gaaaaagatc acaagcctgc cagcatgggt
gctgctcctt gtagccacc catgagaagc aagagacctt aaaggcttcc tatccacca
attacagggg aaaaacgtgt gatgatcctg aagcttacta tgcagcctac aaacagcctt
agtaattaaa acattttata ccaataaaat tttcaaatat tactaactaa tgtagcatta
actaacgatt ggaaactaca tttacaactt caaagctgtt ttatacatag aaatcaatta
cagctttaat tgaaaactgt aaccattttg ataatgcaac aataaagcat cttccaaaaa
aaaaaaaaa aaaaaaaaa aaaaaaa

```

AF208111

cggcgatgtc gctcgtgctg ataagcctgg ccgcgctgtg caggagcgcc gtaccccag
 agccgaccgt tcaatgtggc tctgaaactg ggccatctcc agagtggatg ctacaacatg
 atctaatecc cggagacttg agggacctcc gagtagaacc tgttacaact agtggtgcaa
 caggggacta ttcaattttg atgaatgtaa gctgggtact ccgggcagat gccagcatec
 gcttggtgaa ggccaccaag atttgtgtga cgggcaaaag caacttccag tectacagot
 gtgtgaggtg caattacaca gaggccttcc agactcagac cagacctctt ggtggtaaat
 ggacattttc ctatatcggc ttcctgttag agctgaacac agtctatttc attggggccc
 ataattattc taatgcaaat atgaatgaag atggccttcc catgtctgtg aatttcacct
 caccaggctg cctagaccac ataatgaaat ataaaaaaa gtgtgtcaag gccggaagcc
 tgtgggateg gaacatcact gcttgtaaga agaatgagga gacagtagaa gtgaacttca
 caaccactec cctgggaaac agatacatgg ctcttatcca acacagcact atcatcgggt
 tttctcaggt gtttgagcca caccagaaga aacaaacgag agcttcagtg gtgattccag
 tgactgggga tagtgaagg gctacgggtg aggtaaagt cagtgaagctg ctctggggag
 ggaagggaca tagaagactg ttccatcatt cattgctttt aaggatgagt tctctcttgt
 caaatgcact tctgccagca gacaccagtt aagtggcgtt catgggggtt ctttcgctgc
 agcctccacc gtgctgaggt caggaggccg acgtggcagt tgtggtccct tttgtgtcac
 ttaatggctg ctgaccttcc aaagcacttt ttattttcat tttctgtcac agacactcag
 ggatagcagt accattttac ttccgcaagc ctttaactgc aagatgaagc tgcaaagggg
 ttgaaatggg aaggtttgag ttccaggcag cgtatgaact ctggagaggg gctgccagtc
 ctctctgggc cgcagcggac ccagctggaa cacaggaagt tggagcagta ggtgctcctt
 cactctcag tatgtctctt tcaactctag tttttgaagt ggggacacag gaagtcaggt
 ggggacacag cactcccca aagaataagg aacttccatg cttcattccc tggcataaaa
 agtgntcaaa cacaccagag ggggcaggca ccagccaggg tatgatgggt actaccttt
 tctggagaac catagacttc ccttactaca gggacttgca tgtcctaaag cactggctga
 aggaagccaa gaggatcact gctgctcctt tttttagag gaaatgtttg tgtacgtggg
 aagatatgac ctagcccttt taggtaagcg aactggtag ttagtaacgt gtacaaagtt
 taggttcaga ccccgagggt cttgggcatg tgggtctcgg gtcactgggt ttgactttag
 ggctttgtta cagatgtgtg accaagggga aaatgtgcat gacaacacta gaggtagggg
 cgaagccaga aagaagggaa gttttggtg aagtaggagt cttgggtgaga ttttgctgtg
 atgcatgggt tgaactttct gagectcttg tttttcctca gctgactcca tattttccta
 cttgtggcag cgactgcac cgacataaag gaacagtgtg gctctgccc caaacaggcg
 tccctttccc tctggataac aacaaaagca agccgggagg ctggctgcct ctctcctgc
 tgtctctgct ggtggccaca tgggtgctgg tggcagggat ctatctaag tggaggcacg
 aaaggatcaa gaagacttcc ttttctacca ccacactact gccccccatt aaggttcttg
 tggtttacc atctgaaata tgtttccatc acacaatttg ttacttccat gaatttcttc
 aaaaccattg cagaagttag gtcactcctg aaaagtggca gaaaaagaaa atagcagaga
 tgggtccagt gcagtggctt gccactcaaa agaaggcagc agacaaagtc gtcttcttc
 tttccaatga cgtcaacagt gtgtgcatg gtacctgtg cctttaacct tttctgcagt
 gtgagaactc tcaagacctc ttcccccttg tctactttag agagattgat acaaaagacg
 gccagattca tctgcacaaa tacgtgggtg accacttcat gaaggatgcc actgctttct
 attacaatgc tctcagtgtc tgcaccaagt tgtcagcagg aaaaagatca caagcctgcc
 gtgcagaact tctccatgtc aagcagcagg tagcccaccc atgagaagca agagacctta aaggcttct
 acgatggctg ctgctccttg tagcccaccc atgatcctga agcttactat gcagcctaca
 atcccaccaa ttacaggga gtaattaaaa cattttatata caataaaatt ttcaaataat actaactaat
 gtagcattaa ctaacgattg gaaaactacat ttacaacttc aaagctgttt tatacataga
 aatcaattac agctttaatt gaaaactgta accattttga taatgcaaca ataaagcatc
 ttccaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaa

```

atgtegctcg tgetgctaag cctggccgcg ctgtgcagga gcgcggtacc ccgagagccg
accgttcaat gtggctctga aactgggcca tctccagagt ggatgctaca acatgatcta
atcccgggag acttgagggg cctccgagta gaacctgtta caactagtgt tgcaacaggg
gactattcaa ttttgatgaa tgtaagctgg gtaactccggg cagatgccag catccgcttg
ttgaaggcca ccaagatttg tgtgacgggc aaaagcaact tccagtccta cagctgtgtg
aggtgcaatt acacagaggg cttccagact cagaccagac cctctgggtg taaatggaca
ttttcctata tcggcttccc tgtagagctg aacacagctc atttcattgg ggcccataat
attcctaata caaatatgaa tgaagatggc ccttccatgt ctgtgaattt cacctacca
ggctgcctag accacataat gaaatataaa aaaaagtgtg tcaaggccgg aagcctgtgg
gatccgaaca tcaactgctt taagaagaat gaggagacag tagaagtgaa cttcacaacc
actcccctgg gaaacagata catggtctct atccaacaca gcactatcat cgggttttct
caggtgtttg agccacacca gaagaaacaa acgcgagctt cagtgggtgat tccagtgaact
ggggatagtg aaggtgctac ggtgcagctg actccatatt ttcctacttg tggcagcgac
tgcattccgac ataaagggaac agttgtgctc tgcccacaaa caggcgtccc tttccctctg
gataacaaca aaagcaagcc gggaggctgg ctgcctctcc tctgtctgtc tctgtctgtg
gccacatggg tgetgggtgg agggatctat ctaatgtgga ggcacgaaag gatcaagaag
acttcctttt ctaccaccac actactgccc cccattaagg ttctgtgggt ttaccatct
gaaatatgtt tccatcacac aatttgttac ttcactgaat ttcttcaaaa ccattgcaga
agtgggttca tcttgaaaa gtggcagaaa aagaaaatag cagagatggg tccagtgcag
tggcttgcca ctcaaaagaa ggcagcagac aaagtctgtc tcttctttc caatgacgtc
aacagtgtgt gcgatgggtac ctgtggcaag agcaggggca gtcccagtga gaactctcaa
\ gacctcttc cccttgccct taaccttttc tgcagtgatc taagaagcca gattcatctg
caciaatatc tgggtggtcta ctttagagag attgatacaa aagacgatta caatgctctc
agtgtctgcc ccaagtacca cctcatgaag gatgccactg ctttctgtgc agaacttctc
catgtcaagc agcaggtgtc agcaggaaaa agatcacaag cctgccacga tggctgtgtc
tccttgtagc ccacccatga gaagcaagag accttaaagg gttccttttc ccatcattta
caggggaaaa acgtgtgatg atc

```

AK095091

```

catattagag tctacagata tgcctttctt acagcaatcc tgcaccacca taaaagctac
attttcaata caagattaaa aggtattctg caaaatgtgc aaggttttca tgtctgtgg
ttagctgtga gtgatggctt catgaatttt ttctttttt gactatgggt cttacgtgg
attcatttat cttgaaatgg tgaacaatca cagctgcaga cctcaattt atggtacata
tcaagcaatt tggctttttt tcttgtaatg aaaaaaaaaa gttttttttg ctttttttca
tgacactgct tcttggggagc actgccagca ttactagtgg cacttcgtat gggtcctaag
gtgttattga aggtttacga tattgcacta aacacgaaaa ataccagaga accactggag
atacttttta ctgtgatatg taatttactg gagacaggaa ctgctcgttt ggagatgggt
agcatcacag ggtgttttaa gtogatactt gcaacccttg agctcaccac agtagcaaca
ggaggtggct aggaaattat tcacagcagg acagtacgca ctgcaattaa ttgtatgcag
ttatgattta ataccacatc tttatgctca cgtttctctc aactgtgaat ggtgccatgt
acagttggta tgtgtgtgtt taagttttga taaattttta acttttaata gttaaaatag
ttaactattg gtatggtagg aaatgataaa gttagactagt atctgtatac attttctgca
tttatgacat acctttttct tcattttttt caatatttta attgaaaagt tcatccgagt
ttcatctaag ttttttcaaa gtgatacaaa tctccaaaaa attttccaat atatgtattg
aaaaaatcca ggtgtaagtg gctctgcgca gtccaaacct gtgttgttca agggccaact
gtgtatgaat ccaagcgaaa gcttttctta acacctcata agaactattt tttaaaaaac
aggaactagc atagagtaac catcacaggt aaagtgtaat ttgttatcag ccattctttg
cccatttccg tactggtaga aggtcfaatg gtaaaaaata aaacgggaca gtcagaagat
ctggaagtcc tgacctgctc ttcacctggc atgtgtaatc cagtcatgct cgtatcagtc
tctgtaggag cacttgaagg tattacataa atgctatcta actctgggaa acgccaacat
gtgattgocct ccagaggaat cttcttttaa aaaaaattca aaatgttatt tcttactag
gatgtcttta aagaattata accttaccg tgctccaca ttagatagat ccctgccacc

```

agcaccatg tggccaccag cagagacagc aggaggagag gcagccagcc tcccggcttg
cttttgtctg gaaaaaacia agcttattca cctttggaaa acaaatccac acttatctct
taatttaaaa actaagactt ggtatacttt atagaggttt atttattttt tattattttt
tagttttgag acagagtctc gctttgttgc ctaggctgga gtgcagtggc gcaatctcgg
ttcactgcag cctccgtctc ccgggttcaa gcaatgctgc ctcagcctcc tgagttagctg
ggattacagg catgtgtcac cgcgcccage cactttgtag agatttagat ccctttaaaa
ccatcagtcga gaagctcttt agatagtctg ccaatcatat ctttttccct agagtgtgca
ggtcttgcac tagattctca aaagggatat gggacccagg aagttaagaa cagtcctaaa
atctcttttg cttctttgtc ctgatatgca ccggcatttt cacagtagga actaggggtt
ctgtccagtt tttttgggtc ttttaaggaat taatgttatt ctgggtacaa ctgcttacat
acatagcaca tatagatgac atttttacag gcgctcttgt tagactgaca tacatggagg
atagtccac ccgcctcaca agaacatcag gtaagctcag gcacagagtg cccaggaatc
tgtaaggctt ccgccaagca caagtccagg ctgccagtca cctgggttgt cttcacttta
tttggctgog tctaattgaca ccttccaaact tttgaccca cccctggact gttgtgtaaa
cattgtattt ctocatctgt aatgaaaaag ctaacacatc tctaactcca gagacatttt
ccagaacatg ctgttctcag gcactagtga ggcgggtacca ttattcctca tttgttatcc
aaatgttggc catgtgacca caccaaaagc tcatcctggg ccactgagac tagtaattga
atcagaatat agtgaaatat tcattctcat atatacccag ccatcttaca tctttggctt
ttttcagcag atccttgtgg cactcagaac atccattttg cactgtgtat tttttccct
tctgtgtatc ctgctttgta aagagtcacg agtggtttta caaataaagc ctgttcttac
tcag

BM983744

TTTTTTTTTTTTTTTTCTGAGTAAGAACAGGCTTTATTTGTAAAACCACTCGTGACTCTT
TACAAAGCAGGATACACAGAAGGGAAAAAATACACAGTGCAAAATGGATGTTCTGAGTG
CCACAAGGATCTGCTGAAAAAAGCCAAAGATGTAAGATGGCTGGGTATATATGAGAATGA
ATATTTCACTATATTCTGATTCAATTACCAGTCTCAGTGGCCAGGATGAGCTTTTGCTG
TGGTCACATGGCCAAACATTTGGATAACAAATGAGGAATAATGGTACCSCCTCACTAGTGC
CTGAGAACAGCATGTTCTGAAAAATGTCTCTGGAGTTAGAGATGTGTTAGCTTTTTCATT
ACAGATGGAGAAATACAATGTTTACACAACAGTCCAGGGGTGGGGTCAAAAGTTGGAAGG
TGTCATTAGACGCAGCCAAATAAAGTGAAGACAACCCAGGTGACTGGCAGCCCTGACTTG
TGCCTGGGCGAAGCCTTACAGATTCCTGGGCACTCTGTGCCCTGAGCTTACCTGATGTTCT
TGTGAGGCGGGTGGCACTATCCTCCATGTATGTGAGTCTAACAAGACGGCCTGTAAAAAT
GTCATCTATATGTGCTATGTATGTAAGCAGTTGTACCCAGAATAACATTAATCCTCGTGC
CGAAT

CB305764

TTTTTTTTTTTTTTTTGTTGGGCTGAAGATGCTTTATTATTGCATTATCAAAATGGTTATA
GTTTTCAATTAAACTGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGT
AGTTTCCAATCGTTAGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTAT
AAAATGTTTTAATTACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACA
CGTTTTTTCCCTGTAATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGT
GGGCTACAAGGAGCAGCAGCCATCGTGGCAGGCTTGTGATCTTTTCCCTGCTGACACCTG
CTGCTTGACATGGAGAAGTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTT
GGGGCAGACACTGAGAGCATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCAC
CACGTATTTGTGCAGATGAATCTGGCTTCTTAGATCACTGCAGAAAAGGTTAAAGGCAAG
GGGAAGAGGTCTTGAGAGTTCTCACTGGGACTGCCCTCGCTCTTGCCACAGGTACCATC
GCACACACTGTTNACGTCAATTGGAAAGAAGGAAGACGACTTTGTCTGCTGCCTTCTTTTG
AGTG

BM715988

TGGTTTTTGTTCATTTCTGTTGGATTACAGAAAAAGAATGGGACCCATTTCAG
 GTCTCGATTTCCAAAGGTAAAGATGGAAGGCTGGGCAGACTGGCTTTTGTTACCTGACAT
 GCCGTAGGGTGAGCTTAGAGGAAGAAAGAAAAACAATTTTATTTGGCCAAAACAGAACAA
 ATGCTGAAAAGGAAATCTTGTTTTTCTTAAAGCCAAATAGAAATGATTTGGGTATAAT
 TTAAGAGTCCTTGTTGTACAGATATGGTGAATGATGTAGTTATTAATACTACCAACTT
 AGTCATCAAGCCTCAATTTTCTTTACCTGAAGGATTAAGTGAAAGCTTTTGGAGTTTCAT
 GATGTTTCAGTATGATCAGTTAACCTTAACCTCTGAGCATCCTGAAGCAAATCTAAATAA
 TGCAGCTATTACCACTGGTGGTCCAGGCTCTGGTGAAGCCCTCTGAGCCCAGGAGGAAGA
 GAAAGCATTGTCCAGAGGTAGGAACACAGTCTGGGAGCCCAGAGCTCTGGGAGGAGTGGG
 AAAATGCTGCTTCTGCTGCTTCTAGGCACCTGCTTCCGCCATCTCACTTACCATG
 GCTAGAGATGGGGGTGAGACTGGGGAAGGACACAAGCAGGGAACAGATAAGGGATGGAAA
 TCAGAAGGGAATATAGAAAGAACTCTGGATGTGGAGACATGCCGGTACCTGAGCATTTTG
 TATCAATGGGAGTACCTCT

BM670929

TTTTTTTTTTTTTTTTTTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTACAG
 TTTTCAATTAAAGCTGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTA
 GTTTCCAATCGTTAGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATA
 AAATGTTTTAATTACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACAC
 GTTTTTTTCCCTGTAATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGT
 GGGCTACAAGGAGCAGCAGCCATCGTGGCAGGCTTGTGATCTTTTCTGCTGACACCTG
 CTGCTTGACATGGAGAAGTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTT
 GGGGCAGACACTGAGAGCATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCAC
 CACGTATTTGTGCAGATGAATCTGGCTTCTTAGATCACTGCAGAAAAGGTTAAAGGCAAG
 GGGGAAGAGGTCTTGAGAGTTCTCACTGGGACTTGCCTCGCTCTTGCCACAGGTACCATC
 GCACACACTGTTGACGTCATTGGAAAGAAAGAAGACGACTTTGTCTGCTGCCTTCTT

BI792416

GCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTACAGTTTCAATTAAAGCTGTAA
 TTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGT

BI715216

CACGCGTCCGATTTTATACCAATAAAATTTTCAAATATTGCTAACTAATGTAGCATTAAC
 TAACGATTGGAACTACATTTACAACCTCAAAGCTGTTTTATACATAGAAATCAATTACA
 GCTTTAATTGAAAACGTAAACCATTTTGATAATGCAACAATAAAGCATCTTCAGCCAAAA
 AAAAAA

N56060

AGAAAAAGAAAATAGCAGAGATGGGTCCAGTGCAGTGGCTTGCATAAAAAAGAAGGCAGC
 AGACAAAGTCGTCTTCTTCTTCCAATGACGTCAACAGTGTGTGCGATGGTACCTGTGG
 CAAGAGCGAGGGCAGTCCCAGTGAGAACTCTCAAGACCTCTTCCCCCTTGCCCTTAACC
 TTTTCTGCAGTGATCTAAGAAGCCAGATTCTCTGCACAAATACGTGGTGGTCTACTTTA
 GAGAGATTGATACAAAAGACGATTACAATGCTCTCAGTGTCTGCCCAAGTACCACCTCA
 TGAAGGATGCCACTGCTTCTGTGCAGAACTTCTCCATGTCAAGCAGCAGGTTTCAGCAG
 G

CB241389

TTTTTTTTTTTTTTGTTTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTACAG
 TTTTCAATTAAAGCTGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTA
 GTTCCCAATCGTTAGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATA
 AAATGTTTTTAATTACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACAC
 GTTTTTTCCCTGTAATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTG
 GGTACAAGGAGCAGCAGCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGC
 TGCTTGACATGGAGAAGTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTG
 GGGCAGACACTGAGAGCATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACC
 ACGTATTTGTGCAGATGAATCTGGCTTCTTAGATCACTGCAGAAAAGGTTAAAGGCAAGG
 GGAAGAGGTCTTGAGAGTTCTCACTGGGACTGCCCTCGCTCTTGCCACAGGTACCATCG
 CACACACTGTTGACGTCATTGGAAAGAAGGAAGACGACTTTGTCTGCTGCCTTCTTTTGA
 GTGGCAAGCCACTGCACTGGACCCATCTCTGCTATTTTCTTTTCTNGCACTTTTCAAGG
 ATGACTCACTTCTGCAATGGTTTTTGAGAATTCAAGTGAAGTACAAATGTGTGATGGAACA
 TAT

AV660618

CGCTCGTGCTGCTAAGCCTGGCCGCGCTGTGCAGGAGCGCCGTACCCCGAGAGCCGACCG
 TTCAATGTGGCTCTGAACTGGGCCATCTCCAGAGTGGATGCTACAACATGATCTAATCC
 CCGGAGACTTGAGGGACCTCCGAGTAGAACCTGTTACAACCTAGTGTGCAACAGGGGACT
 ATTCAATTTTGATGAATGTAAGCTGGGTACTCCGGGCAGATGCCACACCAGAAGAAACAA
 ACGCGAGCTTCAGTGGTGATTCCAGTGAAGTGGGATAGTGAAGGTGCTACGGTGCAGCTG
 ACTCCATATTTTCTACTTGTGGCAGCGACTGCATCCGACATAAAGGAACAGTTGTGCTC
 TGCCACAAACAGGCGTCCCTTTCCTCTGGATAACAAC

BX088671

GCTGAGTGTGATGGTGTAAGCCTGTGGTCCCAGCTACTAGGGAGGCTGAGATGGGATTAC
 AGGTGTGAGCCACGGCGCCTGGCCTAAAAGCATCTTTTCTTTAACGCAGAGGTTATGTT
 GTATTATTAGCATAAATGTTTTTTCTGGGAATGCTTATTTACACAGCACAACTACTGAA
 TCTTCTCTGGAATGTGGATCGATTTAGATGGATGACTATTAAATGTGTATATTGTCAG
 ATTATCCTTAAAGGGCCACCTCATGCCTTCTAATTTATGTCTTACGGATAAAAAATCAA
 ATGAAGCATAAAGTAAAACTGTGTCCAGCTTTACAAGTGGACGCTTAGTAATGGCTGAG
 GCAATATGTTTAAATGTAGCAAATTTTACTTATTTGTCATGATCAGTTTTACAGTGCTTG
 TAAGTGCTGGTAATAGAAGATGGACATGGTTTAGGTCAAACTTGGACCAGAAACCAACT
 TCTTTTGAACAGCTCTACCAGNTATAAGAGCAATATG

CB154426

CTGTTGACGTCATTGGAAAGAAGGAAGACGACTTTGTCTGCTGCCTTCTTTTGAGTGGCA
 AGCCACTGCACTGGACCCATCTCTGCTATTTTCTTTTCTGCCACTTTTCAAGGATGACC
 TCACTTCTGCAATGGTTTTGAAGAAATTCAAGTGAAGTAACAAATTGTGTGATGGAAACAT
 ATTTAGATGGGTAAACCACAAGAACCTTAATGGGGGGCAGTAGTGTGGTGGTAGAAAAG
 GAAGTCTTCTTGATCCTTTCTGTGAGAGGAGAAAAGCATTTGTTATCTGTGAACAGCAA
 CAGCAGGCTTTCACTCTGTAAACCATCCCTGACAAATGATCCCTTGCTAGAGAATGTCAG
 CTGAGCACCAAGGGCCTTGTAGTGACAGCAAGGAAAAACATCCTGATGTTCTTTTGAA
 CACATCACCTGAAACACACTGATGCTTAAACCTTAACTTTTTTTTTTTTTGGAGACACAGT
 CTCACTCTGT

CA434589

TTTTTTTTTTTTTTTTTCTGAGTAAGAACAGGCTTTATTTGTAAAACCACTCGTGACTC
TTTACAAAGCAGGATACACAGAAGGGAAAAAATACACAGTGCAAAATGGATGTTCTGAG
TGCCACAAGGATCTGCTGAAAAAGCCAAAGATGTAAGATGGCTGGGTATATATGAGAAT
GAATATTTCACTATATTCTGATTCAATTACCAGTCTCAGTGGCCCAGGATGAGCTTTTGG
TGTGGTCACATGGCCAACATTTGGATAACAAATGAGGAATAATGGTACCGCTCACTAGT
GCCTGAGAACAGCATGTTCTGGAAAATGTCTCTGGAGTTAGAGATGTGTTAGCTTTTTCA
TTACAGATGGAGAAATACAATGTTTACACAACAGTCCAGGGGTGGGGTCAAAAGTTGGAA
G

CA412162

TTTTTTTTTTTTTTTTTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAG
TTTTCAATTAAACTGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTA
GTTTCCAATCGTTAGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATA
AAATGTTTTAATTACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACAC
GTTTTTTCCCTGTAATTGGTGGGATAGGAAGCCTTTAAGGTCCTTGCTTCTCATGGGTG
GGCTACAAGGAGCAGCAGCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGC
TGCTTGACATGGAGAAGTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACGTG
GGGCAGACACTGAGAGCATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACC
ACGTATTTGTGCAGATGAATCTGGCTTCTTAGATCACTGCAGAAAAGGTTAAAGGCAAGG
GGGAAGA

CA314073

TTTTTTTTTTTTTTTTTGAAGGGTCAGGACTTCCAGATCTTCTGACTGTCCCGTTTTT
ATTTTACCATTGAGCCTTCTACCAGTACTGAAATGGGCAAAAGATGGCTGATAACAAAT
TACACTTTACCTGTGATGGTTACTCTATGCTAGTTCTTGTTTTTAAAAAATAGTTCTTA
TGAGGTGTTAAGAAAAGCTTTCGCTTGGATTATACACAGTTGACCCTTGAACAAACACAG
GTTTGGACTGCGCAGAGCCACTTACACCTGGATTTTTTCAATACATATATTGGAAAATTT
TTTGAGATTTGTATCACTTTGAAAAAATTAGATGAAACTCGGATGAACCTTTTCAATTA
AAATATTGAAAAAATGAAGAAAAGGTATGTCATAAATGCAGAAAATGTATACAGATAC
TAGTCTACTTTATCATTTCCCTACCATACCAATAGTTAACTATTTTAACTATTAAAAGTTA
AAAATTTATCAAACTTAAACACACACATACCAACTGTACATGGCACCATTACAGTTGA
GAGAAACGTGAGCATAAAGATGTGGTATTAAATCATAACTGCATACAATTAATTGCAGTG
CGTACTGTCTGCTGTGAATATTTCCCTAGCCCTCGTGCCGAATC

BF921554

GTGGGTGACCGTGGCTTGCCACTCAAAGAAGGCAGCAGACAAAGTCGTCTTCTCTTT
CCAATGACGTCAACAGTGTGTGCGATGGTACCTGTGGCAAGAGCGAGGGCAGTCCCAGTG
AGAACTCTCAAGACCTCTTCCCCCTTGCTTTAACCTTTTCTGCAGTGATCTAAGAAGCC
AGATTCATCTGCACAAATACGTGGTGGTCTACTTTAGAGAGATTGATACAAAAGACGATT
ACAATGCTCTCAGTGTCTGCCCCAAGTACCACCTCATGAAGGATGCCACTGCTTTCTGTG
CATAACTTCTCCATGTCAAGCAGCAGGTGTGAGCAGGAAAAAGATCACAAGCCTGCCACG
ATGGCTGCTGCTCCTTGTAGCCACCCATGAGAAGCAAGAGACCTTAAAGGCTTCTATC
CCACCAATTACAGGGAAAAAAACGTGTGATGATCCTGAAGCCACGGTCAA

BF920093

TAGAGGATCCCGGTGACGGTGGTTCACTGATCATCACACTTTTTCCCTGTAATAGGTGG
GATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAGCCAT
CGTGGCAGGCTTGTGATCTTTTCCCTGCTGACACCTGCTGCTTGACATGGAGAAGTTATG
CACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTGT

AATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGATGAATCT
GGCTTCTTAGATCACTGCAGAAAAGGTTAAAGGCAAGGGGAAGAGGTCTTGAGAGTTCT
CACTGGGACTGCCCTCGCTCTTGCCACAGGTACCATCGCACACACTGTTGACGTCAATGG
AAAGAAGGAAGACGACTTTGTCTGCTGCCTTCTTTTGAGTGGCAAGCCACGGTCAACCCA
CAAGCCACGGTCAACCCAC

AV685699

TCTACGTGGTAAGATATGACCTAGCCCTTTTAGGTAAGCGAACTGGTATGTTAGTAACGT
GTACAAAGTTTAGGTTTACAGCCCCGGGAGTCTTGGGCATGTGGGTCTCGGGTCACTGGTT
TTGACTTTAGGGCTTTGTTACAGATGTGTGACCAAGGGGAAAATGTGCATGACAACACTA
GAGGTAGGGGCGAAGCCAGAAAGAAGGGAAGTTTGGCTGAAGTAGGAGTCTTGCGACTG
CATCCGACATAAAGGAACAGTTGTGCTCTGCCACAAAACAGGCGTCCCTTTCCCTCTGGA
TAACAACAAAAGCAAGCCGGGAGGCTGGCTGCCTCTCCTCCTGCTGTCTCTGCTGGTGGC
CACATGGGTGCTGGTGGCAGGGATCTATCTAATGTGGAGGCACGAAAGGATCAAGAAGAC
TTCCTTTTCTACCAACCACTACTGCCCCCATTAAAGGTTCTTGTGGTTTACCCATCTGA
AATATGTTTCCATCACACAATTTGTTACTTCACTGAATTTCTTCAAAACCATTGCAGAAG
TGAGGTCACTCCTTGAAAGTGGCAGAGTAGCAGAGATGGGTCCAGTGCAGTGGCTTGCCAC
TCGTGCGATGGTCTT

AV650175

GGCACGAGCACTGGCTGAAGGAAGCCAAGAGGATCACTGCTGCTCCTTTNTTCTAGAGGA
AATGTTTGTCTACGTGGTAAGATATGACCTAGCCCTTTTAGGTAAGCGAACTGGTATGTT
AGTAACGTGTACAAAGTTTAGGTTTACAGCCCCGGGAGTCTTGGGCATGTGGGTCTCGGGT
CACTGGTTTGTACTTTAGGGCTNTGTTACAGATGTGTGACCAAGGGGAAAATGTGCATGA
CAACACTAGAGCTGACTCCATATTTTCTACTTGTGGCAGCGACTGCATCCGACATAAAG
GAACAGTTGTGCTCTGCCACANACAGGCGTCCCTTTCCCTCTGGATAACAACATAAGCA
AGCCGGGAGGCTGGCTGCCTCTCCTCCTGCTGTCTCTGCTGGTGGCACATGGGTGCTGGT
GGAGGGATCTATCTAATGTGGAGGCACGGATCAAGAAGACTTNCTTNTCTACCAACACAC
TACTGGCCCCAATAAGGGTCTNGTGGNTACCCCATCTGAATATGTTTATACACAATTTGT
ACTCACTGAATTTCTCAAAACATTGAGAGTGAGGCATCCTGAAAGTGCAGAAAGANATGCN
AATGGTCAGTGCATGCTGCACTAGCAGCATGGACTT

BX483104

GATCCCGCGCAGTGGCCCCGGCGATGTGCTCGTCTGCTAAGCCTGGCCGCGCTGTGCAG
GAGCGCCGTACCCCCGAGAGCCGACGTTCAATGTGGCTCTGAAACTGGGCCATCTCCAGA
GTGGATGCTACAACATGATCTAATCCCCGGGAGACTTGAGGGACCTCCGAGTAGAACCTGT
TACAACCTAGTGTGCAACAGGGGACTATTCAATTTTGATGAATGTAAGCTGGGTACTCCG
GGCAGATGCCAGCATCCGCTTGTGTAAGGCCACCAAGATTTGTGTGACGGGCAAAAGCAA
CTTCCAGTCTTACAGCTGTGTGAGGTGCAATTACACAGAGGCCTTCCAGACTCAGACCAG
ACCCTCTGGTGGTAAATGGACATTTTCTTACATCGGCTTCCCTGTAGAGCTGAACACAGT
CTATTTTATTGGGGCCATAATATTCCTAATGCAAATATGAATGAAGATGGCCCTTCCAT
GTCTGTGAATTTACCTCACCAGGCTGCCTAGACCACATAATGAAATATAAAAAAAGTG
TGTCAGGCCCGGAAGCCTGTGGGATCCGAACATCACTGCTTGTGAAGAAGATGAGGAGAC
AGTAGAAGTGAACCTTACAACCACTCCCTGGGAAACAGATACATGGCTCTTATCCAACA
CAGCACTATCATTCGG

CD675121

GTCTTGCAATTAGATTCTCAAAAGGGATATGGGACCCAGGAAGTTAAGAACAGTCTTAAAA
TCTCTTTGGCTTCTTTGTCTGATATGCACCGGCATTTTCACAGTAGGAAGTGGGTTTC

TGTCCAGTTTTTTTGGTTCTTTAAGGAATTAATGTTATTCTGGGTACAACTGCTTACATA
 CATAGCACATATAGATGACATTTTTACAGGCCGTCTTGTTAGACTGACATACATGGAGGA
 TAGTGCCACCCGCTCACAAGAACATCAGGTAAGCTCAGGCACAGAGTGCCAGGAATCT
 GTAAGGCTTCGCCCACGCACAAGTCAGGGCTGCCAGTCACCTGGGTTGTCTTCACTTTAT
 TTGGCTGCGTCTAATGACACCTTCCAACTTTTGACCCACCCCTGGACTGTTGTGTAAAC
 ATTGTATTTCTCCATCTGTAATGAAAAAGCTAACACATCTCTAACTCCAGAGACATTTTC
 CAGAACATGCTGTTCTCAGGCACTAGTGAGGCGGTACCATTATTCCTCATTGTGTATCCA
 AATGTTGGCCATGTGACCACACCAAAAGCTCATCCTGGGCCACTGAGACTGGTAATTGAA
 TCAGAATATAGTGAAAATATTATTCTCATATATACCCAGCCATCTTACATCTTTGGCTTT
 TTTTCAGCAGATCCTTGTGGCACTCAGAACATCCATTTTGCCTGTGTATTTTTTCCCTT
 CT

BE081436

TGTGTAACCTCTCAAGACCTCTTCCCCCTTGCCTTTAACCTTTTCTGCAGTGATCTAAGAA
 GCCAGATTCTGTCACAAATACGTGGTGGTCTACTTTAGAGAGATTGATACAAAAGACG
 ATTACAATGCTCTCAGTGTCTGCCCCAAGTACCACCTCATGGAGGATGCCACTGCTTTCT
 GTGCAGAACTTCTCCATGTCAAGTAGCAGGTGTGAGCAGGAAAAAGATCACAAGCCTGCC
 ACGATGGCTGCTGCTCCTTGTAGCCACCCATGAGAAGCAAGAGACCTTAAAGGCTTCCT
 ATCCACCAATTACAGGGAAAAAACGTGTGATGAT

AW970151

CTGAAATATGTTTCCATCACACAATTTGTTACTTCACTGAATTTCTTCAAAACCATTGCA
 GAAGTGAGGTCTCCTTGAAAAGTGGCAGAAAAAGAAAATAGCAGAGATGGGTCCAGTGC
 AGTGGCTTGCCACTCAAAAGAAGGCAGCAGACAAAGTCGTCTTCCCTTCTTTCCAATGACG
 TCAACAGTGTGTGCGATGGTACCTGTGGCAAGAGCGAGGGCAGTCCCAGTGAGAACTCTC
 AAGACCTCTTCCCCCTTGCCTTTAACCTTTTCTGCAGTGATCTAAGAAGCCAGATTCATC
 TGCACAAATACGTGGTGGTCTACTTTAGAGAGATTGATACAAAAGACGATTACAATGCTC
 TCAGTGTCTGCCCCAAGTACCACCTCATGAAGGATGCCACTGCTTTCTGTGCAGAACTTC
 TCCATGTCAAGTAGCAGGTGTGAGCAGGAAAAAGATCACAAGCCTGCCACGATGGCTGCT
 GCTCCTTGTAGCCACCCATGAGAAGCAAGAGACCTTAAAGGCTTCTATCCCACCAATT
 ACAGGGAAAAAACGTGTGATGATCCCTGAAGCTTACTATGCAGCCTACANACAGCCTTA
 GTAATAAAACATTTTATCCAATAAAATTTCAAATTTTGTCTTAACTATGTGCATAAACTAC
 GATTGAAAACCTTTTCACT

AW837146

CATTGTGGTTGCAGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTTCCCTGTAATT
 GGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCA
 GCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGGAGAAG
 TTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGCTACTTGGGGCAGACACTGAGAGC
 ATTGTAATCGTCTTTTGTATCAATCTCCCTAAAGTAGACCACCACGTATTTGTGCAGATG
 AATCTGGCTTCTTAGATCACTGCAGAAAAGGTTAAAGGCAAGGGGGAAGAGGTCTTGAGA
 GTTCTCACTGGGACTGCCCTCGCTCTTGCCACAGGTACCATCGCACACACTGTTGACGTC
 ATTGGAAAGAAGGAAGACGACTTTGTCTGCTGCCTTCTTTGAGTGGCAAGCCACTGCAC
 TGGACCCATCT

AW368264

GTGAATAAGCTTTGTTTTTTCCAGACAAAAGCAAGCCAGGAGGCTGGCTGCCTCTCCTCC
 TGCTGTCTCTGCTGGTGGCCACATGGTTGCTGGTGGCAGGGATCTATCTAATGTGGAGGC
 ACGGTAAGGGTTATAATTCTTTAAAGTCATCCTAGTAAGGAAATAACATTTGGAATTTT

TTTTAAAGAAGATTCTCTGGAGGCAATCACCTGTTGGCGTTTCCCAGAGTTAGATAGCA
 TTTATGTAATACCTTCAAGTGCTCCTACAGAGACTGATACGAGCATGACTGGATTACACA
 TGCCAGGTGAAAGCAGGGCCAGGACTTCCAGATCTTCTGACTGTCCCGTTTTTATTTTA
 CCATTGAGCCTTCTACCAGAACTGAAATGGGCAAAAGATGGCTGATAACAAATTACACTT
 TACCTGTGATGGTTACTCTATGCTAGTTCTGTTTTTAAAAAATAGTTCTTATGAGGTG
 TCAAGAAAAGCTTTCGCTTGGATTCTATACACAGTTGACCCTTGAACAACACAG

D25960

GATCCTGAAGCTTACTATGCAGCCTACAAACAGCCTTAGTAATTAAAACATTTTATACCA
 ATAAAATTTTCAAATATTGCTAACTAATGTAGCATTAACTAACGATTGGAACTACATNN
 ACACTTCAAAGCTGTTTTATACATAGAAATCAATTACAGCTTTAATTGAAAACATAAC
 CATTTTGATAATGCAACANTAAAGCATCTTCAGCCAAA

AV709899

GCAACTTCCAGTCTTACAGCTGTGTGAGGTGCAATTACACAGAGGCCTTCCAGACTCAGA
 CCAGACCCTCTGGTGGTAAATGGACATTTTCTATATCGGCTTCCCTGTAGAGCTGAACA
 CAGTCTATTTTCAATTGGGGCCCATATATTCTAATGCAAATATGAATGAAGATGGCCCTT
 CCATGTCTGTGAATTTACCTCACCAGGCTGCCTAGACCACATAATGAAATATAAAAAAA
 AGTGTGTCAAGGCCGGAAGCCTGTGGGATCCGAACATCACTGCTTGTGAAGAAGATGAGG
 AGACAGTAGAAGTGAACCTTCAACACCACTCCCCTGGGAAACAGATACATGGCTCTTATCC
 AACACAGCACTATCATCGGGTTTTCTCAGGTGTTTGAGCCACACCAGAAGAAACAAACGC
 GAGCTTCAGTGGTGATTCCAGTGACTGGGGATAGTGAAGGTGCTACGGTGCAGCTGACTC
 CATATTTTCTACTTGTGGCAGCGACTGCATCCGACATAAAGGAACAGTTGTGCTCTGCC
 CACAAACAGGCGTNCCTTTTCTCTGGATAACAACAAAAGCAAGCCGGGAGGCTTGGCTG
 CTCTCCTTCTGCTGGCCTTTGCTGTGGCCACATTGGTGTGGTGGCAGGGATCTATCTAA
 TGTGGATGCACGTCTCGTGGTTTACCCATCTGAAATATGTTCN

BX431018

ATTTTTCTCTTGTGGCAGCGACTGGCATCCGACATAAAGGAACAGTTGTGCTCTGCCCA
 CAAACAGGCGTCCCTTTCCCTCTGGATAACAACAAAAGCAAGCCGGGAGGCTGGCTGCCT
 CTCCTCCTGTCTGTCTGTCTGGTGGCCACATGGGTGCTGGTGGCAGGGATCTATCTAATG
 TGGAGGCACGAAAGGATCAAGAAGACTTCTTTTCTACCACCACACTACTGCCCCCATT
 AAGGTCTTGTGGTTTACCCATCTGAAATATGTTTCCATCACACAATTTGTTACTTCACT
 GAATTTCTTCAAACCATTCAGAGAGTGAGGTGATCCTTGAAAAGTGGCAGAAAAAGAAA
 ATAGCAGAGATGGGTCCAGTGACGTGGCTTGCCACTCAAAGAAGGCAGCAGACAAAGTC
 GTCTTCCTTCTTTCCAATGACGTCAACAGTGTGTGCGATGGTACCTGTGGCAAGAGCGAG
 GGCAGTCCAGTGAGAACTCTCAAGACCTCTTCCCCCTTGCCCTTTAACCTTTTCTGCAGT
 GATCTAAGAAGCCAGATTCTCTGCACAAATACGTGGTGGTCTACTTTAGAGAGATTGAT
 ACAAAGACGATTACAATGCTCTCAGTGTCTGCCCCAAGTACCACCTCATGAAGGATGCC
 ACTGCTTTCTGTGCAGAACTTCTCCATGTCAAGCAGCAGGTGTCAGCAGGAAAAAGATCA
 CAAGCCTGCCACGATGGCTGCTGCTCCTTGTAGCCCACCCATGAGAAGCAAGAGACCTTA
 AGGCTTCTATCCACCANTACAGNAAAAACGTGTGATGATCCTGAAGCTTACTATGCAG
 CCTACAACAGGCTTAGTATTAAACATTTATACCCATAAATTTTCAAATTGCT

AL535617

TAGGTGACACTATAGAACAAGTTTGTACAAAAAGCAGGCTGGTACCGGTCCGGAATTCC
 CGGGATAGTGGMCCGGCGAKGTCGCTCGTGCTGCTAAGCCTGGCCGCGCTGTGCAGGAGC
 GCCGTACCCGAGAGCCGACCGTTCAATGTGGCTCTGAACTGGGCCATCTCCARAGTGG
 ATGSKACAACATGATCTAATCCCGGGAGACTTGAGGGACCTCCGAGTAGAACCTGTTACA

ACTAGTGTTGCAACAGGGGACTATTCAATTTTGATGAATGTAAGCTGGGTACTCCGGGSA
 GATGCCAGCATCCGCTTGTTGAAGGCCACCAAGATTTGTGTGAMGGGCAAAAGCAACWTC
 CAGTCCTACAGCWGTGTGAGGTAGCAATTACACAGAGAGCACATATCCAGACTCTAGACC
 AGACCCCTCTGGWGGTAAATGGACATTTTCTATATCGGCTTCCCTGTAGAGCTGAACACA
 GTCTATATTCAATTGGGGCCCCAWAATAWWCCTAATGCAAAATATGAATGAAGATGGCCCTTC
 CATGTCTGTGAATTTACCTCACCAGGCTGCCTAGACCACATAATGAAATAWAAAAAAAAA
 GTGTGTCAAGGCCGGAAGCCTGTGGGATCCGAACATCACTGCTTGTGAAGAAGAATGARGA
 GACAGTAGAAGTGAACCTTCAACAACCACTCCCCTGGGAAACAGATAMATKGCTCTTATCCA
 ACACARMACTATCATCGGGTTTTCTCAGGTGTTTGAGCCACACCAGAAGAAACAAACGCG
 AGCTTCAGTGGTGATTCCAGTGACTGGGGATAGTGAAGGTGCTACGGGTGCAGCTGACTCC
 ATATTTTCTACTTGTGGCAGCGWCTGCATCCGACATAAAGGAACAGTTGTGCTCTGCCC
 ACAAACAGGCGTCCCTTTYCCTCTGGATAACAACAAAAGCAACYGGGAGSTGGYTYCT

AL525465

WAATWAKADDRATANHTGAAAACTATAACCATTTNTGATAATNGNAANAATAAAGCATCT
 TCAGCCAAACATCTAGTCTTCCATAGACCATGCATTGCAGTGTACCCAGAWCTGTTTAGC
 TAATATTCTATGTTAATTAATGAATACTAAGAACCCCTCACTGATTCACTCAA
 TAGCATCTTAAGTGAAAAACCTTCTATTACATGCAAAAAATCATTGTTTTTAAGATAACA
 AAAGTAGGGAATAAACAAGCTGAACCCACTTTTACTGGACCAAATGATCTATTATATGTG
 TAACCACTTGTATGATTTGGTATTTGCATAAGACCTTCCCTCTACAACTAGATTATAT
 CTGATTCTTGTACAGGTGCCTTTTAAACATGAACAACAAAATACCCACAACTTGTCTAC
 TTTTGCCTAAAGTTACCTATTAGAGGTCACTGTSAGAGTKCTCAGTTTCTTAGTTACTAT
 TTAASTTTTSATGTTCAAAATGAAAATAATTCTKAAGTKGAAAGSGCTCTTGAAGTAACC
 TTTTTATAAATGAGTTATTATAATGGTTTACTTAAATAAAAVAGAGGGGKTTTTGCGGTG
 GCTCATGCCTCCAATCCCAGCACTTTGGCAAGGCCAAGGCAAAAVGATCGCTCAAGACCA
 GGCTACGTACAAAGCGAGACCTCCATCTCTACAAAAGATTTAAAAAATTAGCTGAGTGT
 GATGGTGTGAGCCTGTGGTCCCAGCTACTAGGGAGGCTGAGATGGGAGGATCACTTGAGC
 CCTGGAGGTCAAGGGTGCAGTAAACGGTGATTGTGCCACTGCACTCCATCCTGGGTGAGA
 GCAGACCCTGTCTAAAAACAAACAAACGAAAAAACCCCCACAGAATGACAGAACATAAAAG
 ATGCACATTTTGTCTTCCAACCTTTTTACTCTTCTAAAAGCATCTTTTTTAAATTTTAA
 ATTTTTTTTTTTTTGAGACAGAGTTTCACTCTGTACACAGGCTGGAGTGMGTGGCGTGA
 CTCGGCTCACTAMAACCTCTGCTCCGGGGTYACSCATCTCCTGCWCAGCTCCTGAGAAAGC
 KGGAYAMAGGMCCACACAAACCAGTAAYTTTATWTTTTGAAAAAGGGTTYACCTGTASMA
 GRAGGCTGAATCCGACMAARTMACCMCCACYCAAADGAGGAWAAGKGRSMGGSCEBGGC
 A

BX453536

TTATGGGGGGCAGTAGTGTGGTGGTAGAAAAGGAAGTCTTCTTGATCCTTTTCTGTCCTCC
 CATTAGATAGATCCCTGCCACCAGCACCCATGTGGCCACCAGCAGAGACAGCAGGAGGAG
 AGGCAGCCAGCCTCCCGGCTTGCTTTTGTGTTATCCAGAGGGAAAGGGACGCTGTTTG
 TGGGCAGAGCACAACTGTTCCCTTATGTGCGATGCAGTCGCTGCCACAAGTAGGAAAAATA
 TGGAGTCAGCTGCACCGTAGCACCTTCACTATCCCCAGTCACTGGAATCACCCTGAAGC
 TCGCGTTTGTCTTCTGCTGTGGCTCAAACACCTGAGAAAACCCGATGATAGTGCTGTG
 TTGGATAAGAGCCATGTATCTGTTTCCAGGGGAGTGGTTGTGAAGTTCACTTCTACTGT
 CTCCTCATTCTTCTTACAAGCAGTGATGTTCCGATCCACAGGCTTCCGGCCTTGACACA
 CTNTNTTTTATATTTCAATTATGTGGTCTAGGCAGCCTGGTGAGGTGAAATTACAGACAT
 GGAAGGGCCATCTTCAATCATATTTGCATTAGGAATATTATGGGCCCCAATGAAATAGAC
 TGTGTTTCAGCTCTACAGGGGAAGCCGATATAGGAAAAATGTCCATTTACCACCAGAGGGTC
 TGGTCTGAGTCTTGAAGGCCTTTTGTGTTATTGCACCTTACACAGCTGTTAGACTGGGAA
 GTTGCTTTTGGCCCGCACACAAATCTTGTGGGCCTTCAACAGCGGATGCTGCCATTTGCC
 CCGAAGTCCCCAGCTCAATTCATTAAAAATTGAATAGGCCCTTGTGGCAACCTAGTTG

GTACAGGGTTTTACTTGGGGGGCCCCCTCTAAGTTTCCCGGGATATAAAACAAAGTGTGG

BX453537

TTATGGGGGGCAGTAGTGTGGTGGTAGAAAAGGAAGTCTTCTTGATCCTTTCTGTCCTCC
ACATTAGATAGATCCCTGCCACCAGCACCCATGTGGCCACCAGCAGAGACAGCAGGAGGA
GAGGCAGCCAGCCTCCCGGCTTGCTTTTGTGTGTATCCAGAGGGAAAGGGACGCCTGTTT
GTGGGCAGAGCACAACTGTTCTTTATGTGGATGCAGTCGCTGCCACAAGTAGGAAAAT
ATGGAGTCAGCTGCACCGTAGCACCTTCACTATCCCCAGTCACTGGAATCACCCTGAAG
CTCGCGTTTGTCTTCTGCTGTGGCTCAAACACCTGAGAAAACCCGATGATAGTGTCTGT
GTTGGATAAGAGCCATGTATCTGTTTCCAGGGGAGTGGTGTGAAGTTCACTTCTACTG
TCTCCTCATTCTTCTTACAAGCAGTGATGTTCCGATCCCACAGGCTTCCGGCCTTGACAC
ACTTTTTTTTATATTTTATTATGTGGTCTAGGCAGCCTGGTGAGGTGAAATTCACAGACA
TGGAAGGGCCATCTTCATTCATATTTGCATTAGGAATATTATGGGCCCCAATGAAATAGA
CTGTGTTTCACTCTACAGGGAAGCCGATATAGGAAAATGTCCATTTACCACCAGAGGGTC
TGGTCTGAGTCTGGAAGGCCTCTGTGTAATTGCACCTCACACAGCTGTAGGACTGGGAGT
TGCTTTTGGCCGTACACAAATCTTGTGGCCTTCAACAAGCGGATGCTGGCATCTGGCGG
GGTACCCAGCTTACATTCATCAAAATTGAATAGTCCCCTTGTGTGCAACACTAGTTTGTA
AACAGGTTCTACTCCGGGGGTCCCCTCAGTCTCCCCG

AW28945

CAAATATGAATGAAGATGGCCCTTCCATGTCTGTGAATTTACCTCACCAGGCTGCCTAG
ACCACATAATGAAATATAAAAAAAGTGTGTCAAGGCCGGAAGCCTGTGGGATCCGAACA
TCACTGCTTGTAAGAAGAATGAGGAGACAGTAGAAGTGAACCTCACAACTCCCTGG
GAAACAGATACATGGCTCTTATCCAACACAGCACTATCATCGGGTTTTCTCAGGTGTTTG
AGCCACACCAGAAGAAACAAACGCGAGCTTCAGTGGTGATTCCAGTGACTGGGGATAGTG
AAGGTGCTACGGTGCAACTGACTCCATATTTTCTACTTGTGGCAGCGACTGCATCCGAC
ATAAAGGAACAGTTGTGCTCTGCCCCACAAACAGGCGTCCCTTTCCTCTGGATAACAAC

AV728939

GCAAATATGAATGAAGATGGCCCTTCCATGTCTGTGAATTTACCTCACCAGGCTGCCTA
GACCACATAATGAAATATAAAAAAAGTGTGTCAAGGCCGGAAGCCTGTGGGATCCGAAC
ATCACTGCTTGTAAGAAGAATGAGGAGACAGTAGAAGTGAACCTCACAACTCCCTGT
GGAAACAGATACATGGCTCTTATCCAACACAGCACTATCATCGGGTTTTCTCAGGTGTTT
GAGCCACACCAGAAGAAACAAACGCGAGCTTCAGTGGTGATTCCAGTGACTGGGGATAGT
GAAGGTGCTACGGTGCACTGACTCCATATTTTCTACTTGTGGCAGCGACTGCATCCGA
CATAAAGGAACAGTTGTGCTCTGCCCCACAAACAGGCGTCCCTTTCCTCTGGATAACAAC

AV727345

GCAAATATGAATGAAGATGGCCCTTCCATGTCTGTGAATTTACCTCACCAGGCTGCCTA
GACCACATAATGAAATATAAAAAAAGTGTGTCAAGGCCGGAAGCCTGTGGGATCCGAAC
ATCACTGCTTGTAAGAAGAATGAGGAGACAGTAGAAGTGAACCTCACAACTCCCTGT
GGAAACAGATACATGGCTCTTATCCAACACAGCACTATCATCGGGTTTTCTCAGGTGTTT
GAGCCACACCAGAAGAAACAAACGCGAGCTTCAGTGGTGATTCCAGTGACTGGGGATAGT
GAAGGTGCTACGGTGCACTGACTCCATATTTTCTACTTGTGGCAGCGACTGCATCCGA
CATAAAGGAACAGTTGTGCTCTGCCCCACAAACAGGCGTCCCTTTCCTCTGGATAACAAC
AAAAGCAAGCCGGGAGGCTGGCTGCCTCTCCTCCTGCTGTCTCTGCTGGTGGCCACATGG
GTGCTGGTGGCAGGGATCTATCTAATGTGGAGGCACGAAAGGATCAAGAAGACTTCCTTT
TTTACCACCACACTACTGTCTCCCATTAAGATCTTGTGGTTTATCCATCTGAAATATTG
TTCCATTACACATATTGGTACCTAAGTGAATTTCTTAAACCATTGCAAATTGAGGTCA

CTCTTGAAAGGGCGTG

Secuencias identificadas como las de CHDH

- 5 >gi|26011703|gb|CA774243.1 |CA774243 en 24a07.x1 Páncreas Fetal Humano 1B, ADNc de Homo sapiens clon IMAGE: 3'

TAAATAACAAAÇACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATTCTTA
CTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAAATGACTCTTTCCTGACATAAATCCT
TTTTTATTAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTTATACTTGGG
AGTGCTCCTTGCACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGTAAGTGT
GCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCCAATGGT
AACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGTCAACAT
TATCCTGGACTCTGTGTCTCTCTCTGTTGGGTCTTGTGGCATCACATCAGGCCAAAATTGC
CAGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCTTTTCCAAAGTCAGTACTTACA
AACTGGCATTCCTACAGGCTGCACCATTTCTTAGTATGTCTGCTTAAGCCTGGTTCAACC
TCTCATCGAATATTAAATTTTTCTTTGTA

- 10 >gi|23527150|gb|BU679327.1|BU679327UI-CF-DU1-aau-i-03-0-UI.s1 UI-CF-DU1 ADNc de Homo sapiens clon UI-CF-DU1-aau-i-03-0-UI 3'

TTTTTTTTTTTTTTCTTACAAAGAAAAATTTAATATTCGATNGAGAGGTTGAACCAGGCT
TAAAGCAGACATACTAGGAAATGGTGCAGCCTGTAAGAATGCCAGTTTGTAAAGTACTGA
CTTTGGAAAAGATCATCGCCTCTATCAGACACTTAGGGTCTGGTCTGGCAATTTTGGCC
TGATGTGATGCCACAAGACCCAACAGAGAGAGACACAGAGTCCAGGATAATGTTGACAG
TGGTGTAGCCCTTTAGGAGAAATGGCGCTCCCTGCGGCTGGTATTAGGTTACCATTTGGCA
CCGAAGGAACCAGGAGGATAAGAATATCCATAATTTAGAGCTGCCCTGGCACAGTACC
TGCCCCGTGCGAGGCTCTCACTGGCAAATGACAGCTCTGTGCAAGGAGCACTCCCAAGT
ATAAAAATTATTACACAGTTTTATTCTGAAGAACATTTGCATTTTAATAAAAAAGGATT
TATGTCAGGAAAGAGTCATTTACAAACCTTGAAGTGTTTTTGCCTGGATCAGAGTAAGAA
TGTCTTAAGAAGAGGTTTGTAAAGGTCTTCATAACANAGTGGTGTGTTATTTACAAAAA
AAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAACCTCGTGCCGAATTCT

- 15 >gi|23274053|gb|BU608029.1|BU608029 UI-CF-FN0-aes-l-02-0-UI.s1 UI-CF-FN0 ADNc de Homo sapiens clon UI-CF-FN0-aes-1-02-0-UI 3'

TTTTTTTTTTTTTTTTGTAAATAACAAACACCACTTTGGTTATGAAGACCTTACAAACCTC
TTCTTAAGACATTCTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAAATGACTCTT

TCCTGACATAAATCCTTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAAT
AATTTTTTATACTTGGGAGTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCG
ACAGGGCAGGTACTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTT
CCTTCGGTGCCAATGGTAACCTAATAACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGC
TACACCACTGTCAACATTATCCTGGACTCTGTGTCTCTCTCTGTTGGGTCTTGTGGCATCA
CATCAGGCCAAAATTGCCAGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCTTTTC
CAAAGTCAGTACTTACAACTGGCATTCTTACAGGCTGCACCATTTCCTAGTATGTCTGC
TTTAAGCCTGGTTCAACCTCTCATCGAATATTAAATTTTTCTTTGTAAGAAAAATTTGAAG
TTGTAGAGCATGGTTTTTTGTTTTCCCTTGTCTTAGGAAAGTTTTAAGATGAAATGTTTTT
CC

>gil_11451784|gb|BF439267.1 |BF439267 nab62a07.x1 Soares_NSF_F8_9W_OT_PA_P_S1_ ADNc de Homo sapiens clon IMAGE:3272340 3'

5

TTTTTTTTTTTTTTGGGCCAAAATTTTAAATAGTATACAGACAACCTGTTAATTTTTTTTTT
TTTTTTTTTTGTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTTTAA
GACATTCTTACTCTGATCCAGGCCAAAACACTTCAAGGTTTGTAAATGACTTTTTCTCTGA
CATAAATCCTTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTT
TATACTTGGGAGTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGG
CAGGTACTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCCTCGG
TGCCAATGGTAACCTAATAACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACC
ACTGTCAACATTATCCTGG

>gil11448468|gb|BF436153.1|BF436153 nab77h10.x1 Soares_NSF_F8_9W_OT_PA_P_S1 ADNc de Homo sapiens clon IMAGE:3273859 3'

TTTTTTTTTGGTCCAAAATTTTAAATAGTATACAGACAACCTGTTAATTTTTTTTTTTTTT
TTTTGTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACAT
TCTTACTCTGATCCAGGCCAAAACACTTCAAGGTTTGTAAATGACTCTTCTCTGACATAA
ATCCTTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTTATAC
TTGGGAGTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGT
ACTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCCTTCGGTGCCA
ATGGTAACCTAATAACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGTC
AACATTATCCTGGACTC

10

>gil10033200|gb|BE672659.1|BE672659 7b71b03.x1NCI_CGAP_Lu24 ADNc de Homo sapiens clon IMAGE: 3233645 3'

TTTGTAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATT
CTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCTGACATAAA
TCCTTTTTTATTAAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTTATACT
TGGGAGTGCTCCTTGCACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACAGGGCAGGT
ACTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCCA
ATGGTAACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGTC
AACATTATCCTGGACTCTGTGTCTCTCTCTGTTGGGTCTTGTGGCATCACATCAGGCCAAA
ATTGCCAGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCTTTTCCAAAGTCAGTAC
TTACAAACTGGCATTCTTACAGGCTGCACCATTTCTAGTATGTCTGCTTTAAGCCTGGTT
CAACC

>gi|6837905|gb|AW341279.1|AW341279 xz97e03.x1NCI_CGAP_Lu24 ADNc de Homo sapiens clon IMAGE:
2872156 3'

TAAATAACTAACACCACTTTTGTATGAAGACCTTACAAACCTCTTCTTAAGACATTCTTAC
TCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCTGACATAAAATCCTT
TTTTTATTAAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTTATACTTGGG
AGTGCTCCTTGCACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGTACTGT
GCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCCAATGGT
AACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGTCAACAT
TATCCTGGACTCTGTGTCTCTCTCTGTTGGGTCTTGTGGCATCACATCAGGCCAAAATTGC
CAGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCTTTTCCAAAGTCAGTACTTACA
AACTGGCATTCTTACAGGCTGCACCATTTCTAGTATGTCTG

5

>gi|5543775|gb|AI869807.1|AI869807wm04b02.x1 NCI_CGAP_Ut4 ADNc de Homo sapiens clon IMAGE:24349233'

GTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATTCTT
ACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCTGACATAAAATCC
TTTTTTATTAAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTTATACTTGG
GAGTGCTCCTTGCACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGTACTGT
GCCAGGGCAGCTCTGAAATATGGATATTCTTACCTCCTGGTTCCTTCGGTGCAAATGGTA
ACCTAATACCAGCCGCAGGGAGCGCCATTTCT

10 >gi|5543768|gb|AI869800.1|AI869800wm04a03.x1 NCI_CGAP_Ut4 ADNc de Homo sapiens clon IMAGE:24349243'

GTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATTCTT
ACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCTGACATAAAATCC
TTTTTTATTAAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTTATACTTGG
GAGTGCTCCTTGCACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGNGCAGGTACTG
TGCCAGGGCAGCTCTGAATTATGGATATTCTTATCCTCCTG

>gi|4610281|gb|AI601252.1|AI601252 ar88c09.x1 colon Barstead HPLRB7 ADNc de Homo sapiens clon IMAGE: 2152336 3'

TTTGTAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATT
CTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCTGACATAAA
TCCTTTTTTATTAAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTTATACT
TGGGAGTGCTCCTTGCACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGT
ACTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCCA
ATGGTAACCTAATAACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGTC
AACATTATCCTGGACTCTGTGTCTCTCTGTTGGGTCTTGTGGCATCACATCAGGCCAAA
ATTGCCAGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCTTTTCCAAAGTCAGTAC
TTACAAACT

5
>gi|4330088|gb|AI467998.1|AI467998tj84e10.x1 Soares_NSF_F8_9W_OT_PA_P_S1 ADNc de Homo sapiens clon IMAGE:2148234 3'

TTTTTTTTTTTTTTTTTGGTCCAAAATTTTAATAGTATACAGACAACCTGTTAATTTTTTT
TTTTTTTTTTTTTTGTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTT
AAGACATTCTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCT
GACATAAATCCTTTTTTATTAAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAAT
TTTTATACTTGGGAGTGCTCCTTGCACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACG
GGGCAGGTACTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCT
TCGGTGCCAATGGTAACCTAATAACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTAC
ACCACTGTCAACATTATCC

10
>gi|4311745|gb|AI459166.1|AI459166 tj65h07.x1 Soares_NSF_F8_9W_OT_PA_P_S1 ADNc de Homo sapiens clon IMAGE:2146429 3'

TTTTTTTTTGGTCCAAAATTTTAATAGTATACAGACAACCTGTTAATTTTTTTTTTTTTT
TTTTTGTAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACAT
TCTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCTGACATAA
ATCCTTTTTTATTAAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTTATAC

TTGGGAGTGCTCCTTGCACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGT
ACTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCCA
ATGGTAACCTAATAACCAGCCGCAGGGAGCGCCATT
T

15
>gi|4222874|gb|AI393327.1|AI393327 tg44a12.x1 Soares_NFL_T_GBC_S1 ADNc de Homo sapiens clon IMAGE: 21116143'

TTTATATTATTCACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATTCTTAC
TCTGATCCAGGCCAAAAACACTTCAAGGTTTGTAAATGACTCTTTCCTGACATAAATCCTT
TTTTATTAAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTTATACTTGGGA
GTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACAGGGCAGGTACTGTG
CCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCCAATGGTA
ACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGTCAACATT
ATCCTGGACTCTGTGTCTCTCTCTGTTGGGTCTTGTGGCATCACATCAGGCCAAAATTGCC
AGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCTTTT

>gi|3933768|gb|AI290994.1|AI290994qm09e05.x1 NCI_CGAP_Lu5 ADNc de Homo sapiens clon IMAGE:18813443'

TAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATTCTTA
CTCTGATCCAGGCCAAAAACACTTCAAGGTTTGTAAATGACTCTTTCCTGACATAAATCCT
TTTTATTAAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTTATACTTGGG
AGTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGTACTGT
GCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCCAATGGT
AACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGTCAACAT
TATCCTGGACTCTGTGTCTCTCTCTGTTGGGTCTTGTGGCATCACATCAGGCCAAAATTGC
CAGACCAGGACCCTAAGTGTCTGATAGA

5

>gi|3844395|gb|AI248998.1|AI248998 qh80f04.x1 Soares_fetal_higado_bazo_1NFLS_S1 ADNc de Homo sapiens clon IMAGE:1851007 3'

TGTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATTCT
TACTCTGATCCAGGCCAAAAACACTTCAAGGTTTGTAAATGACTCTTTCCTGACATAAATC
CTTTTTATTAAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTTATACTTG
GGAGTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGTACT
GTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCCAATG
GTAACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGTCAAC

ATTATCCTGGACTCTGTGTCTCTCTCTGTTGGGTCTTGTGGCATCACATCAGGCCAAAATT
GCCAGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCTTTTCCAAAGTCAGTACTTA
CAAACCTGGCATTCTTACAG

10

>gi|3405022|gb|AI075844.1|AI075844 oz16d08.x1 Soares_fetal_higado_bazo_1NFLS_S1 ADNc de Homo sapiens clon IMAGE:1675503 3'

TTTTTTTTTTGTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTA
AGACATTCTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCTG
ACATAAATCCTTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAAACTGTGTAATAATTT
TTATACTTGGGAGTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGG
GCAGGTACTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCG
GTGCCAATGGTAACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACAC
CACTGTCAACATTATCCTGGACTCTGTGTCTCTCTCTGTGTTGGGTCTTG

>gi|3307963|gb|AI051972.1|AI051972 ow83h10.x1 Soares_fetal_higado_bazo_1NFLS_S1 ADNc de Homo sapiens
clon IMAGE:1653475 3'

5

TTTTTTTTTTGTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAG
ACATTCTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCTGAC
ATAAATCCTTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAAACTGTGTAATAATTTTT
ATACTTGGGAGTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACAGGGC
AGGTACTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGT
GCCAATGGTAACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACCA
CTGTCAACATTATCCTGGACTCTGTGTCTCTCTCTGTGAGTCTTGTGGCATCACATCAGG
CCAAAATTGCCAGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCTT

>gi|3254601|gb|AI033648.1|AI033648ow22e09.x1 Soares_tumor_paratiroideo_NbHPA ADNc de Homo sapiens clon
IMAGE: 1647592 3'

10

TGTAAATAACAAACACCACTTGGTTATGAAGACCTTACAAACCTCTTCTTAAGACATTCT
TACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCTGACATAAATC
CTTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAAACTGTGTAATAATTTTTATACTTG
GGAGTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGTACT
GTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCCAATG
GTAACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGTCAAC
ATTATCCTGGACTC

>gi|3229367|gb|AI015031.1|AI015031 ot30f04.s1 Soares_testículo_NHT ADNc de Homo sapiens clon
IMAGE:1618303 3'

TTTTTCTTACAAAGAAAAATTTAATATTCGATGAGAGGTTGAACCAGGCTTAAAGCAGAC
ATACTAGGAAATGGTGCAGCCTGTAAGAATGCCAGTTTGTAAAGTACTGACTTTGGAAAA
GATCATCGCCTCTATCAGACACTTAGGGTCCTGGTCTGGCAATTTTGGCCTGATGTGATG
CCACAAGACCCAACAGAGAGAGACACAGAGTCCAGGATAATGTTGACAGTGGTGTAGCC
CTTTAGGAGAAATGGCGCTCCCTGCGGCTGGTATTAGGTTACCATTGGCACCGAAGAGAC
CAGGAGGATAAGAATATCCATAATTTAGAGCTGCCCTGGCACAGTACCTGCCCCGTCG
GAGGCTCTCACTGGCAAATGACAGCTCTGTGCAAGGAGCACTCCCAAGTATAAAAATTA
TTACACAGTTTTATTCTG

>gi|3055333|gb|M915941.1 |M915941 on18d06.s1 NCI_CGAP_Lu5 ADNc de Homo sapiens con IMAGE: 155703 3'

TTGTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATT
TTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAAATGACTCTTTCCTGACATAAAT
CCTTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTATACTT
GGGAGTGCTCCTTGCACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGAAGGGGCAGGTA
CTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCCAA
TGGTAACCTAATAACCAGCCGCAGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGTCAA
CATTATCCTGGACTCTGTGTCTCTCTCTGTTGGGTCTTGTGGCATCACATCAGGCCAAAAT
TGCCAGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCTTTTCCAAAGTCAGTACTT
A

5

>gi|2953300|gb|AA861160.1|AA861160 ak36b12.s1Soares_testículo_NHT ADNc de Homo sapiens clon IMAGE: 1408031 3'

TTGTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATT
CTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAAATGACTCTTTCCTGACATAAA
TCCTTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTATACT
TGGGAGTGCTCCTTGCACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGAC

10

>gi| 1782040|gb| AA192157.1 |AA192157 zq02g05.s1 músculo Stratagene 937209 ADNc de Homo sapiens clon MAGE: 628568 3'

GTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATTCTT
ACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAAATGACTCTTTCCTGACATAAATCC
TTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTATACTTGG

GAGTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGTACTG
TGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCCAATGG
TAACCTAATACCAGCCGCAGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGTCAACAT
TATCCTGGGACTCTGTGTCTCTCTCTGTTGGGTCTTGTGGCATCACATCAGGCCAAAATTG
GCCAGACCAGGACCCCAAGTGGTCTGATAGAAGGCGATGATCTTTTCCAAAGTCAGTAC
TTACA

>gi|1230953|gb|N73668.1|N73668 yz78h05.s1 Soares_esclerosis_múltiple_2NbHMSP ADNc de Homo sapiens clon
IMAGE:289209 3'

5

GGTCCAAAATTTTAAATAGTATACAGACAACCTGTTAATTTTTTTTTTTTTTTTTTTGTAA
ATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATTCTTACTCT
GATCCAGGCAAAAACACTTCAAGGTTTGTAAATCGACTCTTTCCTGACATAAATCCTTTT
TTATTAATAATNGCAAAATTGTTCTTCAGAATAAACTGTGTAATAATTTTTATACTTGGG
AGTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGTACTGT
GCCAGGGCAGCTCTGAAATTATGGAAATTCTTATCCCCCTGGTTCCTNCGGTGGCCAATG
GGTAACCTAATACCAGCCCGCGGAAGCGCCAATTTCNCCCAAAGGGGGTAAACCACT
GGTNAAACATTA

>gi|2142013|gb|AA437099.1|AA437099 zv53b09.s1 Soares_testículos_NHT ADNc de Homo sapiens clon IMAGE:
757337 3'

10

CTTACAAAGAAAAATTTAATATTCGATGAGAGGTTGAACCAGGCTTAAAGCAGACATAC
TAGGAAATGGTGCAGCCTGTAAGAATGCCAGTTTGTAAAGTACTGACTTTGGAAAAGATC
ATCGCCTCTATCAGACACTTAGGGTCTGGTCTGGCAATTTTGGCCTGATGTGATGCCAC
AAGACCCAACAGAGAGAGACACAGAGTCCAGGATAATGTTGACAGTGGTGTAGCCCTTT
AGGAGAAATGGCGCTCCCTGCGGCTGGTATTAGGTTACCATTGGCACCGAAGAGACCAG
GAGGATAAGAATATCCATAATTTAGAGCTGCCCTGGCACAGTACCTGCCCCGTCGGAG
GCTCTCACTGGCAAATGACAGCTCTGTGCAAGGAGCACTCCCAAGTATAAAAATTAT

>gi|3836330|gb|AI240933.1 |AI240933 qk01 b11 .x1 NCI_CGAP_Kid3 ADNc de Homo sapiens con IMAGE: 1867677
3'

15

TTTTTTTTTTTTTTTTTTGGTCCAAAATTTTAAATAGTATACAGACAACCTGTTAATTTT
TTTTTTTTTTTTTTTTTGGAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCT
TCTTAAGACATTCTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGGAAATGACTCTT
TCCTGACATAAATCCTTTTTTATTAATAATGCAAAAGGTTCTTCAGAATAAACTGTGTAA
TAATTTTTATACTTGGGAGTGCTCCTTGACAGAGCTGTCATTTGCCAG

>gi|2669971|gb|AA682690.1|AA682690 zj86f07.s1 Soares_fetal_hígado_bazo_1NFLS_S1 ADNc de Homo sapiens
clon IMAGE:461797 3'

AATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTATACTTGGGATGTGCTCC
TTGCACAGAGCTGTCATTTGCCAGTGAGAGCCTCGACAGGCAGGTACTGTGCCAGGGCA
GCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCTGTGCTCAATGGTAACCTAAT
ACCAGCCGCAGGACNCGCCATTTCTCCTAAAGGGCTACACCACTGTCAACATTATC

>gi|1162804|gb|N39597.1|N39597 yy51e04.s1 Soares_esclerosis_múltiple_2NbHMSP ADNc de Homo sapiens clon
IMAGE:277086 3'

5

ACAAAGAAAAATTTAATATTCGATGAGAGGTTGAACCAGGCTTAAAGCAGACATACTAG
GAAATGGTGCAGCCTGTAAGAATGCCAGTTTGTAAAGTACTGACTTTGGAAAAGATCATC
GCCTCTATCAGACACTTAGGGTCCTGGTCTGGCAATTTTGGCCTGATGTGATGCCACAAG
ACCCAACAGAGAGAGACACAGAGTCCAGGATAATGTTGACAGTGGTGTAGCCCTTTAGG
AGAAATGGCGCTCCCTGCGGCTGGTATTAGGTTACCATTGGCACCGAAGAACCAGGAGG
ATAAGAATATCCATAATTTTCAGAGCTTGGCCTGGCACAGTACCTGCCCCGTCGGAGGCTC
TCACTGGGCAAATGGACAGCTCTGTGCAAGGAGCACTCCCAAGTATAANAATTATTACA
CAGTTTTATTCTGAAGAACATTTTGCATTTTAATAAAAAANGGA

>gi|3838507|gb|AI243110.1|AI243110 qh26f06.x1 Soares_NFL_T_GBC_S1 ADNc de Homo sapiens clon IMAGE:
1845827 3'

10

TTTTCTTACAAAGAAAAATTTAATATTCGATGAGAGGTTGAACCAGGCTTAAAGCAGAC
ATACTAGGAAATGGTGCAGCCTGTAAGAATGCCAGTTTGTAAAGTACTGACTTTGGAAAA
GATCATCGCCTCTATCAGACACTTAGGGTCCTGGTCTGGCAATTTTGGCCTGATGTGATG
CCACAAGACCCAACAGAGAGAGACACAGAGTCCAGGATAATGTTGACAGTGGTGTAGCC
CTTTAGGAGAAATGGCGCTCCCTGCGGCTGGTATTAGGTTACCATTGGCACCGAAGGAAC
CAGGAGGATAAGAATATCCATAATTTTCAGAGCTGCCCTGGCACAGTACCTGCCCCGTG
GAGGCTCTCACTGGCAAATGACAGCTCTGTGCAAGGAGCACTC

Secuencias identificadas como las del grupo QPRT

15

BC005060

```

1 caccagccca gacagctgca agtcaccatg gacgctgaag gcctggcgct gctgctgccg
61 cccgtcaccc tggcagccct ggtggacagc tggctccgag aggactgccc agggctcaac
121 tacgcagcct tggtcagcgg ggcaggcccc tcgcaggcgg cgtgtgtggc caaatccctt
181 gggatactgg cagggcagcc tttcttcgat gccatattta cccaactcaa ctgccaaagtc
241 tccgtggttc tcccagagg atcgaagctg gtgcccgttg ccagagtggc cgaggtccgg
301 ggccctgccc actgctgct gctgggggaa cgggtggccc tcaacacgct ggcccgtgct
361 agtggcattg ccagtgtgc cgcgctgca gtggaggcgg ccaggggggc cggctggact
421 gggcacgtgg caggcacgag gaagaccacg ccaggcttcc ggctggtgga gaagtatggg
481 ctccgtggtg gcggggccgc ctgcacgcgc tacgacctgg gagggctggt gatggtgaag
541 gataaccatg tggtgccgc cggtgccgtg gagaaggcgg tgcgggcggc cagacaggcg
601 gctgacttcg ctctgaaggt ggaagtggaa tgcagcagcc tgcaggaggc cgtgcaggca
661 gctgaggctg gtgcgacct tgtcctgctg gacaacttca agccagagga gctgcacccc
721 acggccaccg tgcgaaggc ccagttcccc agtgtggttg tggagccag tgggggcatc
781 accctggaca acctcccca gttctgcggg ccgcacatag acgtcatctc catggggatg
841 ctgaccagg cgccccagc ccttgatttc tccctcaagc tgtttgccaa agaggtggct
901 ccagtgccta aaatccacta gtccataacc ggaagaggat gacaccggcc atgggttaac
961 gtggctcctc aggacctct gggtcacaca tctttagggt cagtggccaa tggggcacat
1021 ttggcactag cttgagcca actctggctc tgccacctgc tgcctctgtg acctgtcagg
1081 gctgacttca cctctgtcga tctcagtttc ctaatctgta aaatgggtct aataaaggat
1141 caaccacatg gggttctgcg gtgataatga gcacatagtg aggggtcagc aaatgtcaga
1201 agttacctgg gacagccggg cagcatggct cacacctgta atcccagcac tttgggaggc
1261 tgaggcggga agatcacttg agttcaggag tttgagacca gcctggccaa catggtgaaa
1321 ccccatctct accaaaaata gaagaattag ctgggtgttg tggcacgcgc ctgtaatccc
1381 agctacttag gaggctgagg caggagaatc gcttgaacct aggaagtgga ggttgcagtg
1441 agctgatggt gccactgcac tcagcctgg gtgatagagc gagactctgt ctccaaagaa
1501 gaaacaaaaa aaaaaaaaaa aaaa

```

BE299670

```

1 gtggccacca ctacgtcga tgcaacctac acttccctggg ttcaagcgat tctcctgcct
61 cagcctccta agtagctggg attacaggcg cgtgccacca caccagcta attcttctat
121 ttttggtaga gatggggttt caccatgttg gccaggctgg tctcaaactc ctgaactcaa
181 gtgatcttcc cgctcagcc tccaaagtgt ctgggattac aggtgtgagc catcgtgccc
241 ggctgtccca ggtaacttct gacatttgcg gacctcac tatgtgctca ttatccgcgc
301 agaaccatgt gtggttgatc ctttattaga cccattttac agattaggaa actgagatga
361 gcagaggtga agtcagccct gacaggtcac aggagcagca ggtggcagag ccagagtggg
421 gctcaagcta gtgccaatg tgccccattg gccactgacc ctaaaagatg tgtgaccagc
481 agggctcctg ggagccacgt taacctatgg ccggtgtcat cctcttccgg tttaggacta
541 gtggattttc gggcactgga gccacctctt tggcaacta gctttgaggg agaaatcaag
601 ggctggggcc gctgggtcag catcccatg gagatgacgt ctatgtgcgg ccgccgaact
661 gggggagggt gtccagggtt atgcccact ggcttcccag ccacactcgg gaaccggggc
721 ttccgcacgg tggccgttgg ggtgcagctc ctctggcttg aagttgtcca gcaggccaag
781 gtcggcacgc tcagctgctg cagctctgct agctgcgtcc cactccactt cgagcaaccg
841 cgctgtttgc cgccaacgct atcaggcacg gggcacctgt atcttacctc ccgccccgct
901 tagggcgggg cccacaggca ttcacgcgac tggtaacctg agccgacgcg cggcaggaga
961 caggcgtgcc acgaggccac cgtgcagtt tcgtcccggt cctc

```

BE273102

```

1 atggacgctg aaggcctggc tgctgctgct gcctgccctg tcacccttgc agcctggtg
61 gacagcgtgg ctccgagagg actgccagt ggctcaacta cgcagccttg gtcagcgggg
121 caggcccttc cgcagtgcgg cgcgttgttg ggcaagatcc cctgggtgta ctggcagtgg

```



```

181 cagccggtttc gttcgaatgcc gatattttacc cagactcaac ttgccaaggt ctccttggtt
241 ccgtccctcg agggcgatcg aagcttggtt gcctgggttg gccagagttg gcctgagtgt
301 cctgtgtggc cgctgccgca cttgccttgc ttggctgggg ggaacggggt tggccgcgtc
361 aagcactgct tgtgccctgg ctgcatgttg tgcatttgcc catgtgcttg cctggcctgc
421 ttgcatgttg tgagtgcctg cccagtgttg tgcctgtgct gtgagctggt gcacgtgtgt
481 gcagtggcga ctgagggaat gacccactgc ccgaggggct tcctggcctg tgttggaacg
541 aacgtatttg ggcgtcctgt gtgggggggg gcgcgcgctc tgcacgggct tacggacct
601 ggggtgagggc tgcgtgcctg cgtgaagggg attaacgcct gttgttttgg cccgcctggt
661 ggcgctgtgt gacagtgcgg tgcctggcctg tgccaacatg gcgtgctgac gttcgggtc

```

BC010033

```

1 cagacagctg caagtcacca tggacgctga aggcctggcg ctgctgctgc cgcccgctac
61 cctggcagcc ctgggtggaca gctggctccg agaggactgc ccagggtcca actacgcagc
121 cttgggtcagc ggggcaggcc cctcgagggc ggcgctgttg gccaaatccc ctgggggtact
181 ggcagggcag cctttcttcg atgccatatt taccacaactc aactgccaaag tctcctggtt
241 cctccccgag ggatcgaagc tgggtgccgtt ggccagagtg gccgaggtcc ggggcccctgc
301 ccaactgcctg ctgctggggg aacgggtggc cctcaacacg ctggcccgtc gcagtggcat
361 tgccagtgtt gccgcgctg cagtggaggc cgcagggggg gccggttga ctgggcacgt
421 ggcagggcac aggaagacca cgcagggctt cggctgtgtg gagaagtatg ggctcctggt
481 gggcgggggc gcctcgacc gctaogacct gggagggctg gtgatggtga aggataacca
541 tgtggtggcc gccggtggcg tggagaaggc ggtgcggcg gccagacagg cggctgactt
601 cgctctgaag gtggaagtgg aatgcagcag cctgcaggag gccgtgcagg cagctgaggc
661 tgggtgccgac cttgtcctgc tggacaactt caagccagag gagctgcacc ccacggccac
721 cgtgctgaag gccagttcc cagtgtgtggc tgtggaagcc agtgggggca tcaccctgga
781 caacctcccc cagttctgcg ggccgcacat agacgtcatc tccatgggga tgctgacca
841 ggcggcccca gcccttgatt tctccctcaa gctgttttgc aaagaggttg ctcagtgcc
901 caaaatccac tagtcctaaa ccggaagagg atgacaccgg ccatgggtta acgtggctcc
961 tcaggacctt ctgggtcaca catotttagg gtcagtggcc aatggggcac atttggtact
1021 agcttgagcc caactctggc tctgccacct gctgctcctg tgacctgtca gggctgactt
1081 cacctctgct catctcagtt tcctaactctg taaaatgggt ctaataaagg atcaaccaca
1141 tgggggttctg cgggtgataat gagcacatag tgaggggtca gcaaatagg gaagttacct
1201 gggcagccg ggcacgatgg ctcacacctg taatcccagc actttgggag cgtgaggcgg
1261 gaagatcact tgagttcagg agtttgagac cagcctggcc aacatggtga aaccccatct
1321 ctacaaaaaa tagaagaatt agctgggtgt ggtggcacgc gcctgtaatc ccagctactt
1381 aggaggctga ggcaggagaa tcgcttgaac ccaggaagtg gaggttgagc tgagctgatg
1441 gtgccactgc actccagcct ggggtgataga gcgagactct gtctccaaag aagaaaaaaa
1501 aaaaaaaaaa a

```

BC018910

```

1 gtcttgagca gccaacacac cagcccagac agctgcaagt caccatggac gctgaaggcc
61 tggcgctgct gctgccgcc gtcacctggt cagccctggt ggacagctgg ctccgagagg
121 actgcccagg gctcaactac gcagccttgg tcagcggggc aggccctcg caggcggcgc
181 tgtgggcaa atcccctggg gtactggcag ggcagccttt cttcgatgcc atattttacc
241 aactcaactg ccaagtctcc tggttcctcc ccgagggatc gaagctggtg ccggtggcca
301 gagtggccga ggtccggggc cctgccact gcctgctgct gggggaacgg gtggccctca
361 acacgctggc ccgctgcagt ggcattgcca gtgctgcgcg cgctgcagtg gagccgcca
421 gggggggcgg ctggactggg cacgtggcag gcacgaggaa gaccacgcca ggcttcggc
481 tgggtggagaa gtatgggctc ctgggtggcg gggccgcctc gcaccgctac gacctgggag
541 ggctggtgat ggtgaaggat aaccatgttg tggccgcggg tggcgtggag aaggcgggtg
601 gggcgggcag acaggcggct gacttcgctc tgaagggtga agtggaatgc agcagcctgc
661 aggaggcgtg gcaggcagct gaggtggtg ccgaccttgt cctgctggac aacttcaagc
721 cagaggagct gcacccacg gccacogtgc tgaaggccca gttcccagat gtggctgtgg
781 aagccagtgg gggcatcacc ctggacaacc tccccagtt ctgcgggcgg cacatagacg
841 tcctctccat ggggatgctg acccaggcgg cccagccctt tgatttctcc ctcaagctgt

```

```

901 ttgccaaga ggtggctcca gtgccccaaa tccactagtc ctaaaccgga agaggatgac
961 accggccatg ggttaacgtg gctcctcagg accctctggg tcacacatct ttagggtcag
1021 tggccaatgg ggcacatttg gcactagctt gagcccaact ctggctctgc cactgtctgc
1081 tcccttgacc tgtcagggtc gacttcacct ctgctcatct cagtttcta atctgtaaaa
1141 tgggtctaataaaagatcaa ccacaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
1201 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

```

BE902622

```

1 agcccagaca gctgcaagtc acatggacgc tgaaggcctg gcgctgctgc tgcgcgccgt
61 caccctggca gccctgggtg acagctggct ccgagaggac tgcccagggc tcaactacgc
121 agccttggtc agcggggcag gccctcgca ggcggcgtg tgggccaaat cccctggggt
181 actggcaggc cagcctttct tcgatgccat atttacccaa ctcaactgcc aagtctcctg
241 gttcctcccc gagggatcga agctgggtgc ggtggccaga gtggccgagg tccggggccc
301 tgccactgc ctgctgctgg gggaacgggt ggccctcaac acgctggccg ctgcagtggc
361 attgccagtg ctgccgccgc tgcagtggag gccgcagggg ggccggctgg actgggcacg
421 tggcaggcac gaggaagacc acgccaggct tccggctggg ggagaagtat gggctcctgg
481 tgggcgggccc gctcgcaccg tacgacctgg gagggctggg gatggtgaag gataaccatg
541 tggtgccgccc cgggtggcgtg gagaaggcgg tgcggggcgg agacaggcgg tgacttcgct
601 ctgaagggtg aagtggaaatg cagcagcctg caggaggccg tgcaggcagt gaggtcgtg
661 cgacttgctc tgctggacaa ttcagccaga ggagctgcac ccagggcacg tgctgaaggc
721 cagtcogagt gtggtgtgga accagggtgg ggatcacctg gaaacctccc caatctgcgg

```

N23182

```

1 gtggttgatc ctttattaga ccattttac agattaggaa actgagatga gcagaggtga
61 agtcagccct gacaggtcac aggagcagca ggtggcagag cagaggttgg gctcaagcta
121 gtgccaatg tgcccatg gccactgacc ctaaagatgt gtgaccaga ggncttgagg
181 agccacgtta acccatggcc ggtgtcatcc tottcgggt taggactagt ggattttggg
241 cactggagcc acctctttgg caaacagctt gaaggagaa atcaagggtc ggggcccct
301 gggtcagcat ccccatggag atgacgtcta tntgcggccc gcagaactgg gggaggttgt
361 ccagggtgat gccccactg gcttcacag ccacactcgg gaactgggccc ttcagcacgg
421 tggccgtg

```

N32648

```

1 agccaacaca ccannccaga cagctgcaag tcaccatgga cgctgaaggc tggcgctgct
61 gctgcgcgcc gtcaccctgg cagccctggg ggacagctgg ctccgagaga ctgccagggg
121 ctcaactacg cagccttggg cagcggggca ggccctcgca ggccggcgtg tgggccaaat
181 cccctggggt actggcaggc cagcctttct tcgatgccat atttacccaa ctcaactgcc
241 aagtctcctg gttcctcccc gagggatcga agctgggtgc ggtgncagag tggcgaaggc
301 ccggggccct gccaaactgcc tgctgctggg ggaacgggtg gccctcaaca cgctggcccc
361 ctgcagtggc attncagtgc tgcgcgcgt g

```

BE735342

```

1 tagtcaggct ggggtgttaat cggcaccgag ggcaagtcac catggacgct gaaggctggc
61 gctgctgctg ccgcccgcac cctggcagcc ctggtggaca gctggctccg agaggactgc
121 ccaggctcaa ctacgcagcc ttggtcagcg gggcaggccc ctgcagcggc cgctgtgagg
181 ccaaatcccc tggggtaactg gcagggcagc ctttcttcga tgccatattt acccaactca
241 actgccaaat ctctgggttc ctccccgagg gatcgaagct ggtgccggtg gccagagtgg
301 ccgaggtccg gggccgcctc gcaccgctac gaactgggag ggtggtgat ggtgaaggat
361 aacctgtgg tggccgcccg tggcgtggag aaggcgggtg gggcggccag acaggcggct
421 gacttcgctc tgaagggtgga agtggaaatgc agcagcctgc aggaggcgt gcaggcagct
481 gaggtggtg ccgaccttgt cctgctggac aacttcaagc cagaggagct gcacccacg
541 gccaccgtgc tgaaggccca gttcccagct gtggctgtgg aagccagtgg gggcatcacc

```

ES 2 399 246 T3

```

601 ctggacaact cccccagttc tggggggccgc acatagacgt catctccatg gggatgctga
661 cccaggcggc ccagcccttg atttctccct caagctgttt tgccacagag gtggtccag
721 tgcccaaaaa tctactagtc taaaccggaa gaggatgaca ccggcatggg taacgtggct
781 cctcagaccc tctgggtcac acatcttagg gtccaggcca atgggcccac ttggccctac
841 ttgagcccaa tctgggtctgc actgctgtct gtgaccgtca gggtagcttc acttctgtct
901 ctctcagttc taacctgtaa cggctctcta agcgtoaccc caaaaaaaaa aaactcggtc
961 ttttcgtcgg cgaaaggctg aaggctgtcc gtccgcacct g

```

BX283118

```

1 gacagctgca agtcaccatg gacgctgaag gcctggcgct gctgctgccc cccgtcacc
61 tggcagccct ggtggacagc tggctccgag aggaactgcc agggctcaac taacgacct
121 tggtcagcgg ggcaggcccc tcgcaggcgg cgctgtgggc caaatccct ggggtactgg
181 cagggcagcc tttcttcgat gccatattta cccaactcaa ctgccaaagtc tctgtgttc
241 tccccgaggg atcgaagctg gtgccggtgg ccagagtggc cgaggtccgg ggccctgcc
301 actgcctgct gctgggggaa cgggtggccc tcaacacgct ggcccgtgc agtggcantg
361 ccagtgtgct cgcgctgca gtggaggcgg ccaggggggc cggctggact gggcacgtgg
421 caggcacgag gaagaccacg ccaggcttcc ggctgggtga gaagtatggg ctctgtgtg
481 gcgggggcgc ctgcaccgc tacgacctgg gaggg

```

BG769505

```

1 ggacgctgaa ggccctggcg tggctgtgcc gcccgtcacc ctggcagccc tgggtggacag
61 ctggctccga gaggactgcc cagggtcaa ctacgcagcc ttggtcagcg gggcaggccc
121 ctgcaggcg gcgctgtggg ccaaatcccc tgggtactg gcagggcagc cttctcttga
181 tgccatattt acccaactca actgccagt ctctgtgttc ctcccagag gatcgaagct
241 ggtgccgggt gccagagtgg ccgaggtccg gggccctgcc cactgcctgc tgtggggga
301 acgggtggcc ctcaacacgc tggcccgtg cagtggcatt gccagtgtg ccgcgctgc
361 agtggaggcc gccagggggg ccgctggac tgggcacgtg gcaggcacga ggaagaccac
421 gccaggcttc cggctgggtg agaagtatgg gctcctgggt ggccggggcc cctcgcaccg
481 ntacgacctg ggagggtgtg tgatggtgaa ggataaccat gtggtggccg ccggtggcgt
541 ggagaaggcg gtgcgggcgg cagacaggcg gctgacttcg ctctgaacgg tggcaagtgg
601 actgcagcag cctgcagcga ggccgtgcag gcagctgagg ctggtgccga ccttgtctg
661 ctggacaact tcaagccaga ggagctgcaa cccacggcc acgtgctgaa ggccagttcc
721 agagtgtggc tgtggacgcc agtggcgcca tcacctggac aacctcccc agttctgagg
781 gccgcacata gaagtcactc tccatggcga tgctgaccca ggccggccca gccctggatt
841 tctctcagc tggttgcaca gaggtgctc agtgcccaa tcaactcatc aacc

```

BG750434

```

1 ggacagctgc aagtcaccat ggacgctgaa ggccctggcg tgctgctgcc gcccgtcacc
61 ctggcagccc tgggtggacag ctggctccga gaggactgcc cagggtcaa ctacgcagcc
121 ttggtcagcg gggcaggccc ctgcaggcgg gcgctgtggg ccaaatcccc tgggtactg
181 gcagggcagc cttctcttga tgccatattt acccaactca actgccagt ctctgtgttc
241 ctcccagagg gatcgaagct ggtgccggtg gccagagtgg ccgaggtccg gggccctgcc
301 cactgcctgc tgetggggga acgggtggcc ctcaacacgc tggcccgtg cagtggcatt
361 gccagtgtg ccgcccgtgc agtggaggcc gccagggggg ccgctggac tgggcacgtg
421 gcaggcacga ggaagaccac gccaggcttc cggctgggtg agaagtatgg gctcctgggt
481 ggccggggcg cctcgcaccg ctacgacctg ggagggtgtg tgatggtgaa cggataacca
541 tgtggtggcc gccggtggcg tggagaaggc ggtgcgggcg gccagaccag gccgctgact
601 tgcctctgaa ggtggaagtg gaatgcagca gcctgcagga ggccgtgcag gcagcttgag
661 gctggtgccg accttgtctc gcttgacaa cttcaaagcc agggaggact ggaccacagg
721 ccaccgtgct gaaggccagt tcccagtggt ggctgtggaa gccagtgggg gcacagcct
781 ggacaagctc cccccagttc ctggaggggc cgcacattaa accgtcatct ccatggggga
841 ttgtgaacc aaggcggacc cagcccttgg atttctctc aagctgttgc cacaggaggt
901 ggtccagtgg ccaaatcca ctagtcctaa acgggaagga tgaccggggc catggttacc
961 gggctctaag acct

```

BG766440

```

1 ggctgcaagt caccatggac gctgaaggcc tggcgctgct gctgccgccc gtcaccctgg
61 cagccctggt ggacagctgg ctccgagagg actgccagag gctcaactac gcagccttgg
121 tcagcggggc agggccctcg caggcgggcg tgtgggcca atcccttgg gtactggcag
181 ggcagccttt cttegatgcc atatttacc aactcaactg ccaagtctcc tggttcctcc
241 ccgagggatc gaagctggtg ccggtggcca gagtggccga ggtccggggc cctgcccact
301 gcctgctgct ggggggaacgg gtggccctca acacgctggc ccgctgcagt ggcattggca
361 gtgctgccgc cgtgcagtg gaggccgcca ggggggcccgt ctggactggg cactggcag
421 gcacgaggaa gaccaogcca ggcttccggc tggaggagaa gtatgggctc ctggggggcg
481 gggccgctc gcaccgctac gacctgggag ggctgggtgat ggtgaaggat aacctgtgg
541 tggccgcccg tggcggtggag acaggcggtg cggcgggcca gacaggcggc ttgacttcgc
601 tctgaagggt gaagcggaat tgcagcagcc tgcaggaggc cgtgcaggca gctgaggctg
661 gtgccgacct tgcctgctg gacaacttca agccagagga gctgaacca cggcaccgtg
721 ctgaaagcca gttcccagat gtggctgtgg aagccagtgg gggcatcacc tggaaacact
781 cccca

```

BG401877

```

1 agccttggtc ctgagcagcc aacacaccag cccagacagc tgcaagtcac catggacgct
61 gaaggcctgg cgtgctgct gccgcccgtc accctggcag ccctgggtgga cagctggctc
121 cgagaggact gccagggctc aactacgcag ccttggctcag cggggcaggc ccttcgcagg
181 cggcgctgtg ggccaaatcc cctggggtac tggcagaggc agcctttctt cgatgccata
241 tttacccaac tcaactgcca agtctcctgg ttctcccccg agggatcgaa gctgggtggag
301 aagtatgggc tcttgggtgg cggagccgac tcgcaccgct acgacctga gagggtgggt
361 gatggtgaag gataacctg tggaggccgc cgggtggcgtg gagaaggcgg tgcggggcggc
421 cagacaggcg gctgacttcg ctctgaagggt ggaagtggaa tgcagcacgc ctggcaggag
481 ggccgtgcag gcagctgagg ctgggtgcgac cttgtcactg ctgagcactt caagccagag
541 aagctgcacc ccagagccac gtgcttgaag cccagtgcca aagtgtggct gtgaaaccag
601 tgcggcttca cctgtgaca cctccccag tactgctgcc ggcattagac gtcacttcca
661 tgggatgctg ccaggatgcc agcctgaatc tccctcagct gtgccaaaag tgtccagctc
721 aatcactatc tacgaaaaga taccggacag gttagggcca agcctggtac acttagacga
781 cccagggcac ttgcatagt cactaggtga cgagtctgag agagaagacc ggaacatata
841 aggaaagacc agtcagaaat ggagagcagc aagaaaggac ggccacatg ctaactct

```

BG337811

```

1 gagccaacac accagcccag acagctgcaa gtcaccatgg acgtgaagg cctggcgctg
61 ctgctgccgc ccgtcaccct ggcagccctg gtggacagct ggctccgaga ggactgcca
121 gggctcaact acgcagcctt ggtcagcggg gcaggcccct cgcaggcggc gctgtgggce
181 aaatcccctg ggtactggc agggcagcct ttcttcgatg ccatatttac ccaactcaac
241 tgccaagtct cctggttctt ccccgaggga tccaagctgg tgcgggtggc cagagtggcc
301 gaggtccggg gccctgccc actgctgctg ctgggggaaac ggggtggcct caacacgctg
361 gcccgctgca gtggcattgc cagtgtgccc gccgctgcag tggaggccgc cagggggggc
421 ggetggactg ggcacgtggc aggcacgagg aagaccacgc caggcttccg gctggtggag
481 aagtatgggc tcttgggtgg cggggccgac tcgcaccgct acgacctggg agggctgggtg
541 atggtgaagg ataacctgt ggtggccgac ggtggcgtgg aagaaggcgg tgcgggcggc
601 cagacaggcg gctgacttcg ctctgaagggt ggaagtggaa tgcagcagac tgcaggaggg
661 cgtgcacgca gctgaggctg gctgccgact tgtcctgggt gacaccttca agccagagga
721 gctggacgca aggcaacgtg ctgaaggaca gttccgagtg ttggtgtgga agccacatgg
781 agggcatatc actggaaaaa ctcccccaaa tctgcggaac ggaactaaac gataagcacc
841 aggcggaaat cagaccccag gaaggcccag acccgaaata ctcccaaga ggatggccaa
901 gagggggcac acgcccacaa cacacaaacc aacgcgaaag aacacggcag gcaaaatggc
961 ccaagacacg ggaccaacct acgcaggggc aaggcaatgg aacatgccaa gagccacgac
1021 ccagcaggac cag

```

BG338063

ES 2 399 246 T3

```

1 gtggacgctg aaggcctggc gctgctgctg ccgcccgta ccttggcagc cctggtggac
61 agctggctcc gagaggactg ccaggggctc aactacgcag ccttggtcag cggggcagggc
121 cctcgcaggg cggcgctgtg ggccaaatcc cctgggggtac tggcagggca gcctttcttc
181 gatgccatat ttacccaact caactgccaa gtctcctggt tcctccccga gggatcgaag
241 ctggtgccgg tggccagagt ggccgaggtc cggggccctg cccactgcct gctgctgggg
301 gaacgggtgg ccctcaacac gctggcccg cgcagtggca ttgccagtgc tggccgcgct
361 gcagtggagg ccgccagggg gccggctgga ctgggcacgt ggcaggcacg aggaagacca
421 cgccaggctt ccggctggtg gagaagtatg ggctcctggt ggggcggggc gcctcgcacc
481 gctacgacct gggagggtct gtgatggtga aggataacca tgtggtggcc gcggtggcgt
541 gagaagcggg gggggcgccc agacaggcgg tgacttcgct ctgaagggtg aagtggaatg
601 cagcagcctg caggaggccg tgcaggcagc tgaggctggt gccgaccttg tcctgctgga
661 caacttcaag ccaggaggag tgcaaccac gccaccgtgc tgaaggccca gttccgagtg
721 tgggtgtgaa cccagtgggg gcatcaccgg gaaaaactcc cccag

```

BG251163

```

1 gctgcaagtc accatggagc ctgaaggcct ggcgctgctg ctgcggcccg tcaccctggc
61 agccctggtg gacagctggc tccgagagga ctgccaggg ctcaactacg cagccttggg
121 cagcggggca gggccctcgc aggcggcgct gtgggccaaa tcccctgggg tactggcagg
181 gcagcctttc ttcgatgcca tatttaccga actcaactgc caagtctcct ggttcctccc
241 cgagggatcg aagctggtgc cgggtggccag agtggccgag gtccggggcc ctgccactg
301 cctgctgctg ggggaacggg tggccctcaa cagctggcc cgctgcagtg gcattgccag
361 tgctgcggcc gctgcagtgg aggcggccag gggggccggc tggactgggc acgtggcagg
421 cagaggaag accacggcag gcttccggct ggtggagaag tatgggctcc tgggtggcgg
481 ggccgcctcg caccgctacg acctgggagg gctggtgatg gtgaaggata accatgtggt
541 ggccgcccgt ggctgggaga aggcggtgcg ggcggcagac aggcggctga cttcgctctg
601 aaggtggaag tggaatgcag cagcctgcag gaggccgtgc aggcagctga ggctggtgcc
661 gacttgctcc tgctggacaa cttcaagcc agaggagctg caccacggc acgtgctgaa
721 ggccagttcc cgatgtggct gtggaaccct ggggctcacc tggaaaactc cccatn

```

BG396079

```

1 gcttggctct gagcagccaa cacaccagcc cagacagctg caagtcacca tggacgctga
61 aggcctggcg ctgctgctgc cgccgtcac cctggcagcc ctggtggaca gctggctccg
121 agaggactgc ccagggtcga actacgcagc cttggtcagc ggggcaggcc cctcgcaggc
181 ggcgctgtgg gccaaatccc ctgggggtact ggcagggcag cctttcttcg atgccatatt
241 tacccaactc aactgccaa gttcctggtt cctccccgag ggatcgaagc tgggtgccgg
301 ggccagagtg gccgaggtcc ggggcccctg cactgcctg ctgctggggg aacgggtggc
361 cctcaacacg ctggcccgtg gcagtggcat tgccagtgc gccgcccgtg cagtggaggc
421 cgccaggggg gccggctgga ctgggcacgt ggcaggcacg aggaagacca cgccaggctt
481 ccggctggtg gagaagtatg ggctcctggt gggcgggggc gctcgacccg ctacgacctg
541 ggagggtggt tgatggtgaa ggataaccat gtggtggccg ccggtggcgt gtagaaggcg
601 gtgcggggcg cagacaggcg gtgacttcgc totgaagggt gaagtggaa gcagcagctg
661 caggaggccg tgcaaggcag tgaggtggtg ccgacttggt ctgctggaaa ttcagcaagg
721 agtgacca

```

BF316915

```

1 ggtcaccatg gacgctgaag gcctggcgct gctgctgccg cccgtcacc cggcagccct
61 ggtggacagc tggctccgag aggactgccc agggctcaac tacgcagcct tggctcagcg
121 ggcaggcccc tgcaggcggg cgctgtgggc caaatccct ggggtactgg cagggcagcc
181 tttcttcgat gccatattta cccaactcaa ctgccaaagt tcctggttcc tccccagggg
241 atcgaagctg gtgcgggtgg ccagagtggc cgaggctcgg ggccctgccc actgcctgct
301 gctgggggaa cgggtggccc tcaacacgct ggcccgtgc agtggcattg ccagtgtgct
361 cgcgctgcag tggaggccgc caggggggcg gctggactgg gcacgtggca ggcacgagga
421 agaccacgcc aggtttccgg ctggtggaga agtatgggct cctggtgggc gggcgctcg
481 caccgtacga ctgggagggg tggtgatggt gaaggataac atgtggtggc gcggtgggg
541 ggagaagggt gcgggctggc agaaggcggt gaacttcgct ctgaagggtg aaggtggaat

```

```

601 gaggagctgt agggaggcgtg caggcagtga ggctgggtgca gacttggttct gtggaaattc
661 aagcagagga gtgacccacg gcacgggtga ggcagttccc agtgtggtgg tggagcagtg
721 gggtcacctg aaacctccca ggctgggcga atgagtgttc catgggatcg acaggggcag
781 cctgatttcc tcaaggt

```

BF313098

```

1 gcagacagct gcaagtcacc atggacgctg aaggcctggc gctgctgctg ccgcccgtca
61 ccctggcagc cctgggtggac agctggctcc gagaggactg cccagggtcc aactacgcag
121 ccttgggtcag cggggcaggc ccctcgacgg cggcgctgtg ggccaaatcc cctggggtac
181 tggcagggca gcctttcttc gatgccatat ttaccaact caactgccaa gtctcctggt
241 tctcccccga gggatogaag ctgggtgccc tgccagagt ggcagaggtc cggggccctg
301 cccactgcct gctgctgggg gaacgggtgg ccctcaacac gctggcccgc tgcagtggca
361 ttgccagtgc tgcgcgcgct gcagtggagg ccgccagggg gccggctgga ctgggcacgt
421 ggcaggcacg aggaagacca cgccaggctt ccggctggtg gagaagtatg ggctcctggt
481 gggcggggcc gtctcgaccc gctacgacct gggagggtcg gtgatggtga aggataacca
541 tgtggtggcc gccgggtggcg tggaaagcag cggctgcgggc ggcaacaggc ggtgacttcc
601 tctgaagggt gaagtggact gcagcaggcc tgcaggaggc cgtgcacggc agctgagggc
661 tgggtgccga ccttgtctgc tgcacacttc aagccaggag gagctgcacc ccacgccacc
721 gtgctgactg cccagtcccg agtgtgggtg tggaaagcag gggggcatca cctggacact
781 cccagttttg cgggcccgcg gacgtctctc ctggggtgtt gccagggcc ccgctgactt
841 cctccactgt tgccagtggc ggccaccttc caccgaggga cggcggttgg gcccgctg
901 c

```

BG257831

```

1 gcagacaggt tgacctgcaa aaaaagacag gggaggcatg gaaacaaagg acccatcgca
61 aatgttttcc atgctgatct ccaaagtgtc ccaccccccag cctggggcct ctggggagcct
121 tggctcctgag cagccaacac accagcccag acagctgcaa gtcaccatgg acgctgaagg
181 cctggcgctg ctgctgcgcg ccgtcacctt ggcagccctg gtggacagct ggctccgaga
241 ggactgcca ggctcaaota cgcagccttg gtcagcgggg caggcccttc gcaggcggcg
301 ctgtgggcca aatcccctgg ggtactggca gggcagcctt tcttcgatgc catatttacc
361 caactcaact gccaagtctc ctggttcttc cccgagggat cgaagctggt gccgggtggc
421 agagtgggcc gaggtccggg gccctgcccc ctgcctgctg ctgggggaac ggggtggaact
481 caacacgctg gccgctgcag tggcattgcc agtgctgccc cgtgacagtg gagggcgcca
541 ggtgggcccg ctggactggc acgtggcagg caccaggaag acacgccagg ttccggctgg
601 tggagaagta tgggtccctt ggggtggcgg ggcagctcga caccggatac gaacctggag
661 gggctggtga tgggtgaagg gatacaccat gtgttggccg cgggtgagtg gagaagcggg
721 ggggggaggc caaacaggcg gtgattgtct gagggtggaag tgaatgcaga cactgagagg
781 cagagcagca gtgaagggtg tgagaactgt ctgaggacaa taacacaaag atgaaca

```

BF312976

```

1 ggacagctgc aagtcaccat ggacgctgaa ggcctggcgc tgctgctgcc gccgctcacc
61 ctggcagccc tgggtggacag ctggctccga gaggactgcc cagggtcaa ctacgcagcc
121 ttggtcagcg gggcaggccc ctgcagggcg gcgctgtggg ccaaattccc tggggtactg
181 gcagggcagc ctttcttoga tgccatattt acccaactca actgccaagt ctctgggtc
241 ctcccagag gatcgaaagt ggtgcgggtg gccagagtgg ccgaggtccg gggccctgcc
301 cactgcctgc tgctggggga acgggtggcc ctcaacacgc tggcccgtcg cagtggcatt
361 gccagtgtcg ccgcccgtgc agtggaggcc gccagggggc cggctggact gggcacgtgg
421 caggcacagag gaagaccacg ccaggcttcc ggctgggtgga gaagtatggg ctctgggtgg
481 ggcggggccg cctgcacccg ctacgacctg ggagggtgg tgatggtgaa ggataccatg
541 tgggtggccg cgggtgcgtgg gaggcgggtg gggcgggccg acaggcggct gacttcgctc
601 tgaaggtgga gtggaatgca gcagcctgtc aggaggggcc tgcaggcagc tgaggctggt
661 gccgaccttt ggtctgctgg taccacttcg agccagagga gctgaccca cgggcaccgt
721 gctgcaggcc ccgttccccg agtgtggctg tggacgcccg tgggggcctc actcttgga

```

781 acctccccgg ttctgcgggc gccctagacg tcctctcctg ggggtgctgac cagggggccc

BG115486

```

1 gtcttgagca gccaacacac cagcccagac agctgcaagt caccatggac gctgaaggcc
61 tggcgctgct gctgccgccc gtcaccctgg cagccctggt ggacagctgg ctccgagagg
121 actgcccagg gctcaactac gcagccttgg tcagcggggc agggccctcg caggcggcgc
181 tgtgggccaa atcccctggg gtactggcag ggcagccttt cttegatgcc atatttacc
241 aactcaactg ccaagtctcc tggttcctcc ccgagggatc gaagctggtg ccggtggcca
301 gagtggcga ggtccggggc cctgccact gcctgctgct gggggaacgg gtggccctca
361 acacgctggc ccgctgcagt ggcattgccg gtgctgccgc cgctgcagtg gaggcggcca
421 gggggggcgg ctggactggg cactggcag gcacgaggaa gaccagcca ggcttcgggc
481 tggtagagaa gtatgggctc ctggtagggc gggccgcctc gcaccgctac gagctgggag
541 ggctggtgat ggtgaaggat aaccatgtgg tggccgcggg tggcgtggag aaggcgggtg
601 gggccggcag acaggcgggt gacttcgctc tgaagggtga agtggaatgc aggagcctgc
661 aggaggcgtg aaggcagtcg aggatggtgg cgacctgtc ctggctggac aaacttcagc
721 cagaggagct ggacaccaca gccaacgggc tgaaggccag gtcccgagag tgcagaggaa
781 ccaaaggggc tacacctggg acactacca gactgcagac gccaaaagc gcatccggcg

```

BG748194

```

1 ggcctggcag aggacaggag ggaggcttgg ggagcctggg aagggccggc tgacagcttg
61 atgggcctc cctccctcca cagtcccacc ccagcctgg ggccctctgg agccttggtc
121 ctgagcagcc aacacaccag ccagacagc tgcaagtcac catggacgct gaaggcctgg
181 cgctgctgct gccgcccgtc accctggcag ccctggtgga cagctggctc cgagaggact
241 gccaggtc aactacgcag ccttggtcag cggggcaggc cctcgcagg cggcgctgtg
301 ggccaaatcc cctggggtac tggcagggca gcctttcttc gatgccatat ttacccaact
361 caactgccaa gtctccttgg tctcccga gggatcgaag ctggtgccgg tggccagagt
421 ggcgagggtc cggggccctg ccactgcct gctgctgggg gaacgggtgg ccctcaacac
481 gctggcccgc tgcagtggca ttgccagtgc tgccgcgct gcagtggagg ccgccagggg
541 ggcggctgg actgggcaog tggcaggcac gagggaagac acgccaggct tccggctggt
601 ggagaagtat gggctcctgg tgggcggggc cgcctgcac cgctacgacc tgggaggg

```

BF980859

```

1 cccagcctg gggcctctgg gagccttggc cctgagcagg ccaacacacc agccagaca
61 gctgcaagtc accatggacg ctgaaggcct ggcgctgctg ctgccgcccg tcacctggc
121 agccctgggt gacagctggc tccgagaggc ctgccaggc tcaactacgc agccttggtc
181 agcggggcag gcccctcgca ggcgcgctg tgggccaat cccctggggt actggcaggg
241 cagcctttct tcgatgccat atttaccaca ctcaactgcc aagtctcctg gttcctccc
301 gagggatcga agctggtgcc ggtggccaga gtggccgagg tccggggccc tgccactgc
361 ctgctgctgg gggaacgggt ggccctcaac acgctggccc gctgcagtgg cattgccagt
421 gctgccgccg ctgcagtggg ggcgcaggg ggggcgggt ggactgggca cgtggcaggc
481 acgaggaaga ccacgccagg cttccggctg gtggagaagt atgggctcct ggtgggcggg
541 gccgcctcgc accgctacga cctgggaggg ctggtgatgg tgaaggataa ccatgtggtg
601 gccgcgggtg gcgtggagaa agcgtgccc ggcggccagc aggcgggtga cttgcctct
661 gaatgtggag tggaaatgaga gctgcagagg cgtgcagagc tgagggtggg caactggtct
721 ggtggacaac tcagcaaggg gctggccacg gcacggctga aggccatccg agtgtggcgt
781 gagccggggc gttccgggac ccgcagtcgg gggcaaaagc cccggggggg gccagggccc
841 ccatttccc caggcgacag ggcgcgcga ccacgaaaag cgcgcgggg cgccccgcac
901 aaggaggcga aacacaccg cggcgcgaca cgggcgacgc gtctcagctt gtccgc

```

Secuencias identificadas como las del grupo HOXB 13

5 BF676461

GGGATTCCCCCGGCTGGGTGGGGAGAGCGAGCTGGGTGCCCCCATAGATTCCCCTGCCCCG
AACCTCATGAGCCGACCCCTCGGCTCCATGGAGCCCGGAAATTATGCCACCTTGGATGGAG
CCAAGGATATCGAAGGCTTGTTGGGAGCGGGAGGGGGCGGAATCTGGTCGCCCCACTCCC
TCTCTGACCAGCCACCCAGCGCGCTACGCTTGATGCCTGTGTCAATATGCCCCCTTGATC
TGCCAGGCTCGGGGAGCGGCCAAAAGCAATGCCACCCCTATGCTCTGGGGGTGCCCAGGG
GACTGTCCCCGGCTCCGTGCCTTATGGTTACTGTGGGGCGGGGTACATACTCCTGCAGAG
TTGTCCCGGAGCTCGTTGAAACCTTGTCGCCGAGGAGAGCCACCCTGGCGGTACCCGGGAA
GACTCCCCAGGGCGGGAAGAGTACCCAGCGGCCCAATGAGTTGTGCTTCTATCGGGATA
TCCGGGACCTACCAGGCCTATGTGCAGGTACTGGACGTGTCTGTGCTGCAGACTCTGGG
TGTCGGTGGAGACCCGACATTGGCTCGCTGTGGCCTGTGGCCGGTACCAGTCTTGGGCT
CTCGGTGTGTGGCTGGACACGCCGGTTGTGTTTCGGGGAGACCGCACCCACCAGGTTCCCT
TTGGGAGGGCCGCTTTGCAGACTCCGGGGGAGGCCCTCTGAGGCGGGGCCTTTTCGGGG
GGGCGAAGAAAGCTTTCCGACGCAGGCGCTTGCGGAGCTGGCGGGACATCGGGACACTTC
ACCCAGCGAAGCGCGGCTTGGGGCCCCCTCTGGGCGCGGTCTCGGTTGACACCGGCGAAGA
GTTTCGGGAGAGGCCCATATCTTCTGGGGAGGGCGTTGCGTCGCCCCCG

BC007092

ggattceccc ggctgggtg gggagagcga gctgggtgcc ccctagattc cccgcccccg
cacctcatga gccgaccctc ggctccatgg agcccgcaa ttatgccacc ttggatggag
ccaaggatat cgaaggcttg ctgggagcgg gagggggcg gaatctggtc gccactccc
ctctgaccag ccaccagcg gcgcctacgc tgatgcctgc tgtcaactat gccccttg
atctgccagg ctccgaggag ccgccaaagc aatgccacc atgccctggg gtgcccagg
ggacgtcccc agctcccggt ccttatggtt actttggagg cgggtactac tctgccgag
tgtcccgagg ctgctgaaa cctgtgccc aggcagccac cctggccgag taccocgag
agactccac gcccgggaa gattaccca gccgcccac tgagtttgcc ttctatccg
gatatccggg aacctaccag cctatggcca gttacctgga cgtgtctgtg gtgcagactc
tggtgctcc tggagaaccg cgacatgact cctgttgcc tgtggacagt taccagtctt
gggtctctgc tgggtgctgg aacagccaga tgtgttgcca gggagaacag aaccaccag
gtcccttttg gaaggcagca tttgcagact ccagcgggca gcacctcct gacgcctgc
cctttcgtcg cgccgcaag aaacgcattc cgtacagcaa ggggcagttg cgggagctgg
agcgggagta tgcggctaac aagttcatca ccaaggaca gaggcgcaag atctcggcag
ccaccagcct ctcgagcgc cagattacca tctggtttca gaaccgccg gtcaaagaga
agaaggttct cgccaaggtg aagaacagcg ctaccctta agagatctcc ttgctgggt
gggaggagcg aaagtggggg tgcctgggg agaccaggaa cotgccaagc ccaggctggg
gccaggact ctgctgagag gccctagag acaacacct tcccaggcca ctggctgtg
gactgttct caggagcggc ctgggtaccc agtatgtgca gggagacgga acccatgtg
acagccact ccaccagggt tcccaaagaa cctggcccag tcataatcat tcactctgac
agtggcaata atcacgataa ccagtactag ctgcatgat cgttagcctc atattttcta
tctagagctc tgtagagcac tttagaaacc gctttcatga attgagctaa ttatgaataa
atttgaaaa aaaaaaaaa aaaaaaaaa aaaaaa

BM462617

ATCCCCCGGCTGGGTGGGGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCA
CCTCATGAGCCGACCCCTCGGCTCCATGGAGCCCGGCAATTATGCCACCTTGGATGGAGCC
AAGGATATCGAAGGCTTGCTGGGAGCGGGAGGGGGCGGAATCTGGTCGCCCCACTCCCCT
CTGACCAGCCACCCAGCGCGCTACGCTTGATGCCTGTGTCAACTATGCCCCCTTGGAT
CTGCCAGGCTCGGCGGAGCCGCCAAAGCAATGCCACCCATGCCCTGGGGTGGCCAGGGG
ACGTCCCCAGCTCCCGTGCCTTATGGTTACTTTGGAGGCGGGTACTACTCCTGCCGAGTG
TCCCGGAGCTCGCTGAAACCTTGTCGCCAGGCAGCCACCCTGGCCGCGTACCCCGCGGAG
ACTCCACGGCCGGGAAGAGTACCCAGCGCCCCACTGAGTTTGCCTTCTATCCGGGA
TATCCGGGAACCTACCAGCCTATGGCCAGTTACCTGGACGTGTCTGTGGTGCAGACTCTG

GGTGCTCCTGGAGAACCGCGACATGACTCCCTGTTGCCTGTGGACAGTTACCAGTCCTGG
GCTCTCGCTGGTGGCTGGAACAGCCAGATGTGTTGCCAGGGAGAACAGAACCCACCAGGT
CCCCTTTTGGAAGGCAGCATTTCAGACTCCAGCGGCAGCACCCCTCTGACGCCTGCGC
CTTTCGT

BG752489

GCAGGCGACTTGCGAGCTGGGAGCGATTTAAAAAGCTTTGGATTCCCCCGGCCTGGGTGG
GGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCATGAGCCGACCCTCG
GCTCCATGGAGCCCGGCAATTATGCCACCTTGGATGGAGCCAAGGATATCGAAGGCTTGC
TGGGAGCGGGAGGGGGGCGGAATCTGGTCGCCCCACTCCCCCTCTGACCAGCCACCCAGCGG
CGCCTACGCTGATGCCTGCTGTCAACTATCCCCCTTGGATCTGCCAGGCTCGGCGGAGC
CGCCAAAGCAATGCCACCCATGCCCTGGGGTGGCCAGGGGACGTCCCCAGCTCCCGTGC
CTTATGGTTACTTTGGAGGCGGGTACTACTCCTGCCGAGTGTCCCGAGCTCGCTGAAAC
CCTGTGCCCAGGCAGCCACCCTGGCCGCGTACCCCGCGGAGACTCCACGGCCGGGGAAG
AGTACCCAGCCGCCCCACTGAGTTTGCCTTCTATCCGGGATATCCGGGAACCTACCAGC
CTATGGCCAGTTACCTGGACGTGTCTGTGGTGCAGACTCTGGGTGCTCCTGGAGAACCGC
GACATGACTCCCTGTTGCCTGTGGACAGTTACCAGTCTTGGGCTCTCGCTGGTGGTGGGA
ACAGCCAGATGTGTTGCCAGGGAGAACAGAAGCCACCAGGTCCCTTTTGAAGGCAGCAT
CTGCAGACTCCAGCGGCAGGACCTCCTGACGCCTGCGGCCTTTCGTGCGAGCGCAAGA
AACGCATTCCGTA

BG778198

GGATTTAAAAAGCTTTGGATTCCCCCGGCCTGGGTGGGGAGAGCGAGCTGGGTGCCCCCT
AGATTCCCCGCCCCCGCACCTCATGAGCCGACCCTCGGCTCCATGGAGCCCGGCAATTAT
GCCACCTTGGATGGAGCCAAGGATATCGAAGGCTTGTCTGGGAGCGGGAGGGGGGCGGAAT
CTGGTCGCCCCACTCCCCCTCTGACCAGCCACCCAGCGGCGCCTACGCTGATGCCTGCTGTC
AACTATGCCCCCTTGGATCTGCCAGGCTCGGCGGAGCCGCCAAAGCAATGCCACCCATGC
CCTGGGGTGGCCAGGGACGTCCCCAGCTCCCGTGCCTTATGGTTACTTTGGAGGCGGGT
ACTACTCCTGCCGAGTGTCCCGGAGCTCGCTGAAACCCTGTGCCAGGCAGCCACCCTGG
CCGCGTACCCCGCGGAGACTCCACGGCCGGGGAAGAGTACCCAGCCGCCCCACTGAGT
TTGCCTTCTATCCGGGATATCCGGGAACCTACCAGCCTATGGCCAGTTACCTGGACGTGT
CTGTGGTGCAGACTCTGGGTGCTCCTGGAGAACCGCGACATGACTCCCTGTTGCCTGTGG
ACAGTTACCAGTCTTGGGCTCTCGCTGGTGGGCTGGAACAGCCAGATGTGTTGCCAGCGC
AGAACAGAACCCACCAGGTCCCTTTTGAAGGCAGCATTTCAGACTCCAGCGGGCAGAA
CCCTCCTGACGCCTGCGCCTTTCGTTGCGGGCGAAAAA

CB050884

AAGAAACGCATTCCGTACAGCAAGGGGCAGTTGCGGGAGCTGGAGCGGGAGTATGCGGCT
AACAAGTTCATCACCAAGGACAAGAGGCGCAAGATCTCGGCAGCCACCAGCCTCTCGGAG
CGCCAGATTACCATCTGGTTTCAGAACCGCCGGGTCAAAGAGAAGAAGGTTCTCGCCAAG
GTGAAGAACAGCGCTACCCCTTAAGAGATCTCCTTGCCCTGGGTGGGAGGAGCGAAAGTGG
GGGTGTCCTGGGGAGACCAGGAACCTGCCAAGCCCAGGCTGGGGCCAAGGACTCTGCTGA
GAGGCCCTTAGAGACAACACCCTTCCAGGCCACTGGCTGCTGGACTGTTCTCAGGAGC
GGCCTGGGTACCCAGTATGTGCAGGGAGACGGAACCCCATGTGACAGCCCACTCCACCAG
GGTTCCTCAAAGAACCTGGCCAGTCATAATCATTCATCCTGACAGTGGCAATAATCACGA
TAACCAGTACTAGCTGCCATGATCGTTAGCCTCATATTTTCTATCTAGAGCTCTGTAGAG
CACTTTAGAAACCGCTTTCATGAATTGAGCTAATTATGAATAAATTTGAAGGCGAAAAA
AAAAACCTCGTGCC

CB050885

ATTTCGGCACGAGGTTTTTTTTTTTCGCCTTCCAAATTATTCATAATTAGCTCAATTCATG
AAAGCGGTTTTCTAAAGTGCTCTACAGAGCTCTAGATAGAAAATATGAGGCTAACGATCAT
GGCAGCTAGTACTGGTTATCGTGATTATTGCCACTGTCAGGATGAATGATTATGACTGGG
CCAGGTTCTTTGGGAACCTGGTGGAGTGGGCTGTACATGGGGTTCCGTCTCCCTGCAC
ATACTGGGTACCCAGGCCGCTCCTGAGGAACAGTCCAGCAACCAGTGGCCTGGGAAGGGT
GTTGTCTCTAGGGGCCTC

BF965191

GGGTGGGGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCATGAGCCGA
CCCTCGGCTCCATGGAGCCCGCAATTATGCCACCTTGGATGGAGCCAAGGATATCGAAG
GCTTGCTGGGAGCGGGAGGGGGCGGAATCTGGTCGCCCACTCCCTCTGACCAGCCACC
CAGCGGCGCTACGCTGATGCCTGCTGTCAACTATGCCCCCTTGGATCTGCCAGGCTCGG
CGGAGCGGCCAAAGCAATGCCACCCATGCCCTGGGGTGCCCCAGGGGACGTCCCCAGCTC
CCGTGCCTTATGGTTACTTTGGAGGCGGGTACTACTCCTGCCGAGTGTCCCGAGCTCGC
TGAAACCTGTGCCAGGCAGCCACCCCTGGCCGCGTAACCCGACGGAGACTCTCACGTGCG
GGGAAGAGTACCCCTAGCGCCCCACATGAGTTTGCCTTCTATCCGGGATATCCGGGACCG
TACCAGCCTATGGCAGTTACCTGGACGTGTCTGTGGTGCCGACTCTGGGTGCTCCTGGAG
AACCGCGGACATGACTCCTTGTTTGTGTGTGCGACGCTCACCAGTCTGGGCTCCTCGTGG
TGGTCGCACTCCCACTTTTGGCCGGGCGACATCCCCCGGGGCCCTTCCGGAACAGCGAC
CTTGCGAGCCCCCGGGGACACACCCCGTAAGCGGCCTATCATCGCTGATAAACCTCATC
AGAGGGCACCGAAAGCCGCGACTCTAACCCCCCACTACGACTCACGACCGCACAGGTAC
TCGAACCGCCCAATATCTGGTTCTAACCCATGGCGCATCTCAGCCGCTAGAGAGCCAACC
AAACGCGCCACGCGCAACCACTACACCACGGCACCCCTTTCATCTCACTCCACGCGG
ATCACTCTTCACCCTCCAGAATCATTCCCTCGCACATCCTACCTATCTCATGCCTCCCA
GTTCACCCCATTCCCTCCCTAATCTCACCCACACATTACGACAGTTCTCACTACGCTT
CGTCCGACCCACATCCTCACCCCAACATTACATACCACTTACCATCACGACCCCCCTT
CTCATCGACTCCTGTCTCATTTCTCAACCACAGTACTACCAGCTCCAACACACCACTCAC
CCAAGCTATCCATCACCTACACGCTTTCACCCTCACCGCTCCCAAGTAATTGAGTAC
TCAAACACAATCTGCTACATACTCATCCCTCCCCCACTCCAGTACAGTCAAACCACCGA
CCAACCTACCTCCGCGCCACCCGCGCGCCCACTCACCGGCCCAACCGCCCGCACAGG
GCACGCACCCCCCGGAACCGCGCGATCCGGCCGTACACACTCTTGGGCGGCACGCAGCT
GAGGACATTCCGCGGGAGCGCCCCACCGTGGGTACGTGGGTGCGGACCCGGCGGGGCGC
GTGCGGCGTGCCTCGCCCGCCCGCCCGGACTGCGACCCAGTCGAG

BU930208

GGGGCTTTGGATTCCCCCGGCCTGGGTGGGGAGAGCGAGCTGGGTGCCCCCTAGATTCCC
CGCCCCCGCACCTCATGAGCCGACCTCGGCTCCATGGAGCCCGCAATTATGCCACCTT
GGATGGAGCCAAGGATATCGAAGGCTTGCTGGGAGCGGGAGGGGGCGGAATCTGGTCGC
CCACTCCCTCTGACCAGCCACCCAGCGGCGCTACGCTGATGCCTGCTGTCAACTATGC
CCCCCTTGGATCTGCCAGGCTCGGCGGAGCGGCCAAAGCAATGCCACCCATGCCCTGGGGT
GCCCCAGGGGACGTCCCCAGCTCCCGTGCCTTATGGTTACTTTGGAGGCGGGTACTACTC
CTGCCGAGTGTCCCGGAGCTCGCTGAAACCTGTGCCAGGCAGCCACCTGGCCGCGTA
CCCCGCGGAGACTCCACGCGCGGGGAAGAGTACCCAGCGCCCCCACTGAGTTTGCCTT
CTATCCGGGATATCCGGGAACCTACCAGCCTATGGCCAGTTACCTGGACGTGTCTGTGGT
GCAGACTCTGGGTGCTCCTGNAGAACCGCGACATGACTCCCTGTTGCCTGTGGACAGTTA
CCAGTCTTGGGCTCTCGCTGGTGGCTTGAACAGCCCAGATGTGTTTGGCCAGGGNAGAA
CACGAACCCCAACCGGTTCCCCCTTTTGGGAAGGGGAGCCATTTTGGCCAGCCTTCCAA
GCGGGGCCAACCAACCCCTCCCTGGACAGGCCCTGGT

AA807966

GCGGCCGCAAGAAACGCATTCCGTACAGCAAGGGGCAGTTGCGGGACTGGAGCGGGAGTA
TGCGGCTAACAAGTTCATCACCAAGGACAAGAGGCGCAAGATCTCGGCAGCCACCAGCCT
CTCGGAGCGCCAGATTACCATCTGGTTTCAGAACC GCCGGGTCAAAGAGAAGAAGGTTCT
CGCCAAGGTGAAGAACAGCGCTACCCCTTAAGAGATCTCCTTGCTGGGTGGGAGGAGCG
AAAGTGGGGGTGTCTTGGGGAGACCAGGAACCTGCCAAGCCCAGGCTGGGGCCAAGGACT
CTGCTGAGAGGCCCCCTAGAGACAACACCCTTCCCAGGCCACTGGCTGCTGGACTGTTCT
CAGGAGCGGCCTGGGTACCCAGTATGTGCAGGGAGACGGAACCCCATGTGACAGCCCAT
CCACCAGGGTTCCCAAAGAACCTGGCCCAGTCATAATCATTTCATCCTGACAGTGGC

AI884491

AGCGGCCGCAAGAAACGCATTCCGTACAGCAAGGGGCAGTTGCGGGAGCTGGAGCGGGAG
TATGCGGCTAACAAGTTCATCACCAAGGACAAGAGGCGCAAGATCTCGGCAGCCACCAGC
CTCTCGGAGCGCCAGATTACCATCTGGTTTCAGAACC GCCGGGTCAAAGAGAAGAAGGTT
CTCGCCAAGGTGAAGAACAGCGCTACCCCTTAAGAGATCTCCTTGCTGGGTGGGAGGAG
CGAAAGTGGGGGTGTCTTGGGGAGACCAGGAACCTGCCAAGCCCAGGCTGGGGCCAAGGA
CTCTGCTGAGAGGCCCCCTAGAGACAACACCCTTCCCAGGCCACTGGCTGCTGGACTGTTCT
CTCAGGAGCGGCCTGGGTACCCAGTATGTGCAGGGAGACGGAACCCCATGTGACAGCCCA
CTCCACCAGGGTTCCCAAAGAACCTGGCCCAGTCATAATCATTTCATCCTGACAGTGGCAA
TAATCACGATAACCAGTACTAGCTGCCATGATCGTTAGCCTCATATTTTCTATCTAGAGC
TCTGTAGAGCAC

AA652388

GCGGCCGCAAGAAACGCATTCCGTACAGCAAGGGGCAGTTGCGGGACTGGAGCGTGAGTA
TGCGGCTAACAAGTTCATCACCAAGGACAAGAGGCGCAAGATCTCGGCAGCCACCAGCCT
CTCGGAGCGCCAGATTACCATCTGGTTTCAGAACC GCCGGGTCAAAGAGAAGAAGGTTCT
CGCCAAGGTGAAGAACAGCGCTACCCCTTAAGAGATCTCCTTGCTGGGTGGGAGGAGCG
AAAGTGGGGGTGTCTTGGGGAGACCAGGAACCTGCCAAGCCCAGGCTGGGGCCAAGGACT
CTGCTGAGAGGCCCCCTAGAGACAACACCCTTCCCAGGCCACTGGCTGCTGGACTGTTCT
CAGGAGCGGCCTGGGTACCCAGTATGTGCAGGGAGACGGAACCCCATGTGACAGCCCACT
CCACCAGGGTTCCCAAAGAACCTGGCC

BF446158

TTTTTTTTTTTTTTTTTTTTTTCGCCTTCCAAATTTATTCATAATTAGCTCAATTTCATGAAAG
CGGTTTCTAAAGTGCTCTACAAAGCTCTAAATAAAAAATATGAGGCTAACGATCATGGCA
GCTAGTACTGGTTATCGGGATTATTGCCACTGTGAGGATGAATGATTATGACTGGGCCAG
GTTCTTTGGGAACCCCTGGTGGAGTGGGCTGTACATGGGGTTCCGTCTCCCTGCACATAC
TGGGTACCCAGGCCGTTCTTGAGGAACAGTCCACCACCCAGTGGCCTGGGAAGGGTGTG
TCTCTAGGGGCCTCTCAACAAAGTCCTTGGCCCCAGCCTGGGCTTGGCAGGTTCTTGGTC
TCCCCAGGACACCCCCACTTTCGCTCCTCCCACCCAGGCAAGGAGATCTCTTAAGGGG

AA657924

GACGCNAGGTATGCGGCTAACAAGTTCATCACCAAGGACAAGAGGCGCAAGATCTCGGCA
GCCACCAGCCTCTCGGAGCGCCAGATTACCATCTGGTTTCAGAACC GCCGGGTCAAAGAG
AAGAAGGTTCTCGCCAAGGTGAAGAACAGCGCTACCCCTTAAGAGATCTCCTTGCTGGG
TGGGAGGAGCGAAAGTGGGGGTGTCTTGGGGAGACCAGGAACCTGCCAAGCCCAGGCTGG
GGCCAAGGACTCTGCTGAGAGGCCCCCTAGAGACAACACCCTTCCCAGGCCACTGGCTGCT
GGACTGTTCTCAGGAGCGGCCTGGGTACCCATGTATGTGCAGGGAGACGGAACCCCATG

TGACAGCCCACTCCACCAGNGTTCTTAAAGAACCCTGGCCAGTCA

AA644637

GCAGGCGACTTGCGAGCTGGGAGCGGTTTAAACGCTTTGGATTCCCCCGGCCTGGGTGG
GGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCATGAGCCGACCCTCG
GTCCATGGACACGGCAATTATGCCACCTTGATGGAGCCAAGGATATCGAAGGCTTGCTG
GGAGCGGGAGGGGGGGCGGAATCTGGTCGCCCCACTCCCCCTGACCAGCCACCCAGCGGCG
CCTACGCTGATGCCTGCTGTCAACTATGCCCCCTTGATCTGCCAGGCTCGGCGGACTCT
NAAAGCATATGCCACCCNATGCCCTGGGGTGCCCCAGGGGAACGTCCCCAGCTCCCGTGC
CTTATGGTT

BF222357

GCGGCCGCAAGAAACGCATTCCGTACAGCAAGGGGCAGTTGCGGGAGCTGGAGCGGGAGT
ATGCGGCTAACAAGTTCATCACCAGGACAAGAGGCGCAAGATCTCGGCAGCCACCAGCC
TCTCGGAGCGCCAGATTACCATCTGGTTTCAGAACCGCCGGGTCAAAGAGAAGAAGGTTT
TCGCCAAGGTGAAGAACAGCGCTACCCCTTAAGAGATCTCCTTGCCTGGGTGGGAGGAGC
GAAAGTGGGGGTGTCTGGGGAGACCAGGAACCTGCCAAGCCCAGGCTGGGGCCAAGGAC
TCTGCTGAGAGGCCCCCTAGAGACAACACCCTTCCCAGGCCACTGGCTGCTGGACTGTTC
TCAGGAGCGGCCTG

AA527613

GTCGACGAACAGCGCTACCCCTTAAGAGATCTCCTTGCCTGGGTGGGAGGAGCGAAAGTG
GGGGTGTCTGGGGAGACCGGGAACCTGCCAAGCCCAGGCTGGGGCAAGGACTCTGCTGAG
AGGCCCCCTAGAGACAACACCCTTCCCAGGCCACTGCTGCTGGACTGTTCTCAGGAGCGG
CCTGGGTACCCAGTATGTGCAGGGAGACGGAACCCCATGTGACAGCCCACTCCACCAGGG
TTCCCAAAGAACCTGGCCCAGTCATAATCATTATCTGACAGTGGCAATAATCACGATA
ACCAGTACTCAGCTGCCATGATCGTTAGCCTCATATT

AA533227

GCGTCGACCCCTTGAAGAGATCTCCTTGCCTGGGTGGGAGGAGCGAAAGTGGGGGTGTCC
TGGGGAGACCAGGAACCTGCCAAGCCCAGGCTGGGGCCAAGGACTCTGCTGAGAGGCCCC
TAGAGACAACACCCTTCCCAGGCCACTGGCTGCTGGACTGTTCTCAGGAGCGGCCTGGG
TACCCAGTATGTGCAGGGAGACGGAACCCCATGTGACAGCCCACTCCACCAGGGTTCCCA
AAGAACCTGGCCCAGTCATAATCATTATCTGACAGTGGCAATAATCACGATAACCAGT
ACTAGCTGCCATGATCGTTAGCCTCATATTTCTATCTAGAGCTCTGTAGAGCACTTGTA
GAAACCGCTTTCATGAATTGAGCTAATTATGAATAGATTTGGAAGGGGAAAAAAGTGGAA
AAAGTTTTCGCCAAAGTGGGTCGTTTACGTCG

AA456069

CTCCCTGGCAACACATCTGGCTGTTCCAGCACCAGCGAGACCCAAGACTGGTAACCTGTCC
ACAGGCAACAGGGAGTCATGTCGCGGTTCTCCAGGAGCACCCAGAGTCTGCACCACAGAC
ACGTCCAGGTAACCTGGCCATAGCTGAGTAGGTTCCCGGATATCCCGGATAGAAGGCAAAC
TCAGTGGGGGGCTGGGGTACTCTCCCCGGCCGTGGAGAGTCTCCGCGGGGTACGGCCCC
AGGGTGGCTGCCTGGGCATCAGGGTTTCAGCGAGCTCCGGGACACTCGGCAGGAGTAGTA
CCCGCCTCCAAAGTAACCATAGGCACGGGAGCTGGGGACGTCCCTGGGGCACCCAG

AA455572

TTTAAACGCTTTGGATTCCCCCGGCCTGGGTGGGGAGAGCGAGCTGGGTGCCCCCTAGA
 TTCCCCGCCCCCGCACCTCATGAGCCGACCTCGGTCCATGGAGCCGGCGAATTATGCCA
 CCTTGGATGGAGCCAAGGATATCGAAGGCTTGCTGGGAGCGGGAGGGGGCGGAATCTGG
 TCGCCCACTCCCCTCTGACCAGCCACCCAGCGGCGCTACGTGATGCCTGCTGTCAACTAT
 GCCCTTGGATCTGCCAGCTCGCGGAGCCAAAGCAATGCCACCCATGCCCTGGGGTGCCCC
 AGGTGACGTCCCCAGCTCCCGTGCCCTTATGGTTACTTTGGAGGCGGGTACTACTCCTGCC
 GAGTGTCCCGAGCTCGCTGAAACCCTGTGCCAGGCAGCCACCCCTGGCCGCGTACCCCG
 CGATGACTCCACGCGCCGGGAAGAGTACCCAGCCGCCCCACTGAGTTTGCT

BX117624

CAGGCGACTTGCGAGTCTGGGAGCGATTAAACGCTTTGGATTCCCCCGGCCTGGGTGG
 GGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCATGAGCCGACCTCG
 GCTCCATGGAGCCCGCAATTATGCCACCTTGGATGGAGCCAAGGATATCGAAGGCTTG
 TGGGAGCGGGAGGGGGCGGAATCTGGTCGCCCACTCCCCTCTGACCAGCCACCCAGCGG
 CGCCTACGCTGATGCCTGCTGTCAACTATGCCCCCTTGGATCTGCCAGGCTCGGCGGAGC
 CGCCAAAGCAATGCCACCCATGCCCTGGGGTGCCCCAGGGGACGTCCCAGCTCCCGTGC
 CTTATGGTTACTTTGGAGGCGGGTACTACTCCTGCCGAGTGTCCCGGAGCTCGCTGAAAC
 CCTGTGCCAGGCAGCCACCCCTGGCCGCGTACCCCGCGGAGACTCCACGGCCGGGGAAG
 AGTACCCAGCCGCCCCACTGAGTTTGCTTCTATCCGGGATATCCGGGAACCTACCAGC
 CTATGGCCAGTTACCTTGGACGTGTCTGTGGTGACACTCTGGGTGCTCCTGGAGAACC
 CGACATGACTCCCTGNTGCCCTGTGGACAGTTACCACTCTGGGCTCTCGCTGGTGGCTGG
 AACAGCCAGATGTGTTGNCAGGGAGAACAGAACCCACCAGGTCCCTTTTGAAGGCAGAT
 TTGCAGACTNCAGCGGGCA

BQ673782

AGGCAGCCACCCTGGCCGCGTACCCCGCGGAGACTCCACGGCCGGGGAAGAGTACCCCA
 GCCGCCCCACTGAGTTTGCTTCTATCCGGGATATCCGGGAACCTACCAGCCTATGGCCA
 GTTACCTGGACGTGTCTGTGGTGACACTCTGGGTGCTCCTGGAGAACCAGCAGATGACT
 CCCTGTTGCTGTGGACAGTTACCACTCTGGGCTCTCGCTGGTGGCTGGAACAGCCAGA
 TGTGTTGCCAGGGAGAACAGAACCCACCAGGTCCCTTTTGAAGGCAGCATTGTCAGACT
 CCAGCGGGCAGCACCCCTCCTGACGCCTGCGCCTTTCGTGCGGCGCGCAAGAAACGCATTC
 CGTACAGCAAGGGGAGTTGCGGGAGCTGGAGCGGGAGTATGCGGCTAACAAAGTTTCATCA
 CCAAGGACAAGAGGCGCAAGATCTCGGCAGCCACCAGCCTCTCGGAGCGCCAGATTACCA
 TCTGGTTTTCAGAACCAGCCGGGTCAAAGAGAAGAAGGTTCTCGCCAAGGTGAAGAACAGCG
 CTACCCCTTAAGAGATCTCCTTGCTGGGTGGGAGGAGCGAAAGTGGGGGTGTCTGGGG
 AGACCAGGAACCTGCCAAGCCCCAGGCTGGGGCCAAGGACTCTGCTGAGAGGCCCTAGAG
 GACAACACCCTTCCAGGCCACTGGCTGCTGGACTGTTCTCAGGAGCGGCCTGAGTACC
 CCGTATGTGCAGGGGAGACGGAACCCCTGTGACCAGCCCCCTCCACCCGTGGTCTCCC
 AGATAACCTGGCCCCCACTCATAAATCATTTCTTCCCGGGCCGGGGCCAATCATTTCCC
 GAACTACCCCGGTACCTTATACAATTAGATTGACATGAATCCTCTCGGGGCGATTCCCT
 ATGGCGCTGAGGCCCCCTCACACCT

AI814453

GGGTGCTGTCTCTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACCTGGTGCGTTCTGT
 TCTCCCTGGCAACACATCTGGCTGTTCCAGCCACCAGCGAGAGCCCAAGACTGGTAACTG
 TCCACAGGCAACAGGGAGTCATGTGCGGTTCTCCAGGAGCACCCAGAGTCTGCACCACA
 GACACGTCCAGGTAACCTGGCCATAGGCTGGTAGGTTCCTGGATATCCCGGATAGAAGGCA
 AACTCAATGGGGCGGCTGGGGTACTCTTCCCCGGCCGTGGGAGTCTCCGCGGGGTACGCG
 GCCAGGGTGGCTGCCTGGGCACAGGGTTTCAGCGAGCTCCGGGACACTCGGCAGGAGTAG
 TACCCGCTCCAAAGTAACCATAAGGCACGGGAGCTGGGGACGTCCCTGGGGCACCCCA

NGGCATGGGTGGCATTGCTTTGGCGGCTCCGCCGAGCCTGGCAGATCCAAGGGGGCATAG
TTGACAGCAGGCATCAGCGTAGGCGCCGCTGGGTGGCTGGTCAAAGGGAGTGGCGACCA
NATTCGCCCCCCTCCCGCTTCCCAG

AI417272

GGGTGCTGCCCGCTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACCTGGTGGGTCTGT
TCTCCCTGGCAACACATCTGGCTGTTCCAGCCACCAGCGAGAGCCAGGACTGGTAACTG
TCCACAGGCAACAGGGAGTCATGTCGCGGTTCTCCAGGAGCACCCAGAGTCTGCACCACA
GACACGTCCAGGTAACCTGGCCATAGGCTGGTAGGTTCCTCGGATATCCCGGATAGAAGGCA
AACTCAGTGGGGCGGCTGGGGTACTCTTCCCCCGCGTGGGAGTCTCCGCGGGGTACGCGG
CCAGGGTGGCTGCCTGGGCACAGGGTTTCAGCGAGCTCCGGGACACTCGGCAGGAGTAGT
ACCCGCCTCCAAAGTAACCATAAGGCACGGGAGCTGGGGACGTCCCTGGGGCACCCAG
GGCATGGGTGGCATTGCTTTGGCGGCTCCGCCGAGCCTGGCAGATCCAAGGNGGCATAGT
TGACAGCAGGCATCAGCGTANGCGCCGCTGGGTGGCTGTCAAGAGG

AA535663

TCGACGTTACCTGGACGTGTCTGTGGTGCAGACTCTGGGTGCTCCTGGAGAACCGCGACA
TGACTCCCTGTTGCCTGTGGACAGTTACCAGTCTTGGGCTCTCGCTGGTGGCTGGAACAG
CAGATGTGTTGCCAGGGAGAACAGAACCACCAGGTCCCTTTTGGAAGGCAGCATTTGCA
GACTCCAGCGGGCAGCACCTCCTGACGCCTGCGCCTTTCGTGCGGGCCGCAAGAAACGC
ATTCCGTACAGCAAGGGGCAGTTGCGGGACTGGAGCGGGAGTATGCGGCTAACAAAGTTCA
TCACCAAGGACAAGAGGCGCAAGATCTCGGCAGCCACCAGCCTCTCGGAGCGCCAGATTA
CCATCTGGTTTCAGAACCGCCGGGTCAAAGAGAAGAAGGTTCTCGCCAAGGTGAAGAACA
GCGCTACCCCTTAAGAGATCTCCTTGCCTGGGTGGGAGGAGCGAAAGTGTG

AI400493

GTCAGGAGGGTGCTGCCCGCTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACCTGGTGG
GTTCTGTTCTCCCTGGCAACACATCTGGCTGTTCCAGCCACCAGCGAGAGCCAGGACTG
GTAAGTGTCCACAGGCAACAGGGAGTCATGTCGCGGTTCTCCAGGAGCACCCAGAGTCTG
CACCACAGACACGTCCAGGTAACCTGGCCATAGGCTGGTAGGTTCCTCGGATATCCCGGATA
GAAGGCAAACTCAGTGGGGCGGCTGGGGTACTCTTCCCCGCGCGTGGGAGTCTCCGCGGG
GTACGCGGCCAGGGTGGCTGCCTGGGCACAGGGTTTCAGCGAGCTCCGGGACACTCGGCA
TGAGTAGACCCGCTTCCAAGTAACCATAAGGCACGGGAGCTGGTAACGTCCCTGGGGC
ACCCANGGCCATGGGTGCATTGCTTTGGCGGCTCCGCCGAGCCCTGCAGATCCAAGGTG
GGCATATTGACAGCAGGCATTCACGTATGCGCCCCCTGGGTGGCTGTCATATTGGGGATT
GCGAC

AW779219

GCAGGCGTCAGGAGGGTGCTGCCCGCTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACC
TGGTGGGTTCTGTTCTCCCTGGCAACACATCTGGCTGTTCCAGCCACCAGCGAGAGCCCA
AGACTGGTAACTGTCCACAGGCAACAGGGAGTCATGTCGCGGTTCTCCAGGAGCACCCAG
AGTCTGCACACAGACACGTCCAGGTAACCTGGCCATAGGCTGGTAGGTTCCTCGGATATCC
CGGATAGAAGGCAAACTCAGTGGGGCGACTGGGGTACTCTTCCCGGCCGCTGGGGAGTCTC
CGCGGGGTACGCGGCCAGGGTGGCTGCCTGGGCACCAGGGTTTCAGCGAGCTCCGGGA
CACTCNGCAGGAAANTAGTACCCGCCCTCCCAAAGTAACCATAAGCACCGGACTGNGGGNN
GGACGTCCCTGGGGCAC

AA594847

GCGACCGGACGAAAAGGAGGCGTCAGGAGGGTGCTGCCCCGCTGGAGTCTGCAAATGCTGCC
TTCAAAAGGGACCTGGTGGGTTCTGTTCTCCCTGGCAACACATCTGGCTGTTCCAGCAC
CAGCGAGACCCAAGACTGGTAACTGTCCACAGGCAACAGGGAGTCATGTCGCGGTTCTCC
AGGAGCACCAGAGTCTGCACCACAGACACGTCCAGGTAAGTGGCCATAGCTAGGTAGGT
TCCCGGATATCCCGGATAGAAGGCAAACTCAGTGGGGCGACTGGGGTACTCTTCCCGGC
CGTGGGAGTCTCCGCGGGGTACGCCCATGGGTGGCTGCCTGGGCACAGGGTTTCAGCGAG
CTCCGGGACA

AI150430

GCAGGCGTCAGGAGGGTGCTGCCCCGCTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACC
TGGTGGGTTCTGTTCTCCCTGGCAACACATCTGGCTGTTCCAGCCACCAGCGAGAGCCCA
AGACTGGTAACTGTCCACAGGCAACAGGGAGTCATGTCGCGGTTCTCCAGGAGCACCAG
AGTCTGCACCACAGACACGTCCAGGTAAGTGGCCATAGGCTGGTAGGTTCCCGGATATCC
CGGATAGAAGGCAAACTCAGTGGGGCGACTGGGGTACTCTTCCCGGCCCGTGGGAGTCTC
CGCGGGGTACCGCGCCAGGGTGGCTGCCTGGGCACAGGGTTTCAGCGAGCTCCGGGACAC
TCGGCAGGAGTAGTACCCGCTCCAAAGTAACCATAAGGCACGGGAGCTGGATGCGTCCC
CTAGGGCACCCCATGGCATGGGTGGCATTGCTTTGGCGGCTCCGCCGAGCCTGGCAGATC
CAAGGAGGCACTGTT

AA494387

GGGTGCTGCCCCGCTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACCTGGTGGGTTCTGT
TCTCCCTGGCAACACATCTGGCTGTTCCAGCCACCAGCGAGACCCAAGACTGGTAACTGT
CCACAGGCAACAGGGAGTCATGTCGCGGTTCTCCAGGAGCACCAGAGTCTGCACCACAG
ACACGTCCAGGTAAGTGGCCATAGGCTGGTAGGTTCCCGGATATCCCGGATAGAAGGCAA
ACTCAGTGGGGCGGCTGGGGTACTCTTCCCGGCCCGTGGGAGTCTCCGCGGGGTACCGGT
CCAGGGTGGCTGCCTGGGCACAGGGTTTCAGCGAGCTCCGGGACACTCGGCAGGAGTAGT
ACCCGCTCCAAAGTAACCATAAGGCACGGGAGCTGGGGACGTCCCTG

AA662643

GGGTGCTGCCCCGCTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACCTGGTGGGTTCTGT
TCTCCCTGGCAACACATCTGGCTGTTCCAGCCACCAGCGAGACCCAAGACTGGTAACTGT
CCACAGGCAACAGGGAGTCATGTCGCGGTTCTCCAGGAGCACCAGAGTCTGCACCACAG
ACACGTCCAGGTAAGTGGCCATAGGTGGTAGGTTCCCGGATATCCCGGATAGAAGGCAA
CTCAGTGGGGCGGCTGGGGTACTCTTCCCGGCCCGTGGGAGTCTCCGCGGGGTACGCGC
CAGGGTGGCTGCCTGGGCACAGGGTTTCAGCGAGCTCCGGGACA

AI935940

GGGTGCTGCCCCGCTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACCTGGTGGGTTCTGT
TCTCCCTGGCAACACATCTGGCTGTTCCCTGCCACCAGCGAGAGCCCAAGACTGGTAACTG
TCCACAGGCAACAGGGAGTCATGTCGCGGTTCTCCAGGAGCACCAGAGTCTGCACCACA
GACACGTCCAGGTAAGTGGCCATAGGCTGGTAGGTTCCCGGATATCCCGGATAGAAGGCAA
AACTCAGTGGGGCGGCTGGGGTACTCTTCCCGGCCCGTGGGAGTCTCCGCGGGGTACGCG
GCCAGGGTGGCTGCCTGGGCACAGGGTTTCAGCG

AA532530

GGGTGCTGCCCCGCTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACCTGGTGGGTTCTGT
TCTCCCTGGCAACACATCTGGCTGTTCCAGCCACCAGCGAGACCCAAGACTGGTAACTGT

CCACAGGCAACAGGGAGTCATGTCGCGGTTCTCCAGGAGCACCCAGAGTCTGCACCACAG
ACACGTCCAGGTAACCTGGCCATAGGTNGGTAGGTTCCCGGATATCCCGGATAGAAGGCAA
ACTCAGTGGGGCGGCTGGGCTACTCTTCCCCGGCCGTGGGAGTCTCCG

AA857572

CTCCCTGGCAACACATCTGGCTGTTCCAGCACCAGCGAGAGCCAAGACTGGTAACTGTCC
ACAGGCAACAGGGAGTCATGTCGCGGTTCTCCAGGAGCACCCAGAGTCTGCACCACAGAC
ACGTCCAGGTAACCTGGCCATAGGTGCGTAGGTTCCCGGATATCCCGGATAGAAGGCAAAC
TCAGTGGGGCGACTGGGGTACTCTTCCCCGGCCGTGGGAGTCTCCCGGGGTACGGCNAC
AGGGTGGCTGCCTGGGCACAGGGTTTCAGCGAGCTCCGGGACACTCGGCAGGAGTAGTAN
CCGCCTCAAAGTAACCATAANGCACGGGAGCTGGGGACGTCCC

AI261980

ACGAAAGGCGCAGGCGTCAGGAGGGTGCTGCCCGCTGGAGTCTGCAAATGCTGCCTTCCA
AAAGGGACCTGGTGGGTCTGTTCTCCCTGGCAACACATCTGGCTGTTCCAGCCACCAGC
GAGAGCCCAAGACTGGTAACTGTCCACAGGCAACAGGGAGTCATGTCGCGGTTCTCCAGG
AGCACCCAGAGTCTGCACCACAGACAGTCCAGGTAACCTGGCCATAGGCTGGTAGGTTCC
CGGATATCCCGGATAGAAGGCAAACCTCAGTGGGGCGACTGGGGTACTCTTCCCCGGCCCG
GGGAGTCTCCGCGGGGTACGCGGCCAGGGTGGCTGCCTGGGCACAGGGTTTCAGCGAGCT
CCGGGACACTCGGCGGAGNTAGTACCCGCCTCCAAAGTAACCATAAGGCACGGGAGCTGG
GGAACCGTCCCCTGGGGCACC

BE888751.1

GAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCATGAGCCGACCCCTCGGCT
CCATGGAGCCCGCAATTATGCCACCTTGGATGGAGCCAAGGATATCGAAGGCTTGCTGG
GAGCGGGAGGGGGGGCGGAATCTGGTCGCCCCTCCCCCTGACCAGCCACCCAGCGGCGC
CTACGCTGATGCCTGCTGTCAACTATGCCCCCTTGGATCTGCCAGGCTCGGCGGAGCCGC
CAAAGCAATGCCACCCATGCCCTGGGGTGCCCCAGGGACGTCCCCAGCTCCCGTGCCTTA
TGGTTACTTTGGAGGCGGGTACTACTCCTGCCGAGTGTCCCGGAGCTCGCTGAAACCCCTG
TGCCAGGCAGCCACCCCTGGCCGCGTACCCCGCGGAGACTCCACGGCCGGGGAAGAGTA
CCCCAGCCGCCCCACTGAGTTTGCCTTCTATCCGGGATATCCGGGAACCTACCAGCCTAT
GGCCAGTTACCTGGACGTGTCTGTGGTGACAGTCTGGGTGCTCCTGGAGAACCAGGACA
TGACTCCCTGTTGCCCTGTGGACAGTTACAGTCTTGGGCTCTCGCTGGTGGCTGGAACAG
CCAGATGTGTTGCCAGGGAGAACAGAACCCACCAGGTCCCTTTTGGAAAGGCAGCATTTG
CAGACTCCAGCGGCAGGACCTCCTGAACGCCTGCGCCTTTCGTGCGGGCGTCTAAAGTAA
TCCTCGAGG

AI378797

GCGGCCGCGGCCACCACCAACTGCTCGCCACCGACCCCACTACTCGCCACCGACCCGCT
GCTCGGAGCTTCGGTTCTGCGGGTTGTCCAGACTTCAGGCCTGTGCGCTCAATCGTGGAG
AATGCGCCGGCAGGCCCCCCCCACCCCAAGGTGCAGGAAGGACCAGCACGAACCCG
CTGGCTTTGCTGCGCGGCCAGGAGATGAGTCCCACCGGGCACTGAGCCAGGTACAGGAC
ATCAGAGAATGAACACAGAGGCAGAGGCCCTCATGTCCCTCTCAGAGTCCCGGCTCTGCA
NAGAGCCCGTCTGTCTCCAGCTTCCAGAATTCCGCACTGTGAATCTGTCTACGTGGACTG
GGAAAACAGGGTTGGCACCCTCTGCCACTCCGTTTGTGCTGGGAAGGGCTAAGTATGC
AAGGCTACAAACATCTACTTCACTGGGATCCCAAATGCTCAACAAACCATGACCTGCTNT
GGTCAGAACCACAGAAATATT

AA234220

GCAGGCGACTTGCGAGCTGGGAGCACTTTAAACGCTTTGGATTCCCCCGGCCTGGGTGG
GGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCATGAGCCGACCCTCG
GCTCCATGGAGCCTGGCATATTATGCCACCTTGGTATGGAGCCAAGGATATCGAAGGCTT
GCTGGGAGCGGGAGGGGGGCGGAATCTGGTGGCCACTCCCCTCTGACCAGCCACCCAGC
GGCGCCTACGCTGATGCCTGCTGTCAACTATGCCCCCTTGA

AA236353

GCCCCGTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACCTGGTGGGTTCTGTTCTCCCT
GGCAACACATCTGGCTGTTCCAGCCACCAGCGAGACGCCAAGACTGGTAACTGTCCACAG
GCAACAGGGAGTCATGTCGCGGTTCTCCAGGAGCACCAGAGTCTGCACCACAGACACGT
CCAGGTAACCTGGCCATAGGTNGGTAGGTTCCCGGATATCCCGGATAGAAGGCAAACTCAG
TGGGGCGGCTGGGGTACTCTTCCCCGGCCGTGGGAGTCTCCGCGGGGTACGCGCACAGGG
TGGCTGCCTGGGCACAGGGTTTCAGCGAGCTCCGGGACACTCGGCAGGAGTAGTACCCGC
CTCCAAAGTAACCATAAGGCA

AA588193

AACTGCTCGCCACCGACCCCACTACTCGCCACCGACCCGCTGCTCGGAGCTTCGGTTCG
CGGGTTGTCCAGACTTCAGGCCTGTGCGCTCAATCGTGGAGAATGCGCCGGCAGCCCCCA
CCCCCAGCCTAAGGTGCAGGAAGGACCAGCACGAACCCGCTGGCTTTGCTGCGCGGCCAG
GAGATGAGTCCCACCGGCACTGAGCCAGGTACAGGACATCAGAGAATGAACACAGAGG
CAGAGGCCCTCATGTCCCTCTCAGAGTCCCGGCTCTGCAAAGAGCCCGTCTGTCTCCAGC
TTCCAGAATTCGCACTGTGAATCTGTCTACGTGGACTGGGAAAACAGGGTTGGCACCAC
TCTGCCACTCCGTTTGTGCCTGGGAAGGGCTAAGTATGCAAGGCT

AI821103

GATCCCTTTGCAGGGAAGCTTTCTCTCAGACCCCCCTCCATTACACCTCTCACCCCTGGTA
ACAGCAGGAAGACTGAGGAGAGGGGAACGGGCAGATTCGTTGTGTGGCTGTGATGTCCGT
TTAGCATTTTTCTCAGCTGACAGCTGGGTAGGTGGACAATTGTAGAGGCTGTCTCTTCCT
CCCTCCTTGTCCACCCCATAGGGTGTACCCACTGGTCTTGGAAGCACCCATCCTTAATAC
GATGATTTTTCTGTGCTGTGAAAATGAAGCCAGCAGGCTGCCCCCTAGTCAGTCTTCCTT
CCAGAGAAAAAGAGATTGAGAAAAGTGA

AI821851

TTTTTTTTTTTTTTTTTTTTCTTTTTCACTTTCTCAAATCTCTTTTTCTCTGGAAGGAAG
GACTGACTAGGGGCAGCCTGCTGGCTTCATTTTCACACGACAAAAAATCATCGTATTAA
GGATGGGTGCTTCCAAAACCAGTGGGTACACCCATATGGGGGGGACAAGGAGGGAGGAAGA
GACAGCCTCTACAATTGTCCACCTACCCAGCTGTGAGCTGAGAAAAATGCTAAACGGACA
TCACAGCCACACAACGAATCTGCCCGTTCCCCTCTCCTCAGTCTTCCTGCTGTTACCAGG
GTGAGAGGTGTAATGGAAGG

AA635855

TTTTTTTTTTTTTTTTTTTTCTTTTTCACTTTCCCAAATCTCTTTTTCTCTGGAAGGAAG
GACTGACTAGGGGCAGCCTGCTGGCTTCATTTTCACACGACAGAAAAATCATCGTATTAA
GGATGGGTGCTTCCAAAGACCAGTGGGTACACCCATATGGGGTGGACACAGGAGGGAGGAAG
AGACAGCCTCTACAATTGTCCACCTACCCAGCTGTGAGCTGAGAAAAATGCTAAACGGAC
ATCACAGCCACACAACGAATCTGCCCGTTCCCCTCTCCTCAGTCTTCCTGCTGTTACCAG

GGTGAGAGGTGTAATGGAAGG

AI420753

GCGGCCGCGGCCACCACTGCTCGCCACCGACCCCACTACTCGCCACCGACCCGCT
GCTCGGAGCTTCGGTTCGCGGGTTGTCCAGACTTCAGGCCTGTGCGCTCAATCTTGAG
AATGCGCCGGCAGGCCCCCAGCCTAAGGTGCAGGAAGGACCAGCACGAACCCG
CTGGCTTTGCTGCGCGCCAGGAGATGAGTCCCACCGGGCACTGAGCCCAGGTACAGGAC
ATCAGAGAATGAACACAGAGGCAGAGGCCCTCATGTCCCTCTCAGAGTCCCGGCTCTGCA
AAGAGCCCGTCTGTCTCCAGCTTCCAGAATCCGCACTGTGAATCTGTCTACGT

BG180547

CACGCGTCGATCCCAGTGAAGTAGATGTTTGTAGCCTTGCACTTAGTCCTTCCCAGGC
ACAAACGGAGTGGCAGAGTGGTGCCAACCTGTTTTCCAGTCCACGTAGACAGATTAC
AGTGCGGAATTCTGGAAGCTGGAGACAGACGGGCTCTTTCAGAGCCGGGACTCTGAGAG
GGACATGAGGGCCTCTGCCTCTGTGTTCACTCTCTGATGCTGTACCTGGGCTCAGTGC
CCGGTGGGACTCATCTCCTGGCCGCGCAGCAAAGCCAGCGGGTTCGTGCTGGTCTTCCT
GCACCTTAGGCTGGGGTGGGGGCTGCGGCGCATTCTCCACGATTGAGCGCACAGGC
CTGAAGTCTGGACAACCCGAGAACCGAAGCTCCGAGCAGCGGTCGGTGGCGAGTAGTG
GGGTGGTGGCGAGCAGTTGGTGGTGGG

AA468306

TCGACCTCGCCAAGGTGAAGAACAACGCTACCCCTTAAGAGATCTCCTTGCCTGGGTGGG
AGGAGCGAAAGTGGGGTGTCTGGGGAGACCAGGAACCTGCCAAGCCCAGGCTGGGGCC
AAGGACTCTGCTGAGAGGCCCTAGAGACAACACCTTCCCAGGCCACTGGCTGCTGGAC
TGTTCTCAGGAGCGGCCTGGGTACCCAGTATGTGCAGGGAGA

AA468232

TTTTTTACTGGTTATCGTGGTTATTGCCACTGTGAGGATGAATGATTATGACTGGGCCAG
GTTCTTTGGGAACCTGGTGGAGTGGGCTGTACATGGGGTTCCGTCTCCCTGCACATAC
TGGGTACCCAGGCCGCTCCTGAGGAACAGTCCAGCAG

CB050115

GGCCACCACTGCTCGCCACCGACCCCACTACTCGCCACCGACCCGCTGCTCGGAG
CTTCGGTTCGCGGGTTGTCCAGACTTCAGGCCTGTGCGCTCAATCGTGGAGAATGCGCC
GGCAGGCCCCCAGCCTAAGGTGCAGGAAGGACCAGCACGAACCCGCTGGCTTT
GCTGCGCGCCAGGAGATGAGTCCCACCGGGCACTGAGCCCAGGTACAGGACATCAGAGA
ATGAACACAGAGGCAGAGGCCCTCATGTCCCTCTCAGAGTCCCGGCTCTGCAAAGAGCCC
GTCTGTCTCCAGCTTCCAGAATCCGCACTGTGAACCTCGTGCC

CB050116

GGCACGAGGTTTACAGTGCGGAATTCTGGAAGCTGGAGACAGACGGGCTCTTTGCAGAGC
CGGGACTCTGAGAGGGACATGAGGGCCTCTGCCTCTGTGTTTCAATCTCTGATGTCTGTGTA
CCTGGGCTCAGTGCCCGGTGGGACTCATCTCCTGGCCGCGCAGCAAAGCCAGCGGGTTG
TGCTGGTCTTCTGACCTTAGGCTGGGGGTGGGGGGCCTGCCGCGCATTTCTCCACGA
TTGAGCGCACAGGCCTGAAGTCTGGACAACCCGAGAACCGAAGCTCCGAGCAGCGGGT
GGTGGCGAGTAGTGGGGTGGTGGCGAGCAGTTGGTGGTGGGCC

AA661819

GCTGCTCGGAGCTTCGGTTCTGCGGGTTGTCCAGACTTCAGGCCTGTGCGCTCAATCGTG
GAGAATGCGCCGGCAGCCCCACCCAGCCTAAGGTGCAGGAAGGACCAGCACGAACCC
GCTGGCTTTGCTGCGCGGCCAGGAGATGAGTCCCACCGGCACTGAGCCAGGTACAGGACA
TCAGAGAATGAACACAGAGGCAGAGGCCTCATGTCCCTCTCAGAGTCCCGGCTCTGCAA
GAGCCGTACTGTCTCCAGCTTCCAGAATTCCGCACTGTGAATCTGTCTACGTGGACTGGG
AAAC

CF146837

CACGAGGATTTTCTATCTAGAGCTCTGTAGAGCACTTTAGAAACCGCTTTTATGAATTGA
GCTAATTATGAATAAATTTGGAAGGCGATCCCTTTGCAGGGAAGCTTTCTCTCAGACCCC
CTTCCATTACACCTCTCACCTGGTAACAGCAGGAAGACTGAGGAGAGGGGAACGGGCAG
ATTGCTTGTGTGGCTGTGATGTCCGTTTAGCATTTTTCTCAGCTGACAGCTGGGTAGGTG
GACAATTGTAGAGGCTGTCTCTTCCCTCCCTCCTTGTCCACCCCATAGGGTGTACCCACTG
GTCTTGGAACACCCATCCTTAATACGATGATTTTCTGTCTGTGAAAATGAAGCCAGC
AGGCTGCCCCCTAGTCAGTCTTCCCTCCAGAGAAAAAGAGATTTGAGAAAGTGCCTGGGT
AATTCACCATTAATTTCTCCCCCAAACCTCTCTGAGTCTTCCCTTAATATTTCTGGTGGT
TCTGACCAAAGCAGGTATGGTTTGTGAGCATTGCGGATCCAGTGAAGTAGATGTTTG
TAGCCTTGCTACTTAGCCCTTCCCAGGCACAAACGGAGTGGCAGAGTGGTGCCAACCCCT
GTTTTCCAGTCCACGTAGACAGATTACAGTGCGGAATTCTGGAAGCTGGAGACAGACG
GGCTCTTTGCAGAGCCGGGACTCTGAG

CF146763

CACGAGGATTTTCTATNCTAGAGCTCTGGTAGAGCACTTTANAAACCGCTTTTATGAATT
GAGCTAATTATGAATAAATTTGGAAGGCGATCCCTTTGCAGGGAAGCTTTCTCTCAGACC
CCCTTCCATTACACCTCTCACCTGGTAACAGCAGGAAGACTGAGGAGAGGGGAACGGGC
AGATTGCTTGTGTGGCTGTGATGTCCGTTTAGCATTTTTCTCAGCTGACAGCTGGGTAGG
TGGACAATTGTAGAGGCTGTCTCTTCCCTCCCTCCTTGTCCACCCCATAGGGTGTACCCAC
TGGTCTTGGAACACCCATCCTTAATACGATGATTTTCTGTCTGTGAAAATGAAGCCA
GCAGGCTGCCCCCTAGTCAGTCTTCCCTCCAGAGAAAAAGAGATTGAGAAAGTGCCTGGG
TAATTCACCATTAATTTCTCCCCCAAACCTCTCTGAGTCTTCCCTTAATATTTCTGGTGG
TTCTGACCAAAGCAGGTATGGTTTGTGAGCATTGCGGATCCAGTGAAGTAGATGTTT
GTAGCCTTGCTACTTAGCCCTTCCCAGGCACAAACGGAGTGGCAGAGTGGTGCCAACCC
TGTTTTCCAGTCCACGTAGACAGATTACAGTGCGGAATTCTGGAAGCTGGAGACAGAC
GGGCTCTTTGCAGAGCCGGGACTCTGA

CF144902

CACGAGGGAAGCCAGCAGGCTGCCCCCTAGTCAGTCCTTCCTTCCAGAGAAAAAGAGATTT
GAGAAAGTGCCCTGGGTAATTCACCATTAATTTCTCCCCCAAACCTCTCTGAGTCTTCCCT
TAATATTTCTGGTGGTTCTGACCAAAGCAGGTCATGGTTTGTGAGCATTGTTGGGATCCCA
GTGAAGTAGATGTTTGTAGCCTTGACATACTTAGCCCTTCCCAGGCACAAACGGAGTGCGA
GAGTGGTGCCAAACCCTGTTTTCCAGTCCACGTAGACAGATTACAGTGCGGAATTCTGG
AAGCTGGAGACAGACGGGCTCTTTCAGAGCCGGGACTCTGAGAGGGACATGAGGGCCTC
TGCTCTGTGTTTATTCTCTGATGTCCTGTACCTGGGCTCAGTGCCCGGTGGGACTCATC
TCCTGGGCGCGCAGCAAAGCCAGCGGGTTCGTGCTGGTCCTTCCTGCACCTTA

CF141511.1

CACGAGGCCTGGTAACAGCAGGAAGACTGAGGAGAGGGGAACGGGCAGATTGTTTGTGTG
GCTGTGATGTCCGTITAGCATTTTTCTCAGCTGACAGCTGGGTAGGTGGACAATTGTAGA
GGCTGTCTCTTCTCCCTCCTTGTCACCCCATAGGGTGTACCCACTGGTCTTGAAACA
CCCATCCTTAATACGATGATTTTTCTGTGCTGTGAAAATGAAGCCAGCAGGCTGCCCCTA
GTCAGTCCTTCTTCCAGAGAAAAAGAGATTTGAGAAAGTGCTGGGTAATTCACCATTA
ATTTCTCCCCCAAACCTCTCTGAGTCTTCCCTTAATATTTCTGGTGGTTCTGACCAAAGC
AGGTCATGGTTTGTGAGCATTGTTGGATCCCAGTGAAGTAGATGTTTGTAGCCTTGACATA
CTTAGCCCTTCCCAGGCACAAACGGAGTGGCAGAGTGGTGCCAACCCTGTTTTCCAGTC
CACGTAGACAGATTACAGTGCGGAATTCTGGAA

CF139563.1

CACGAGGTCTTCCCTTAATATTTCTGGTGGTTCTGACCAAAGCAGGTCATGGTTTGTGTA
GCATTTGGGATCCCAGTGAAGTAGATGTTTGTAGCCTTGACATACTTAGCCCTTCCCAGGC
ACAAACGGAGTGGCAGAGTGGTGCCAACCCTGTTTTCCAGTCCACGTAGACAGATTAC
AGTGCGGAATTCTGGAAGCTGGAGACAGACGGGCTCTTTCAGAGCCGGGACTCTGAGAG
GGACATGAGGGCCTCTGCCTCTGTGTTTATTCTCTGATGTCCTGTACCTGGGCTCAGTGC
CCGGTGGGACTCATCTCTGGCCGCGCAGCAAAGCCAGCGGGTTCGTGCTGGTCCTTCCT
GCACCTTAGGCTGGGGGTGGGGGGCCTGCCGGCGCATTCTCCACGATTGAGCGCACAGGC
CTGAAGTCTGGACAACCCGCAGAACCGAAGCTCCGAGCAGCGGGTGGTGGCGAGTA

CF139372

CACGAGGATTTCTGGTGGTTCTGACCAAAGCAGGTCATGGTTTGTGAGCATTGTTGGGATC
CCAGTGAAGTAGATGTTTGTAGCCTTGACATACTTAGCCCTTCCCAGGCACAAACGGAGTG
GCAGAGTGGTGCCAACCCTGTTTTCCAGTCCACGTAGACAGATTACAGTGCGGAATTCT
TGGAAGCTGGAGACAGACGGGCTCTTTCAGAGCCGGGACTCTGAGAGGGACATGAGGGC
CTCTGCCTCTGTGTTTATTCTCTGATGTCCTGTACCTGGGCTCAGTGCCCGGTGGGACTC
ATCTCTGGCCGCGCAGCAAAGCCAGCGGGTTCGTGCTGGTCCTTCCTGCACCTT

CF139319

CACGAGGAAGGCGATCCCTTTGCAGGGAAGCTTTCTCTCAGACCCCCTTCCATTACACCT
CTCACCCCTGGTAACAGCAGGAAGACTGAGGAGAGGGGAACGGGCAGATTGTTTGTGTGGC
TGTGATGTCCGTTTAGCATTTTTCTCAGCTGACAGCTGGGTAGGTGGACAATTGTAGAGG
CTGTCTCTTCTCCCTCCTTGTCACCCCATAGGGTGTACCCACTGGTCTTGAAACACC
CATCCTTAATACGATGATTTTTCTGTGCTGTGAAAATGAAGCCAGCAGGCTGCCCCTAGT
CAGTCCTTCTTCCAGAGAAAAAGAGATTTGAGAAAGTGCTGGGTAATTCACCATTAAT
TTCCTCCCCCAAACCTCTCTGAGTCTTCCCTTAATATTTCTGGTGGTTCTGACCAAAGCAG
GTCATGGTTTGTGAGCATTGTTGGATCCCAGTGAAGTAGATGTTTGTAGCCTTGACATACT
TAGCCCTTCC

CF139275

CACGAGGTGGATTCCCCCGGCCTGGGTGGGGAGAGCGAGCTGGGTGCCCCCTAGATTCCC
CGCCCCCGCACCTCATGAGCCGACCCTCGGCTCCATGGAGCCCGGCAATTATGCCACCTT
GGATGGAGCCAAGGATATCGAAGGCTTGCTGGGAGCGGGAGGGGGGCGGAATCTGGTCGC
CCACTCCCCCTCTGAGCAGCCACCCAGCGGCGCCTACGCTGATGCCTGCTGTCAACTATGC
CCCCTTGGATCTGCCAGGCTCGGCGGAGCCGCCAAAGCAATGCCACCCATGCCCTGGGGT
GCCCCAGGGGACGTCCCCAGCTCCCGTGCCTTATGGTTACTTTGGAGGCGGGTACTACTC
CTGCCGAGTGTGCGGGAGCTCGCTGAAACCTGTGCCAGGCA

CF122893

CACGAGGATTTTCTATCTAGAGCTCTGTAGAGCACTTTAGAAACCGCTTTTCATGAATTGA
GCTAATTATGAATAAATTTGGAAGGCGATCCCTTTGCAGGGAAGCTTTCTCTCAGACCCC
CTTCCATTACACCTCTCACCCTGGTAACAGCAGGAAGACTGAGGAGAGGGGAACGGGCAG
AFTCGTTGTGTGGCTGTGATGTCCGTTTAGCATTTTTCTCAGCTGACAGCTGGGTAGGTG
GACAATTGTAGAGGCTGTCTCTTCCCTCCCTCCTTGTCACCCCATAGGGTGTACCCACTG
GTCTTGGAACACCCATCCTTAATACGATGATTTTTCTGTCGTGTGAAAATGAAGCCAGC
AGGCTGCCCCCTAGTCAGTCCTTCCTTCCAGAGAAAAAGAGATTTGAGAAAGTGCCTGGGT
AATTCACCATTAAATTTCTCCCCAAACTCTCTGAGTCTTCCCTTAATATTTCTGGTGGT
TCTGACCAAAGCAGGTGATGGTTTGTGAGCATTGGGATCCCAGTGAAGTANATGTTTG
TAGCCTTGCACTTAGCCCTT

AI972423

CATTTTCACACGACTGTAAATCATCGTATTAAAGGATGGGTGCTTCCAAGACCAGTGGGT
ACACCCTATGGGGTGGACAAGGAGGGAGGAAGAGACAGCCTCTACAATTGTCCACCTACC
CAGCTGTCAGCTGAGAAAAATGCTAAACGGACATCACAGCCACACAACGAATCTGCCCCGT
TCCCCCTCTCCTCAGTCTTCCCTGCTGTTACCAGGGTGAGAGGTGTAATGGAAGGGGGTCTG
AGAGAAAGCTTCCCTGCAAAGGGATCGCCTTCCAAATTTATTCATAATTAGCTCAATTCA
TGAAAGCGGTTTCTAAAGTGCTCTACAGAGCTCTAGATAGAAAATATGAGGCTAACGATC
ATGGCAGCTAGTACTGGTTATCGTGATTATTGCCACTGTCAGGATGAATGATTATGACTG
GGCCAGGTTCTTTGGGAACCCCTGGTGGAGTGGGCTGTCACATG

AI918975

TGCAGCTAGTACTGGTTATCGTGATTATTGCCACTGTCAGGATGAATGATTATGACTGGG
CCAGGTTCTTTGGGAACCCCTGGTGGAGTGGGCTGTCACATGGGGTTCCGTCTCCCTGCAC
ATACTGGGTACCCAGGCCGCTCCTGAGGAACAGTCCAGCACAGGGTTTCAGCGAGCTCCG
GGACACTCGGCCTCGTGC

AI826991

TTTTTTTTTTTTTTTTTTTTCTTTTTCACTTTCTCAAATCTTTTTCTCTGGAAGGAAG
GACTGACTAGGGGCAGCCTGCTGGCTTCATTTTTCACACCACAAAAAATCATCGTATTAA
GGATGGGTGCTTCCAAAACCAAGTGGGTACACCCTATGGGGTGGACAAGGAGGGAGGAAAA
AACAGCCTCTACAATTGTCCACCTACCCAGCTGTCAGCTGAAAAAATGCTAAACGGACA
TCACAGCCACACAACGAATCTGCCGTTCCCTCTCCTCAGTCTTCCCTGCTGTTACCAGG
GTGAAAGGTGTAATGGAAGG

AI686312

ACCGACCCCACTACTTGCCACCGACCCGCTGCTCGGAGCTTCGGTTCTGCGGGTTGTCCA
 GACTTCAGGCCTGTGCGCTCAATCGTGGAGAATGCGCCGGCAGGCCCCCCACCCCCAGCC
 TAAGGTGCAGGAAGGACCAGCACGAACCCGCTGGCTTTGCTGCGCGGCCAGGAGATGAGT
 CCCACCGGGCACTGAGCCCAGGTACAGGACATCAGAGAATGAACACAGAGGCAGAGGCC
 TCATGTCCCTCTCAGAGTCCCGGCTCTGCAAAGAGCCCGTCTGTCTCCAGCTTCCAGAAT
 TCCGCACTGTGAATCTGTCTACGTGGACTGGGAAAACAGGGTTGGCACCCTCTGCCACT
 CCGTTTGTGCCTGGGAAGGGCTAAGTATGCAAGGCTACAAACATCTACTTCACTGGGATC
 C

AI655923

TTTTTTTTTTTTTTTTCCCTGCAAAGGGATCGCCTTCCAAATTTATTCATAATTAGCTCAA
 TTCATGAAAGCGGTTTCTAAAGTGCTCTACAGAGCTCTAGATAGAAAATATGAGGCTAAC
 GATCATGGCAGCTAGTACTGGTTATCGTGATTATTTGCCACTGTGAGGATGAATGATTATG
 ACTGGGCCAGGTTCTTTGGGAACCCCTGGTGGAGTGGGCTGTACATGGGGTTCCGTCTCC
 CTGCACATACTGGGTACCCAGGCCGCTCCTGA

CF146922

CACGAGGCGACTTGCGAGCTGGGAGCGATTTAAAACGCTTTGGATTCCCGGGCCTGGGTG
 GGGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCATGAGCCGACCCCTC
 GGCTCCATGGAGCCCGCAATTATGCCACCTTGGATGGAGCCAAGGATATCGAAGGCTTG
 CTGGGAGCGGGAGGGGGGCGGAATCTGGTTCGCCCACTCCCTCTGACCAGCCACCCAGCG
 GCGCCTACGCTGATGCCTGCTGTCAACTATGCCCCCTTGGATCTGCCAGGCTCGGCGGAG
 CCGCCAAAGCAATGCCACCCATGCCCTGGGGTGCCCCAGGGGACGTCCCCAGCTCCCGTG
 CCTTATGGTTACTTTGGAGGCGGGTACTACTCCTGCGGAGTGTCCCGGAGCTCGCTGAAA
 CCCTGTGCCCAGGCAGCCACCCTGGCCGCGTACCCCGCGGAGACTCCACGGCCGGGGAA
 GAGTACCCAGCCGCCCCACTGAGTTTGCCTTCTATCCGGGATATCCGGGAACCTACCAG
 CCTATGGCCAGTTACCTGGACGTGTCTGTGGTGACAGCTCTGGGTGCTCCTGGAGAACGC
 GACATGACTCCCTGTTGCCTGTGGACAGTTACCAGTCTTGGGCTCTCGCTGGTGGCTGGA
 ACAGCCAGATGTGTTGCCA

BF476369

GCGGCCGCGGCCACCACCAACTGCTCGCCATTGACCCGACTACTCGCCACCGACCCGC
 TGCTCGGAGCTTCGGTTCTGCGGGTTGTCCAGACTTCAGGCCTGTGCGCTCAATCGTGGA
 GAATGCGCGGCAGGCCCCCCACCCCCAGCCTAAGGTGCAGGAAGGACCAGCACGAACCC
 GCTGGCTTTGCTGCGCGGCCAGGAGATGAGTCCACCGGGCACTGAGCCCAGGTACAGGA
 CATCAGAGAATGAACACAGAGGCAGAGGCCCTCATGTCCCTCTCAGAGTCCCGGCTCTGC
 AAAGAGCCCGTCTGTCTCCAGCTTCCAGAATTCCGCACTGTGAATCTGTCTACGTGGACT
 GGGAAAACAGGGTTGGCACCCTCTGCCACTCC

BF057410

GCGGCCGCGGCCACCACCAACTGCTCGCCACCGACCCCACTACTCGCCACCGACCCGCT
 GCTCGGAGCTTCGGTTCTGCGGGTTGTCCAGACTTCAGGCCTGTGCGCTCAATCGTGAG
 AATGCGCCGGCAGGCCCCCCACCCCCAGCCTAAGGTGCAGGAAGGACCAGCACGAACCCG
 CTGGCTTTGCTGCGCGGCCAGGAGATGAGTCCACCGGGCACTGAGCCCAGGTACAGGAC
 ATCAGAGAATGAACACAGAGGCAGAGGCCCTCATGTCCCTCTCAGAGTCCCGGCTCTGCA
 AAGAGCCCGTCTGTCTCCAGCTTCCAGAATTCCGCACTGTGAATCTGTCTACGTGGACTG
 GGAAAACAGGGTTGGCACCCTCTGCCACTCCGTTTGTGCCTGGGAAGGGCTAAGTATGC
 AAGGCTACAAACATCTACTTCACTGGGATCCCAAATGCTCAACAAACCATGACCTGCTNT
 GGTGAGAACCACAGAAATATTAA

BE645544

GCGGCCGCGGCCACCACTGCTCGCCACCGACCCCACTACTCGCCACCGACCCGCT
 GCTCGGAGCTTCGGTTCTGCGGGTTGTCCAGACTTCAGGCCTGTGCGCTCAATCGTGGAG
 AATGCGCCGGCAGGCCCCCAGCCCTAAGGTGCAGGAAGGACCAGCACGAACCCG
 CTGGCTTTGCTGCGCGGCCAGGAGATGAGTCCACCGGGCACTGAGCCCAGGTACAGGAC
 ATCAGAGAATGAACACAGAGGCAGAGGCCCTCATGTCCCTCTCAGAGTCCCGGCTCTGCA
 AAGAGCCCGTCTGTCTCCAGCTTCCAGAATTCCGCACTGTGAATCTGTCTACGTGGACTG
 GGAAAACAGGGTTGGCACCCTCTGCCACTCCGTTTGTGCCTGGGAAGGGCTAAGTATGC
 AAGGCTACAAACATCTACTTCACTGGGATCC

BE645408

TCCTCCCTCTAAGAAAGGCGCAAGCGTCAAGAGGGTGCTGCCCGCTGGTTTCTGCAAATG
 CTGCCTTCAAAAAGGACCTGGTGGGTTCTGTTCTCCCTGGCAACACATCTGGCTGTTCC
 AGCCACCAGCGAGAGCCCAAGACTGGTAAGTGTCCACAGGCAACAGGGAGTCATGTGCGG
 GTTCTCCAGGAGCACCCAGAGTCTGCACCACAGACACGT

BE388501

TTAATACGATGATTTTTCTGTCGTGTGAAAATGAAGCCAGCAGGCTGCCCCCTAGTCAGTC
 CTTCCCTCCAGAGAAAAAGAGATTTGAGAAAGTGCCTGGGTAATTCACCATTAATTTCCCT
 CCCCCAACTCTCTGAGTCTTCCCTTAATATTTCTGGTGGTTCAGACCAAAGCAGGTCA
 GGTTCGTTGAGCATTGTTGGGATCCAGTGAAGTAGATGTTGTAGCCTTGACATACTTAGCC
 CTTCCCAGGCACAAACGGAGTGGCAGAGTGGTGCCAAACCTGTTTTCCAGTCCACGTAG
 ACAGATTCACAGTGCAGGAATCTGGAAGCTGGAGACAGACGGGCTCTTTCAGAGCCGGG
 ACTCTGAGAGGGACATGAGGGCCTCTGCCTCTGTGTTCAATCTCTGATGTCTGTACCTG
 GGCTCAGTGCCCGGTGGGACTCATCTCCTGGCCGCGCAGCAAAGCCAGCGGGTTCGTGCT
 GGTCTTCCCTGCACCTTAGGCTGGGGGTGGGGGGCCTGCCGGCGCATTCTCCACGATTGA
 GCGCACAGGCCTGAAGTCTGGACAACCCGAGAACCGAAGCTCCGAGCAGCGGGTCGGTG
 GCGAGTAGTGGGGGTGGTGGCGAACAAAGTGGTGGTGGGCGGGGCCGATAACTCGAGG
 ACTTTCCTCCCGGAGCAGTCCCTAAAAACCGGGGGCGC

CF147366

GACGAGGACAATTGTAGAGGCTGTCTCTTCCCTCCCTCCTGTACCCCATAGGGTGTACC
 ACTGGTCTTGAAGCACCCATCCTTAATACGATGATTTTTCTGTCGTGTGAAAATGAAGC
 CAGCAGGCTGCCCCTAGTCAGTCTTCCCTCCAGAGAAAAAGAGATTTGAGAAAGTGCCT
 GGGTAATTCACCATTAATTTCTCCCCAACTCTCTGAGTCTTCCCTTAATATTTCTGG
 TGGTCTGACCAAAGCAGGTCAAGTTTGTGAGCATTGTTGGGATCCAGTGAAGTAGATG
 TTTGTAGCCTTGACATACTTAGCCCTTCCAGGCACAAACGGAGTGGCAGAGTGGTGCCAA
 CCCTGTTTTCCAGTCCACGTAGACAGATTCACAGTGCAGGAATCTGGAAGCTGGAGACA
 GACGGGCTCTTTCAGAGCCGGGACTCTGAGAGGGACATGAGGGCCTCTGCCTCTGTGTT
 CAATCTCTGATGTCTGTACCTGGGCTCAGTGCCCGGTGGGACTCATCTCCTGGCCGCGC
 AGCAAAGCCAGCGGGTTCGTGCTGGTCTTCTCTGC

CF147143

CACGAGGCGACTTGCGAGCTGGGAGCGATTTAAAACGCTTTGGATTCCCCCGGCCTGGGT
 GGGGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCATGAGCCGACCCT
 CGGCTCCATGGAGCCCGGCAATTATGCCACCTTGGATGGAGCCAAGGATATCGAAGGCTT
 GCTGGGAGCGGGAGGGGGGCGGAATCTGGTCGCCCCACTCCCCTCTGACCAGCCACCCAGC
 GCGGCCTACGCTGATGCCTGCTGTCAACTATGCCCCCTTGGATCTGCCAGGCTCGGCGGA
 GCCGCCAAAGCAATGCCACCCATGCCCTGGGGTGCCCCAGGGGACGTCCCCAGCTCCCCGT
 GCCTTATGGTTACTTTGGAGGCGGGTACTACTCCTGCCGAGTGTCCCGGAGCTCGCTGAA
 ACCCTGTGCCCAGGCAGCCACCCTGGCCGCGTACCCCCGCGGAGACTCCACGGCCGGGGA
 AGAGTACCCAGCCGCCCCACTGAGTTTGCTTCTATCCGGGATATCCGGGAACCTACCAG
 CCTATGGCCAGTTACCTGGACGTGTCTGTGGTGACAGTCTGGGTGCTCCTGGAGAACGC
 GACATGACTCCCTGTTGCCTGTGGACAGTTACCAATCTTGGGCTCTCGCTGGTGGCTGGA
 ACAGCCAGATGTGTTGCCAGGGAG

BT007410

atggagcccg	gcaattatgc	caccttggat	ggagccaagg	atatcgaagg	cttgcctggga
gcgggagggg	ggcggaatct	ggtcgccac	tccctctga	ccagccacc	agcggcgcct
acgctgatgc	ctgctgtcaa	ctatgcccc	ttggatctgc	caggctcggc	ggagccgcca
aagcaatgcc	accatgccc	tggggtgccc	caggggacgt	ccccagctcc	cgtgccttat
ggttactttg	gaggcgggta	ctactcctgc	cgagtgtccc	ggagctcgct	gaaacctgt
gcccaggcag	ccacctggc	cgcgtacccc	gcggagactc	ccacggccgg	ggaagagtac
cccagccgcc	ccactgagtt	tgccttctat	ccgggatata	cgggaacctc	ccagcctatg
gccagttacc	tggacgtgtc	tgtggtgcag	actctgggtg	ctcctggaga	accgcgacat
gactccctgt	tgcctgtgga	cagttaccag	tcttgggctc	tcgctggtgg	ctggaacagc
cagatgtgtt	gccagggaga	acagaacca	ccaggtccct	tttgggaagg	agcatttgca
gactccagcg	ggcagcacc	tcctgacgcc	tgcgccttcc	gtcgcgcccg	caagaaacgc
attccgtaca	gcaaggggca	ggtgcgggag	ctggagcggg	agtatgcggc	taacaagtcc
atcaccaagg	acaagaggcg	caagatctcg	gcagccacca	gcctctcgga	gcgccagatt
accatctggt	ttcagaaccg	ccgggtcaaa	gagaagaagg	ttctcgccaa	ggtgaagaac
agcgtacccc	cttag				

BC007092


```

ggattccccc ggccctgggtg gggagagcga gctgggtgcc ccctagattc cccgcccccg
cacctcatga gccgacctc ggctccatgg agcccggcaa ttatgccacc ttggatggag
ccaaggatat cgaaggcttg ctgggagcgg gagggggggc gaatctggtc gccactccc
ctctgaccag ccaccagcg gcgcctacgc tgatgcctgc tgtcaactat gccccttgg
atctgccagg ctcgcgggag cgcgcaaagc aatgccacc atgccctggg gtgccccagg
ggacgtcccc agctcccgtg ccttatgggtt actttggagg cgggtactac tctgcccag
tgtcccgagg ctcgctgaaa cctgtgccc aggcagccac cctggccgcg taccgcggg
agactccac ggccggggaa gagtaccca gcgccccac tgagtttgcc ttctatccg
gatatccggg aacctaccag cctatggcca gttacctgga cgtgtctgtg gtgcagactc
tgggtgctcc tggagaaccg cgacatgact ccctgttgcc tgtggacagt taccagtctt
gggctctcgc tgggtggctgg aacagccaga tgtgttgcca gggagaacag aaccaccag
gtcccttttg gaaggcagca tttgcagact ccagcgggca gcaccctcct gacgcctgcg
cctttcgtcg cggcgcaag aaacgcattc cgtacagcaa ggggcagttg cgggagctgg
agcgggagta tgcggctaac aagttcatca ccaaggacaa gaggcgcaag atctggcag
ccaccagcct ctcgagcgc cagattacca tctggtttca gaaccgccc gtcaaagaga
agaagggtct cgccaagggt aagaacagcg ctaccctta agagatctcc ttgctgggt
gggaggagcg aaagtggggg tgtcctgggg agaccaggaa cctgccaaag ccaggctggg
gccaaaggat ctgctgagag gcccttagag acaacacct tcccaggcca ctggctgctg
gactgttctt caggagcggc ctgggtaccc agtatgtgca gggagacgga acccatgtg
acagcccact ccaccagggt tcccaaagaa cctggcccag tcataatcat tcatcctgac
agtggcaata atcacgataa ccagtactag ctgccatgat cgttagcctc atattttcta
tctagagctc tgtagagcac tttagaaacc gctttcatga attgagctaa ttatgaataa
atttggaana aaaaaaaaaa aaaaaaaaaa aaaaaa

```

U57052

```

cgggtgcccc ctagattccc cgcccccgca cctcatgagc cgaccctcgg ctccatggag
cccggcaatt atgccacctt ggatggagcc aaggatatcg aaggcttgct gggagcggga
ggggggcgga atctggctgc cactccct ctgaccagcc acccagcggc gcctaogctg
atgctgctg tcaactatgc ccccttggat ctgccaggct cggcgagcgc gccaaagcaa
tgccacccat gccctggggg gcccagggg acgtcccag ctcccgtgcc ttatgggttac
tttggaggcg ggtactactc ctgccgagtg tcccggagct cgtgaaacc ctgtgcccag
gcagccaccc tggccgcgta cccgcggag actcccacgg ccggggaaga gtacccacgc
cgccccactg agtttgctt ctatccggga tatccgggaa cctaccacgc tatggccagt
tacctggacg tgtctgtggt gcagactctg ggtgctcctg gagaaccgcg acatgactcc

```

```

ctgttgccctg tggacagtta ccagtcttgg gctctcgtcg gtggctggaa cagccagatg
tgttgccagg gagaacagaa cccaccagg cctttttgga aggcagcatt tgcagactcc
agcgggcagc accctcctga cgcctccgcc tttcgtcgcg gccgcaagaa acgcattccg
tacagcaagg ggcagttagc ggagctggag cgggagtatg cggctaacaa gttcatcacc
aaggacaaga ggcgcaagat ctgggcagcc accagcctct cggagcgcca gattaccatc
tggtttcaga accgcgggt caaagagaag aaggttctcg ccaagggtgaa gaacagcgct
accccttaag agatctcctt gcctgggtgg gaggagcgaa agtgggggtg tctgggggag
accaggaacc tgccaagccc aggctggggc caaggactct gctgagaggc ccctagagac
aacacc

```

U81599

tcctaatacgc actcactata gggctcgagc ggccgcccgc gcaggtcgaa tgcaggcgac
 ttgcgagctg ggagcgattt aaaacgcctt ggattccccc ggcctgggtg gggagagcga
 gctgggtgcc ccctagattc cccgccccgc cacctcatga gccgaccctc ggctccatgg
 agccccgcaa ttatgccacc ttggatggag ccaaggatat cgaaggcttg ctgggagcgg
 gagggggggc gaatctggtc gcccactccc ctctgaccag ccaccagcgc gcgcctacgc
 tgatgcctgc tgtcaactat gcccccttgg atctgccagg ctcggcggag ccgccaagc
 aatgccaccc atgccctggg gtgccccagg ggacgtcccc agtccccgtg ccttatgggt
 actttggagg cgggtactac tectgcagag tgtcccgag ctgctgaaa ccctgtgccc
 aggcagccac cctggccgcg taccgcgcgc agactccac ggccggggaa gagtacccca
 gtcgccccac tgagtttgcc ttctatccgc gatatccggg aacctaccac gctatggcca
 gttacctgga cgtgtctgtg gtgcagactc tgggtgctcc tggagaaccg cgacatgact
 ccctgttgcc tgtggacagt taccagtctt gggctctcgc tgggtgctgg aacagccaga
 tgtgttgcca gggagaacag aaccaccag gtcccttttg gaaggcagca tttgcagact
 ccagcgggca gcacctcct gacgcctgcg cctttcgtcg cggccgcaag aaacgcatto
 cgtacagcaa ggggcagttg cgggagctgg agcgggagta tgcggctaac aagtccatca
 ccaaggacaa gaggcgcaag atctcggcag ccaccagcct ctcgagcgc cagattacca
 tctggtttca gaaccgcgcg gtcaaagaga agaaggttct cgccaagggtg aagaacagcg
 ctaccctta agagatctcc ttgcctgggt gggaggagcg aaagtggggg tgtcctgggg
 agaccagaaa cctgccaaag ccaggctggg gccaaaggact ctgctgagag gcccttagag
 acaacacct tcccaggcca ctggctgctg gactgttct caggagcggc ctgggtaccc
 agtatgtgca gggagacgga accccatgtg acaggccac tccaccaggg ttcccaaaga
 acctggccca gtcataatca ttcactctca cagtggcaat aatcacgata accagt

CB120119

ATTTTCTGTGCTGTGAAAATGAAGCCAGCAGGCTGCCCCCTAGTCAGTCCTTCCTTCCAG
 AGAAAAAGAGATTTGAGAAAGTGCCCTGGGTAAATTCACCATTAATTTCCCTCCCCAAACTC
 TCTGAGTCTTCCCTTAATATTTCTGGTGGTCTGACCAAAGCAGGTCATGGTTTGTGAG
 CATTTGGGATCCAGTGAAGTAGATGTTTGTAGCCTTGCATACTTAGCCCTTCCCAGGCA
 CAAACGGAGTGGCAGAGTGGTGCCAACCCTGTTTTCCAGTCCACGTAGACAGATTCA
 GTGCGGAATTCTGGAAGCTGGAGACAGACGGGCTCTTTGCAGAGCCGGGACTCTGAGAGG
 GACATGAGGGCCTCTGCCTCTGTGTTTATTCTCTGATGTCCTGTACCTGGGCTCAGTGCC
 CGTGGGACTCATCTCCTGGCTGCGCAGCAAAGCCAGCGGGTTCGTGCTGGTCTCTCCTG
 CACCTTAGGCTGGGGGTGGGGGGCCT

CB125764

ATTTTCTGTGCTGTGAAAATGAAGCCAGCAGGCTGCCCCCTAGTCAGTCCTTCCTTCCAG
 AGAAAAAGAGATTTGAGAAAGTGCCCTGGGTAAATTCACCATTAATTTCCCTCCCCAAACTC
 TCTGAGTCTTCCCTTAATATTTCTGGTGGTCTGACCAAAGCAGGTCATGGTTTGTGAG
 CATTTGGGATCCAGTGAAGTAGATGTTTGTAGCCTTGCATACTTAGCCCTTCCCAGGCA

CAAACGGAGTGGCAGAGTGGTGCCAACCCTGTTTTCCAGTCCACGTAGACAGATTCA
 GTGCGGAATTCTGGAAGCTGGAGACAGACGGGCTCTTTGCAGAGCCGGGACTCTGAGAGG
 GACATGAGGGCCTCTGCCTCTGTGTTTATTCTCTGATGTCCTGTACCTGGGCTCAGTGCC
 CGTGGGACTCATCTCCTGGCTGCGCAGCAAAGCCAGCGGGTTCGTGCTGGTCTCTCCTG
 CACCTTAGGCTGGGGGTGGGGGGGCTGCCGGCGATTCTCCACGATTGAGCGCACAGG
 CCTGAAGTCTGGACAACCCGAGAACCGAAGCTCCGAGCAGCGGGTTCGTGGCGAGT

AU098628

ATTTAAACGCTTTGGATTCTTTCTGCTGCGTGCGGAGAGCGAGCTGGGTGCCCCCTAG
ATTCCCCGCCCCCGCACCTCATGAGCCGACCTCGGCTCCATGGAGCCCGGCACCTTATGC
CACCTTGGATGGAGCCAAGGATATCGAAGGCTTGCTGGGAGCGGGAGGGGGCGGAATCT
GGTCGCCCCACTCCCCCTCTGACCAGCCACCCAGCGGCGCCTACGCTGATGCCTGCTGTCAA
TTATGCCCCCTTGCACTCTGCCAGGCTCGGCGGAGCCGCCAAAGCAATGCCACCCATGCCC

CB126130

ATTTTTCTGTCTGTGAAAATGAAGCCAGCAGGCTGCCCCCTAGTCAGTCCTTCCTTCCAG
AGAAAAAGAGATTGAGAAAGTGCCTGGGTAAATTCACCATTAAATTCCTCCCCAAACTC
TCTGAGTCTTCCCTTAATATTTCTGGTGGTTCTGACCAAAGCAGGTCATGGTTTGTGAG
CATTTGGGATCCCACTGAAGTAGATGTTTGTAGCCTTGCACTTAGCCCTTCCAGGCA
CAAACGGAGTGGCAGAGTGGTGCCAACCTGTTTTCCAGTCCACGTAGACAGATTACA
GTGCGGAATTCGGAAGCTGGAGACAGACGGGCTCTTTGCAGAGCCGGGACTCTGAGAGG
GACATGAGGGCTCTGCCTCTGTGTTTATTCTCTGATGTCTGTACCTGGGCTCAGTGCC
CGGTGGGACTCATCTCTGGCTGCGCAGCAAAGCCAGCGGGTTCTGTCTGCTGCTCCTCTG
CACCTTAGGCTGGGGGTGGGGGGCCTGC

BI023924

AGGCCGCACCCAGTCTTAAGGTGCAGTGAAGGACAGCACGAACCCGCTGTGCTTTGCTGC
GCGGCAGGAGATGAGTCCCACCGGGCACTGAGCCAGGTACAGGACATCAGAGAATGAAC
ACAGAGGCAGAGGCCCTCATGTCCCTCTCAGAGTCCCGGCTCTGCAAAGAGCCCGTCTGT
CTCCAGCTTCCAGAATCCGCACTGTGAATCTGTCTACGTGGACTGNGAAACAGGGTTG
GCACCACTCTGCCACTCCGTTTGTGCCTNGGGGCGGGCAGAGGG

BM767063.1

AAAAACGCTTTGGATTCCCCCGGCTGGGTGGGAGAGCGAGCTGGGTGCCCCCTAGATT
CCCCGCCCCCGCACCTCATGAGCCGACCTCGGCTCCATGGAGCCCGGCAATTATGCCAC
CTTGGATGGAGCCAAGGATATCGAAGGCTTGCTGGGAGCGGGAGGGGGCGGAATCTGGT
CGCCCACTCCCCCTGACCAGCCACCCAGCGGCGCCTACGCTGATGCCTGCTGTCAACTA
TGCCCCCTTGGATCTGCCAGGCTCGGCGGAGCCGCCAAAGCAATGCCACCCATGCCCTGG
GGTGCCCCAGGGGACGTCCCCAGCTCCCGTGCTTATGGTTACTTTGGAGGCGGGTACTA
CTCCTGCCGAGTGTCCCGGAGCTCGCTGAAACCTGTGCCAGGCAGCCACCTTGCCCGC
GTACCCCGCGGAGACTCCCACGGCCGGGGAAGAGTACCCAGCCGCCCACTGAGTTTGC
CTTCTATCCGGGATATCCGGGAACCTACCAGCCTATGGCCAGTTACCTGGACGTGTCTGT
GGTGCACTCTGGGTGCTCCTGGAGAACCGGACATGACTCCCTGTTGCCTGTGGACAG
TTACCAGTCTTGGGCTCTCGCTGGTGGCTGGAACAGCCAGATGTGTTGCCA

BM794275

GCAGACTCTGGGTGCTCCTGGAGAACCGGACGTGACTCCCTGTTGCCTGTGGACAGTTA

CCACTCTTGGGCTCTCGCTGGTGGCTGGAACAGCCAGATGTGTTGCCAGGGAGAACAGAA
CCCACCAGGTCCCTTTTGAAGGCAGCATTTCAGACTCCAGCGGGCAGCACCTCCTGA
CGCCTGCGCCTTTCGTGCGGGCGCAAGAAACGCATTCCGTACAGCAAGGGGCAGTTGCG
GGAGCTGGAGCGGGAGTATGCGGCTAACAAAGTTCATCACCAGGACAAGAGGCGCAAGAT
CTCGGCAGCCACCAGCCTCTCGGAGCGCCAGATTACCATCTGGTTTCAGAACC GCCGGGT
CAAAGAGAAGAAGGTTCTCGCCAAGGTGAAGAACAGCGCTACCCCTTAAGAGATCTCCTT
GCCTGGGTGGGAGGATCTAAAGTGGGGGTGTCTGGGGAGACCAGGAACCTGCCAAGCCC
AGGCTGGGGCCAAGGACT

BQ363211

ACGCTGCACTGCGTTTCAAAGAGAAGAAGGTTCTCGCCAAGGTGAAGAACAGCGCTACCC
CTTAAGAGATCTCCTTGCTTGGGTGGGAGGAGCGAAAGTGGGGGTGTCCTGGGGAGACCA
GGAACCTGCCATCACCAGGCTGGGCCCAAGGACTCTGCTGAGAGGCCCTAGAGACAACA
CCCTTCCCAGGCCATTGCTTGCTGGACTGTGCCTCAGGAGCGGCCTGGGTACC

BM932052

GAGTTTTCCAATTTCCAAAGAAAAATTTAGGTTTCTGCAAGCCGTGACATATGTGTGTGC
ACTGGGATGGGTAAATGTGTGTGTGTGTGTGTGTATGCGCATGTATTGGGAGTGGGGGCA
GAAACGTGTTTCCAGAATTTGCCTGTAGAATCTAAAAGAGTGGCCAAGAGTCTGGAAATG
CATGAAGACTGGACGTATGTGATGGTGGGCAAAGGCCTGACTGTGTGTGGTGTGTGGGTA
TGTTTGAGATTGCGGGGTGTGAGAGCAGTGATGGGTGAGGGTGGCCTTCAGGAGCCAAG
GCTGATCGGTGGTGAGAGAACAAGCCGGAAGCCAGGGTGCTGTCTTGGTATGCTTTGGAG
GAACAGGATTGCACGTGCGCCTGTAGGGTGACCTGTGTGCACCTGTGAGATGACTTAGCT
TGGGGCTTGCAAGGCCTGGGTCTGCATGGGTGGGTATCTGACCATGCCTTTTCTCCCTC
CCTTTCAGCCCGCGCAGACTCCAGCGGGCAGCACCTCCTGACGCCTGCGCCTTTCGTC

AA357646.1

CCGGCCTGGGTGGGGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCAT
GAGCCGACCCTCGGCTCCATGGAGCCCGGCAATTATGCCACCTTGGATGGAGCCAAGGAT
ATCGAAGGCTTCTGCGAGCGGAGCGGGGCGCAATCTGCTGCGCACTCCCTCTGACC

AW609525

ACCGCGGGTCAAATTTATTATAATTAGCTCAATCATGAAAGCGGTTCTAAAGTGCTCTA
CAGAGCTCTAGATAGAAAATATGAGGCTAACGATCATGGCAGCTAGTACTGGTTATCGTG
ATTATGGCCACTGTCAGGATGAATGATAATGACTGGGCCAGGTCCTTTGGAAACCTGGT
GGAGTGGGCTGTACATGGGGTCCCGTCTCCCTGCACATACTGGGTACCCAGGCCGCTCC
TGAGGAACAGTCCAGCAGCCAGTGGCCTGGGAAGGGTGTGGTCTCTAGGGGCCTCTCAGC
AGAGTCTTTGGCCCCAGCCTGGGCTTGGCAGGTCCCTGGTCTCCCAGGACACCCCCACT
TTCGCTCCTCCCACCCAGGCAAGGAGATCTCTTAAGGGGTAGCGCTGTTCTTCACCTTGG
CGAGAACCTTCTTCTTTGAACCGGCGGTGCGGCGTGGGTACCGAGC

CB126919

ATTTTCTGTGCTGTGAAAATGAAGCCAGCAGGCTGCCCCCTAGTCAGTCCTTCTTCCAG
AGAAAAAGAGATTTGAGAAAGTGCCCTGGGTAAATCACCATTAAATTTCTCCCCAACTC
TCTGAGTCTTCCCTTAATATTTCTGGTGGTCTGACCAAAGCAAGTCATGGTTTGTGAG
CATTTGGGATCCAGTGAAGTAGATGTTTGTAGCCTTGCATACTTAGCCCTTCCCAGGCA

CAAACGGAGTGGCAGAGTGGTGCCAACCTGTTTTCCAGTCCACGTAGACAGATTCA
GTGCGGAATTCTGGAAGCTGGAGACAGACGGGCTCTTTGCAGAGCCGGGACTCTGAGAGG
GACATGAAGGCCTCTGCCTCTGTGTTTATTCTCTGATGTCTGTACCTGGGCTCAGTGCC
CGGTGGGACTCATCTCTGGCTGCGCAGCAAAGCCAGCGGTTCTGTGCTGGT

AW609336

CCAACGAGAAGAAGGTTCTCGCAAGGTGAAGAACAGCGCTACCCCTTAAGAGATCTCCTT
GCGTGGGTGGGAGGAGCGAAAGTGGGGGTGTCTGGGGAGACCAGGAACCTGCCAGCCCA
GGCTGAGGCCAAGGACTCTGCTGAGAGGCCCTAGAGACAACACCCTTCCCAGGCCACTG
GATGCTGAACTGTCCCTCAGGAGCGGCCTGGGTACCCAGTATGTGCAGGGAGACGGAACC
CCATGTGACAGCCCACTCCACCAGGGTTCCCAAAGAACCTGGCCCCAGTCATAATCATTC
ATCCTGACAGTGGCAATAATCACGATAACAGTACTAGCTGCCATGATCGTAAGCCTCAT
ATTTGCTATCTAGAGCTCTGTAGAGCACTTTAGAAACCGCTTTCATGAATTGAGCTAATT
ATGACTCAATTTGAACCGGCGTCCGGCGTG

AW609244

ACGCGCACCGCGGTCAAGAGAAGAAGGTTCTCGCAAGGTGAAGAACAGCGCTACCCCTTA
AGAGATCTCCTTGCCTGGGTGGGAGGAGCGAAAGTGGGGGTGTCTGGGGAGACCAGGAA
CCTGCCAAGCCCAGGCTGTGGCCAAGGACTCTGCTGAGAGGCCCTATGAGACAACACCC
TCCCCAGGCCACTGGCTGCTGGGACTGTTCTCAGGAGCGGCCTGGGTACCCGAGTAATG
TGCAGGGGAGACGGAACCCCATGTGACAGCCCACTCCACCAGGGTTCCCAAAGAACCCT
GGCCCAGTCATAATCATTCATCCTGACAGTGGCAATAATCACGATAACAGTACTAGCTG
CCATGATCGTAAGCCTCATATTTGCTATCTAGAGCTCTGTAGAGCCCTTTAGAAACCGCT
TTCATGAATGGAGCTAAATTATGAATACATTTGAACCGGCGATCCGACGTGA

BF855145

CTAGAGGATCCCGGAAGCAACTGCAACAGGTTCCCAAAGAACCGGGCCAGTCATAATCAT
TCATCCTGACAGGGCAATAATCACGATAACAGTACTAGCTGCCATGATCGTTAGCCTCA
TATTTCTATCTAGAGCTCTGTAGAGCACTTTAGAAACCGCTTTCATGAATGGAGCTAAT
TATGAATAAATTTGGAAGGCGATCCCTTGGCAGGGAAGCTTTCTCTCAGACCCCCCTTCCA
TTACACCTCTCACCTGGTAACAGCAGGAAGACTGAGGAGAGGGGAACGGGCAGATTTCGT
GGTGTGCACTGTGCTTCCG

AU126914

GAGCGAATGCAGGCGACTTGCGAGCTGGGAGCGATTTAAAACGCTTTGGATTCCCCCGGC
CTGGGTGGGGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCATGAGCC
GACCCTCGGCTCCATGGAGCCCGGCAATTATGCCACCTTGGATGGAGCCAAGGATATCGA
AGACTTGCTGGGAGCGGGAGGGGGCGGAATCTGGTCGCCCCACTCCCCTCTGACCAGCCA
CCCAGCGGCGCTACGCTGATGCCTGCTGTCAACTATGCCCCCTTGGATCTGCCAGGCTC
GGCGGAGCCGCCAAAGCAATGCCACCCATGCCCTGGGGTGCCCCAGGGGACGTCCCCAGC
TCCCGTGCCCTTATGTTACTTTGGAGGCGGGTNCTACTCCTGCCGAGTGTCCCGGAGCTC
GCTGAAACCCTGTGCCCCANNCANCCACCCTGGCCGCGTN

CB126449

CTCTGCCTCTGTGTTCAATCTCTGATGTCTGTACCTGTGCTCAGTGCCCGGTGGGACTC
ATCTCCTGGCTGCGCAGCAAAGCCAGCGGGTTCTGTCTGGTCTTCTCTGCACCTTCGGCT
GGGGGTGGGGGGCTGCCGGCGCATTCTCCACGATT

AW582404

ACGCTGCACCGCCGGTCCAAGAGAAGAAGGTTCTCGCCAAGGTGAAGAACAGCGCTACCC
CTTTAAGAGATCTCCTTGCTGGGGTGGGAGGAGCGAAAGTGGGGGTGTCTGGGGAGACCA
GGAACCTGCCAGCCCCAGGCTGGGCCCAAGGACTCTGCTGAGAGGCCCTAGAGACAACA
CCCTTCCCAGGCCACTGTCTGCTGGACTGTTCTCAGGAGCGGCCTGGGTACNCAGTATG
TGCAGGGAGACGGAACCCCATGTGACAGCCCACTCCACCAGGGTCCCAAAGAACCTGGC
CCAGTCATAATCATTCATCTGACAGTGGCAATAATCACGATAACCAGTACTAGCTGCCA
TGATCGTTAGCCTCATATTTCTATCTAGAGCTCTGTAGAGCACTTTAGAAACCGCTTTC
ATGAATTGAGCTACTTATGAATCACTTTGAACCGGCGGTGCGGCGTG

BX641644

GGGGGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCATGAGCCGACCC
TCGGCTCCATGGAGCCCGCAATTATGCCACCTTGGATGGAGCCAAGGATATCGAAGGCT
TGCTGGGAGCGGGAGGGGGCGGAATCTGGTGGCCCACTCCCTCTGACCAGCCACCCAG
CGGCGCCTACGCTGACGCTGCTGTCAACTATGCCCCCTTGGATCTGCCAGGCTCGGCGG
AGCCGCCAAAGCAATGCCACCCATGCCCTGGGGTGCCCCAGGGGACGTCCCCAGCTCCCG
TGCCTTATGGTTACTTTGGAGGCGGGTACTACTCCTGCCGAGTGTCCCGGAGCTCGCTGA
AACCTGTGCCCAGGCAGCCACCTGGCCGCGTACCCCGCGGAGACTCCACGGCCGGGG
AAGAGTACCCAGCCGCCCCACTGAGTTTGCCTTCTATCCGGGATATCCGGGAACCTACC
AGCCTATGGCCAGTTACCTGGACGTGTCTGTGGTGACACTCTGGGTGCTCCTGGAGAAC
CGCGACATGACTCCCTGTTCCTGTGGACAGTTACCAGTCTTGGGCTCTCGCTNGTGGCT
GGAACAGCCAGATGTGTTGCCAGGGAGAACAGAACCACCCAGGTCCCTTTTGAAGGCAG
CATTTG

[0187] Secuencias de la Tabla 4 no descritas anteriormente

AW006861 (Clon IMAGE ID::2497262)

GCTGAGTTCTGAAGCTTCTGAGTTCTGCAGCCTCACCTCTGAGAAAACCTCTTTTCCACC
AATACCATGAAGCTCTGCGTGAAGTGTCTGTCTCTCCTCATGCTAGTAGCTGCCTTCTGC
TCTCTAGCGCTCTCAGCACCAATGGGCTCAGACCTCCACCGCCTGCTGCTTTTCTTAC
ACCGCGAGGAAGCTTCTCGCAACTTTGTGGTAGATTACTATGAGACCAGCAGCCTCTGC
TCCCAGCCAGCTGTGGTATTCCAAACCAAAAGCAAGCAAGTCTGTGCTGATCCCAGT
GAATCCTGGGTCCAGGAGTACGTGTATGACCTGGAAGTGAAGTGAAGTGTCTCAGAGACAG
GAAGTCTTCAGGAAGGTACCTGAGCCCGGATGCTTCTCCATGAGACACATCTCCTCCA
TACTCAGGACTCCTCTCCGAGTTCTGTCCCTTCTCTTAATTTAATCTTTTATGTGC
CGTGTTATTGTATTAGGTGTCATTTCCATTATTTATATTAGTTTAGCCAAAGGATAAGTG
TCCCCTATGGGGATGGTCCACTGTCACTGTTTCTCTGCTGTTGCAAATACATGGATAACA
CATTTGATTCTGTGTGTTTTTATAATAAACTTTAAATAAAATGCAAAAAAAAAAAAAA
AAAA

X59770

GCCACGTGCTGCTGGGTCTCAGTCTCCACTTCCCGTGTCTCTGGAAGTTGTCAGGAGC
AATGTTGCGCTTGTACGTGTTGGTAATGGGAGTTTCTGCCTTCACCCTTCAGCCTGCGGC
ACACACAGGGGCTGCCAGAAGCTGCCGGTTTCGTGGGAGGCATTACAAGCGGGAGTTTCAG
GCTGGAAGGGGAGCCTGTAGCCCTGAGGTGCCCCAGGTGCCCTACTGGTTGTGGGCCTC
TGTGAGCCCCCGCATCAACCTGACATGGCATAAAAATGACTCTGCTAGGACGGTCCCAGG

AGAAGAAGAGACACGGATGTGGGCCCAGGACGGTGCTCTGTGGCTTCTGCCAGCCTTGCA
 GGAGGACTCTGGCACCTACGTCTGCACTACTAGAAATGCTTCTTACTGTGACAAAATGTC
 CATTGAGCTCAGAGTTTTTGAGAATACAGATGCTTTCCTGCCGTTTATCTCATACCCGCA
 AATTTTAACCTTGTCAACCTCTGGGGTATTAGTATGCCCTGACCTGAGTGAATTCACCCG
 TGACAAAACCTGACGTGAAGATTCAATGGTACAAGGATTCTCTTCTTTTGGATAAAGACAA
 TGAGAAATTTCTAAGTGTGAGGGGGACCACTCACTTACTCGTACACGATGTGGCCCTGGA
 AGATGCTGGCTATTACCGCTGTGTCTTGACATTTGCCCATGAAGGCCAGCAATACAACAT
 CACTAGGAGTATTGAGCTACGCATCAAGAAAAAAAAAAGAAGAGACCATTCTGTGATCAT
 TTCCCCCTCAAGACCATATCAGCTTCTCTGGGGTCAAGACTGACAATCCCGTGTAAAGGT
 GTTCTGCGGAACCGGCACACCCCTTAACCACCATGCTGTGGTGGACGGCCAATGACACCCA
 CATAGAGAGCGCCTACCCGGGAGGCCGCGTGACCGAGGGGCCACGCCAGGAATATTGAGA
 AAATAATGAGAACTACATTGAAGTGCCATTGATTTTGTATCCTGTCAACAAGAGAGGATTT
 GCACATGGATTTTAAATGTGTTGTCCATAATACCCCTGAGTTTTTTCAGACACTACGCACCAC
 AGTCAAGGAAGCCTCCTCCACGTTCTCCTGGGGCATTTGTGCTGGCCCCACTTTCACTGGC
 CTTCTTGGTTTTTGGGGGAATATGGATGCACAGACGGTGCAAACACAGAACTGGAAAAGC
 AGATGGTCTGACTGTGCTATGGCCTCATCATCAAGACTTTCATCCTATCCCAAGTGA
 TAAATGGAATGAAATAATTCAAACACAAAAAAAAAAAAAAAAAAAAA

AB000520

GGATCCAAGCTATTGTCTGCCCATGGCTTCCCATCTCAGGACGCTCTCTGGCCGCTATC
 ATCCCAGCAGTGGAGTTTCAAGCCACTACTCTGAACCAAGCCGAGGTGGCTGCTATGGGAC
 TGAAGCCATGAATGGTGCAGGCCCTGGCCCGCCGAGCCGCCCCGGTCCCAGTCCCGGT
 CCCGGTCCCGGACTGGCGGCAGTTCTGCGAGCTGCATGCGCAGGCGGCCCGCTGGACTT
 TGGCGACAAGTTCTGCCGTTTCTGCGGGACAACCCAGCTTACGACACGCCCGACGCCGG
 CGCCTCCTTCTCCCGCCACTTCGCGGCCAACTTCTTGACGCTCTTGGCGAGGAGGTGCG
 CCGCGTGTGGTGGCTGGGCGGACGACTCGGGGCGCGGCCGTGAGCGCAGAGGCCATGGA
 GCCGGAGCTCGCGGACACCTCTGCACTCAAGGCGGCGTCTTACGGCCACTCGCGGAGCTC
 GGAGGACGTGTCCACGCACGCGGCCACCAAGGCCCGGCTTCGCAAGGGCTTCTCGCTGCG
 CAACATGAGCCTGTGCGTGGTGGACGGCGTGCGCGACATGTGGCACCGGCGCGCCTCGCC
 CGAGCCCGACGCGGCAGCTGCCCCGCGCACCGCCGAGCCCCGCGACAAGTGGACGCGGCG
 CCTGAGGCTGTGCGGACGCTGGCTGCCAAGGTGGAGCTGGTGGACATTCAACGCGAGGG
 GCGCTGCGCTTTCATGGTGGCCGACGACGCGGCCGCGGGCTCCGGGGGCTCGGCTCAGTG
 GCAGAAGTGCCGCTGCTCCTGCGCAGGGCTGTGGCCGAGGAACGCTTCCGCCTGGAGTT
 CTTGCTGCCGCCCCAAAGCCTCCAGGCCCAAGGTGAGCATCCCACTGTGAGCCATCATTTGA
 GGTCCGCACCAACCATGCCCCTGGAATGCCAGAGAAAGGATAACACATTCTGCTCAAGGT
 AGAGAATGGAGCCGAATACATCTTGGAGACCATCGACTCTCTGCAGAAGCACTCGTGGGT
 AGCTGACATCCAGGGCTGCGTGGACCCCGGTGACAGTGAGGAAGACACCGAGCTCTCCTG
 TACCCGAGGAGGCTGTCTGGCCAGCCGCGTGGCCTCCTGCACTGTGAGCTCCTGACTGA
 TGCAGTCGACCTGCCCCGCCCCCAGAGACGACAGCCGTGGGTGAGTGGTGACAGCCCC
 CCACAGCCGAGGTGAGATGCCGTGAGAGAATCCCTGATCCACGTCCCGCTAGAGACCTT
 TCTGCAGACCCCTGGAATCCCCGGGCGGCAGCGGCAGTGACAGCAATAACACAGGGGAACA
 GGGTGCAGAGACGGATCCCGAGGCTGAACCCGAGCTGGAGCTATCCGACTACCCATGGTT
 CCACGGGACACTGTCCCGGGTCAAGGCTGCTCAACTGGTTCTGGCAGGGGGGCCCCGGAA
 CCACGGCCTCTTCTGATCCGCCAAAGTGAGACTCGGCCTGGGGAGTACGTGCTGACCTT
 CAACTTCCAGGGCAAGGCCAAGCACCTGCGCCTGTCCCTGAACGGCCACGGCCAGTGTC
 CGTACAGCATCTGTGGTTCCAGTCTGTGCTTGACATGCTCCGCCACTTCCACACACACCC
 CATCCCACTGGAGTCAGGGGGCTCGGCCGACATCACCTTCGAGCTATGTGCGGGCCCA
 GGACCCCCCACCAGAGCCGGGCCCCACGCCCCCTGCCGCGCCCGCTCCCCGGCCTGCTG
 GAGCGACTCGCCCGCCAGCACTACTTCTCCAGCCTCGCCGCGCGCCGCTGCCCGCCTGC
 CTCGCCCTCCGACGCGCGCGCGCCTCCTCGTCTTCCGCCTCGTCTGCTCTGCGCGCTC
 GGGGCCCCCCCCCGCGCCCCGTGAGGGCCAGCTCAGCGCGCGGAGCCGAGCAACAG
 CGCCGAGCGCCTGCTGGAGGCCGTGGCCGCCACCGCCCGGAGGAGCCCCCGAGGCGCG

GCCCCGCGCGCGCGCGCGCGTGGAGAACCAGTACTCCTTCTACTAGCCCGCGGCGCGCGCC
CGGGTGGGACACGCCAAGCTCTTCAGTGAAGACACGATGTTATTAAGCCTGTTTATAGG
GACTGCAAAA

AI820604 (Clon IMAGE ID: 1605108)

GATTCCAGCACGGGCTTCGCAGACTGCAGGACACAGAGGCACGCGTGACATCATGTCTT
CTAAGGAATTTGAACACTGTTGAGAAGACTGTGTACAAGAGAGATGTGCCATGTCAGCCT
TGCAAGGGACAGCGTGAAAACCTACCCATCTCCGGTCACCAAGTTGCAGGAGGCCAGGAGC
CAGGAGGGGAAACCGCTCAGTTTGCAAAACGTCGCTTCCACAAGCCTGATGGCTGAAAAC
GCTCACTGTACCTTGAACACAGCTTTACCTACAGCTTCTGAGATAAACTGCTGCAACTCT
GGGACCCACGATGCCTATCACAGTGGCTCATCAATGGAACCTGCCGGCTCCCAACCCCTTC
CTAGGGCCCATGAACTCTCTGAAAAGAGGAACAGAAATATTTCTCCTTTTTGTAAATCT
TTAACCTTCCCTTTGTTCTTCATGTACACGCTGAACTGCAATTCTTCTCCCAAATAAAA
CATTAAATTTAAAAAA

5

AI087057 (Clon IMAGE ID:1671188)

GGCCCCGGAGGGAGAGTAACCCGGCCCATCCATCCGTCCGCCGGTTCTTGGGGAACTACT
TTCAGGGGCTTCTTGCCGTCCCCTCATCAGCTCTGTGCGAACCCCTCTGTCCGCAGCCATT
GAGGAGACCCTGCCCCCTGGACCCTGACCACATATAGATTGAGGCCGAGGAGTGGCTGCC
CTGTCCCTTTTATGACAGCCCGCAGAAGCCCCGGGGTGAGGCATGGAGGAGGCAGGCGAC
AGCTGACAGGGACCCTGTTGGCCTCCAGCATGTCCAGCCAGCCGGGCAGGATTTCTCTGC
TTCTGGCTGGCAGCCAGGAACCTGAGTATGACAATGTTGTACTAAAGAAAGGCCCAAAGTG
ACAGAGGCAGCAGAGGGATGGTCCACCGCCCTTGGCTTCTGCTGGTGACTCCTCCTGGC
CACTGCATCAGAAGAACCCTCCTCTGCCCTTCTGGAGCCCGAGGCCTGGCCTGTCTTCGT
TGGGGCTGATAAATTGCCCTCTCCAGGGCCTGCTGGGTGAGTCACCATCCCAAAGCAGGA
AGGGTGCCCTGGAGAGAACCACCCTCCTCCTACTCTTTTTCCACTTCCTCCTCTTTCTTT
CCCCAGCTGAGGAGGAACCTGGGGCATTTAGGGCAGAGGACAAAAGGATGTCAGCAATTG
CTTGGGCTGCTTGGCTATGCAAGCCTCCTGCCTGCTGATGGCCACTTCAGGGACAGCCTG
GGCCCAGGCACCCAGGGGGATGGCGGCAGCTTCTGCACCTTTCAGATTTCTTGGTGGCA
TTAAAGCATTTTTCAGAACAAAAA

AJ272267

GGCGGGCCTGGACGGCCGCGTGCTGTACTGGCCACGCGCGCGCTCTGGGGTGGCTCCTC
ATCCCTCAATGCCATGGTCTACGTCCGTGGGCACGCGGAGGACTACGAGCGCTGGCAGCG
CCAGGGCGCCCGCGGCTGGGACTACGCGCACTGCCCTGCCCTACTTCCGCAAGGCGCAGGG
CCACGAGCTGGGCGCCAGCCGGTACCGGGGCGCCGATGGCCCGCTGCGGGTGTCCCGGGG
CAAGACCAACCACCCGCTGCACTGCGCATTCTTGAGGGCCACGCAGCAGGCCGGCTACCC
GCTCACCGAGGACATGAATGGCTTCCAGCAGGAGGGCTTCGGCTGGATGGACATGACCAT
CCATGAAGGCAAACGGTGGAGCGCGGCCTGTGCCCTACCTGCACCCAGCACTGAGCCGCAC
CAACCTCAAGGCCGAGGCGGAGACGCTTGTGAGCAGGGTGCTATTTGAGGGCACCCGTGC
AGTGGGCGTGAGTATGTTAAGAATGGCCAGAGCCACAGGGCTTATGCCAGCAAGGAGGT
GATTCTGAGTGGAGGTGCCATCAACTCTCCACAGCTGCTCATGCTCTCTGGCATCGGGAA
TGCTGATGACCTCAAGAACTGGGCATCCCTGTGGTGTGCCACCTACCTGGGGTTGGCCA
GAACCTGCAAGACCACCTGGAGATCTACATTGAGCAGGCATGCACCCGCCCTATCACCCCT
CCATTGAGCACAGAAGCCCTGCGGAAGGTCTGCATTGGTCTGGAGTGGCTCTGGAAATT
CACAGGGGAGGGAGCCACTGCCCATCTGGAAACAGGTGGGTTCATCCGCAGCCAGCCTGG
GGTCCCCACCCGGACATCCAGTTCCATTTCTGCCATCCCAAGTGATTGACCACGGGCG
GGTCCCCACCCAGCAGGAGGCTTACCAGGTACATGTGGGGCCCATGCGGGGCACGAGTGT
GGGCTGGCTCAAACTGAGAAGTGCCAATCCCCAAGACCACCCTGTGATCCAGCCCAACTA
CTTGTCAACAGAACTGATATTGAGGATTTCCGTCTGTGTGTGAAGCTCACCAGAGAAAT

TTTTGCACAGGAAGCCCTGGCTCCGTTCCGAGGGAAAGAGCTCCAGCCAGGAAGCCACAT
 TCAGTCAGATAAAGAGATAGATGCCTTTGTGCGGGCAAAAGCCGACAGCGCCTACCACCC
 CTCGTGCACCTGTAAAGATGGGCCAGCCCTCCGATCCCACTGCCGTGGTGGATCCGCAGAC
 AAGGGTCCTCGGGGTGGAACCTCAGGGTCGTGATGCCTCCATCATGCCTAGCATGGT
 CAGCGGCAACCTGAACGCCCCACAATCATGATCGCAGAGAAGGCAGCTGACATTATCAA
 GGGGCAGCCTGCACTCTGGGACAAAGATGTCCCTGTCTACAAGCCCAGGACGCTGGCCAC
 CCAGCGCTAAGACAGTTGCTGCTGGAGGATGACCAGGGAAGCCCCCTGATAAGCCAAGAG
 GGCCAGCACAGCCCTTGCTCCAGGCTCCTGCCCTGAACTATCTAGCACACTAGGACCCA
 GGTGGTACCTTACTCAGTGGCTGAGAATTGGATAAAGTCTTKGGGAAATGAGACAAGTAC
 TGGGCAGTGAATCCAGCTCCTTTTCCCCAGCCTTCCCTGTGGGCCATTTGGGGAAGGCC
 AGCATTYCAGCCTGAGATGTTCCCTCCCTGCCTCCTGGGGGGGCARAAGGGVTAGGWTGGT
 TAACTCCTGCCGCATCCTTCCCTGCCTCCTGGAGGGACAGAAGGGGAGGATGGTTAACTC
 CTGCCGCATCCTTTTCTTGTGTTACGTGGCATTCTCTAACCAGGGCAGTGGTTCCTT
 CCCAGGCCATGCACAGAGGCTGGGTGCCTGCCAGACCCACGGAGGGTTCCGAAGGAAGG
 GGCATCCTCCTTCTTGAGCTGCAAGCTTTAGCTGAGGCAGTAAGTCACACAGTAGTTAGT
 TCAGCCTGGGCTGGCACATAAGTCCCCAGTGTCCCTGTTGAGAGGGGAAAGTTGCCTGCT
 GGTGTA AAAACTGGCTTTTCTTTCTCGCTGCCTAATTTCACTCTCAGAGTGAGGCAGGT
 AACTGGGGCTCCACTGGGTCACTCTGAGAGGGTTGTGGCTCTGGTTCTTATTAAACCAGG
 GCCAGGTGCAGGGCTCACACCTGTAATCCAGCACTTTGGGAAGGTCACTTGAGCTCAGG
 AGTTCAAGACCAGCCTGGGCAACATAGTGAGACCTTGTCTCTGGA AAACAATTAGCTGGG
 CATGGTGGTACACACCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCGGGAGGATGGCTT
 AGCCAGGAGGTTGAGGCTCCTGTGAACCTGATGGCACCCTGCACTCCAGCCTGGGTG
 ACAGGGTGAGACCCTGTCTCAAAAAAAA

N30081 (Clon IMAGE ID: 258695)

CCGCCGTTGNCAAAGGGCCAGAATATGGGCCATGGACNATCTCCATGCCTGGGGAAATT
 CCCTCGGGTCCTTTGGNTAACNCCTTATAGAAAGGTAATGNCATGGAGTCTCTACAGGG
 NGCACAAGGTGGACTAATTGATACGAAGAGCCCTGTAAATATGTGGGCAGCGGCAGATTT
 TGACCAATTGGACCGAACTGTATTTGACACAGCGCAATATCTGGAACCTGGTTGGTCAAAA
 ACCTGCTTGTCTTGTAAATTTCTCTGTCCAAGGACATGGAATCTCTCTAATTTTAC
 TTCAAATTTCCCTTTCTTTCAATTTCTCTAAAAACGTTAAATAAGAAAGAAGATTGTAAAG
 CCAGCATTGTAAGCCTAAGTATTGAAAGTCTTTGACAATTTCTGAAATCAGACTTGACAT
 CTTTCCCCCGCCTTGCAAATTTCTTGAAGAAATAAGAAGCTACATGTAAGCATCATATG
 TTATTAAATTACAATGAGAACTCTCACTCAATCTTGACCAGAGCAGACTCTTAACTTGG
 AAGCAGAGTCCCTCTAAAGGTAACCTTTGTGGTCACTCAATATTGTATTGGCATTGTCAT
 ATTAATAGACATTTCACTAGCATTT

5 AI700363 (Clon IMAGE ID: 2327403)

TGGCCCGCGGTGCGGGTGGGATCCTAGCCCTGTCTCCTCTCCTGGGAAGGAGTGAGGGTG
 GGACGTGACTTAGACACCTACAAATCTATTTACCAAAGAGGAGCCCGGACTGAGGGAAA
 AGGCCAAAGAGTGTGAGTGCATGCGGACTGGGGGTTTCAAGGGAAGAGGACGAGGAGGAGG
 AAGATGAGGTGATTTTCTGATTTAAAAAATCGTCCAAGCCCCGTGGTCCAGCTTAAGGT
 CCTCGGTTACATGCGCGCTCAGAGCAGGTCACTTTCTGCCTTCCACGTCTCCTTCAAG
 GAAGCCCCATGTGGGTAGCTTTCAATATCGCAGGTTCTTACTCCTCTGCCTCTATAAGCT
 CAAACCCACCAACGATGGGCCTGTGGGAGGGGGCAAGATAGATGAGGGGGAGCGGCATGGT
 AGTTTCAAGCGCAGATGGGCCTGTGGGAGGGGGCAAGATAGATGAGGGGGAGCGGCATGGT
 GCGGGGTGACCCCTTGGAGAGAGGAAAAAGGCCACAAGAGGGGCTGCCACCGCCACTAAC
 GGAGATGGCCCTGGTAGAGACCTTTGGGGGTCTGGAACCTCTGGACTCCCATGCTCTAA
 CTCCCACACTCTGCTATCAGAACTTAACTTGAGGATTTTCTCTGTTTTTCACTCGCAA
 TAAATTCAGAGCAACAAAAA

AL117406

CAATAGGCCGGCTTTTGAACGTCTTCGAGGGGACTTGGAACAGCTGGACCAGCTCTTGC
 CCATCTTTTCAGAGCAGTTCCTGGTCCTGTCTTAATGGTGATCGCCGTCCTGTTGATTG
 TCAGTGTGCTGTCTCCATATATCCTGTTAATGGGAGCCATAATCATGGTTATTTGCTTCA
 TTTATTATATGATGTTCAAGAAGGCCATCGGTGTGTTCAAGAGACTGGAGAACTATAGCC
 GGTCTCCTTTATTCTCCACATCCTCAATTCTCTGCAAGGCCTGAGCTCCATCCATGTCT
 ATGGAAAACTGAAGACTTCATCAGCCAGTTTAAGAGGCTGACTGATGCGCAGAATAACT
 ACCTGCTGTTGTTTCTATCTTCCACACGATGGATGGCATTGAGGCTGGAGATCATGACCA
 ACCTTGTGACCTTGGCTGTTGCCCTGTTTCGTGGCTTTTGGCATTTCCTCCACCCCCCTACT
 CCTTTAAAGTCATGGCTGTCAACATCGTGCTGCAGCTGGCGTCCAGCTTCCAGGCCACTG
 CCCGATTGGCTTGGAGACAGAGGCACAGTTCACGGCTGTAGAGAGGATACTGCAGTACA
 TGAAGATGTGTGTCTCGGAAGCTCCTTTACACATGGAAGGCACAAGTTGTCCCCAGGGGT
 GGCCACAGCATGGGGAAATCATATTTACAGGATTATCACATGAAATACAGAGACAACACAC
 CCACCGTGCTTACGGCATCAACCTGACCATCCCGGGCCACGAAGTGGTGGGCATCGTGG
 GAAGGACGGGCTCTGTAGGTTTTTACTGAGCACCTACTATGTGCCTGGGAACCGAAAGGG
 AAGTCCTCCTTGGGCATGGCTCTCTTCCGCTGGTGGAGCCCATGGCAGGCCGGATTCTC
 ATTGACGGCGTGGACATTTGCAGCATCGGCCTGGAGGACTTGCGGTCCAAGCTCTCAGTG
 ATCCCTCAAGATCCAGTGCTGCTCTCAGGAACCATCAGATTCAACCTAGATCCCTTTGAC
 CGTCACACTGACCAGCAGATCTGGGATGCCTTGGAGAGGACATTCTTGACCAAGGCCATC
 TCAAAGTTCCCCAAAAAGCTGCATACAGATGTGGTGGAAAAACGGTGGAACTTCTCTGTG
 GGGGAGAGGCAGCTGCTCTGCATTGCCAGGGCTGTGCTTCGCAACTCCAAGATCATCCTT
 ATCGATGAAGCCACAGCCTCCATTGACATGGAGACAGACACCCTGATCCAGCGCACAAATC
 CGTGAAGCCTTCCAGGGCTGCACCGTGCTCGTCATTGCCACCGTGTCAACCACTGTGCTG
 AACTGTGACCACATCCTGGTTATGGGCAATGGGAAGGTGGTAGAATTTGATCGGCCGGAG
 GTACTGCGGAAGAAGCCTGGGTCAATTGTTTCGAGCCCTCATGGCCACAGCCACTTCTTCA
 CTGAGATAAGGAGATGTGGAGACTTCATGGAGGCTGGCAGCTGAGCTCAGAGGTTACAC
 AGGTGCAGCTTCGAGGCCACAGTCTGCGACCTTCTTGTGGAGATGAGAACTTCTCCT
 GGAAGCAGGGGTAAATGTAGGGGGGGTGGGGATTGCTGGATGGAAACCCTGGAATAGGCT
 ACTTGATGGCTCTCAAGACCTTAGAACCCAGAACCATCTAAGACATGGGATTCACTGAT
 CATGTGGTTCTCCTTTAACTTACATGCTGAATAATTTTATAATAAGGTAAAAGCTTATA
 GTTTTCTGATCTGTGTTAGAAGTGTGCAATGCTGTACTGACTTTGTAAATATAAAAC
 TAAGGAAAACTCAAAAAAAAAAAAA

M92432

CCCACAGGGGGACCGGCCCTGTGACCCCTCACCGGGGCGGTGGGCCCGAGCCCCGGACTT
 CCTTAAGCCGGCAATGACCGCCTGCGCCCGCGAGCGGGTGGGCTTCCGGACCCCGGGCT
 CTGCGGTCCCGCGTGGTGGGCTCCGTCCCTGCCCCGCTCCCCCGGGCCCTGCCCCGGCT
 CCCGCTCCTGCTGCTCCTGCTTCTGCTGCAGCCCCCGCCCTCTCCGCCGTGTTACGGT
 GGGGTCTTGGGCCCCCTGGGCTTGGGACCCCATCTTCTCTCGGGCTCGCCCGGACCTGGC
 CGCCCGCTGGCCCGCCCGCCCTGAACCGGACCCCGGCTGGCAGGCGGTCCCCGCTT
 CGAGGTAGCGCTGCTGCCCGAGCCTTGC CGGACGCCGGCTCGCTGGGGCCGTGTCTC
 CGCGCTGGCCCGCGTGTGCGGCCTCGTGGGTCCGGTGAACCCTGCGGCCTGCCGGCCAGC
 CGAGCTGCTCGCCGAAGAAGCCGGGATCGCGCTGGTGCCCTGGGGCTGECCTGGACGCA
 GGCGGAGGGCACACCGGCCCTGCCGTGACCCCGCGCGGATGCCCTCTACGCCCTGCT
 TCGCGATTTCGGCTGGGCGCGCGTGGCCCTGGTCAACGCCCCCGAGGACCTGTGGGTGGA
 GGCGGACGCTCACTGTCCACGGCACTCAGGGCCCCGGGGCTGCCGTGCTCGCTCGTGAC
 TTCCATGGAGCCCTTGGACCTGTCTGGAGCCCGGAGGCCCTGAGGAAGGTTCCGGACGG
 GCCAGGGTCACAGCAGTGATCATGGTGATGCACTCGGTGCTGCTGGGTGGCGAGGAGCA
 GCGCTACCTCCTGGAGGCCGAGAGGAGCTGGGCCTGACCGATGGCTCCCTGGTCTTCT
 GCCCTTCGACACGATCCACTACGCCTTGTCCCAGGCCCGAGGCCTTGGCCGCACTCGC
 CAACAGCTCCAGCTTCGAGGGCCACGATGCCGTGCTCACCTCACGCGCCACTGTCC

CTCTGAAGGCAGCGTGCTGGACAGCCTGCGCAGGGCTCAAGAGCGCCGCGAGCTGCCCTC
 TGACCTCAATCTGCAGCAGGTCTCCCCACTCTTTGGCACCATCTATGACGCGGTCTTCTT
 GCTGGCAAGGGGCGTGGCAGAAGCGCGGGCTGCCGAGGTGGCAGATGGGTGTCCGGAGC
 AGCTGTGGCCCGCCACATCCGGGATGCGCAGGTCCCTGGCTTCTGCGGGGACCTAGGAGG
 AGACGAGGAGCCCCCATTCGTGCTGCTAGACACGGACGCGGCGGGAGACCGGCTTTTTCG
 CACATACATGCTGGATCCTGCCCGGGGCTCCTTCTCTCCGCCGTACCCGGATGCACTT
 CCCGCGTGGGGGATCAGCACCCGGACCTGACCCCTCGTGCTGGTTCGATCCAAACAACAT
 CTGCGGTGGAGGACTGGAGCCGGGCCCTCGTCTTTCTTGGCTTCCTCCTGGTGCTTGGGAT
 GGGGCTGGCTGGGGCCTTCCTGGCCCATTTATGTGAGGCACCGGCTACTTCACATGCAAAT
 GGTCTCCGGCCCCAACAGATCATCCTGACCGTGGACGACATCACCTTTCTCCACCCACA
 TGGGGGCACCTCTCGAAAGGTGGCCCAGGGGAGTGGATCAAGTCTGGGTGCCCGCAGCAT
 GTCAGACATTCGACGCGGCCCCAGCCAACTTTGGACAGCCCCAACATTGGTGTCTATGA
 GGGAGACAGGGTTTGGCTGAAGAAATTCAGGGGATCAGCACATAGCTATCCGCCACG
 AACCAAGACGGCCTTCTCCAAGCTCCAGGAGCTCCGGCATGAGAACGTGGCCCTCTACCT
 GGGGCTTTTCTGGCTCGGGGAGCAGAAGGCCCTGCGGCCCTCTGGGAGGGCAACCTGGC
 TGTGGTCTCAGAGCACTGCACGCGGGGCTCTCTTCAGGACCTCCTCGCTCAGAGAGAAAT
 AAAGCTGGACTGGATGTTCAAGTCTCTCCTCCTGCTGGACCTTATCAAGGGAATAAGGTA
 TCTGCACCATCGAGGCGTGGCTCATGGGCGGCTGAAGTCACGGAACCTGCATAGTGGATGG
 CAGATTCTGTAAGTCAAGATCACTGACCACGGCCACGGGAGACTGCTGGAAGCACAGAAGGT
 GCTACCGGAGCCTCCAGAGCGGAGGACCAGCTGTGGACAGCCCCGGAGCTGCTTAGGGA
 CCCAGCCCTGGAGCGCCGGGGAACGCTGGCCGGCGACGTCTTTAGCTTGGCCATCATCAT
 GCAAGAAGTAGTGTGCCGAGTGGCCCTTATGCCATGCTGGAGCTCACTCCCGAGGAAGT
 GGTGCAGAGGGTGGGAGCCCCCTCCACTGTGTGCGCCCTTGGTGTCCATGGACCAGGC
 ACCTGTGAGTGTATCCTCCTGATGAAGCAGTGTGCTGGGCAGAGCAGCCGGAACCTTCGGCC
 CTCATGGACCACACCTTCGACCTGTTCAAGAACATCAACAAGGGCCGGAAGACGAACAT
 CATTGACTCGATGCTTCGGATGCTGGAGCAGTACTTAGTAACCTGGAGGATCTGATCCG
 GGAGCGCACGGAGGAGCTGGAGCTGGAAAAGCAGAAGACAGACCGGCTGCTTACACAGAT
 GCTGCCCTCCGTCTGTGGCTGAGGCCCTGAAGACGGGGACACCACTGGAGCCCGAGTACTT
 TGAGCAAGTGACACTGTACTTTAGTGACATTGTGGGCTTCACCACCATCTCTGCCATGAG
 TGAGCCCATTTAGAGTTGTGGACCTGCTCAACGATCTCTACACACTCTTTGATGCCATCAT
 TGGTTCCACGATGTCTACAAGGTGGAGACAATAGGGGACGCCTATATGGTGGCCCTCGGG
 GCTGCCCCAGCGGAATGGGCAGCGACACGCGGCAGAGATCGCCAACATGTCACTGGACAT
 CCTCAGTGCCGTGGGCACTTTCCGCATGCGCCATATGCCTGAGGTTCCCGTGCGCATCCG
 CATAGGCCCTGCACTCGGGTCCATGCGTGGCAGGCGTGGTGGGCCTCACCATGCCGCGTA
 CTGCTGTGTTGGGGACACGGTCAACACCGCCTCGCGCATGGAGTCCACCGGGCTGCCTTA
 CCGCATCCACGTGAACCTTGAGCACTGTGGGGATTCTCCGTGCTCTGGACTCGGGCTACCA
 GGTGGAGCTGCGAGGCCGACCGAGCTGAAGGGCAAGGGCGCCGAGGACACTTTCTGGCT
 AGTGGGCAGACGCGGCTTCAACAAGCCCATCCCCAAACCGCCTGACCTGCAACCGGGGTC
 CAGCAACCACGGCATCAGCCTGCAGGAGATCCCACCGAGCGGCGACGGAAGCTGGAGAA
 GGCGCGGCCGGGCCAGTTCTCTTGAGAAGTGAGGCCCGGCCCGGACAGGGTCTGGGCCC
 TGCTCCCTGTCCCATCTGCAGTGGACCCACGGCACCCCTTTGAGGAGGTGGGGTGAAC
 TGCTCCTTGGCAGGGATTGTGACACTGCATTGCTGGGCTGTGTTCTCGGGCTCTTCTG
 GACCTTGCACCGTGGATACCAGGCCATGTGCCATGGTATTTGGGTCTGGGAGGGTGGGT
 GAAATAAAGGCATACTGTCTT

AL050227

CTTTCACAGAAAGAAAGTAACAGGCATAATTCTGTGATGAGGCTGGGATTGTTTTTAA
 GAGGAGAGATAATAACTTCATATTTTAAAGTGCCAGTAGCCTAATATGTGAAACAGATC
 AGAATCTGTTGTGTAGTAAGTCTGCTTTGTTGAAGAATTTATTATGGGAGTAAAGATAAG
 AAGGAAAGAGATCACCATCAGAAACAAGTCAGCCTTTTCATGCTTTTTTGAGCATTTTTG
 GAGATGATTCCACTTCTCAAGTTATTATCATTTGTGCATCTCTCAATGCTATTGTTAAA
 TGCTTTAGAATTAGAATATTTTGATCCTTTAATTAAAGTAAGCCAAACGTCTAGGCCAAA

ACAGCCAATCATTAACTTTAATAGTAATTCAAATATAGATTTCTCATACAGTTTTCCAT
GTCTGTAGAAATCAAAGTTGTAATGTTAAGCAGAGGGAAATGCGTGTGATTTACTAATAC
ACTTCAACGTTCTACTTTTGAAAGGATACTCATGTGGGTGGGGCAGAGAACATAGAAAAA
GATATGATGGAAAACCTGTCCATTTTCTACCTGTTAACCTTCATCATTTTGTGCAGGCCC
TGGAAGCAAAGAGAGGAAGGGACCGACTGCATTTATCTTTGAACACTTGAGCATCAGTAG
TACTACTGAGTGGCCAGGGGTCTTGTCTGTCAAAGCAAATGATAAGTTCCTCAGGCCAT
TATTGACTGCTGAACTCTCTTCCCTTCCCACTCTTCCCTTGAAAGAGAAAAAATACTTTG
CCTTCTTGCTCTCCTTATCAAATGTTTTTGTACAAATAGTGTAAGCCTGTTTAAGCAAAC
CAATTAAAAATAGGCACTGATTATTTTGATCTGTTTGTAAACAAATGAATGTAAGTACTATT
TACATGGTGTGCCTAGGAGGAGCTGAAATCATTGGCACTTTAATCCATATGTAAAGATC
AGTATCAAAAGCATAGTGTCTTCCCTCTCCTCCTCAGCATCCATCTCTATATACTTGA
TTAAATGGAAAAGTCTCTTTTATCACCTCTATGTAAAGTTTTATGGGTAGTTATCGTCAG
TGTATTTAAATATATCTTCTAGTATGTTTTAAAGGCTGGTCTTCAATACTGTGGAGACAA
AAAATAAAAGAGCGTATGAAAAGTACGTTAGACTTTTGCTGGCATTCAAGTCATGGCTAG
TCTGTGTATTTAATAAATGTGTGTTATTTATGTCGTGTTTGTCAATGGAATAAAGTTG
AATATTCTGAAAAAAAAAAAAAA

AW613732 (Clon IMAGE ID: 2953502)

CCTANAAGTNCCATTTTGGCAAGGATAAACTCCCATGACAANCTCCCANTACTGCATGTG
AATGAATAAGAAACAAGAANTGACCACACCAAAGCCTCCCTGGCTGGTGTACANGGGAT
CAGGTCCACAGTGGTGCAGATTCAACCACCACCCAGGGAGTGCTTGCAGACTCTGCATAG
ATGTTGCTGCATGCGTCCCATGTGCCTGTCAGAAATGGCAGTGTTTAATTCTCTTGAAAGA
AAGTTATTTGCTCACTATCCCCAGCCTCAAGGAGCCAAGGAAGAGTCATTACATGGAAG
GTCCGGGACTGGTCAGCCACTCTGACTTTTCTACCACATTAAATTCTCCATTACATCTCA
CTATTGGTAATGGCTTAAGTGTAAGAGCCATGATGTGTATATTAAGCTATGTGCCACAT
ATTTATTTTGTAGACTCTCCACAGCATTTCATGTCAATATGGGATTAATGCCTAAACTTTGT
AAATATTGTACAGTTTGTAAATCAATGAATAAAGGTTTTGAGTGTAATAAAAAAAAAAAAAA
AAAAAA

5 BC007783 (Clon IMAGE ID: 4308472)

GGCACGAGGGCAAAGAGTAGTCAGTCCCTTCTTGGCTCTGCTGACACTCGAGCCACATT
CCATCACCTGCTCCCAATCATGCAGGTCTCCACTGCTGCCCTTGCCGTCTCTCTGCAC
CATGGCTCTCTGCAACCAGGTCTCTCTGCACCACTTGCTGCTGACACGCCGACCGCCTG
CTGCTTCAGCTACACCTCCCGGCAGATTCCACAGAATTTATAGCTGACTACTTTGAGAC
GAGCAGCCAGTGCTCCAAGCCAGTGTCATCTTCTTAACCAAGAGAGGCGGCGAGGTCTG
TGCTGACCCCAAGTGAGGAGTGGGTCCAGAAATACGTCACTGACCTGGAGCCGAGTGCCCTG
AGGGGTCCAGAAGCTTCGAGGCCCAGCGACCTCAGTGGGCCCAGTGGGGAGGAGCAGGAG
CCTGAGCCTTGGGAACATGCGTGTGACCTCCACAGCTACCTCTTCTATGGACTGGTTATT
GCCAAACAGCCACACTGTGGGACTCTTCTTAACTTAAATTTTAATTTATTTATACTATTT
AGTTTTTATAATTTATTTTGTATTTACAGTGTGTTTGTGATTGTTTGTCTCTGAGAGTTC
CCCCTGTCCCCTCCACCTTCCCTCAGAGTGTGTCTGGTGACAACCGAGTGGCTGTTCATCG
GCCTGTGTAGGCAGTCATGGCACCAAGCCACCAGACTGACAAATGTGTATCAGATGCTT
TTGTTACAGGGCTGTGATCGGCCTGGGGAAATAATAAAGATGTTCTTTTAAACGGTAAAAA
AAAA

X81896

AGAAAACTATTTTCTAAATATTAACACTGAAAATGTTTTGTTAGCTTTTCCTTCTTCTC
TCCAGAAGAAACATGGATAGATGATAGCTGTTTCATTGTTTGTGTTTGTCAAGCATATTC
ACTTTCTCCTTGTCTCTGATTCTGAGCAAAGGGCCTCAGACTCTGAACTTCCCTCAAG
TGCCGTTGTTATGTGAACTCTTCATTGAGATTCAGAGAGGTTCTCATGCTCCCCCCCC

CTCCTTATTTGTAGCAATCGTAGCAACTAATTCCTACTAAGTACAAGGGAGTTTTTTTACAC
 TCCTCCATTTTTATAGCATCTGCATTTTTTTTTTTTGTAGGTACATGTATACACCTGCC
 TGAGTATAAATACTCTCTCTACCTAATAATAACATCAACCAACATCTTTTCCAAATTAGG
 GCCACAGAACAGCAACATTTGTCTGACAGTAGTATAAAGAATAATGATAGCTCTATCCTT
 AAGAAGTATTTCTTTCTTTTATATAGTCCCCTTAGGGTTTAAAAACCATATTGATCAA
 CTAGAAAGAAAAATATGAAAAGAGAAAAATATTTTAATTTAAAAATTGTAATACATTGAT
 TTATAAAATGCCTTCTCTGATACTTTTGAAACAGATGTGAAAAACAGAAAAAGAAAAAT
 TGTCTGAAATGTTTTATTTTGCAAAACAGTGCAATAGAATCTAGTTATGCCTTCATCACTG
 TTGACAGTAAATACTGACAGCCCCCTGCAGTGTGTTAGTTTTAGATCACTCTGTTTTAGT
 TGAGAGAAATGTTTTATATCATGGTTTTATATGAATACAAATTATTTCTCAAAGATTTA
 TAGCACACACTATTCTCAGGAATCTGTATTACATGAATGCTGCTTATATATTTTTCATAT
 TCTAAGTTGTCTTTTCAAGCAAATAACTAATATATATGTGCATGCAGTCTGCCTTGACAA
 GTTGTTCCTAAGCTGAAGAGCTTTCACTGTACAATGTGTGGAAAATCACCATAGATCATGG
 CTGAAATAGTTTGTAAATTGTCTGAGTCTGTGCACGTACTTTTAGATAAAATGCTGCTGAG
 TGAATGCTGATGAGATACAACCTCTGAATGCTGCACATTCTTCCAAAATGATCCTTAGC
 ACAATCTATTGTATGATGGAATGAATAGAAAACCTTTTCACTCAATAAATTATTATTTGA
 TATGGTAAAAAATAA

BC004960 (Clon IMAGE ID: 3632495)

CCCAAGGTTGTTATATCTTCATGTCCTCATTTCTTAGGGAGGTACCTTCAGAACCAATAG
 TGACCCCTAAGCTTCTCTGGTGGTGGGTTCCATGAAAGGCAAAGGAGTGTGAGAGAGGAGT
 GGATGGTCAACCTCCACTGCCATGGTAACATGGGTGCTGGCTGATGGGAGCAGAAAATA
 ATTTAGTGAAAGTCTGTGGGGGCAGTCAAGATGTCTGAGAAAACCTGGCGAGCCAGCTG
 CTGAAACAGGGACAAGGAAGCCTCCGTGGCTGGAGCCCAAATCACACTGCAGACCCAGA
 CACCGTGACCACCACCATGGACTCCAGAGAGAGCAGCTTATAGTACTCAATCAGCTGCCA
 CTACCACCATCCAGAACACCAGATGTTGTAGCCATGGCTGCAGCAGGAATGGATGTCCCA
 CTGTCCCTGCTCCTCGGTGTGACTTGCTCCCAAGTTCAGGGCAGGTCCATCTGATTGGCT
 GAGTCTGGAATGTCTGCCTGTGCCTCAGCTGTGAGGGAGGCAGGGAAAGTAAGCCTTTTC
 AGCTTCTGTCTGGGAGGTGGGCTCTGCCTCTACCAAGAATCAAAGGGTGGAGGATCTT
 CAAACACAGGAAAAGAACCCTGGATCCTGGCACCCTCCAAATTTTCAGAGTCCATTTTCAGAG
 CATAAGAAATTGAGGGTCCAAGATCATTCATGTAAGAAGTTTAGAGGGGGAAGAAAAGAA
 TGATAAACGAAAAGAACAGCAATAGTAAAGGATCTTTTCTTTGTTTCAGTAAGATGAAGA
 GGCCTGAGCAGTTTCGTGGAGGGGAAGAAAACAGGAAAACCTCTTCAAAGACAAAAAGCT
 GGCATGCAATCTCTCTCTGTAGCAGGACAGAACTGTCTAAAGACAAGACCCCTTTGGCC
 AAAATAAAGGAACCTGAAACATTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
 AAAAAAAAAAAAAAAAAAAAAAAAAAACCTCGGG

AK027250

AAATTTAATTAATTATAAACTCAGTCTCTTGGTTGCACCAGCCACATTTTCAGATGCTCAA
 TAGCCACATGTGGCTAGTGGGTACCATATTGGACAGGGCAGCTATAGAATATTTCCATCA
 TTGCAGAAAGTTCTATTGGATAGTACCATAATCTTTTATAGTAACTTGGAATACTATT
 TGATATTAGATGTTAGACCACAAAAAGAAGAAAAATGTTAGGACTATTTTCAGATATAAAA
 AGGAACTGAATTGTGACATAATTAGCATCTTACATTCATACAGTTGAATACCTTATGCT
 GTGACAACCATAGTTAATCATTTTCAGTGTCTGTTCAACATACATACCTATCAGCAGTGTGT
 TTAGACCAGGGGTCTGCAAACTTTCTGTGAATGGACAAAGAGTAAATACTTTAGTAAATG
 TCTTAGGCTTTGTGGCCTACATGATCTTTGTGCAAGTACTCAACTCTGCCATTATAGAG
 TTAAAGCAGCCATACACAATATATAAAACAAATGGGCATAGTTGTATTTTCAGTAAACTT
 TATTTACAAAGACAGGCGGTAGGCCAGATTTGGCTTGCATGCTGTAGAGCTGTGGTCTAA
 ATTTTATTCATAGACTTTCTTTGCAAATACAGTGTGAGTATTGTTCCATTTACAGTATTA
 TTATTTTTTAGATACCTGGTTTTTAGATTCTTGCCTGGTAACTTTTACTGAAAATACAA
 GAATTCGTACTGCATTTGCATCTCCGAGATTAGGGAGCACCTGTGAGGATATGTTGTTTC

TATCAGGGTTACTTCTGTTGACTACCTCTTAGATTTTGATACAGTTATATTGTTGAGTTT
 CATTTTCATATATTCTTGTAGTGTCTGCTTGCCCTGTGACTTCTGGTAAAATAAAATAAGC
 CTTTGAAAATATTTTAGCATGGTATTTAACAATTTCTAAATATTATGGCATTTTGACATA
 TTTTAGTCAGCGAAGACATCTGCCCTTTGGTGTTTCTACTTGCTTATGATTGAGATTTT
 ACAAGCCCTTCAAACCTCCGTTTAAAGGAATTTATGTAAAACATTAACTTTAATAAAT
 AGTGTTTTTCACAGATCAGATCATTATACTTGGAACTTCTAAATCATGCAATTTCTGAATA
 AGGACATAAGGCTAGATTCAATTTTCTTAATAGAGAAAAAGGAAATTTCTGATTATCAC
 TTTTCTAGTTGATAAGTAGGATTCAAAACGTTTGATATGTAAGTATTTATATAAGACTAA
 TGTAAATTTAAAGTTCTGTATPATTTGTGATTAAATCATACAGAAATTCAGGAACCTGATCAGA
 AGTGAGATTCTTTTCCACATCTGGTTAATGTAGTGAGTTGACACCCTGTGGGTGGTAAAG
 CATTATAAACATTTTCTTGAACCATGATTATACACATCTGTGTTATAAGGGAGGCTT
 GAGTACATATAACCAATGAAGAGATATTCAGCATTGTCTATTTGATAAGGAATTAAATGT
 CCTAGTGATTATAAAGTAAACACAGACCAATTTGCAAATGATCTTCAATGTTAAGCAC
 TTGCTCTAAGATTAAAATTCCTTTTCTTTTAAAGGTTAAGGGTGTGTACGTATGGCAGTG
 ATGCTATGTTGAGATTAACTTATGTATTGAGGAAAAATTTGAAGTTTATTTTTCGATGA
 ATAAGGCTGTCAAATGATTTAGTATAGATTAAATGACATCTTTTATAGAAATATTAAAGTG
 AGTATTCCTCATTATGTATCATTTCTGATAATTAGAGTGCTAATTTGAATGTTAGATAA
 TGTTTCCACATCTATACCTATTTCTTTCTAGGGCACTTCTGACCCTGGGGCTTGGGGATG
 GCCTTTAGGCCACAGTAGTGTCTGTGTTAAGTTCACTAAATGTGTATTTAATGAGAAACA
 TTCTATGTAAAAATGTGTGTATGTGAACGTATGCATACATTTTATTGTGCACCTGTAC
 ATTGTGAAGAAGTAGTTTGGAAATTTGTAAAGCACAAACCATAAAAGAGTGTGGAGTTAT
 TAAATGATGTAGCACAAATGTAATGTTTAGCTTATAAAAGGTCCTTTCTATTTTCTATGG
 CAAAGACTTTGACACTTGAAAAATAAAACCAATATTTGATTTATTTTGTAAAGTATTTAG
 GATATTTATTTTAAATAAATGATTGTCCATTATCAATAAAAAAAAAAAAAAAAAAAAA

Secuencias de la Tabla 5 no descritas anteriormente

NM_014298

GTCCTGAGCAGCCAACACACCAGCCCAGACAGCTGCAAGTCACCATGGACGCTGAAGGCC
 TGGCGCTGCTGCTGCCGCCCGTCACCCTGGCAGCCCTGGTGGACAGCTGGCTCCGAGAGG
 ACTGCCCAGGGCTCAACTACGCAGCCTTGTTGTCAGCGGGGCAGGCCCTCGCAGGCGGCGC
 TGTGGGCCAAATCCCCCTGGGGTACTGGCAGGGCAGCCTTTCTTCGATGCCATATTTACCC
 AACTCAACTGCCAAGTCTCCTGGTTCTCCCCGAGGGATCGAAGCTGGTGCCGGTGGCCA
 GAGTGGCCGAGGTCCGGGGCCCTGCCCACTGCCTGCTGCTGGGGGAACGGGTGGCCCTCA
 ACACGCTGGCCCGCTGCAGTGGCATTGCCAGTGTCTGCCGCCGCTGCAGTGGAGGCCGCCA
 GGGGGGCCGGCTGGACTGGGCACGTGGCAGGCACGAGGAAGACCACGCCAGGCTTCCGGC
 TGGTGGAGAAGTATGGGCTCCTGGTGGGCGGGGCGCCTCGCACCGCTACGACCTGGGAG
 GGCTGGTGATGTTGAAGGATAACCATGTGGTGCCCCCGGTGGCGTGGAGAAGGCGGTGC
 GGGCGGCCAGACAGGCGGCTGACTTCGCTCTGAAGGTGGAAGTGAATGCAGCAGCCTGC
 AGGAGGTCTCCAGGCAGCTGAGGCTGGCGCCGACCTTGTCCTGCTGGACAACCTCAAGC
 CAGAGGAGCTGCACCCACGGCCACCGCGCTGAAGGCCAGTTCCCGAGTGTGGCTGTGG
 AAGCCAGTGGGGGCATCACCTGGACAACCTCCCCAGTTCTGCGGGCCGCACATAGACG
 TCATCTCCATGGGGATGCTGACCCAGGCGGTCCAGCCCTTGATTTCTCCCTCAAGCTGT
 TTGCCAAAGAGGTGGCTCCAGTGCCCAAAATCCACTAGTCCTAAACCGGAAGAGGATGAC
 ACCGGCCATGGGTAAACGTGGCTCCTCAGGACCCTCTGGGTACACATCTTTAGGGTCAG
 TGAACAATGGGGCACATTTGGCACTAGCTTGAGCCCAACTCTGGCTCTGCCACCTGCTGC
 TCCTGTGACCTGTCAGGGCTGACTTCACCTCTGCTCATCTCAGTTTCTTAATCTGTAAAA
 TGGGTCTAATAAAGGATCAACCAAAAAAAAAAAAAAAAAAAAA

AF033199

CGGGGCATGCTGCTTCCCTTCACCTTCCACCATGATTGTAAGTTTCCTGAGGCCTCCCA
 GGTGTGCTTCTGTACAGCCTGTGGAATGTTACCAAAGACGTTGGAAGAGGTGGCTATGGG
 ACATCACCTGGGAGAAGTGAAGCAAATGGACACTGTTTCTGAGAGTCCATATACAGAAACA
 TACTTGGAAAAATATAGAAACCTGGTTTGTCTAGATGGGAAGCTTGCAGCTGGGGCCAAG
 ACATCAAGAGTAGAGCAGCAGGACATTTCAAAGAAGATTAACTCAAAGATTAGAGATGG
 AAGAAGTTGCAAAGAGAAAGTCTGTACCGGAAGAAATCTGGAATCTAGAGGCCAGTTTA
 AGAATCAGCAGCTAAACAAGGAGAATAATCTAGGGCAAGAGATAGCTACCTGCACAAAAA
 TTCTTACCAGAAAAAGAGACATAGAATCTAATGAATTTGTGAAAAATTTTACTGTAAGAT
 CAATACTTGTTCGAGAACAGATAGATCTTATGGAAGAGAATTGTCATAAATATGGTACAT
 GTTGAAAGATGCTCAAACAAACTCAGATTTAATTATACAAAGAAAGTATGATGGAAAAA
 AAAAAACCTTGTAATATAGTGAATGTGGGAGAACCTTCAGAGGCCACATCACTCTTGTT
 CAGCATCAAATAACTCATTGTGGAGAGAGACCCTGTAAATGTACTGAGTGTAGAAAGGGA
 TTTAATCAGAGTTCCCACTTAAGAAATAATCAGAGAAAACTCTTTCAGGAGAAAAGCCC
 TACAAATGCAGTGAGTGTGGGAAGGCCTTCAGTTATTGCTTAGTTCTTAATCAACACCAG
 AGAATTCACAGTGGAGAGAAACCTTATGAGGGTACTGAATGTGGCAAGACATTCATTTCAG
 TCGTACATACCTTACTCAGCATCAAAGAATTCACACACTGGTGAGAAGCCCTATACATGT
 CTTGAATGTGGGAAGGCTTTTTAGTCAGAACACACATCTTACTCTACATCAGAGAATCCAT
 ACTGGAGAGAAACCTTATGAATGCAATGAATGTGGTAGGTCTTTTAGTCAGACTGCACAT
 CTTACTCAACATCAAAGAATGTATACAGGAGAAAACTCTATGAATGTAATGAATGTGAG
 AAAGCCTTCCATGATCACTCAGCTCTTATTCAACATCATATTGTCCATACTGCAGAGAAA
 CCTTATGATATCATGACTGGGAAAACTTTTCAGTTACTGTTTCAGACCTCATTCAACATCAG
 AGAATGCACACTGGAGAGAAACCATACAAATGCAATGAATGTGGGAATGCCTTTAGTGAT
 TGTTTCATCCCTTATTCAGCATCAAAGAACTCACACTGGAGAAGAGCCTTATGAATGTAAG
 CAATGTGGAAAAGCCTTTAGCAGAAGCACATACCTTACTCAACATCAGAGAAGTCACGCA
 GGAGAGAAACAGTATAAATGCAATGAATGTGAGAAAACCTTTCAGCCTGAGTTTCATTCTT
 ACACAGCATATGAGGGTTTCAGACTGGAGAAAAACCTTACAAATATAATGAATATGGAAAA
 GCTTTTAGTGACTGCTCAGGACATTTTCAGAGAAGTCACTGGAGAGAAGCCCTGTGAA
 TGTAATGACTGTGGGAAACCTTTTCAGTTTCTGTTTCAGCCCTAATTCAACATAAGAGAATT
 CATACCAGAAAGAAGCCCTGACTGTACCTTCATACCAGTAAATGCCTGACTGTGGAAAA
 GCCTTCAGTGATTGGTTAGCACTTGTTCACATCAGATAACTCAACACTGGAGAAAAACC
 GTATAAATGTACTGAATGTGGAAAAGCCTTCAGTTGGAGTACAGACCTCAAAAATCACCA
 GAAAACTCATACTAGTGAAAAATCCTATAAATGTAATGAATGTAGAAAGGCCTTTAGTTA
 CTGCTCTGGTCTTATTCAATGTCAGGTCACTCATACTATAGAAAAACCTTATGAATACGG
 TAAATGTGGCAAAGCCTTTAGGCAGAGGACAGACCTTAAAAAACATCAGAAAATGCATAC
 CGAAGAGAAACCTATGAATGTAATGAATGTGGGAAAGCCTTTAGCCAGAGCACATATCT
 TACAAAACACCAAAAAATTCTAGTGAAGAGAAATCAAATATACATACTGAGTGTGGGGA
 AACCATTAGACAAAACCTCTCTTTTACAACAATAAAACCTCACACTGGAGAGTTCTCTG
 AATGCCTTAAGAATTTGGTTAATATGGAGACCCTTCCAGGGAAACAGAAGGAGGATCGT
 GAAAACCGTTGACTACTTGAATGATCACATGGTTTAGTGAGAGAGCATGATTCTGGGTT
 TTTAAAGTCATGGATCTCAATCTCAGCTCCTATTACTAACTAGATCTTTTACTTTGGGGT
 AAGTCACTTCATATCTTAGGCCTTAATTTCTCATCTGAAAACTGGAAGGCCTGACTTG
 ACTTGTGAGCTTAAGATCCTCAATTATTATTTACTAGGAATTCAAGTTTCTATAGAT
 GTGGTTTCAGAATTGTGACTTATTTATTGTACATCAGGTGTGATTCAAGAGTGAGCTTGTA
 GTAGTTATTAAGGAGTCAATAAGATATGATATAAAAAAAAAAAAAAAAAA

Al688494 (Clon IMAGE ID: 2330499)

CATTTTCATCTTCATTGGATAGTGTTACATAGTAATATATTTATGTTTTCTTTAATCATT
 TCATAACTTGGAAAATACTAACATAGTCAAACCTCTAGGGTAGGTGATACATGAGTTTCT
 GTAGTAATCTGGTTGGAGACATGTTGTAATCTGTATATATATGTACATTTATCCCATGC
 ATGTTATGCCTAAACTAAGACGGATACCCCTGAATTAAGAGGTGCTGTTATACATTGACC
 AGGCTTAAGAATATCTCTTTAAAGTGTGTCGACATTTAATTGACCTTTGGAAGTTCAATC
 TGTTAATCATACTCAAAGTGCTAAAGCTATGGTTGACTGCTCTGGTGTCTTTATATTCAT

TCGTGCTTTAGCATATAAATTCCTTCAGCATAATTGCTACTTATTTAGCAAGAGTTTCCTT
 TATTTGAAAATGTGAGTTGTGCTTGTATTTTTGTGTCTTCTTTCTTTCTTTCTTTT
 AAACCTTTGCTTCAGGCTGGGTAGTGGTAGAGGTTTGAATTAAATGTTTTCTGTGAGTA
 AAAAAAAAAA

AL157459

GAGCGAGCCCAGCAGCTTGCCCTTGACAGGTGGGGGCTGGCTGGGGCCTTAATGTGAAAA
 GACAGTGGCAGGCAGCTGGAGTAGAGCGAGCCCAGCAGCCCTAAAAGGCTGCCTTCATGG
 CCATCTAGCCCCAGTTCAGGGCAGCATCCATAGCCCACAAGCCAGCGTGGGTGGGGCGGG
 GGTGGTCCCACAGCTGGGTTCACCTGAAGAGCCTCCGTGCCTCGGAGCAGGAGAGGCAG
 GCTATGGCTGTCACCCTCCCTCCTGCCTGTGTCCCAGTGAGAACTGACCTGAGTCCCTT
 CCAAACCCAGACCCACCTCCTGCCCCAGGCCCACTGAAGCATGTTCCATTTCTAAAAAGC
 CCAGAGTTTCAGTGTGTCCCAAGGAAAACCCAAAGTGGAGGTGCTCAGGTCCAGGGGAGTC
 CAGTGGGCAGGACCCCTTGGCAGGCAAGCCCCCTCCCTTCACTCCCAGGACCTACCTTCTGC
 TAGTAAAGGACTGGCTTCATTCTAATTATGGCCCCACAGACTGCCCCGGAGACCTGGAGGA
 CAGCAGTGTCTGGCATTGGGTGTCCATGGGCCCGTCTGCCGGCTCTGCCTGTGCTGCAAG
 TGTGGCCGTGGGTCCAGCCAACAACCTCCCTACGTCTGTGTGGGGCCCTGCCCAAGTGG
 ATGAGGCATTCCCTTGAGGAGTATCATTTTCCCTGACAATCCCCATCACCTTTAGGGGTTT
 CCTGCTTGGCTCCTTTCCAGCTGAAAACTAGACCTGTGCCATTGGGGAAGCTGGACAAA
 GTCTAGGGGGCCCCGCTGGTAGAGGGTCCCGGAAGCTGGATCTGTGAGCCTCGGCCCTG
 AGGCCCCCTGTTAACTCAAGACTGTGAGCTGCCTCTAGGTGGTCACGTCTGGGAGCTAGCT
 TGTATGGCTTCTGACCAGTATCAGGATTTCTGTTCTGAGAGCAGCGTGGGCAGCAAGGCA
 GGGCAGCCCAGAGGTGGCAGCGGCAGGCAATCTGGTCACTAGGTCTTTGTGATGCCAAAA
 ATAAAAGAGGGTGGGTGGGTGCTTTCTGTTCTCTGATTGGATGGAGTCCGCCAGCAGG
 CATGGGGCTACATTCAGTGCCTGACTATAGGGAGGCACTCCTGATTCCATGGAGCAGCC
 CGGACTTTGAGAATGGGCTCTGGTTTGGGGGGCAGGCGTACCAGACTGCAAGACCCCC
 CAGTACCTCACCGTGCCAAATAGGAAGAGGTGGCCTTGGTGTAGCCAAATGGATCTTTTT
 AACAGTGTGCCTTTGGGGAGGGACCCATGTCCATGGCTTCGTTGAGGGCCATCCATATGC
 CAGCTGGGGGCCAGCCCACAGTGGCCATATTGGCTGCAGCAGGAATGGTGGCCACCTCGG
 CGAATTGAAGGGCTAAGAGTCCAGATAGCTAGGCCAGAGCTGGAAGCAGACAGTAAGGG
 GAAGAGCTGCTCCACAGGAGAGGGAGAGATTCCAGCTCACTGCGCAGCCTGGGAGGAGG
 CGTGGATCCTGGCAGCTGAGCCTCAGGCACCAGCCTCCCTGTGCTCGACAGCAAAGTCT
 TGACTCCTTCTGCTGAGCACTGTGCTACCTTCACTGCTCCAAAGCCAGACTAACAGCTC
 TCCAAGCCCTTGGGGTGA CTGCGCTTCAGGAGCTGTTGGAGAAATGAGGATGTCTGTCC
 CTGTCTGCCTGGGCAGGCAGATTCTCCCCAGCAGCCGGGTCTCTCCAGACCCTGATTC
 GGTGCCTTTCTGTTTACCAGCTACTTCAATCCCAAAGTTTGAATCTGCAGATACCTTACT
 CCCAGCCACTTTGCCTTCTTACTGTGTTGTGTGTTTTCTGGTGCTTCAAGAGCGTGTG
 CAGGGCAAGTGCCGTCACTGGGAACTGCACCAGATGCTCAGACTTGGTTGTCTTATGTTT
 ACCAATAAATAAAAGTAGACTTTTTCTATTTTTATTGCTGCTATTTGTGTGTGTGTTG
 TGTTTGTGTAGCTAGGTATCTGGCACTTCTGACGATGCATTGTTGCTTTTTTCCCGAAGG
 TCCCGCAGGAACTGTGGCAATGGTGTGTGTGTGAAATGGTGTGTTAACCGCGTTTTGTTT
 GCTCCTGTATTGAATAGGAAGCAGTGGCCAGTCTGTCTTCTTAGAGATGTTAGCATATT
 TTTATATGTATATATTTTGTACCAAAAAAGAGTGTTCCTTGTTTTGGTTACACTCGAAAT
 TCTGACCTAGCTGGAGAGGGCTCTGGGCCGAGAGCTTCACTAAGGGGAGACTTCAGGGG
 AGGATCAAGCTTTGAACCAAAGCCAATCACTGGCTTGATTTGTGTTTTTTAATTAAAAAA
 AAAATCATTTCATGTATGCCACTTCTAAAAA

BC002480 (Clon IMAGE ID: 3350037)

GGCAGAGGCTGAGACCGGTGCGCCGCGCTAGTGGCCGCTCTTCCGCGGGCTAGCGGG
 CGGTGGGGGCGCCAGCAGCGCGGAAGGCGGGCACGCGGGCCATGGCTCCCTGGGCGGAGG
 CCGAGCACTCGGCGCTGAACCCGCTGCGCGCGGTGTGGCTCACGCTGACCGCCGCTTCC

TGCTGACCCTACTGCTGCAGCTCCTGCCGCCCCGGCCTGCTCCCGGGCTGCGCGATCTTCC
 AGGACCTGATCCGCTATGGGAAAAACCAAGTGTGGGGAGCCGTGCGGCCCCGCGCCTGCC
 GAGCCTTTGATGTCCCAAGAGATATTTTCCCACTTTTATATCATCTCAGTGCTGTGGA
 ATGGCTTCCTGCTTTGGTGCCTTACTCAATCTCTGTTCCCTGGGAGCACCTTTTCCAAGCT
 GGCTTCATGGTTTGCTCAGAATTCTCGGGGCGGCACAGTTCCAGGGAGGGGAGCTGGCAC
 TGTCTGCATTCTTAGTGCTAGTATTTCTGTGGCTGCACAGCTTACGAAGACTCTTCGAGT
 GCCTCTACGTCAGTGTCTTCTCCAATGTCATGATTCACGTCGTGCAGTACTGTTTTGGAC
 TTGTCTATTATGTCCTTGTGGCCTAACTGTGCTGAGCCAAGTGCCAATGGATGGCAGGA
 ATGCCTACATAACAGGGAAAAATCTATTGATGCAAGCACGGTGGTTCCATATTCTTGGGA
 TGATGATGTTTCATCTGGTCATCTGCCCATCAGTATAAGTGCCATGTTATTCTCGGCAATC
 TCAGGAAAAATAAAGCAGGAGTGGTCATTCACTGTAACCACAGGATCCCATTTGGAGACT
 GGTTTGAATATGTTTCTTCCCCTAACTACTTAGCAGAGCTGATGATCTACGTTTCCATGG
 CCGTCACCTTTGGGTTCACAACTTAACTTGGTGGCTAGTGGTGACAAATGTCTTCTTTA
 ATCAGGCCCTGTCTGCCTTTCTCAGCCACCAATTCTACAAAAGCAAATTTGTCTCTTACC
 CGAAGCATAGGAAAGCTTTCTTACCATTTTGTGTTTAAAGTTAACCTCAGTCATGAAGAAT
 GCAAACCAGGTGATGGTTTCAATGCCTAAGGACAGTGAAGTCTGGAGCCCCAAAGTACAGT
 TTCAGCAAAGCTGTTTGAAACTCTCCATTCCATTTCTATACCCACAAGTTTCACTGAA
 TGAGCATGGCAGTGCCACTCAAGAAAAATGAATCTCCAAAGTATCTTCAAAGAATAAATAC
 TAATGGCAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

REIVINDICACIONES

1. Un método para determinar el riesgo de recurrencia de cáncer en un sujeto que padece cáncer de mama ER+ en caso de tratamiento con tamoxifeno, comprendiendo dicho método ensayar una muestra de células de cáncer de mama ER+, procedente de dicho sujeto, para la expresión de los ARNm de HOXB13 e IL17BR humanos para determinar una proporción de los niveles de expresión de HOXB13 e IL17BR, en el que dicha proporción indica la ausencia de recurrencia de cáncer en dicho sujeto en caso de tratamiento con tamoxifeno.
2. El método de la reivindicación 1, en el que dicha proporción indica la ausencia de recurrencia de cáncer por metástasis o un resultado de supervivencia.
3. El método de acuerdo con una cualquiera de las reivindicaciones 1 y 2, en el que dicha muestra de células de cáncer de mama se obtiene mediante una técnica mínimamente invasiva o se selecciona entre biopsia realizada con aguja gruesa, biopsia excisional, una muestra obtenida por lavado ductal, una muestra obtenida por aspiración con aguja fina o células microdisseccionadas procedentes de dicha muestra.
4. El método de una cualquiera de las reivindicaciones 1 a 3, en el que dicho ensayo para la expresión comprende la detección de ácidos nucleicos preparados por amplificación de ARNm procedente de dicha muestra de células de cáncer de mama o la detección de ácidos nucleicos de dicha muestra de células de cáncer de mama por PCR cuantitativa.
5. El método de una cualquiera de las reivindicaciones 1 a 3, en el que dicho ensayo se realiza por hibridación con un polinucleótido que comprende las secuencias de al menos 15 nucleótidos de la región 3' no traducida, de la región codificante, o de la región 5' no traducida, de los genes de HOXB13 e IL17BR humanos.
6. El método de una cualquiera de las reivindicaciones 1 a 3, en el que dicho ensayo para la expresión comprende ensayar la inactivación o metilación del gen de HOXB13.
7. El método de la reivindicación 4, en el que dicho ensayo es para determinar los niveles de expresión de HoxB 13 e IL17BR y comprende realizar PCR en tiempo real y determinar una proporción de los niveles de expresión de los ARN de HoxB 13 e IL17BR, expresada como una ΔCt de los valores de Ct observados para los niveles de expresión de los ARN de HoxB 13 e IL17BR.

Fig. 1

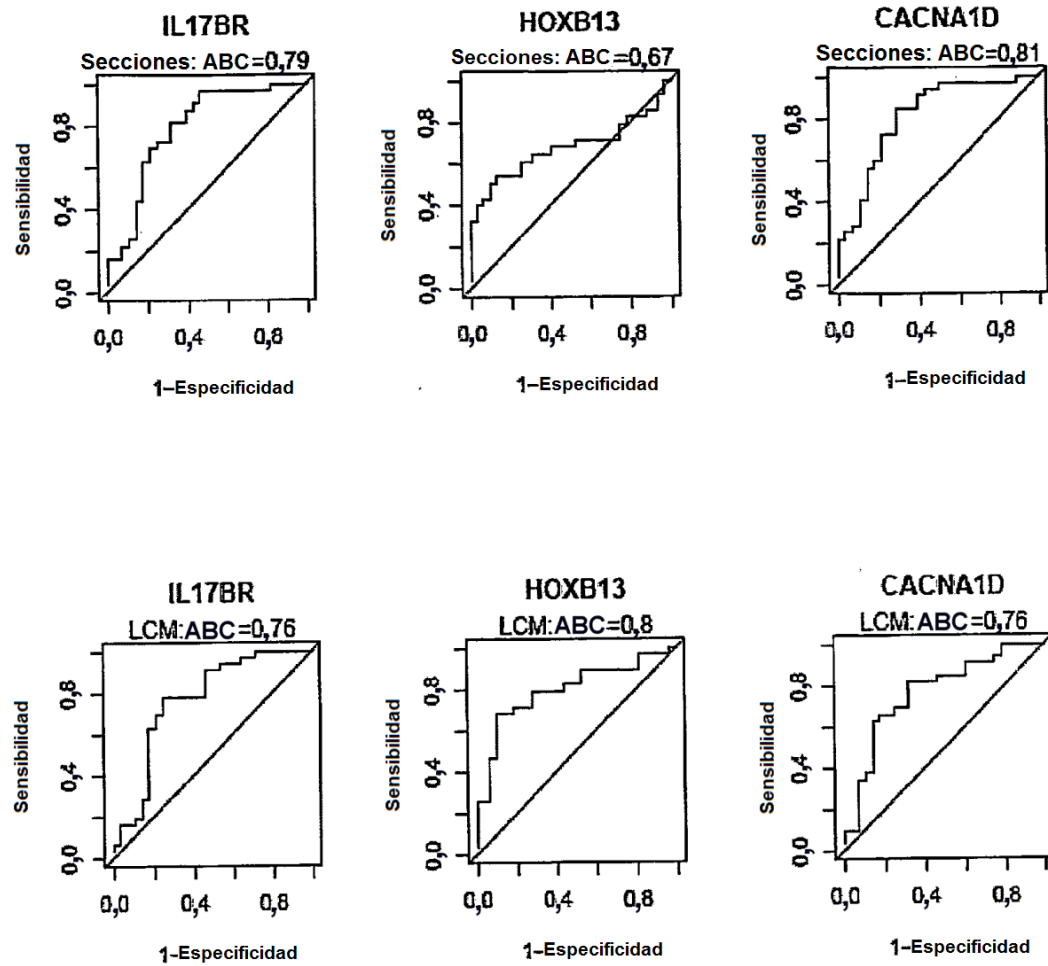


Fig. 2

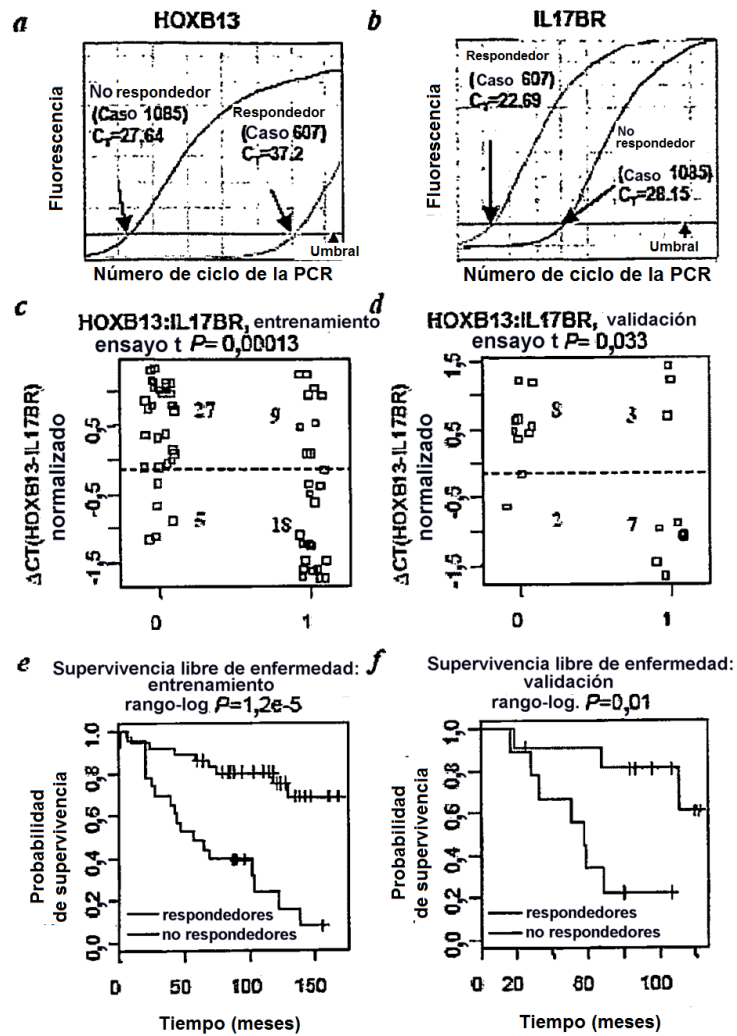


Fig. 3

Extremo del gen CHDH conocido

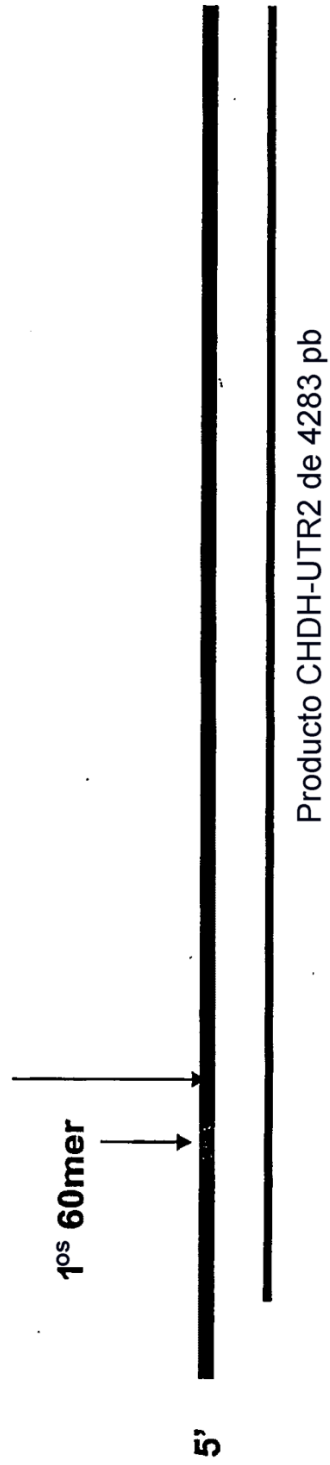


Fig. 4

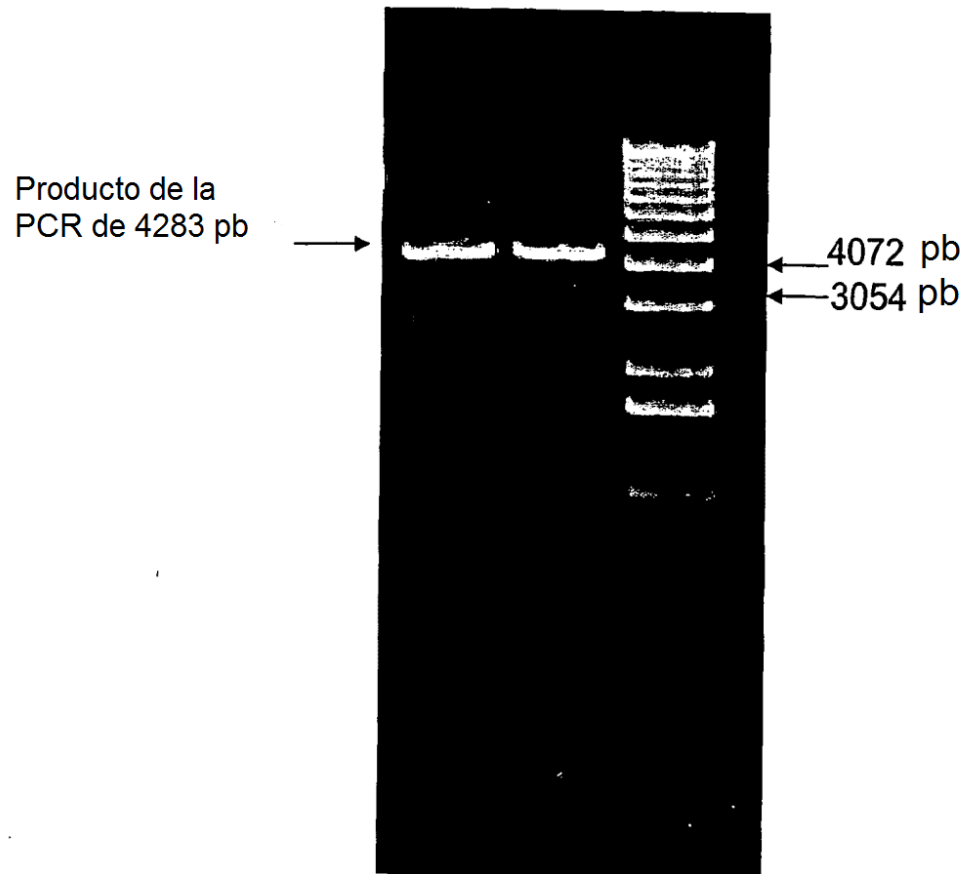


Fig. 5

```

gi|23274053|gb|BU608029.1|BU608029-.....GGAAAAACAATTTTCATCTTTAAAAACTTTTCCTTAAGACAAGGGA
gi|2669971|gb|AA682690.1|AA682690- is in gi|23274053|
gi|3838507|gb|AI243110.1|AI243110+ is in gi|23274053|
gi|3229367|gb|AI015031.1|AI015031+ is in gi|23274053|
gi|2142013|gb|AA437099.1|AA437099+ is in gi|3229367|
gi|23527150|gb|BU679327.1|BU679327+
gi|3405022|gb|AI075844.1|AI075844- is in gi|23527150|
gi|1782040|gb|AA192157.1|AA192157- is in gi|23527150|
gi|3307963|gb|AI051972.1|AI051972- is in gi|23527150|
gi|4610281|gb|AI601252.1|AI601252- is in gi|23527150|
gi|3055333|gb|AA915941.1|AA915941- is in gi|4610281|
gi|3844395|gb|AI248998.1|AI248998- is in gi|23527150|
gi|10033200|gb|BE672659.1|BE672659- is in gi|23527150|
gi|4222874|gb|AI393327.1|AI393327- is in gi|10033200|
gi|26011703|gb|CA774243.1|CA774243- is in gi|23527150|
gi|1162804|gb|N39597.1|N39597+ is in gi|26011703|
gi|3933768|gb|AI290994.1|AI290994- is in gi|26011703|
gi|6837905|gb|AW341279.1|AW341279- is in gi|26011703|
gi|11448468|gb|BF436153.1|BF436153-
gi|2953300|gb|AA861160.1|AA861160- is in gi|11448468|
gi|5543768|gb|AI869800.1|AI869800- is in gi|11448468|
gi|5543775|gb|AI869807.1|AI869807- is in gi|11448468|
gi|1230953|gb|N73668.1|N73668- is in gi|11448468|gb|
gi|3254601|gb|AI033648.1|AI033648- is in gi|11448468|
gi|4311745|gb|AI459166.1|AI459166- is in gi|11448468|
gi|11451784|gb|BF439267.1|BF439267-
gi|4330088|gb|AI467998.1|AI467998-
gi|3836330|gb|AI240933.1|AI240933-
Posiciones .....GGAAAAACAATTTTCATCTTTAAAAACTTTTCCTTAAGACAAGGGA
consenso      . . . . . : . . . . . :

```

Posiciones
consenso

gi 23274053 gb BU608-	TTAAAGCAGACATAC	TAGGAAATGGTGCAGCCTGTAAAGAAATGCCAGTTTGTAAAGTACTGA
gi 3838507 gb AI2431+	TTAAAGCAGACATAC	TAGGAAATGGTGCAGCCTGTAAAGAAATGCCAGTTTGTAAAGTACTGA
gi 3229367 gb AI0150+	TTAAAGCAGACATAC	TAGGAAATGGTGCAGCCTGTAAAGAAATGCCAGTTTGTAAAGTACTGA
gi 2142013 gb AA4370+	TTAAAGCAGACATAC	TAGGAAATGGTGCAGCCTGTAAAGAAATGCCAGTTTGTAAAGTACTGA
gi 23527150 gb BU679+	TTAAAGCAGACATAC	TAGGAAATGGTGCAGCCTGTAAAGAAATGCCAGTTTGTAAAGTACTGA
gi 1782040 gb AA1921-		TGTAAGTACTGA
gi 4610281 gb AI6012-		AGTTTGTAAAGTACTGA
gi 3055333 gb AA9159-		TAAGTACTGA
gi 3844395 gb AI2489-		CTGTAAAGAAATGCCAGTTTGTAAAGTACTGA
gi 10033200 gb BE672-	TTAAAGCAGACATAC	TAGGAAATGGTGCAGCCTGTAAAGAAATGCCAGTTTGTAAAGTACTGA
gi 26011703 gb CA774-	TTAAAGCAGACATAC	TAGGAAATGGTGCAGCCTGTAAAGAAATGCCAGTTTGTAAAGTACTGA
gi 1162804 gb N39597+	TTAAAGCAGACATAC	TAGGAAATGGTGCAGCCTGTAAAGAAATGCCAGTTTGTAAAGTACTGA
gi 6837905 gb AW3412-		CAGACATAC
		TAGGAAATGGTGCAGCCTGTAAAGAAATGCCAGTTTGTAAAGTACTGA
consenso	TTAAAGCAGACATAC	TAGGAAATGGTGCAGCCTGTAAAGAAATGCCAGTTTGTAAAGTACTGA
	.	:
gi 23274053 gb BU608-	CTTTGGAAAAGATCATCGCCT	-CTATCAGAC-ACCTAGGGTCCCTGGTCTGGC-AAATTTTG
gi 3838507 gb AI2431+	CTTTGGAAAAGATCATCGCCT	-CTATCAGAC-ACCTAGGGTCCCTGGTCTGGC-AAATTTTG
gi 3229367 gb AI0150+	CTTTGGAAAAGATCATCGCCT	-CTATCAGAC-ACCTAGGGTCCCTGGTCTGGC-AAATTTTG
gi 2142013 gb AA4370+	CTTTGGAAAAGATCATCGCCT	-CTATCAGAC-ACCTAGGGTCCCTGGTCTGGC-AAATTTTG
gi 23527150 gb BU679+	CTTTGGAAAAGATCATCGCCT	-CTATCAGAC-ACCTAGGGTCCCTGGTCTGGC-AAATTTTG
gi 1782040 gb AA1921-	CTTTGGAAAAGATCATCGCCT	-CTATCAGAC-ACCTAGGGTCCCTGGTCTGGC-AAATTTTG
gi 3307963 gb AI0519-		AAGATCATCGCCT
gi 4610281 gb AI6012-	CTTTGGAAAAGATCATCGCCT	-CTATCAGAC-ACCTAGGGTCCCTGGTCTGGC-AAATTTTG
gi 3055333 gb AA9159-	CTTTGGAAAAGATCATCGCCT	-CTATCAGAC-ACCTAGGGTCCCTGGTCTGGC-AAATTTTG
gi 3844395 gb AI2489-	CTTTGGAAAAGATCATCGCCT	-CTATCAGAC-ACCTAGGGTCCCTGGTCTGGC-AAATTTTG
gi 10033200 gb BE672-	CTTTGGAAAAGATCATCGCCT	-CTATCAGAC-ACCTAGGGTCCCTGGTCTGGC-AAATTTTG
gi 4222874 gb AI3933-		AAAAGATCATCGCCT
gi 26011703 gb CA774-	CTTTGGAAAAGATCATCGCCT	-CTATCAGAC-ACCTAGGGTCCCTGGTCTGGC-AAATTTTG
gi 1162804 gb N39597+	CTTTGGAAAAGATCATCGCCT	-CTATCAGAC-ACCTAGGGTCCCTGGTCTGGC-AAATTTTG
gi 3933768 gb AI2909-		TCTATCAGAC
gi 6837905 gb AW3412-	CTTTGGAAAAGATCATCGCCT	-CTATCAGAC-ACCTAGGGTCCCTGGTCTGGC-AAATTTTG
consenso	CTTTGGAAAAGATCATCGCCT	-CTATCAGAC-ACCTAGGGTCCCTGGTCTGGC-AAATTTTG
	.	:

gi 23274053 gb BU608-	GCCTGATGTGATGCCACAAGACCCCAACAGAGAGACACAGAGTCC - AGGATAATGTTGA
gi 2669971 gb AA6826-	GATAATGTTGA
gi 3838507 gb AI2431+	GCCTGATGTGATGCCACAAGACCCCAACAGAGAGACACAGAGTCC - AGGATAATGTTGA
gi 3229367 gb AI0150+	GCCTGATGTGATGCCACAAGACCCCAACAGAGAGACACAGAGTCC - AGGATAATGTTGA
gi 2142013 gb AA4370+	GCCTGATGTGATGCCACAAGACCCCAACAGAGAGACACAGAGTCC - AGGATAATGTTGA
gi 23527150 gb BU679+	GCCTGATGTGATGCCACAAGACCCCAACAGAGAGACACAGAGTCC - AGGATAATGTTGA
gi 3405022 gb AI0758-	CPAGA CCCAACAGAGAGACACAGAGTCC - AGGATAATGTTGA
gi 1782040 gb AA1921-	GCCTGATGTGATGCCACAAGACCCCAACAGAGAGACACAGAGTCCCAGGATAATGTTGA
gi 3307963 gb AI0519-	GCCTGATGTGATGCCACAAGACTCAACAGAGAGAGACACAGAGTCC - AGGATAATGTTGA
gi 4610281 gb AI6012-	GCCTGATGTGATGCCACAAGACCCCAACAGAGAGAGACACAGAGTCC - AGGATAATGTTGA
gi 3055333 gb AA9159-	GCCTGATGTGATGCCACAAGACCCCAACAGAGAGAGACACAGAGTCC - AGGATAATGTTGA
gi 3844395 gb AI2489-	GCCTGATGTGATGCCACAAGACCCCAACAGAGAGAGACACAGAGTCC - AGGATAATGTTGA
gi 10033200 gb BE672-	GCCTGATGTGATGCCACAAGACCCCAACAGAGAGAGACACAGAGTCC - AGGATAATGTTGA
gi 4228874 gb AI3933-	GCCTGATGTGATGCCACAAGACCCCAACAGAGAGAGACACAGAGTCC - AGGATAATGTTGA
gi 26011703 gb CA774-	GCCTGATGTGATGCCACAAGACCCCAACAGAGAGAGACACAGAGTCC - AGGATAATGTTGA
gi 1162804 gb N39597+	GCCTGATGTGATGCCACAAGACCCCAACAGAGAGAGACACAGAGTCC - AGGATAATGTTGA
gi 3933768 gb AI2909-	GCCTGATGTGATGCCACAAGACCCCAACAGAGAGAGACACAGAGTCC - AGGATAATGTTGA
gi 6837905 gb AW3412-	GCCTGATGTGATGCCACAAGACCCCAACAGAGAGAGACACAGAGTCC - AGGATAATGTTGA
gi 11448468 gb BF436-	GAGTCC - AGGATAATGTTGA
gi 3254601 gb AI0336-	GAGTCC - AGGATAATGTTGA
gi 11451784 gb BF439-	CCAGGATAATGTTGA
gi 4330088 gb AI4679-	GGATAATGTTGA
Consenso	GCCTGATGTGATGCCACAAGACCCCAACAGAGAGAGACACAGAGTCC - AGGATAATGTTGA

gi 23274053 gb BU608-	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGCTCCCTGCGCTGGTATTAGGTTACC-ATT
gi 2669971 gb AA6826-	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGNGTCTGCGGCTGGTATTAGGTTACC-ATT
gi 3838507 gb AI2431+	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGCTCCCTGCGGCTGGTATTAGGTTACC-ATT
gi 3229367 gb AI0150+	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGCTCCCTGCGGCTGGTATTAGGTTACC-ATT
gi 2142013 gb AA4370+	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGCTCCCTGCGGCTGGTATTAGGTTACC-ATT
gi 23527150 gb BU679+	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGTCCCTGCGGCTGGTATTAGGTTACC-ATT
gi 3450522 gb AI0758-	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGTCCCTGCGGCTGGTATTAGGTTACC-ATT
gi 1782040 gb AA1921-	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGCT-CCCTGCGGCTGGTATTAGGTTACC-ATT
gi 3307963 gb AI0519-	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGCTCCCTGCGGCTGGTATTAGGTTACC-ATT
gi 4610281 gb AI6012-	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGTCCCTGCGGCTGGTATTAGGTTACC-ATT
gi 3055333 gb AA9159-	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGCT-CCCTGCGGCTGGTATTAGGTTACC-ATT
gi 3844395 gb AI2489-	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGTCCCTGCGGCTGGTATTAGGTTACC-ATT
gi 10033200 gb BE6672-	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGTCCCTGCGGCTGGTATTAGGTTACC-ATT
gi 4222874 gb AI3933-	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGTCCCTGCGGCTGGTATTAGGTTACC-ATT
gi 26011703 gb CA774-	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGTCCCTGCGGCTGGTATTAGGTTACC-ATT
gi 1162804 gb N39597+	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGTCCCTGCGGCTGGTATTAGGTTACC-ATT
gi 3933768 gb AI2909-	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGTCCCTGCGGCTGGTATTAGGTTACC-ATT
gi 6837905 gb AW3412-	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGTCCCTGCGGCTGGTATTAGGTTACC-ATT
gi 11448468 gb BF436-	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGTCCCTGCGGCTGGTATTAGGTTACC-ATT
gi 5543775 gb AI8698-	AGAAATGGCGCTCCCTGCGGCTGGTATTAGGTTACC-ATT GGCTGGTATTAGGTTACCCATT
gi 1230953 gb N73668-	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGTCCCTGCGGCTGGTATTAGGTTACC-ATT
gi 3254601 gb AI0336-	AAATGGCGTCCCTGCGGCTGGTATTAGGTTACC-ATT
gi 4311745 gb AI4591-	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGTCCCTGCGGCTGGTATTAGGTTACC-ATT
gi 11451784 gb BF439-	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGTCCCTGCGGCTGGTATTAGGTTACC-ATT
gi 4330088 gb AI4679-	CAGTGGTGTAGCCCTTTAGGAGAAATGGCGTCCCTGCGGCTGGTATTAGGTTACC-ATT
Consenso	

gi	23274053	gb	BU608-	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	2669971	gb	AA6826-	GAGCACA	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	3838507	gb	AI2431+	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	3229367	gb	AI0150+	G-GCACC	GAAAGAGACC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	2142013	gb	AA4370+	G-GCACC	GAAAGAGACC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	23527150	gb	BU679+	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	3405022	gb	AI0758	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	1782040	gb	AA1921-	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	3307963	gb	AI0519-	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	4610281	gb	AI6012-	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	3055333	gb	AA9159-	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	3844395	gb	AI2489-	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	10033200	gb	BE6672-	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	4222874	gb	AI3933-	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	26011703	gb	CA774-	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	1162804	gb	N39597+	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	3933768	gb	AI2909-	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	6837905	gb	AW3412-	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	11448468	gb	BF436-	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	5543768	gb	AI8698-	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA		
gi	5543775	gb	AI8698-	T-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	1230953	gb	N73668-	GGCCAC	CGNAGGAAAC	CAGGAGGATAAGAAATTTCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	3254601	gb	AI0336-	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	4311745	gb	AI4591-	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	11451784	gb	BF439-	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
gi	4330088	gb	AI4679-	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA
Consenso						
gi	23274053	gb	BU608-	G-GCACC	GAAAGGAAAC	CAGGAGGATAAGAAATATCCATAAATTTTCAGAGCT-GCCCTGGGCACA

gi 23274053 gb BU608 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 2669971 gb AA6826 -	GTACCTGCC - TGTG - AGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCACAT
gi 3838507 gb AI24314 -	GTACCTGCCCGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 3229367 gb AI01504 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 2142013 gb AA43704 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 34527150 gb BU6794 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 23505022 gb AI0758 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 1782040 gb AA1921 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 3307963 gb AI0519 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 4610281 gb AI6012 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 3055333 gb AA9159 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 3844395 gb AI2489 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 10033200 gb BE672 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 4222874 gb AI3933 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 26011703 gb CA774 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 1162804 gb N395974 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 3933768 gb AI2909 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 6837905 gb AW3412 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 11448468 gb BF436 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 2953300 gb AA8611 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 5543768 gb AI8698 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 5543775 gb AI8698 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 1230953 gb N73668 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 3254601 gb AI0336 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 4311745 gb AI4591 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 11451784 gb BF439 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 4330088 gb AI4679 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T
gi 3836330 gb AI2409 -	GTACCTGCCCCGTGTCGGAGGCTCTCACTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T

CTGG - CAAATG - ACAGCTCTGTGCAAGGAGCAC - T

Consenso

gi 23274053	gb BU608-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 2669971	gb AA6826-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 3838507	gb AI2431+	C
gi 3229367	gb AI0150+	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 2142013	gb AA4370+	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 23527150	gb BU679+	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 3405022	gb AI0758-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 1782040	gb AA1921-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 3307963	gb AI0519-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 4610281	gb AI6012-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 3055333	gb AA9159-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 3844395	gb AI2489-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 10033200	gb BE6672-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 4222874	gb AI3933-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 26011703	gb CA774-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 1162804	gb N39597+	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 3933768	gb AI2909-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 6837905	gb AW3412-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 11448468	gb BF436-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 2953300	gb AA8611-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 5543768	gb AI8698-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 5543775	gb AI8698-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 1230953	gb N73668-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 3254601	gb AI0336-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 4311745	gb AI4591-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 11451784	gb BF439-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 4330088	gb AI4679-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
gi 3836330	gb AI2409-	CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA
Consenso		
		CCCAAGTATAAAAAATTATTACACAGTTTTTATTCTGAAGAACA-TTTTGC-ATTTTAAATAA

gi	23274053	gb	BU608-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	23527150	gb	BU679+	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	3405022	gb	AI0758-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	1782040	gb	AA1921-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	3307963	gb	AI0519-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	4610281	gb	AI6012-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	3055333	gb	AA9159-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	3844395	gb	AI2489-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	10033200	gb	BE672-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	4222874	gb	AI3933-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	26011703	gb	CA774-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	1162804	gb	N39597+	AAAA
gi	3933768	gb	AI2909-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	6837905	gb	AW3412-	AAAAAGGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	11448468	gb	BF436-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	2953300	gb	AA8611-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	5543768	gb	AI8698-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	5543775	gb	AI8698-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	1230953	gb	N73668-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	3254601	gb	AI0336-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	4311745	gb	AI4591-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	11451784	gb	BF439-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	4330088	gb	AI4679-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
gi	3836330	gb	AI2409-	AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTCCAAACCTTGAAGTGTGTTTGGCCTGGAT
Consenso				AAAA-GGATTATGTAGTTCAGGAAAGAGTC-ATTTACAAACCTTGAAGTGTGTTTGGCCTGGAT
				. : . . . : . . . :

gi 23274053 gb BU608 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAACCAAAGTGGTGTGTTGT
gi 23527150 gb BU679+	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 3405022 gb AI0758 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 1782040 gb AA1921 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 3307963 gb AI0519 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 4610281 gb AI6012 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 3055333 gb AA9159 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 3844395 gb AI2489 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 10033200 gb BE672 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 4222874 gb AI3933 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 26011703 gb CA774 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 3933768 gb AI2909 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 6837905 gb AW3412 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 11448468 gb BF436 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 2953300 gb AA8611 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 5543768 gb AI8698 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 5543775 gb AI8698 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 1230953 gb N73668 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 3254601 gb AI0336 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 4311745 gb AI4591 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 11451784 gb BF439 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 4330088 gb AI4679 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
gi 3836330 gb AI2409 -	CAGAGTAAAGAAATGCTTTAAGAAAGAGGTTTGTAAAGTCTTATAAC - AAGTGGTGTGTTGT
Consenso	

gi	23274053	gb	BU608-	TATTACAAAAA
gi	23527150	gb	BU679+	TATTACAAAAA
gi	3405022	gb	AI0758-	TATTACAAAAA
gi	1782040	gb	AA1921-	TATTACAAAAA
gi	3307963	gb	AI0519-	TATTACAAAAA
gi	4610281	gb	AI6012-	TATTACAAA
gi	3055333	gb	AA9159-	TATTACAA
gi	3844395	gb	AI2489-	TATTACA
gi	10033200	gb	BE672-	TATTACAAA
gi	26011703	gb	CA774-	TATTTA
gi	3933768	gb	AI2909-	TATTTA
gi	6837905	gb	AW3412-	TATTTA
gi	11448458	gb	BF436-	TATTACAAAAA
gi	2953300	gb	AA8611-	TATTACAAA
gi	5543768	gb	AI8698-	TATTAC
gi	5543775	gb	AI8698-	TATTAC
gi	1230953	gb	N73668-	TATTACAAAAA
gi	3254601	gb	AI0336-	TATTACA
gi	4311745	gb	AI4591-	TATTACAAAAA
gi	11451784	gb	BF439-	TATTACAAAAA
gi	4330088	gb	AI4679-	TATTACAAAAA
gi	3836330	gb	AI2409-	TATTACAAAAA
Consenso				TATTACAAAAA

gi 11448468 gb BF436 -	GGACCAAAAAAAAAA
gi 1230953 gb N73668 -	GGACC
gi 4311745 gb AI4591 -	GGACCAAAAAAAAAA
gi 11451784 gb BF439 -	GGCCCAAAAAAAAAA
gi 4330088 gb AI4679 -	GGACCAAAAAAAAAA
gi 3836330 gb AI2409 -	GGACCAAAAAAAAAA
Consensus	
	GGACCAAAAAAAAAA

Fig. 6

GGAAAAACATTTTCATCTTAAACTTTTCCTAAGACAAGGGAAAAACAAAAACCATGCTCTA
 CAACTTCAAATTTTTCTTACAAAGAAAAATTTAATATTGATGAGAGGTTGAACCAGGCT
 TAAAGCAGACATACTAGGAAATGGTGCAGCCTGTAAGAATGCCAGTTTGTAACTACTGAC
 TTTGGAAAAGATCATCGCCTCTATCAGACACTTAGGGTCCTGGTCTGGCAATTTTGGCCT
 GATGTGATGCCACAAGACCAACAGAGAGAGACACAGAGTCCAGGATAATGTTGACAGTG
 GTGTAGCCCTTTAGGAGAAATGGCGCTCCCTGCGGCTGGTATTAGGTTACCATTTGGCACC
 GAAGGAACCAGGAGGATAAGAATATCCATAATTTTCAAGAGCTGCCCTGGCACAGTACCTGC
 CCCGTCCGAGGCTCTCACTGGCAAATGACAGCTCTGTGCAAGGAGCACTCCCAAGTATAA
 AAATTATTACACAGTTTTATTCTGAAGAACATTTTGCATTTTAATAAAAAAGGATTTATG
 TCAGGAAAAGAGTCATTTACAAACCTTGAAGTGTTTTTGCCTGGATCAGAGTAAGAATGTC
 TTAAGAAGAGGTTTGTAAAGGTCTTCATAACAAAGTGGTGTTTGTATTACAAAAAAA
 AAAAAAAAAAATTAACAGGTTGTCTGTATACTATTAAAAATTTTGGACCAAAAAAAAAA
 AAAAAAAA

Fig. 7

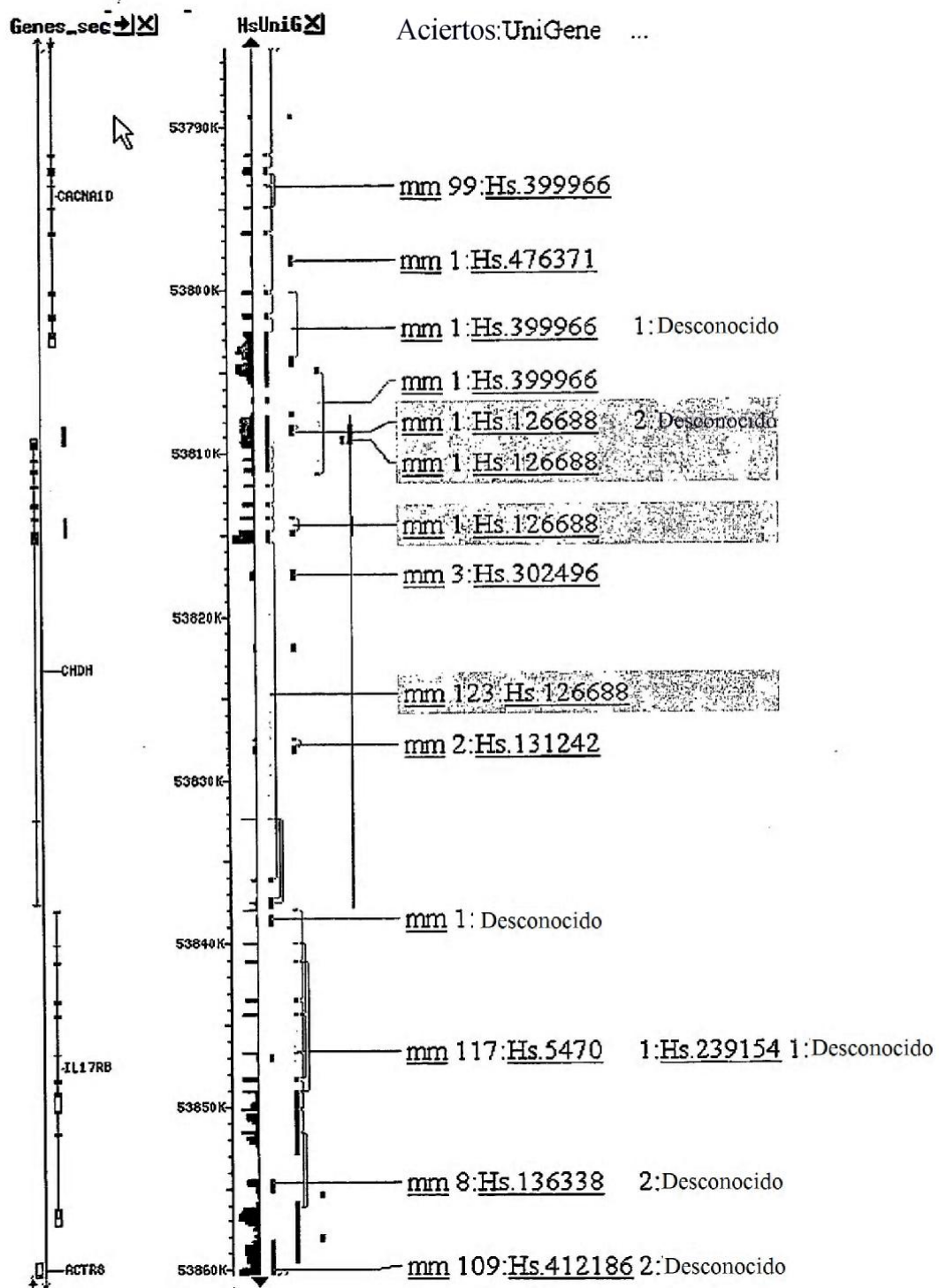
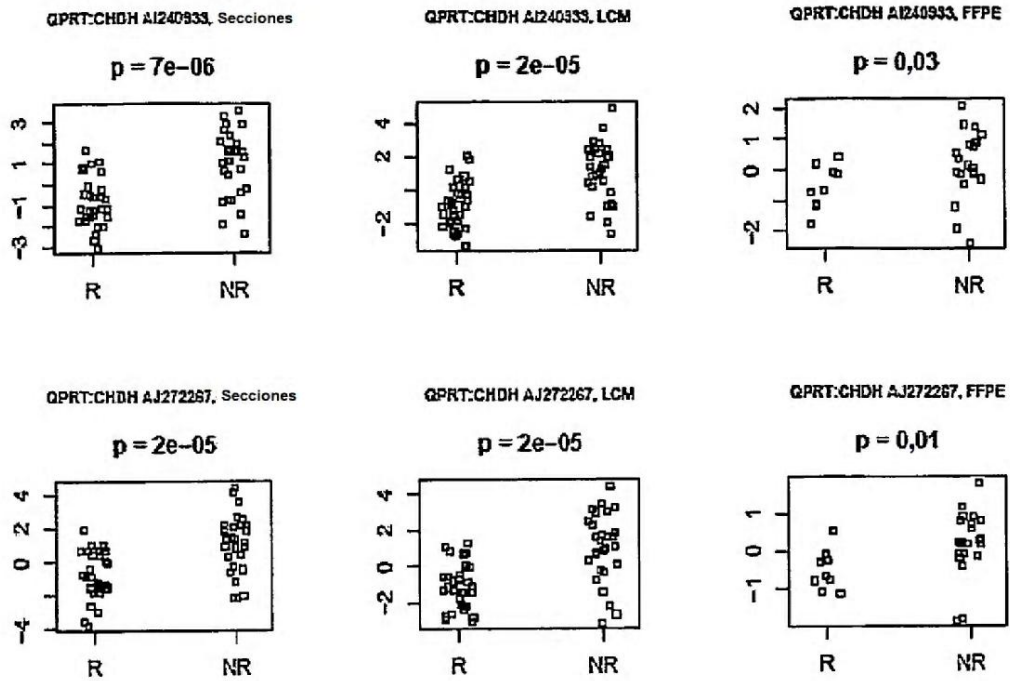


Fig. 8

Parte A



Parte B

