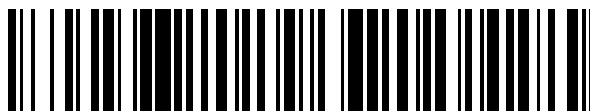


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 550 652**

51 Int. Cl.:

C12Q 1/68

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **02.06.2006 E 06771963 (3)**

97 Fecha y número de publicación de la concesión europea: **12.08.2015 EP 1899484**

54 Título: **Identificación de tumores y tejidos**

30 Prioridad:

03.06.2005 US 687174 P

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

11.11.2015

73 Titular/es:

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

72 Inventor/es:

**ERLANDER, MARK G. y
MA, XIAO-JUN**

74 Agente/Representante:

PONS ARIÑO, Ángel

Observaciones :

Véase nota informativa (Remarks) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 550 652 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

Identificación de tumores y tejidos

5 Campo de la invención

La presente invención se refiere al uso de expresión genética para clasificar tumores humanos. La clasificación se realiza mediante el uso de perfiles, o patrones de expresión genética, de aproximadamente 5 a 49 secuencias expresadas que se correlacionan con tumores que surgen a partir de ciertos tejidos así como que se correlacionan con ciertos tipos de tumor. La invención también proporciona el uso de aproximadamente 5 a 49 secuencias genéticas específicas, cuya expresión se correlaciona con fuente de tejido y tipo de tumor en diversos cánceres. Los perfiles de expresión genética, ya sea realizados en formatos de expresión de ácidos nucleicos, expresión de proteínas, u otros formatos de expresión, se pueden usar para determinar una muestra que contiene células como que contiene células tumorales de un tipo de tejido o a partir de un origen de tejido para permitir una identificación más precisa del cáncer y por lo tanto el tratamiento del mismo así como el pronóstico del sujeto a partir del que se obtuvo la muestra.

Antecedentes de la invención

El documento de patente WO 03/041562 informa de métodos para la clasificación de tipos de enfermedades basándose en clasificadores algorítmicos usados para realizar grandes conjuntos de datos. Golub *et al.*, (Science 1999 286: 531-7) informa de un enfoque para la clasificación del cáncer basándose en el control de la expresión genética mediante micromatrices de ADN y aplica el enfoque a leucemias agudas humanas. El documento de patente WO 2005/059109 describe la identificación de una firma molecular que comprende genes se asocian muy de cerca con mutaciones en el supresor de tumor PTEN. El conjunto de la matriz de U133 del genoma humano de GeneChip de Affymetrix (HG-U133A) es una matriz de oligonucleótidos de alta densidad que consiste en dos matrices de GeneChip y que contiene conjuntos de sondas que representan más de 39.000 transcritos derivados de aproximadamente 33.000 genes humanos.

30 Sumario de la invención

La presente invención se refiere a un método para clasificar una muestra que contiene células como que contiene células tumorales de un tipo de tejido, comprendiendo dicho método: determinación de los niveles de expresión de 5 a 49 secuencias transcritas de células en una muestra que contiene células obtenidas de un sujeto humano, y comparación de los niveles de expresión con los niveles de expresión de las mismas secuencias transcritas en una pluralidad de tipos de tejido de tumor conocidos, y clasificación de la muestra como que contiene células tumorales de un tipo de tejido de la pluralidad de tipos de tejido tumoral conocidos, en el que las cinco a 49 secuencias transcritas no se seleccionan basándose en sus valores de correlación, o una clasificación basada en los valores de correlación, con los tipos de tejido tumoral. La invención proporciona de este modo la capacidad para clasificar tumores en las condiciones del mundo real afrontadas por hospitales y otros laboratorios que realizan ensayos en muestras clínicas de FFPE. Las muestras pueden ser de una muestra de tumor primario o de un tumor que ha surgido de una metástasis de otro tumor. Como alternativa, la muestra puede ser una muestra citológica, tal como, pero no se limita a, células en una muestra de sangre. En algunos casos de una muestra de tumor, los tumores pueden no haber experimentado clasificación mediante técnicas de patología tradicionales, se pueden haber clasificado inicialmente pero se desea una confirmación, o se han clasificado como un "carcinoma de origen primario desconocido" (CUP) o "tumor de origen desconocido" (TUO) o "tumor primario desconocido". La necesidad de confirmaciones particularmente relevante a la vista de las estimaciones de mala clasificación de un 5% a un 10 % usando técnicas convencionales. Por lo tanto, se puede visualizar que la invención proporciona un medio para la identificación del cáncer, o CID.

En un primer aspecto de la invención, la clasificación se realiza mediante el uso de perfiles, o patrones de expresión genética de 5 a 49 secuencias expresadas. Los perfiles de expresión genética, ya sea realizados en expresión de ácidos nucleicos, expresión de proteínas, u otros marcadores de expresión genética, se pueden usar para determinar una muestra que contiene células como que contiene células tumorales de un tipo de tejido o de un origen de tejido para permitir una identificación más precisa del cáncer y por lo tanto un tratamiento del mismo así como el pronóstico del sujeto a partir del que se obtuvo la muestra.

En algunas realizaciones, la invención se usa para su clasificación entre al menos 34 o al menos 39 tipos de tumor con una precisión significativa en una instalación clínica. La invención se basa en parte en el descubrimiento sorprendente e inesperado de que 5 a 49 secuencias expresadas en el genoma humano son capaces de clasificarse entre al menos 34, o al menos 39, tipos de tumor, así como subconjuntos de esos tipos de tumor, de una forma significativa. Indicado de forma diferente, la invención se basa en parte en el descubrimiento de que no es necesario usar un aprendizaje supervisado para identificar secuencias genéticas que se expresan en correlación con diferentes tipos de tumor. Por lo tanto, la invención se basa en parte en el reconocimiento de que cualquiera de 5 a 49 secuencias expresadas, incluso una colección aleatoria de secuencias expresadas, tiene la capacidad de clasificación, y de este modo se pueden usar para clasificar, una célula como que es una célula tumoral de un tejido

u origen tisular. Además, se necesitan secuencias relativamente poco expresadas para su clasificación entre diferentes tipos de tumor. La proporción de sus fuentes expresadas con respecto al número de tipos de tumor que se pueden clasificar, basándose en los niveles de expresión de las secuencias, varía de 1:2 a 5:2 o superior como se demuestra en el presente documento.

En otro aspecto, la invención proporciona la clasificación de una muestra que contiene células como que contiene una célula de tumor de un tipo u origen de tejido mediante la determinación de los niveles de expresión de 5 a 49 secuencias transcritas a continuación la clasificación de la muestra que contiene células como que contiene una célula tumoral de una pluralidad (dos o más) tipos de tumor. Para la clasificación entre 34 y 39 tipos de tumor, y subconjuntos de los mismos, se pueden usar tan pocas como cualquiera de 5 secuencias expresadas para proporcionar clasificación de una manera significativa. Se descubrió que no era necesario que las secuencias expresadas fueran las de los niveles de expresión con las que se correlacionaban de forma evidente o elevada (directamente, o indirectamente a través de correlación con otra secuencia expresada) con cualquiera de los tipos de tumor. Por lo tanto, la invención proporciona, en otra realización más, el uso de los niveles de expresión genética, cuyos niveles de expresión no están correlacionados fuertemente con la clasificación real de la muestra de tumor en particular, como una de las 5 a 49 secuencias transcritas. Todos los genes seleccionados pueden ser tales no correlacionados, o solamente una parte de los genes pueden estar uno correlacionados, por lo general al menos un 90 %, 85 %, 75 %, 50 % o un 25 %, así como partes que entran dentro de los intervalos creados mediante el uso de cualquiera de dos de los ejemplos puntuales mencionados anteriormente como puntos finales de un intervalo.

La invención se pone en práctica mediante la determinación de los niveles de expresión de secuencias genéticas en la que no es necesario que las secuencias se hayan seleccionado basándose en una correlación de sus niveles de expresión con los tipos de tumor a clasificar. Por lo tanto, como un ejemplo limitante, no es necesario que las secuencias genéticas se seleccionen basándose en sus valores de correlación con tipos de tumor o una clasificación basada en los valores de correlación. Además, la invención se puede poner en práctica con el uso de niveles de expresión genética que no se correlacionan necesariamente con uno o más de otros niveles de expresión genética usados para la clasificación. Por lo tanto, en algunas realizaciones, la capacidad del nivel de la expresión de una secuencia expresada para funcionar en la clasificación no es redundante con (es independiente de) la capacidad de al menos otro nivel de expresión genética usado para la clasificación.

La invención se puede aplicar para identificar el origen de un cáncer en un paciente en una gran diversidad de casos que incluyen, pero no se limitan a, identificación del origen de un cáncer en una instalación clínica. En algunas realizaciones, la identificación se realiza mediante la clasificación de una muestra que contiene células conocida porque contiene células cancerígenas, pero se desconoce el origen de esas células. En otras realizaciones, la identificación se realiza mediante la clasificación de la muestra que contiene células como que contiene una o más células cancerígenas seguido de identificación del origen u orígenes de esa célula o células cancerígenas. En otras realizaciones, la invención se pone en práctica con una muestra de un sujeto con historia previa de cáncer, y la identificación se realiza mediante la clasificación de la célula ya sea como que es cáncer a partir de un origen previo de cáncer o un nuevo origen. Algunas realizadas en sanciones incluyen aquellas en las que se encontraron múltiples cánceres en el mismo órgano o tejido y la invención se usa para determinar el origen de cada cáncer, así como si los cánceres tienen el mismo origen.

La invención también se basa en parte en el descubrimiento de que los niveles de expresión de secuencias genéticas en particular se pueden usar para clasificar entre tipos de tumor con mayor precisión que los niveles de expresión de un grupo aleatorio de secuencias genéticas. En una realización, la invención proporciona el uso de niveles de expresión de 5 a 49 secuencias expresadas de un primer conjunto de 74 secuencias expresadas en el genoma humano para clasificar entre al menos 39 tipos de tumor con una precisión significativa. Por lo tanto, la invención proporciona la identificación y uso de patrones de expresión genética (o perfiles o "firmas") basados en las 5 a 49 secuencias expresadas como correlacionadas con al menos los 39 tipos de tumor. La invención también proporciona el uso de 5 a 49 de las 74 de estas secuencias expresadas para clasificar entre subconjuntos de los 39 tipos de tumor. La proporción de secuencias expresadas con respecto al número de tipos de tumor, de 2 a 39, que se pueden clasificar basándose en los niveles de expresión de los intervalos de las secuencias de aproximadamente 1:2 a aproximadamente 5:2 con mayor precisión que el uso de un grupo aleatorio de secuencias expresadas. Dependiendo del número de tipos de tumor, las precisiones que varían de aproximadamente un 75 % a un 95 % se pueden conseguir fácilmente.

En otra realización, la invención proporciona el uso de niveles de expresión de 5 a 49 secuencias expresadas de un segundo conjunto de 90 secuencias expresadas en el genoma humano para clasificar entre al menos 39 tipos de tumor, o subconjuntos de los mismos, con precisión significativa. 38 de las secuencias en este segundo conjunto están presentes en el primer conjunto de 74 secuencias. Los niveles de expresión de las 5 a 49 secuencias en el segundo conjunto se pueden usar de la misma manera que se describe para el primer conjunto de 74 secuencias. Dependiendo del número de tipos de tumor, se pueden conseguir precisiones que varían de aproximadamente un 75 % a aproximadamente un 95 %.

La invención también se basa en parte en el descubrimiento de que el uso de 5 a 49 secuencias expresadas para clasificar entre 53 tipos de tumor, que incluyen (pero no se limitan a) los 34 y 39 tipos que se describen en el

presente documento, se limitaba al número de muestras disponibles de algunos tipos de tumor. Como se indica en lo sucesivo en el presente documento, la precisión se une al número de muestras disponibles de cada tipo de tumor de modo que la capacidad para clasificar tipos de tumor adicionales se consigue fácilmente mediante la aplicación del aumento de números de cada tipo de tumor. Por lo tanto, aunque la invención se ejemplifica mediante el uso en la

5 clasificación entre 34 o 39 tipos de tumor así como subconjuntos de las 34 o 39, también se pueden usar de 5 a 49 secuencias expresadas para clasificar entre todos los tipos de tumor con la inclusión de muestras de los tipos de tumor adicionales. Por lo tanto, la invención también proporciona la clasificación de un tumor como que es un tipo más allá de los 34 o 39 tipos que se describen en el presente documento.

10 La invención se basa en los niveles de expresión de las secuencias genéticas en un conjunto de células tumorales conocidas de diferentes tejidos y de diferentes tipos de tumor. Estos perfiles de expresión genética (de secuencias genéticas en los diferentes células/tipos de tumor conocidos), ya sea incluidos en expresión de ácidos nucleicos, expresión de proteínas, u otros formatos de expresión, se pueden comparar con los niveles de expresión de las mismas secuencias en una muestra de tumor desconocido para identificar la muestra como que contiene un tipo en particular y/o un origen en particular o tipo de célula en particular. La invención proporciona, tal como en una

15 instalación clínica, las ventajas de una identificación más precisa de un cáncer y por lo tanto el tratamiento del mismo así como el diagnóstico, incluyendo supervivencia y/o probabilidad de aparición del cáncer después del tratamiento, del sujeto a partir del que se obtuvo la muestra.

20 La invención se basa adicionalmente en parte en el descubrimiento de que el uso de 5 a 49 secuencias expresadas como se describe en el presente documento como capaz de clasificar entre dos o más tipos de tumor elimina de forma necesaria y eficaz uno o más tipos de tumor desde la consideración durante la clasificación. Esto refleja la falta de la necesidad de genes seleccionados con niveles de expresión que se correlaciona en gran medida los tipos de tumor dentro del intervalo del sistema de clasificación. Indicado de forma diferente, la invención se puede poner

25 en práctica con una pluralidad de genes cuyos niveles de expresión no se correlacionan en gran medida con ninguno de los tipos de tumor individuales o tipos múltiples en el grupo de tipos de tumor que se están clasificando. Esto se produce como contraste con respecto a otros enfoques que se basan en la selección y los de genes muy correlacionados, que probablemente no "excluyen" otros tipos de tumor en oposición a la "inclusión" de un tipo de tumor basado en la correlación positiva.

30 La clasificación de la muestra tumoral como uno de los posibles tipos de tumor que se describen en el presente documento para la exclusión de otros tipos de tumor se realiza por supuesto basándose en un nivel de confianza como se describe a continuación. Cuando el nivel de confianza es bajo, o se prefiere un aumento en el nivel de confianza, la clasificación se puede realizar simplemente al nivel de un origen de tejido o tipo de célula en particular para el tumor en la muestra. Como alternativa, y cuando una muestra de tumor no se clasifica rápidamente con un solo tipo de tumor, la invención permite la clasificación de la muestra como uno de unos cuantos posibles tipos de tumor que se describen en el presente documento. Esto proporciona de forma ventajosa la capacidad de reducir el número de posibles tipos de tejidos, tipos de células, y tipos de tumor desde los que se consideran para la selección y administración de terapia al paciente a partir de que se obtuvo la muestra.

40 Por lo tanto, la invención proporciona un medio no subjetivo para la identificación del agente de tejido y/o tipo de tumor de uno o más cánceres de un sujeto afectado. Cuando anteriormente se ha usado una interpretación subjetiva para determinar la fuente de tejido y/o tipo de tumor, así como el diagnóstico y/o tratamiento del cáncer basándose en esa determinación, la presente invención proporciona patrones de expresión genética objetivos, que se pueden

45 usar solos o en combinación con criterios subjetivos para proporcionar una identificación más precisa de la clasificación del cáncer. La invención se aplica de forma particularmente ventajosa muestras de tumores secundarios o con metástasis, pero también se puede usar con la invención cualquier muestra que contiene células (incluyendo una muestra de tumor primario) para la que se determina preferentemente la fuente de tejido y/o tipo de tumor mediante criterios objetivos. Por supuesto, la última determinación de la clase se puede realizar basándose en una combinación de criterios objetivos y no objetivos (o subjetivos/parcialmente subjetivos).

50 La invención incluye su uso como parte del cuidado clínico o médico de un paciente. Por lo tanto, además del uso de un perfil de expresión de genes como se describe en el presente documento para someter a ensayo una muestra que contiene células de un sujeto afectado con cáncer para determinar la fuente de tejido y/o tipo de tumor del

55 cáncer, el perfil también se puede usar como parte de un método para determinar el diagnóstico del cáncer del sujeto. La clasificación del tumor/cáncer y/o el diagnóstico se puede usar para seleccionar o determinar o alterar el tratamiento terapéutico para dicho sujeto. Por lo tanto, los métodos de clasificación de la invención se pueden dirigir hacia el tratamiento de la enfermedad, que se diagnostica totalmente o en parte basándose en la clase. Dado el diagnóstico, para tratar el cáncer se puede usar la administración de un agente o terapia antitumoral apropiados, o la detención o la alternancia de un agente o terapia antitumoral.

Otros métodos clínicos incluyen los implicados en la provisión de cuidados médicos a un paciente basándose en una clasificación como se describe en el presente documento. En algunas realizaciones, los métodos se refieren a la provisión de servicios de diagnóstico basándose en niveles de expresión de secuencias genéticas, con o sin ilusión

65 de una interpretación de los niveles para la clasificación de células de una más. En algunas realizaciones, el método de provisión de un servicio diagnóstico de la invención va precedido por una determinación de una necesidad del

servicio. En otras realizaciones, el método incluye actuaciones en el control del de la realización del servicio así como actuaciones en la solicitud o recepción de reembolso por la realización del servicio.

- 5 Los detalles de una o más organizaciones de la invención se exponen en las figuras adjuntas y la distinción que sigue a continuación. Otras características, objetos, y ventajas de la invención serán evidentes a partir de las figuras y la descripción detallada, y a partir de las reivindicaciones.

Definiciones

- 10 Como se usa en el presente documento, un "gen" es un polinucleótido que codifica un producto específico, ya sea ARN o tenga una naturaleza proteica. Se observa que más de un polinucleótido puede ser capaz de codificar un producto específico. El término incluye alelos y polimorfismos de un gen que codifica el mismo producto, o un análogo funcionalmente asociado del mismo (incluyendo ganancia, pérdida, o modulación de la función), basándose en la ubicación del cromosoma y la capacidad para recombinación durante la mitosis normal.

- 15 Como se usa en el presente documento, una "secuencia" o "secuencia genética" es una molécula de ácido nucleico o polinucleótido formada por un orden específico de bases de nucleótidos. El término incluye la ordenación de las bases que codifican un producto específico (es decir "región de codificación"), ya sea ARN o con una naturaleza proteica. Se observa que más de un polinucleótido puede ser capaz de codificar un producto específico. También se observa que pueden existir alelos y polimorfismos de las secuencias genéticas humanas y may se pueden usar en la práctica de la invención para identificar el nivel o niveles de expresión de las secuencias genéticas o un alelo o polimorfismo del mismo. La identificación de un alelo o polimorfismo depende en parte en la ubicación del cromosoma y la capacidad de recombinación durante la mitosis.

- 25 Los términos "correlacionar" o "correlación" o equivalentes de los mismos se refieren a una asociación entre expresión de uno o más genes y otro suceso, tal como, pero no se limita a, fenotipo o característica fisiológicos, tal como tipo de tumor.

- 30 Un "polinucleótido" es una forma polimérica de nucleótidos de cualquier longitud, ya sea ribonucleótidos o desoxirribonucleótidos. Este término se refiere solamente a la estructura primaria de la molécula. Por lo tanto, este término incluye ADN y ARN bi y monocatenario. También incluye tipos conocidos de modificaciones que incluyen marcas conocidas en la técnica, metilación, "protecciones", sustitución de uno o más de los nucleótidos de origen natural con un análogo, y modificaciones internucleótidos tales como uniones sin carga (por ejemplo, fosforotioatos, fosforditioatos, etc.), así como formas no modificadas del polinucleótido.

- 35 El término "amplificar" se usa en su sentido amplio para hacer referencia a que la creación de un producto de amplificación se puede realizar de forma enzimática con ADN o ARN polimerasas. "Amplificación", como se usa en el presente documento, se refiere generalmente al proceso de producción de múltiples copias de una secuencia deseada, en particular las de una muestra. "Múltiples copias" hace referencia al menos 2 copias. No es necesario que una "copia" haga referencia necesariamente a complementariedad de secuencias o identidad con la secuencia molde perfectas. Por lo general se conocen en la técnica algunos métodos para amplificación de ARNm, e incluyen PCR de transcripción inversa (RT-PCR) y PCR cuantitativa (o Q-PCR) o PCR en tiempo real. Como alternativa, el ARN se puede marcar directamente como el ADNc correspondiente mediante métodos conocidos en la técnica.

- 45 Por "que corresponde", se hace referencia a que una molécula de ácido nucleico comparte una cantidad sustancial de identidad de secuencia con otra molécula de ácido nucleico. Cantidad sustancial se refiere al menos un 95 %, normalmente al menos un 98 % y más habitualmente al menos un 99 %, y la identidad de la secuencia se determina usando el algoritmo BLAST, como se describe en Altschul *et al.* (1990), J. Mol. Biol. 215: 403-410 (usando el ajuste por defecto publicado, es decir los parámetros $w = 4$, $t = 17$).

- 50 Una "micromatriz" es una matriz lineal o de dos dimensiones o tridimensional (y fase sólida) de regiones específicas, cada una de las cuales tiene un área definida, formada en la superficie de un soporte sólido tal como, pero no se limita a, membranas de vidrio, plástico, o sintéticas. La densidad de las regiones específicas en una micromatriz se determina mediante los números totales de polinucleótidos inmovilizados a detectar en la superficie de un solo soporte en fase sólida, tal como de al menos aproximadamente $50/\text{cm}^2$, al menos aproximadamente $100/\text{cm}^2$, o al menos aproximadamente $500/\text{cm}^2$, hasta aproximadamente $1.000/\text{cm}^2$ o superior. Las matrices pueden contener menos de aproximadamente 500, aproximadamente 1000, aproximadamente 1500, aproximadamente 2000, aproximadamente 2500, o aproximadamente 3000 polinucleótidos inmovilizados en total. Como se usa en el presente documento, una micromatriz de ADN es una matriz de sondas de oligonucleótidos o polinucleótidos colocadas en un chip u otras superficies usadas para su hibridación con polinucleótidos amplificados o clonados a partir de una muestra. Dado que se conoce la posición de cada grupo de sondas en particular en la matriz, las identidades de los polinucleótidos de la muestra se pueden determinar basándose en su unión a una posición en particular en la micromatriz. Como una alternativa para el uso de una micromatriz, en la práctica de la invención se puede usar una matriz de cualquier tamaño, que incluye una colocación de una o más posiciones de una colocación de dos dimensiones o tridimensional en una fase sólida para detectar la expresión de una sola secuencia genética. En algunas realizaciones, una micromatriz para uso con la presente invención se puede preparar mediante técnicas

fotolitográficas (tales como síntesis de sondas de ácidos nucleicos en la superficie del extremo en la posición 3') o mediante síntesis de ácidos nucleicos seguido de deposición en una superficie sólida.

Dado que la invención se basa en la identificación de la expresión genética, algunas realizaciones de la invención determinan la expresión mediante hibridación de ARNm, o una versión amplificada o clonada del mismo, de una muestra celular con respecto a un polinucleótido que es único para una secuencia genética en particular. Los polinucleótidos de este tipo 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 consecutivas de una secuencia genética que no se encuentra en otras secuencias genéticas. El término "aproximadamente" como se ha usado en la frase anterior se refiere a un aumento o disminución de 1 a partir del valor numérico indicado. Otras realizaciones son 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 bases consecutivas de una secuencia que no se encuentra en otras secuencias genéticas. El término "aproximadamente" como se ha acusado en la frase mencionada anteriormente se refiere a un aumento o disminución de un 10 % a partir del valor numérico indicado. Por supuesto que los polinucleótidos más largos pueden contener faltas de coincidencias menores (por ejemplo, mediante la presencia de mutaciones) que no influyen en la hibridación con los ácidos nucleicos de una muestra. Tales polinucleótidos también se pueden denominar sondas de polinucleótidos que son capaces de hibridarse con secuencias de los genes, o porciones únicas de los mismos, que se describen en el presente documento. Tales polinucleótidos se pueden marcar para ayudar en su detección. Las secuencias pueden ser las de ARNm codificado por los genes, el ADNc correspondiente a tales ARNm, y/o versiones amplificadas de tales secuencias. En algunas realizaciones de la invención, las sondas de polinucleótidos se inmovilizan en una matriz, otros dispositivos de soporte sólido, o en aplicaciones puntuales individuales que localizan las ondas.

En otras realizaciones de la invención, toda o parte de una secuencia genética se fue amplificar y detectar mediante métodos tales como la reacción en cadena de la polimerasa (PCR) y variaciones de la misma, tal como, pero no se limitan a, PCR cuantitativa (Q-PCR), PCR por transcripción inversa (RT-PCR), y PCR en tiempo real (que incluye, como un medio para medir las cantidades iniciales de ARNm, copias para cada secuencia en una muestra), opcionalmente RT-PCR en tiempo real o Q-PCR en tiempo real. Tales métodos usarían uno o dos cebadores que son complementarios con partes de una secuencia genética, en los que los cebadores se usan para cebar la síntesis de ácidos nucleico. Los ácidos nucleicos recién sintetizados se marcan opcionalmente y se pueden detectar directamente o mediante hibridación con un polinucleótido de la invención. Los ácidos nucleicos recién sintetizados se pueden poner en contacto con polinucleótidos (que contienen secuencias) de la invención en condiciones que permiten su hibridación. Algunos métodos adicionales para detectar la expresión de ácidos nucleicos expresados incluyen ensayos de protección de ARNsa, que incluyen y liquidaciones en fase líquida, e hibridación de células *in situ*.

Como alternativa, y en realizaciones adicionales de la invención, la expresión genética se puede determinar mediante análisis de proteína expresada en una muestra de células de interés mediante el uso de uno o más anticuerpos específicos para uno o más epítomos de productos genéticos individuales (proteínas), o fragmentos proteolíticos de los mismos, en dicha muestra celular o en un fluido del organismo de un sujeto. La muestra celular puede ser una de células epiteliales de cáncer de mama enriquecidas a partir de la sangre de un sujeto, tal como mediante el uso de anticuerpos marcados frente a marcadores de superficie celular seguido de clasificación celular activada con fluorescencia (FACS). Tales anticuerpos se pueden marcar para permitir su detección después de su unión con el producto genético. Algunas metodologías de detección adecuadas para uso en la práctica de la invención incluyen, pero no se limitan a, inmunohistoquímica de muestras que contienen células o tejidos, ensayos de inmunoabsorción ligados a enzimas (ELISA) que incluyen ensayos de sándwich de anticuerpos de muestras de tejidos que contienen células o muestras de sangre, espectroscopía de masas, y inmuno-PCR.

Los términos "marca" o "marcado" se refieren a una composición capaz de producir una señal detectable indicativa de la presencia de la molécula marcada. Algunas marcas adecuadas incluyen radioisótopos, cromóforos de nucleótidos, enzimas, sustratos, moléculas fluorescentes, restos quimioluminiscentes, partículas magnéticas, restos bioluminiscentes, y similares. Como tal, una marca es cualquier composición detectable por medios espectroscópicos, fotoquímicos, bioquímicos, inmunoquímicos, eléctricos, ópticos o químicos.

El término "soporte" se refiere a soportes convencionales tales como perlas, partículas, varillas, fibras, filtros, membranas y soportes de silano o silicato tales como placas de vidrio.

"Expresión" y "expresión genética" incluyen transcripción y/o traducción de material de ácido nucleico.

Como se usa en el presente documento, la expresión "que comprende" y sus afines se usan en su sentido inclusivo; es decir, equivalente a la expresión "que incluye" y sus afines correspondientes.

Algunas condiciones que "permiten" que se produzca un suceso o condiciones que son "adecuadas" para que se

produzca un suceso, tales como hibridación, extensión de la hebra, y similares, o condiciones "adecuadas" son condiciones que no evitan que tales sucesos se produzcan. Por lo tanto, estas condiciones permiten, aumentan, facilitan, y/o conducen al suceso. Tales condiciones, se conocen en la técnica y se describen en el presente documento, dependen de, por ejemplo, la naturaleza de la secuencia de nucleótidos, temperatura, y condiciones del tampón. Estas condiciones también dependen de qué suceso se desea, tal como hibridación, escisión, extensión la transcripción de la hebra.

"Mutación en la secuencia", como se usa en el presente documento, se refiere a cualquier alteración de la secuencia en la secuencia de un gen de interés que se desvela en el presente documento en comparación con una secuencia de referencia. Una mutación en la secuencia incluye cambios de un solo nucleótido, o alteraciones de más de un nucleótido en una secuencia, debido mecanismos tales como sustitución, supresión o inserción. Como se usa en el presente documento el polimorfismo de un solo nucleótido (SNP) también es una mutación de la secuencia como se usa el presente documento. Dado que la presente invención se basa en el nivel relativo de expresión genética, en la práctica de la invención también se pueden someter a ensayo algunas mutaciones en regiones de no codificación de genes como se desvelan el presente documento.

"Detección" o "que detecta" incluyen cualquier medio de detección, incluyendo la determinación directa e indirecta del nivel de expresión genética y cambios en el mismo.

A menos que se defina de otro modo, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado como normalmente lo entiende un experto habitual en la materia a la que pertenece la presente invención.

Breve descripción de las figuras

La Figura 1 muestra una representación de la aptitud para la capacidad para usar los niveles de expresión de subconjuntos de un conjunto de 100 secuencias genéticas expresadas para clasificar entre 39 tipos de tumor y subconjuntos de los mismos. Se usaron algunos niveles de expresión de combinaciones aleatorias de 5, 10, 15, 20, 25, 30, 35, 40, 45, y 49 (se toma muestra de cada una 10 veces) de las 100 secuencias con datos de tipos de tumor y a continuación se usaron para predecir conjuntos aleatorios del ensayo de muestras de tumor (se toma muestra de cada una 10 veces) que varían de 2 a 39 tipos. Una representación de los números de tipos de tumor (eje x) con respecto a las precisiones de predicción (eje y) para resultados que gustan de 5 a 49 genes se muestran como ejemplos no limitantes. Los datos del uso de 5 genes dan como resultado la curva más cercana al eje x mientras que los datos del uso de 49 genes dan como resultado la curva más alejada del eje x. Generalmente, la precisión aumenta con números más elevados de secuencias genéticas, en las que de 30 a 49 secuencias genéticas (las tres curvas más alejadas del eje x) proporcionan aproximadamente el mismo nivel de precisión.

La Figura 2 muestra una presentación alternativa de los datos usados con respecto a la Figura 1. Se muestra una representación de los números de secuencias genéticas usadas, que varían de 5-49 (y en el eje x), con respecto a las precisiones de predicción (eje y) para diversos números representativos de tipos de tumor. Líneas representadas, desde la parte superior a la inferior, son los resultados de 2, 10, 20, 30, y 39 tipos de tumor, respectivamente.

La Figura 3 proporciona un análisis adicional de la aptitud para usar los niveles de expresión de subconjuntos de un conjunto de 100 secuencias genéticas seleccionadas de forma aleatoria expresadas para clasificar entre 39 tipos de tumor. Los datos usados con las Figuras 1 y 2 se presentan en una representación del número de tipos de tumor frente al número de secuencias genéticas usadas con precisiones de predicción de un 55-70 % se muestran como ejemplos no limitantes. Generalmente, la precisión aumenta con números más elevados de secuencias genéticas.

La Figura 4 muestra una representación de la aptitud para la capacidad para usar los niveles de expresión de partes de un primer conjunto de 74 secuencias genéticas expresadas para clasificar entre 39 tipos de tumor y subconjuntos de los mismos. Se usaron niveles de expresión de combinaciones aleatorias de 5, 10, 15, 20, 25, 30, 35, 40, 45, y 49 (se toma muestra de cada una 10 veces) de las 74 secuencias con datos de tipos de tumor y a continuación se usaron para predecir conjuntos aleatorios de ensayo de muestras de tumor (se toma muestra de cada una 10 veces) que varían de 2 a 39 tipos. Una representación de números de tipos de tumor frente a precisiones de predicción para resultados usando de 5 a 49 genes se muestra como ejemplos no limitantes. Las líneas representadas, de la parte superior a la inferior, son los resultados de 49, 40, 30, 20, 10, y secuencias genéticas, respectivamente.

La Figura 5 muestra una presentación alternativa de los datos usados con respecto a la Figura 4. Se muestra una representación de números de secuencias genéticas usadas, que varía de 5-49, frente a precisiones de predicción para diversos números representativos de tipos de tumor. Las líneas representadas, de la parte superior a la inferior, son los resultados de 2, 10, 20, 30, y 39 tipos de tumor, respectivamente.

La Figura 6 es análoga a la Figura 3 excepto en la presentación de los datos usados con las Figuras 4 y 5.

La Figura 7 muestra una representación de la aptitud para la capacidad para usar los niveles de expresión de subconjuntos de un conjunto de 90 secuencias genéticas expresadas para clasificar entre 39 tipos de tumor y

subconjuntos de los mismos. Se usaron niveles de expresión de combinaciones aleatorias de 5, 10, 15, 20, 25, 30, 35, 40, 45, y 49 (se toma muestra de cada una 10 veces) de las 90 secuencias con datos de tipos de tumor y a continuación se usaron para predecir conjuntos de ensayo aleatorios usados para predecir conjuntos de ensayo aleatorios de muestras de tumor (se toma muestra de cada una 10 veces) que varían de 2 a 39 tipos. Se muestra una representación de números de tipos de tumor frente a precisiones de predicción para resultados usando de 5 a 49 genes como ejemplos no limitantes. Las líneas representadas, de la parte superior a la inferior, son los resultados de 49, 40, 30, 20, 10, y secuencias genéticas, respectivamente.

La Figura 8 muestra una presentación alternativa de los datos usados con respecto a la Figura 7. Se muestra una representación de números de secuencias genéticas usadas, que varían de 5-49, frente a precisiones de predicción para diversos números representativos de tipos de tumor. Las líneas representadas, de la parte superior a la inferior, son los resultados de 2, 10, 20, 30, y 39 tipos de tumor, respectivamente.

La Figura 9 es análoga a las Figuras 3 y 6 excepto en la presentación de los datos usados con las Figuras 7 y 8.

Las Figuras 10A-10D muestran un "árbol" que clasifica tipos de tumor incluidos en el presente documento así como tipos de tumor conocidos adicionales. Se construyó principalmente de acuerdo con "Cancer, Principles and Practice of Oncology, (DeVito, Hellman y Rosenberg), 6^a edición". Por lo tanto, comenzando con un "tumor de origen desconocido" (o "tuo"), las primeras posibilidades son que se trata ya sea de un origen de célula germinal o de célula no germinal. Si se trata del primero, entonces puede ser de origen en ovario o testículo. Dentro de los de origen en el testículo, el tumor puede ser de origen seminoma o de "otro" origen.

Si el tumor es de un origen en célula no germinal, entonces es cualquiera de un origen epitelial o no epitelial. Si es el primero, entonces es cualquiera de origen escamoso o no escamoso. Los tumores de origen escamoso tienen origen en el cuello uterino, esófago, laringe, pulmón, o piel. Los tumores de origen no escamoso tienen origen en la vejiga urinaria, mama, carcinoide de intestino, colangiocarcinoma, digestivo, riñón, hígado, pulmón, próstata, sistema reproductor, células cutáneas-basales, o tiroides-folicular-papilar. Entre los de origen digestivo, los tumores tienen origen en el intestino delgado y grueso, adenocarcinoma de estómago, conducto biliar, esófago, vesícula biliar, y páncreas. Los tumores de origen en el esófago pueden ser de cualquiera de los tipos de esófago de Barrett o adenocarcinoma. De los tumores con origen en el sistema reproductor, pueden tener su origen en el tipo de adenocarcinoma de cuello uterino, tumor endometrial, u ovario. Los tumores con origen en el ovarios son de los tipos transparente, seroso, mucinoso, y endometriode.

Si el tumor no tiene origen epitelial, entonces tiene origen en la glándula adrenal, cerebro, GIST (tumor del estroma gastrointestinal), linfoma, meningioma, mesotelioma, sarcoma, melanoma de piel, o medular de tiroides. Entre los linfomas, son de tipo de linfocitos B, de Hodgkin, o linfocitos T. Entre los sarcomas, son de los tipos leiomioma, osteosarcoma, sarcoma de tejido blando, MFH de tejido blando (histiocitoma fibroso maligno), sarcoma sinovial de tejido blando, sarcoma de Ewing de tejido blando, fibrosarcoma de tejido blando, y rhabdomyosarcoma de tejido blando.

Descripción detallada de los modos para poner en práctica la invención

La presente invención proporciona métodos para el uso de información de la expresión genética para clasificar tumores de una manera más objetiva de lo posible con técnicas de patología convencionales. Por lo tanto, en un primer aspecto, la invención proporciona un método para clasificar una muestra que contiene células como que contiene células tumorales de un tipo de tejido, comprendiendo dicho método:

determinar los niveles de expresión de 5 a 49 secuencias transcritas de células en una muestra que contiene células obtenidas de un sujeto humano, y

comparar los niveles de expresión con los niveles de expresión de las mismas secuencias transcritas en una pluralidad de tipos de tejido de tumor conocidos, y

clasificar la muestra como que contiene células tumorales de un tipo de tejido de la pluralidad de tipos de tejido tumoral conocidos,

en el que las cinco a 49 secuencias transcritas no se seleccionan basándose en sus valores de correlación, o una clasificación basada en los valores de correlación, con los tipos de tejido tumoral.

Como se usa en el presente documento, "una pluralidad" se refiere al estado de dos o más.

La clasificación se basa en una comparación de los niveles de expresión de las 5 a 49 secuencias transcritas en las células de la muestra para sus niveles de expresión en muestras tumorales conocidas y/o muestras conocidas que no son tumorales. Como alternativa, la clasificación se basa en una comparación de los niveles de expresión de las 5 a 49 secuencias transcritas para la expresión de secuencias de referencia en las mismas muestras, con respecto a, o basándose en, la misma comparación en muestras tumorales conocidas y/o muestras no tumorales conocidas.

Por lo tanto, como un ejemplo no limitante, los niveles de expresión de las secuencias genéticas se puede determinar en un conjunto de muestras tumorales conocidas para proporcionar una base de datos frente a las que se comparan los niveles de expresión detectados o determinados en una muestra que contiene células de un sujeto. El nivel o niveles de expresión de la secuencia o secuencias genéticas en una muestra también se comparan con el

- 5 nivel o niveles de expresión de dicha secuencia o secuencias en células normales o no cancerígenas, preferentemente a partir de la misma muestra un sujeto. Como se describe a continuación y en realizaciones de la invención usando Q-PCR o Q-PCR en tiempo real, se pueden usar los niveles de expresión se pueden comparar con los niveles de expresión de genes de referencia en la misma muestra o una proporción de los niveles de expresión.
- 10 En la práctica, el método usa una proporción, de secuencias transcritas con respecto al número de tipos de tumor clasificados, que varía de de aproximadamente 1:2 a aproximadamente 5:2 o más elevada. Indicado de forma diferente, la proporción del número de niveles de expresión necesaria con respecto al número de tipos de tumor que se pueden clasificar basándose en esos niveles, varía de aproximadamente 1:2 a aproximadamente 1:1 a aproximadamente 3:2 a aproximadamente 2:1 a aproximadamente 5:2 o más elevada. Esto se refleja mediante la
- 15 capacidad para usar tanto como aproximadamente 20 niveles de expresión para clasificar entre 39 tipos de tumor (véase la Figura 6). Por lo tanto, y basándose en los datos como se muestra en las Figuras 1-9, la invención se puede poner en práctica con 5 a 49 secuencias genéticas dentro de la proporción de genes evaluados con respecto a tumores clasificados.
- 20 La selección de 5 a 49 secuencias genéticas a usar puede ser aleatoria, o mediante selección basándose en diversos criterios. Como un ejemplo no limitante, las secuencias genéticas se pueden seleccionar basándose en el aprendizaje no supervisado, incluyendo técnicas de agrupamiento. Como otro ejemplo no limitante, la selección se puede realizar para reducir o eliminar la redundancia con respecto a su capacidad para clasificar el tipo de tumor. Por ejemplo, algunas secuencias genéticas se seleccionan basándose en la falta de correlación entre su expresión y
- 25 la expresión de una u otras secuencias genéticas más usadas para la clasificación. Esto se consigue mediante la evaluación del nivel de expresión de cada secuencia genética en el conjunto de datos de expresión la correlación, a través de la pluralidad de muestras, con el nivel de expresión de cada gen en el conjunto de datos para producir una matriz de correlación de coeficientes de correlación. Estas determinaciones de la correlación se pueden realizar directamente, entre la expresión de cada par de secuencias genéticas, o indirectamente, sin comparación directa
- 30 entre los valores de expresión de cada par de secuencias genéticas.

Se puede usar una diversidad de metodologías de correlación en la correlación de datos de expresión de secuencias genéticas individuales dentro del conjunto de datos. Algunos ejemplos no limitantes incluyen métodos paramétricos y no paramétricos así como metodologías basadas en información mutua y enfoques no lineales. Los ejemplos no

35 limitantes de enfoques para métodos incluyen correlación de Pearson (o r de Pearson, también denominada correlación lineal o de producto-momento) y correlación de coseno. Algunos ejemplos no limitantes de métodos no paramétricos incluyen correlación de R de Spearman (orden de clasificación), correlación de Tau de Kendall, y la estadística Gamma. Cada metodología de correlación se puede usar para determinar el nivel de correlación entre las expresiones de secuencias genéticas individuales en el conjunto de datos. La correlación de todas las secuencias

40 con todas las demás secuencias es lo que se prevé considerar más fácilmente como una matriz. Usando correlación de Pearson como un ejemplo la limitante, el coeficiente de correlación r en el método se usa como el indicador del nivel de correlación. Cuando se usan otros métodos de correlación, se puede usar el coeficiente de correlación análogo a r , junto con el reconocimiento de niveles de correlación equivalentes que corresponden a r que son de o de aproximadamente 0,25 o que son de o aproximadamente 0,5.

- 45 El coeficiente de correlación se puede seleccionar como se desee para reducir el número de secuencias genéticas correlacionadas con diversos números. En algunas realizaciones de la invención que usan r , el valor del coeficiente seleccionado puede ser de aproximadamente 0,25 o superior, aproximadamente 0,3 o superior, aproximadamente 0,35 o superior, aproximadamente 0,4 o superior, aproximadamente 0,45 o superior, o aproximadamente 0,5 o superior. La selección de un valor del coeficiente se refiere a que cuando la expresión entre secuencias genéticas en el conjunto de datos se correlaciona con ese valor o uno más elevado, éstas posiblemente no están incluidas en un subconjunto de la invención. Por lo tanto, en algunas realizaciones, el método comprende la exclusión o eliminación (que no se usa para la clasificación) de una o más secuencias genéticas que se expresan en correlación, por encima de un coeficiente de correlación deseado, con otra secuencia genética en el conjunto de datos de tipos de tumor. Sin embargo, se señala que puede haber situaciones de secuencias genéticas que no se correlacionan con ninguna otra secuencia genética, en cuyo caso no se eliminan necesariamente de su uso en la clasificación.

Por lo tanto, en la práctica de la invención se pueden usar los niveles de expresión de las secuencias genéticas, en los que más de aproximadamente un 10 %, más de aproximadamente un 20 %, más de aproximadamente un 30 %, más de aproximadamente un 40 %, más de aproximadamente un 50 %, más de aproximadamente un 60 %, más de aproximadamente un 70 %, más de aproximadamente un 80 %, o más de aproximadamente un 90 % de los niveles no se correlaciona con el de otra de las secuencias genéticas usadas. La correlación entre niveles de expresión se puede basar en un valor inferior a aproximadamente 0,9, aproximadamente 0,8, aproximadamente 0,7, aproximadamente 0,6, aproximadamente 0,5, aproximadamente 0,4, aproximadamente 0,3, o aproximadamente 0,2.

65 La capacidad para clasificar entre clases con exclusión de los niveles de expresión de algunas secuencias genéticas está presente porque la expresión de las secuencias genéticas en el subconjunto se correlaciona con la expresión

de las secuencias genéticas excluidas del subconjunto. De modo que no se perdió información alguna porque la información basada en la expresión de las secuencias genéticas excluidas todavía se representa mediante secuencias retenidas en el subconjunto. Por lo tanto, la expresión de las secuencias genéticas del subconjunto tiene un contenido de información relevante con las propiedades y/o características (o fenotipo) de una célula. Esto tiene aplicación e importancia para la clasificación de clases de tipos de tumores adicionales no incluidos como parte del conjunto de datos de expresión genética original que se puede clasificar mediante el uso de un subconjunto de la invención porque se basa en la redundancia de la información entre la expresión de secuencias en el subconjunto y secuencias expresadas en esas clases adicionales. Por lo tanto, la invención se puede usar para clasificar células como que son de un tipo de tumor mas allá de la pluralidad de clases conocidas usadas para generar el conjunto de datos de expresión genética original.

También se puede usar la selección de secuencias genéticas basándose en la reducción de la correlación de la expresión con respecto a un tipo de tumor en particular. Esto también refleja un descubrimiento de la presente invención, basándose en la observación de que los niveles de expresión que eran los que estaban más correlacionados con uno o más tipos de tumor no era necesariamente de un valor mayor en la clasificación entre diferentes tipos de tumor. Esto se refleja tanto en la capacidad para usar secuencias genéticas seleccionadas de forma aleatoria para la clasificación así como el uso de secuencias en particular, como se describe en el presente documento, que no se expresa con la correlación más significativa con uno o más tipos de tumor. Por lo tanto, la invención se puede poner en prácticas sin selección de secuencias genéticas basándose en los valores de P más significativos o una clasificación basándose en la correlación de la expresión genética y uno o más tipos de tumor. Por lo tanto, la invención se puede poner en práctica sin el uso de metodologías basadas en clasificación, tales como el ensayo de H de Kruskal-Walles.

Las secuencias genéticas usadas en la práctica de la invención pueden incluir las que se ha observado que se expresan en correlación con tipos de tumor en particular, tal como la expresión del receptor de estrógenos, que se ha observado que se expresa en correlación con algunos cánceres de mama y ovario. En algunas realizaciones de la invención, sin embargo, la invención se pone en práctica con el uso del nivel de expresión de al menos una secuencia genética que anteriormente no se ha identificado como que está asociada con ninguno de los tipos de tumor que se están clasificando. Por lo tanto, la invención se puede poner en práctica sin ninguna de las secuencias genéticas que anteriormente se han asociado poco relacionado con la expresión en los 2 o más (hasta 39 o más) tipos de tumor con los que se puede clasificar una muestra que contiene células.

Aunque la invención se describe principalmente con respecto a sujetos humanos, también se pueden usar muestras de otros sujetos. Todo lo que es necesario es la capacidad para evaluar los niveles de expresión de secuencias genéticas en una pluralidad de muestras de tumor conocidas de modo que se pueden comparar los niveles de expresión en una muestra desconocida o de ensayo. Por lo tanto, la invención se puede aplicar a muestras de cualquier organismo para los que están disponibles una pluralidad de secuencias expresadas, y una pluralidad de muestras de tumor conocidas. Un ejemplo la limitante es la aplicación de la invención a muestras de ratón, basándose en la disponibilidad del genoma de ratón para permitir la detección de secuencias de murino expresadas y la disponibilidad de muestras de tumor de ratón conocidas para obtener muestras conocidas. Por lo tanto, la invención se contempla para uso con otras muestras, que incluyen las de mamíferos, primates, y animales usados en ensayos clínicos (tales como ratas, ratones, conejos, perros, gatos, y chimpancés) como ejemplos no limitantes.

Aunque la invención se pone en práctica fácilmente con el uso de muestras que contienen células, en la práctica de la invención se puede usar cualquier muestra que contiene ácido nucleico que se puede someter al ensayo para niveles de expresión genética. Sin limitar la invención como una muestra de la invención puede ser una de la que se sospecha o se sabe que contiene células tumorales. Como alternativa, una muestra de la invención puede ser una "muestra de tumor" o "muestra que contiene tumor" o "muestra que contiene células tumorales" de tejido o fluido aislado de un individuo del que se sospecha que está afectado con, o en riesgo de desarrollar, cáncer. Los ejemplos no limitantes de muestras para uso con la invención incluyen una muestra clínica, tal como, pero no se limita a, una muestra fija, una muestra recién preparada, o una muestra congelada. La muestra puede ser un aspirado, una muestra citológica (incluyendo sangre u otro fluido corporal), o una muestra de ensayo de tejido, que incluye al menos alguna información con respecto al contenido *in situ* de las células en la muestra de ensayo, siempre y cuando las células o ácidos nucleicos apropiados estén disponibles para determinación de los niveles de expresión genética. La invención se basa en parte en el descubrimiento de que los resultados obtenidos con secciones de tejido congelado se pueden aplicar de forma válida a la situación con muestras de tejido o células fijos y se puede extender a muestras recién preparadas.

Algunos ejemplos no limitantes de muestras fijas incluyen las que se fijan con formalina o formaldehído (incluyendo muestras de FFPE), con Boudin, glutaldehído, acetona, alcoholes, o cualquier otro agente fijador, tal como los que se usan para fijar muestras de células o tejidos para inmunohistoquímica (IHC). Otros ejemplos incluyen agentes fijadores que precipitan ácidos nucleicos y proteínas asociados con células. Dadas las posibles complicaciones de la manipulación de muestras de ensayo de tejido congeladas, tal como la necesidad de mantener su estado congelado, la invención se puede poner en práctica con muestras no congeladas, tales como muestras fijas, muestras recién preparadas, que incluyen células de sangre o de otro fluido o tejido corporal, y muestras tratadas mínimamente. En algunas aplicaciones de la invención, la muestra no se ha clasificado usando técnicas convencionales de patología,

tales como, pero no se limitan a, ensayos basados en inmunohistoquímica.

En algunas realizaciones de la invención, la muestra se clasifica como que contiene una célula tumoral de un tipo seleccionado entre los siguientes 53, y subconjuntos de los mismos: Adenocarcinoma de Mama, Adenocarcinoma de
 5 cuello uterino, Adenocarcinoma de Esófago, Adenocarcinoma de Vesícula Biliar, Adenocarcinoma de Pulmón, Adenocarcinoma de Páncreas, Adenocarcinoma de Intestino Delgado-Grueso, Adenocarcinoma de Estómago, Astrocitoma, Carcinoma de Células Basales de la Piel, Colangiocarcinoma de Hígado, Adenocarcinoma de Células
 10 Transparentes de Ovario, Linfoma Linfocitos B Grandes y Difusos, Carcinoma Embrionario de Testículos, Carcinoma Endometriode de Útero, Sarcoma de Ewings, Carcinoma Folicular de Tiroides, Tumor del Estroma Gastrointestinal, Tumor de Células Germinales de Ovario, Tumor de Células Germinales de Testículo, Glioblastoma Multiforme, Carcinoma Hepatocelular de Hígado, Linfoma de Hodgkin, Carcinoma de Pulmón, Leiomioma, Liposarcoma, Carcinoma Lobular de Mama, Histiocitoma Fibroso Maligno, Carcinoma Medular de Tiroides, Melanoma, Meningioma, Mesotelioma de Pulmón, Adenocarcinoma Mucinoso de Ovario, Miodisplasia, Tumor de Intestino
 15 Neuroendocrino, Oligodendroglioma, Osteosarcoma, Carcinoma Papilar de Tiroides, Feocromocitoma, Carcinoma de Células Renales de Riñón, Rhabdomyosarcoma, Seminoma de Testículo, Adenocarcinoma Seroso de Ovario, Carcinoma Microcítico de Pulmón, Carcinoma de Células Escamosas de Cuello Uterino, Carcinoma de Células Escamosas de Esófago, Carcinoma de Células Escamosas de Laringe, Carcinoma de Células Escamosas de Pulmón, Carcinoma de Células Escamosas de Piel, Sarcoma Sinovial, Linfoma de Linfocitos T, y Carcinoma de Células Transicionales de Vejiga.

En otras realizaciones de la invención, la muestra se clasifica como que contiene una célula tumoral de un tipo seleccionado entre los siguientes 34, y subconjuntos de los mismos: adrenal, cerebro, mama, carcinoide-intestinal, cuello uterino (células escamosas), colangiocarcinoma, endometrio, células germinales, GIST (tumor del estroma gastrointestinal), riñón, leiomioma, hígado, pulmón (adenocarcinoma, macrocítico), pulmón (microcítico),
 25 pulmón (escamoso), linfoma (linfocitos B), Linfoma (Hodgkins), meningioma, mesotelioma, osteosarcoma, ovario (células transparentes), ovario (células serosas), páncreas, próstata, piel (células basales), piel (melanoma), intestino grueso y delgado; tejido blando (liposarcoma); tejido blando (MFH o Histiocitoma Fibroso Maligno), tejido blando (Sarcoma sinovial), testículo (seminoma), tiroides (folicular-papilar), tiroides (carcinoma medular), y vejiga urinaria.

En otras realizaciones de la invención, la muestra se clasifica como que contiene una célula tumoral de un tipo seleccionado entre los siguientes 39, y subconjuntos de los mismos: glándula adrenal, cerebro, mama, carcinoide-intestinal, adenocarcinoma de cuello uterino, escamoso de cuello uterino, endometrio, vesícula biliar, ovario de células germinales, GIST, riñón, leiomioma, hígado, adenocarcinoma macrocítico de pulmón, pulmón
 35 microcítico, escamoso de pulmón, linfoma de linfocitos B, linfoma de Hodgkin, linfoma de linfocitos T, meningioma, mesotelioma, osteosarcoma, células transparentes de ovario, células serosas de ovario, páncreas, próstata, células basales de piel, melanoma de piel, escamoso de piel, intestino grueso y delgado, liposarcoma de tejido blando, MFH de tejido blando, sarcoma sinovial de tejido blando, adenocarcinoma de estómago, distintos de testículo (o no seminoma), seminoma de testículo, folicular y papilar de tiroides (folicular-papilar tiroides-folicular-papilar, medular de tiroides, y vejiga urinaria).

Los métodos de la invención también se pueden aplicar para clasificar una muestra como que contiene células como que contiene una célula tumoral de un tumor de un subconjunto de cualquiera de los conjuntos mencionados anteriormente. El tamaño del subconjunto normalmente será pequeño, formado por dos, tres, cuatro, cinco, seis,
 45 siete, ocho, nueve, o diez de los tipos de tumor que se han descrito anteriormente. Como alternativa, el tamaño del subconjunto puede ser cualquier número entero hasta el tamaño total del conjunto. Por lo tanto, algunas realizaciones de la invención incluyen una clasificación entre 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, o 52 de los tipos mencionados anteriormente. En algunas realizaciones, el subconjunto estará formado por tipos de tumor que son del
 50 mismo tipo de tejido u órgano. Como alternativa, el subconjunto estará formado por tipos de tumor de diferentes tejidos u órganos. En algunas realizaciones, el subconjunto incluirá uno o más tipos seleccionados entre glándulas adrenales, cerebro, carcinoide-intestinal, adenocarcinoma de cuello uterino, escamoso de cuello uterino, vesícula biliar, ovario de células germinales, GIST, leiomioma, hígado, meningioma, osteosarcoma, células basales de piel, escamoso de piel, liposarcoma de tejido blando, MFH de tejido blando, sarcoma sinovial de tejido blando, distintos de testículo (o no seminoma), seminoma de testículo, folicular y papilar de tiroides, y medular de tiroides.

La clasificación entre subconjuntos de los tipos de tumor mencionados anteriormente se demuestra con los resultados que se muestran en las Figuras 1-9, en las que los niveles de expresión de unas pocas como
 60 aproximadamente 5 o más secuencias genéticas se pueden usar para clasificar entre muestras aleatorias de 2 tipos de tumor entre aquéllas en el conjunto de los 39 enumerados anteriormente. Se pueden usar algunos niveles de expresión de tan pocos como de aproximadamente 20 a 49 para clasificar entre los 39 tipos de tumor con grados variables de precisión. La invención se puede poner en práctica con los niveles de expresión de aproximadamente 10 o más, aproximadamente 15 o más, aproximadamente 20 o más, aproximadamente 25 o más, aproximadamente 30 o más, aproximadamente 35 o más, aproximadamente 40 o más, o aproximadamente 45 o más para 49
 65 secuencias transcritas tal como se encuentran en el "transcriptoma" humano (parte transcrita del genoma). La invención también se puede poner en práctica con niveles de expresión de aproximadamente 10-20 o más,

aproximadamente 20-30 o más, aproximadamente 30-40 o más, aproximadamente 40-50 o más, o 49 secuencias transcritas. En algunas realizaciones de la invención, los genes transcritos se pueden escoger de forma aleatoria o pueden incluir todas o algunas de las secuencias genéticas específicas que se desvelan en el presente documento. Como se demuestra en el presente documento, se puede realizar una clasificación con precisiones de

aproximadamente un 55 %, aproximadamente un 60 %, aproximadamente un 65 %, aproximadamente un 70 %, aproximadamente un 75 %, aproximadamente un 80 %, aproximadamente un 85 %, aproximadamente un 90 %, o aproximadamente un 95 % o superior mediante el uso de la presente invención.

En otras realizaciones, los niveles de expresión genética de otras secuencias genéticas se pueden determinar junto con las determinaciones de los niveles de expresión que se han descrito anteriormente para uso en la clasificación. Un ejemplo no limitante de esto se observa en el caso de una plataforma basada en micromatriz para determinar la expresión genética, en la que también se mide la expresión de otras secuencias genéticas. Cuando esos otros niveles de expresión no se usan en la clasificación, se pueden considerar los resultados de "exceso" de secuencias transcritas y no es crítico para la práctica la invención. Como alternativa, y cuando esos otros niveles de expresión se usan en la clasificación, éstos se encuentran dentro del alcance de la invención, en la que la descripción del uso de números de secuencias en particular no excluye necesariamente el uso de niveles de expresión de secuencias adicionales. En algunas realizaciones, la invención incluye el uso de un nivel o niveles de expresión de una o más secuencias genéticas en "exceso" secuencias genéticas, tal como las que pueden proporcionar información redundante con respecto a una o más de otras secuencias genéticas usadas en un método de la invención.

Dado que la clasificación de la muestra cómo que contiene células de uno de los tipos de tumor mencionados anteriormente de forma inherente también clasifica el origen del sitio del tejido u órgano de la muestra, los métodos de la invención se pueden aplicar a la clasificación de la muestra de tumor como que es de un sitio de tejido u órgano en particular del paciente. Esta solicitud de la invención es particularmente útil en casos en los que la muestra es de un tumor que es el resultado de metástasis por otro tumor. En algunas realizaciones de la invención, la muestra de cómo se clasifica como es uno de los siguientes 24: Adrenal, Vejiga, Hueso, Cerebro, Mama, Cuello Uterino, Endometrio, Esófago, Vesícula Biliar, Riñón, Laringe, Hígado, Pulmón, Ganglio Linfático, Ovario, Páncreas, Próstata, piel, Tejido Blando, Intestino Delgado/Grueso, Estómago, Testículo, Tiroides, y Útero.

Aunque la invención también proporciona clasificación como uno de los tipos de tumor mencionados anteriormente basándose en comparaciones con los niveles de expresión de secuencias en los 39 tipos de tumor, es posible que se desee un nivel de confianza más elevado en la clasificación. Si se prefiere un aumento en la confianza de la clasificación, la clasificación se puede ajustar para identificar la muestra de tumor como que tiene un origen en particular o tipo de célula como se muestra en la Figura 10. Por lo tanto, se puede realizar un aumento de la confianza como intercambio para una disminución en la especificidad como para un tipo de tumor mediante la identificación del origen o del tipo de célula.

La clasificación de una muestra que contiene células como que tiene una célula tumoral de uno de los 39 tipos de tumor mencionados anteriormente de forma inherente también clasifica el origen del sitio del tejido u órgano de la muestra. Por ejemplo, la identificación de la muestra como que es de tipo de célula escamosa de tumor escamoso de cuello uterino clasifica necesariamente el tumor como que tiene su origen en el cuello uterino, (y por lo tanto con un origen epitelial en lugar de no epitelial) como se muestra en la Figura 10. También se refiere a que el tumor no tenía necesariamente un origen en células germinales. Por lo tanto, los métodos de la invención se pueden aplicar para la clasificación de la muestra de tumor como que son de un sitio de tejido u órgano en particular de un sujeto o paciente. Esta aplicación de la invención es particularmente útil en casos en los que la muestra es de un tumor que es el resultado de metástasis por otro tumor.

La práctica de la invención para clasificar una muestra que contiene células como que tiene una célula tumoral de uno de los tipos mencionados anteriormente se realiza mediante el uso de un algoritmo de clasificación apropiado que usa aprendizaje supervisado para aceptar 1) los niveles de expresión de las secuencias genéticas en una pluralidad de tipos de tumor conocidos como un conjunto de entrenamiento y 2) los niveles de expresión de los mismos genes en una o más células de una muestra para clasificar la muestra como que tiene células de uno de los tipos de tumor. Un análisis adicional de esto se proporciona en la sección de Ejemplos en el presente documento. Los niveles de expresión se pueden proporcionar basándose en las señales en cualquier formato, incluyendo expresión de ácidos nucleicos o expresión de proteínas como se describe en el presente documento.

Como sería evidente para el experto en la materia, el intervalo de clasificación se ve influido por el número de tipos de tumor así como el número de muestras para cada tipo de tumor. Pero dada una serie de muestras adecuadas del intervalo total de tumores humanos como se proporciona el presente documento, la invención se aplica fácilmente para la clasificación de esos tipos de tumor así como tipos adicionales.

Los ejemplos no limitantes de algoritmos de clasificación que se pueden usar en la práctica de la invención incluyen algoritmos de aprendizaje supervisado, algoritmos de aprendizaje con máquina, análisis de discriminación lineal, algoritmos de selección de atributos, y redes neurales artificiales (ANN). En realizaciones preferentes de la invención, se usa un algoritmo de clasificación basado en la distancia, tal como el algoritmo vecinal más cercano a k (KNN), o máquina de vector de soporte (SVM).

El uso de KNN se realiza en algunas realizaciones de la invención y se analiza adicionalmente como un ejemplo no representativo. El KNN se puede usar para analizar los datos de expresión de los genes en un "conjunto de entrenamiento" de muestras de tumor conocidas se incluyen los 39 de los tipos de tumor que se describen en el presente documento. El conjunto de datos de entrenamiento se puede comparar a continuación con los datos de expresión para los mismos genes en una muestra que contiene células. Los niveles de expresión de los genes en la muestra se comparan a continuación con el conjunto de datos de entrenamiento a través de KNN para identificar las muestras tumorales con los patrones de expresión más similares. Como un ejemplo no limitante, los cinco "vecinos más cercanos" se pueden identificar y los tipos de tumor de los mismos usados para clasificar la muestra de tumor desconocido. Por supuesto que se pueden usar otros números de "vecinos más cercanos". Algunos ejemplos no limitantes incluyen menos de 5, aproximadamente 7, aproximadamente 9, o aproximadamente 11 o más "vecinos más cercanos".

Como un ejemplo hipotético, si los cinco "vecinos más cercanos" de una muestra desconocida son cuatro linfomas de linfocitos B y un linfoma de linfocitos T, entonces la clasificación de la muestra como que es de un linfoma de linfocitos B se puede realizar con gran precisión. Esto se ha usado con una precisión de un 84 % una precisión más elevada, tal como de un 90 %, tal como se describe en los Ejemplos.

La capacidad de clasificación se puede combinar con la naturaleza inherente del esquema de clasificación para proporcionar un medio para aumentar la confianza de la clasificación tumoral en ciertas situaciones. Por ejemplo, si los cinco "vecinos más cercanos" de una muestra son tres tumores de células transparentes de ovario y dos tumores serosos de ovario, la confianza se puede aumentar simplemente tratando los tumores como si fueran de origen de ovario y se trata al sujeto o paciente (a partir de que se obtuvo la muestra) en consecuencia. Véase la Figura 10. Se trata de un ejemplo de compensación de la especificidad a favor del aumento de la confianza. Esto proporciona el beneficio añadido de abordar la posibilidad de que la muestra desconocida era de un tumor mucinoso o endometriode. Por supuesto, el experto con experiencia es libre de tratar el tumor como uno o ambos de estas dos posibilidades más probables y proceder de acuerdo con esa determinación.

Dado que el linaje del desarrollo de células tumorales en ciertos tipos de tumor (por ejemplo, células germinales) puede ser complejo y puede implicar múltiples tipos celulares, puede parecer que la Figura 10 está sobreesimplificada. Sin embargo, sirve como una buena base para relación a una histopatología conocida y sirve como un "árbol de guía" para analizar y relación a firmas de expresión genéticas asociadas a tumores.

La naturaleza inherente del esquema de clasificación también proporciona un medio para aumentar la confianza de la clasificación tumoral en casos en los que los "vecinos más cercanos" son ambiguos. Por ejemplo, si los cinco "vecinos más cercanos" eran uno de vejiga urinaria, uno de mama, uno de riñón, uno de hígado, y uno de próstata, la clasificación simplemente puede ser la de un tumor de células escamosas. Tal determinación se puede realizar con una confianza significativa y el sujeto o paciente a partir de que se obtuvo la muestra se puede tratar en consecuencia. Sin quedar ligado por la teoría, y ofrecido solamente para mejorar la comprensión de la invención, los últimos dos ejemplos reflejan las similitudes en la expresión genética de células de un tipo celular y/u origen de tejidos similares.

Realizaciones de la invención incluyen el uso de los métodos y materiales que se describen en el presente documento para identificar el origen de un cáncer de un paciente. Por lo tanto, dada una muestra que contiene células tumorales, el origen del tejido de las células tumorales se identifica mediante el uso de la presente invención. Un ejemplo no limitante se encuentra en el caso de un objeto con un ganglio linfático inflamado que contiene células cancerígenas. Las células pueden ser de un tejido órgano que drena en el ganglio linfático o puede provenir de otra fuente de tejido. La presente invención se puede usar para clasificar las células como que son de un tipo de tumor o tejido (u origen) en particular que permite la identificación de la fuente de las células cancerígenas. En un ejemplo no limitante alternativo, la muestra (tal como la que un ganglio linfático) contiene células, que primero se someten a ensayo mediante el uso de la invención para clasificar al menos una célula como que es una célula tumoral de un tipo u origen de tejido. A continuación esto se usa para identificar la fuente de las células cancerígenas en la muestra. Ambos de estos son ejemplos del uso ventajoso de la invención para ahorrar tiempo, esfuerzo y costes en el uso de otros ensayos de diagnóstico de cáncer.

En otras realizaciones, la invención se pone en práctica con una muestra de un sujeto con una historia anterior de cáncer. Como un ejemplo no limitante, se puede encontrar una muestra que contiene células (del ganglio linfático o de cualquier parte) del sujeto que contiene células cancerígenas de modo que la presente invención se puede usar para determinar si las células son del mismo tejido o de uno diferente del cáncer anterior. Esta aplicación de la invención también se puede usar para identificar un nuevo tumor primario, tal como el caso en el que se encuentran nuevas células cancerígenas en el hígado de un sujeto que anteriormente presentaban cáncer de mama. La invención se puede usar para identificar las nuevas células cancerígenas como que son el resultado de metástasis del cáncer de mama anterior (o de otro tipo de tumor, ya se haya identificado anteriormente o no) o como una nueva aparición primaria de cáncer de hígado. La invención también se aplicará muestras de un tejido u órgano en las que se encuentran múltiples cánceres para determinar el origen de cada cáncer, así como si los cánceres tienen el mismo origen.

- Aunque la invención se puede poner en práctica con el uso de niveles de expresión de un grupo aleatorio de secuencias genéticas expresadas, la invención también proporciona secuencias genéticas a modo de ejemplo para uso en la práctica de la invención. La invención incluye un primer grupo de 74 secuencias genéticas de las que aproximadamente de 5 a 49 se pueden usar en la práctica de la invención. Las 5 a 49 secuencias genéticas se pueden usar junto con la determinación de niveles de expresión de secuencias adicionales siempre y cuando los niveles de expresión de las secuencias genéticas del conjunto de 74 se usen en la clasificación. Un ejemplo limitante de tales realizaciones de la invención es en el que la expresión de aproximadamente 5 a 49 de las 74 secuencias genéticas se mide junto con los niveles de expresión de una pluralidad de otras secuencias, tal como mediante el uso de una plataforma basada en micromatriz usada para realizar la invención. Cuando estos otros niveles de expresión no se usan en la clase, se pueden considerar los resultados de un "exceso" de secuencias transcritas y no es crítico para la práctica de la invención. Como alternativa, y cuando esos otros niveles de expresión se usan en la clasificación, éstos están dentro del alcance de la invención, en la que el uso de las secuencias que se han descrito anteriormente no excluye necesariamente el uso de niveles de expresión de secuencias adicionales.
- 15 Las secuencias de ARNm que corresponden a un conjunto de 74 secuencias genéticas para uso en la práctica de la invención se proporcionan en el Apéndice 1 adjunto (Listado de Secuencias) junto con información de identificación adicional. El listado de la información de identificación, que incluye números de referencia y otra información, se proporciona mediante lo siguiente.
- 20 >Hs.73995_ARNm_1 gi|190403|gb|M60502.1|PERFILHUM ARNm de profilagrina humana, extremo en la posición 3' poliA = 1
>Hs.75236_ARNm_4 gi|14280328|gb|AY033998.1| Homo sapiens poliA = 3
>Hs.299867_ARNm_1 gi|4758533|ref|NM_004496.1| factor nuclear 3 de hepatocitos de Homo sapiens, alfa (HNF3A),
25 ARNm poliA = 3
>Hs.285401_contig1
AI147926|AI880620|AA768316|AA761543|AA279147|AI216016|AI738663|N79248|AI
684489|AA960845|AI718599|AI379138|N29366|BF002507|AW044269|R34339|R66326
|H04648|R67467|AI523112|BF941500 poliA = 2 poliA = 3
30 >Hs.182507_ARNm_1 gi|15431324|ref|NM_002283.2| queratina de Homo sapiens, pelo, básico, 5 (KRTHB5),
ARNm
poliA = 3
>Hs.292653_contig1
AI200660|AW014007|AI341199|AI692279|AI393765|AI378686|AI695373|AW292108|T10352|R44346|AW47040
8|AX380925|BF93 8983|AW003704|H08077|F03 856|H08075|F
35 08895|AW468398|AI865976|H22568|AI858374|AI216499 poliA = 2 poliA = 3
>Hs.97616_ARNm_3 gi|12654852|gb|BC001270.1|BC001270 clon de Homo sapiens MGC:5069 IMAGE:3458016
poliA = 3
>Hs.123078_ARNm_3 gi|14328043|gb|BC009237.1|BC009237 clon de Homo sapiens MGC:2216
40 IMAGE:2989823
poliA = 3
>Hs.285508_contig1 AW194680|BF939744|BF516467 poliA = 1 poliA = 1
>Hs.183274_contig1
BF437393|BF064008|BF509951|AW134603|AI277015|AI803254|AA887915|BF054958|
45 AI004413|AI393911|AI278517|AW612644|AI492162|AI309226|AI863671|AA448864|
AI640165|AA479926|AA461188|AA780161|IBF591180|AI918020|AI758226|AI291375|
BF001845|BF003064|AI337393|AI522206|BE856784|BF001760|AI280300 FLAG=1 poliA = 2 WARN poliA = 3
>Hs.334841_ARNm_3 gi|14290606|gb|BC009084.1|BC009084 clon de Homo sapiens MGC:9270
50 IMAGE:3853674
poliA = 3
>Hs.3321_contig1
AI804745|AI492375|AA594799|BE672611|AA814147|AA722404|AW170088|D11718|BG
153444|AI680648|AA063561|BE219054|AI590287|R55185|AI479167|AI796872|AI01
8324|AI701122|BE218203|AA905336|AI681917|BI084742|A2480008|AI217994|AI40 1468 poliA = 2 poliA = 3
55 >Hs.306216_single1 AW083022 poliA = 1 poliA = 2
>Hs.99235_contig1 AA456140|AI167259|AA450056 poliA = 2 poliA = 3
>Hs.169172_ARNm_2 gi|2274961|emb|AJ000388.1|HSCANPX ARNm de Homo sapiens para proteasa de tipo
calpaína
CANPX poliA = 3
60 >Hs.351486_ARNm_1 gi|16549178|dbj|AK054605.1|AK054605 ADNc de FLJ30043 fis de Homo sapiens, clon
3NB692001548 poliA = 0
>Hs.153504_contig2
BE962007|AW016349|AW016358|AW139144|AA932969|AI025620|AI688744|AI865632|
AA854291|AA932970|AU156702|AI634439|AA152496|AI539557|AI123490|AI613215|
65 AI318363|AW105672|AA843483|AI366889|AW181938|AI813801|AI433695|AA934772|
N72230|AI760632|BE858965|AW058302|AI760087|AI682077|AA886672|AI350384|AW

243848|AW300574|BE466359|AI859529|AI921588|BF062899|BE855597|BE617708 poliA = 2 poliA = 3
 >Hs.199354_single1 AI669760 poliA = 1 poliA = 2
 >Hs.162020_contig1 AW291189|AA505872 poliA = 2 poliA = 3
 >Hs.30743_ARNm_3 gi|18201906|ref|NM_006115.2| antígeno en melanoma expresado preferentemente de
 5 Homo sapiens
 (PRAME), ARNm poliA = 3
 >Hs.271580_contig1
 AI632869|AW338882|AW338875|AW613773|AI982899|AW193151|BE206353|BE208200| AI811548|AW264021
 poliA = 2 poliA = 3
 10 >Hs.69360_ARNm_2 gi|14250609|gb|BC008764.1|BC008764 clon de Homo sapiens MGC:1266 IMAGE:3347571
 poliA = 3
 >Hs.30827_contig1 H07885|N39347|W85913|AA583408|W86449 poliA = 2 poliA = 3
 >Hs.211593_contig2
 BF592799|AI570478|AA234440|R40214|BE501078|AW593784|AI184050|AI284161|W7
 15 2149|AW780437|AI247981|AW241273|H60824 poliA = 2 poliA = 3
 >Hs.155097_ARNm_1 gi|15080385|gb|BC011949.1|BC011949 clon de Homo sapiens MGC:9006
 IMAGE:3863603
 poliA = 3
 >Hs.5163_ARNm_1 gi|15990433|gb|BC015582.1|BC015582 clon de Homo sapiens MGC:23280 IMAGE:4637504
 20 poliA = 3
 >Hs.55150_ARNm_1 gi|17068414|gb|BC017586.1|BC017586 clon de Homo sapiens MGC:26610
 IMAGE:4837506
 poliA = 3
 >Hs.170177_contig3
 25 AI620495|AW291989|AA780896|AA976262|AI298326|BF111862|AW591523|AI922518|
 AI480280|BF589437|AA600354|AI886238|AA035599|H90049|BF112011|N52601|AI57
 0965|AI565367|AW768847|H90073|BE504361|N45292|AI632075|AA679729|AW168052
 |AI978827|AI968410|AI669255|N45300|AI651256|AI698970|AI521256|AW078614|A
 I802070|AI885947|AI342534|AI653624|AW243936|T16586|R15989|AI289789|AI871 636|AI718785|AW148847
 30 poliA = 2 poliA = 3
 >Hs.184601_ARNm_5 gi|4426639|gb|AF104032.1|AF104032 Homo sapiens poliA = 2
 >Hs.351972_single1 AA865917 poliA = 2 poliA = 3
 >Hs.5366_ARNm_2 gi|15277845|gb|BC012926.1|BC012926 clon de Homo sapiens MGC:16817 IMAGE:3853503
 poliA = 3
 >Hs.18140_contig1
 35 AI685931|AA410954|T97707|AA706873|AI911572|AW614616|AA548520|AW027764|BF
 511251|AI914294|AW151688 poliA = 1 poliA = 1
 >Hs.133196_contig2
 BF224381|BE467992|AW137689|AI695045|AW207361|BF445141|AA405473 poliA = 2 WARN poliA = 3
 40 >Hs.63325_ARNm_5 gi|15451939|ref|NM_019894.1| proteasa transmembrana de Homo sapiens, serina 4
 (TMPRSS4), ARNm poliA = 3
 >Hs.250692_ARNm_2 gi|184223|gb|M95585.1|HUMHLF ARNm de factor de leucemia hepática (HLF) humana,
 completo
 cds poliA = 3
 45 >Hs.250726_single4 AW298545 poliA = 2 poliA = 3
 >Hs.79217_ARNm_2 gi|16306657|gb|BC001504.1|BC001504 clon de Homo sapiens MGC:2273 IMAGE:3505512
 poliA = 3
 >Hs.47986_ARNm_1 gi|13279253|gb|BC004331.1|BC004331 clon de Homo sapiens MGC:10940
 IMAGE:3630835
 50 poliA = 3
 >Hs.94367_ARNm_1 gi|10440200|dbj|AK027147.1|AK027147 ADNc de Homo sapiens: FLJ23494 fis, clon
 LNG01885 poliA = 3
 >Hs.49215_contig1
 BI493248|N66529|AA452255|BI492877|AW196683|AI963900|BF478125|AI421654|BE 466675 poliA = 1 poliA = 1
 55 >Hs.281587_contig2
 R61469|R15891|AA007214|R61471|AI014624|N69765|AW592075|H09780|AA709038|A
 I335898|AI559229|F09750|R49594|H11055|T72573|AA935558|AA988654|AA826438| AI002431|AI299721
 poliA = 1 poliA = 2
 >Hs.79378_ARNm_1 gi|16306528|ref|NM_003914.21 ciclina A1 de Homo sapiens (CCNA1), ARNm poliA = 3
 60 >Hs.156469_contig2
 AI341378|AI670817|AI701687|AI335022|AW235883|AI948598|AA446356 poliA = 2 poliA = 3
 >Hs.6631_ARNm_1 gi|7020430|dbj|AK000380.1|AK000380 ADNc de Homo sapiens FLJ20373 fis, clon
 HEP19740
 poliA = 3
 65 >Hs.155977_contig1 AI309080|AI313045 poliA = 1 WARN poliA = 1
 >Hs.95197_ARNm_4 gi|5817138|emb|AL110274.1|HSM800829 ARNm de Homo sapiens; ADNc

DKFZp564I0272
 (del clon DKFZp564I0272) poliA = 3
 >Hs.48956_contig1 N64339|AI569513|AI694073 poliA = 1 poliA = 1
 5 >Hs.118825_ARNm_10 gi|1495484|emb|X96757.1|HSSAPKK3 ARNm de H. sapiens para quinasa de quinasa
 MAP poliA = 3
 >Hs.135118_contig3
 AI683181|AI082848|AW770198|AI333188|AI873435|AW169942|AI806302|AW340718| BF196955|AA909720
 poliA = 1 poliA = 2
 10 >Hs.171857_ARNm_1 gi|13161080|gb|AF332224.1|AF332224 ARNm de proteína de testículo de Homo sapiens,
 cds parcial
 poliA = 3
 >Hs.18910_ARNm_3 gi|12804464|gb|BC001639.1|BC001639 clon de Homo sapiens MGC:1944 IMAGE:2959372
 poliA = 3
 15 >Hs.194774_ARNm_1 gi|16306633|gb|BC001492.1|BC001492 clon de Homo sapiens MGC:1774
 IMAGE:3510004
 poliA = 3
 >Hs.127428_ARNm_2 gi|16306818|gb|BC006537.1|BC006537 clon de Homo sapiens MGC:1934
 IMAGE:2987903
 poliA = 3
 20 >Hs.126852_contig1
 AI802118|BF197404|BF224434|AA931964|AW236083|AI253119|AW614335|AI671372|
 AI793240|AW006851|AI953604|AI640505|AI633982|AW195809|AI493069|AW058576| AW293622 poliA = 2
 poliA = 3
 25 >Hs.28149_ARNm_1 gi|14714936|gb|BC010626.1|BC010626 clon de Homo sapiens MGC:17687
 IMAGE:3865868
 poliA = 3
 >Hs.35453_ARNm_3 gi|7018494|emb|AL157475.1|HSM802461 ARNm de Homo sapiens; ADNc DKFZp761G151
 (del clon DKFZp761G151); cds parcial poliA = 3
 >Hs.180570_contig1
 30 R08175|AA707224|AA699986|R11209|W89099|T98002|AA494546 poliA = 2 poliA = 3
 >Hs.196270_ARNm_1 gi|11545416|gb|AF283645.1|AF283645 cromosoma 8 map 8q21 de Homo sapiens poliA =
 3
 >Hs.9030_ARNm_3 gi|12652600|gb|BC000045.1|BC000045 clon de Homo sapiens MGC:2032 IMAGE:3504527
 poliA = 3
 35 >Hs.1282_ARNm_3 gi|4559405|ref|NM_000065.1| componente 6 del complemento Homo sapiens (C6), ARNm
 poliA = 1
 >Hs.268562_ARNm_2 gi|15341874|gb|BC013117.1|BC013117 clon de Homo sapiens MGC:8711
 IMAGE:3882749
 poliA = 3
 40 >Hs.151301_ARNm_3 gi|16041747|gb|BC015754.1|BC015754 clon de Homo sapiens MGC:23085
 IMAGE:4862492
 poliA = 3
 >Hs.111_contig1 AA946776|AW242338|H24274|AI078616 poliA = 1 poliA = 2
 >Hs.150753_contig1 AI123582|AI288234 poliA = 0 poliA = 0
 45 >Hs.82109_ARNm_1 gi|14250611|gb|BC008765.1|BC008765 clon de Homo sapiens MGC:1622 IMAGE:3347793
 poliA = 3
 >Hs.44276_ARNm_2 gi|12654896|gb|BC001293.1|BC001293 clon de Homo sapiens MGC:5259 IMAGE:3458115
 poliA = 3
 50 >Hs.2142_ARNm_4 gi|13325274|gb|BC004453.1|BC004453 clon de Homo sapiens MGC:4303 IMAGE:2819400
 poliA = 3
 >Hs.180908_contig1 AA846824|AW611680|AA846182|AA846342|AA846360 poliA = 2 poliA = 3
 >Hs.89436_ARNm_1 gi|16507959|ref|NM_004063.2| cadherina 17 de Homo sapiens, cadherina LI (hígado-
 intestino)
 (CDH17), ARNm poliA = 1
 55 >Hs.151544_ARNm_8 gi|3153107|emb|AL023657.1|HSDSHP ADNc de SH2D1A de Homo sapiens, conocido
 anteriormente como
 DSHP poliA = 3
 >Hs.1657_contig4
 AW473119|AA164586|AI540656|AI758480|AI810941|AI978964|AI675862|AI784397|
 60 AW591562|AW514102|AI888116|AI983175|AI634735|AI669577|AI202659|AI910598|
 AI961352|AI565481|AI886254|AI538838|AA291749|AW571455|AI370308|AI274727|
 AW473925|AW514787|AI273871|AW470552|AI524356|AI888281|AW089672|AI952766|
 AW440601|AX654044|AW438839|AI972926 poliA = 2 poliA = 3
 >Hs.35984_ARNm_1 gi|6049161|gb|AF133587.1|AF133587 cromosoma 22 map 22q11.2 de Homo sapiens poliA
 65 = 3
 >Hs.334534_ARNm_2 gi|17389403|gb|BC017742.1|BC017742 Homo sapiens, clon IMAGE:4391536, ARNm

poliA = 3

>Hs.60162_ARNm_1 gi|10437644|dbj|AK025181.1|AK025181 ADNc de Homo sapiens: FLJ21528 fis, clon COL05977 poliA = 3

- 5 Como lo entendería un experto en la materia, la detección de la expresión de cualquiera de las secuencias identificadas anteriormente, o las secuencias proporcionadas en los Apéndices adjuntos (Listado de Secuencias) se puede realizar mediante la detección de la expresión de cualquier parte o fragmento apropiados de estas secuencias. Preferentemente, las partes son lo suficientemente grandes como para contener secuencias únicas con respecto a otras secuencias expresadas en una muestra que contiene células. Además, la persona experta
- 10 reconocería que las secuencias desveladas representan una hebra de una molécula bicatenaria y cualquier hebra se puede detectar como un indicador de la expresión de las secuencias desveladas. Esto continúa porque las secuencias desveladas se expresan como moléculas de ARN en células que se convierten preferentemente en moléculas de ADNc para facilitar la manipulación y la detección. Las moléculas de ADNc resultantes pueden tener las secuencias del ARN expresado así como las de la hebra complementaria a las mismas. Por lo tanto, se puede
- 15 detectar cualquiera de la hebra de la secuencia del ARN o la hebra complementaria. Por supuesto, también es posible detectar el ARN expresado sin conversión en ADNc.

En algunas realizaciones de la invención, los niveles de expresión de las secuencias genéticas se mide mediante la detección de las secuencias expresadas en una muestra que contiene células como que se indican a los siguientes oligonucleótidos, que corresponden a las secuencias mencionadas anteriormente como se indica con los números de referencia que se proporcionan.

>AF133587
CCCGGATCGCCATCAGTGTCTCGAGTTCAAACCCTGAGCCCTTCATTACCTCTGTGAG

25 >BC017742
TGCCCTTGCTCTGTGTCTCATCTCAGTCATTTGACTTAGAAAAGTGCCCTTCAAAGGACCCCT
>BF437393
GGAGGGAGGGCTAATTATATATTTTGTGTTCTCTATACTTTGTTCTGTTGTCTGCGCC

30 >AI620495
CAGTTTGGATTGTATAATAACGCCAAGCCAGTTGTAGTCGTTTGAGTGCAGTAATGAAA
>AK000380
AAATCAGAGTAACCCCTTCTGTATTGAGTGCAGTGTTTTTACTCTTTTCTCATGCACAT

>BC009237
TGCCTGGCACAAGAAGGAAGAATATAAATGATAGTTGACTCGTCTGTGGAAGAACTTA

35 >BC008765
AGTCTTTTGTCTTTTGGCAAACTCTACTTAATCCAATGGGTTTTTCCCTGTACAGTAGAT
>BC001504
GGTTACTGTGGGTGGAATAGTGGAGGCCCTTCAACTGATTAGACAAGGCCCGCCACATCT

40 >NM_019894
TAAAATGCACTGCCCTACTGTTGGTATGACTACCGTTACCTACTGTTGTCATTGTTATTA
>BF224381
TTCTCTTTTGGGGGCAAACACTATGTCCTTTTCTTTTCTAGATACAGTTAATTCCTGGA

>AL157475
AAGACCCACACCCTGTAGCAATACCAAGTGCTATTACATAATCAATGGACGATTTATACT

45 >AY033998
AGTGTGCAAGTTTCCTTTAAACCAACAAAGCCCACAAGTCCTGAATTTCCATTCTTA
>H07885
GTCAGTGTATAGCAGCTGTGATTTACAAGGAAGGGTGCTGCAGGGGGACCTGGTTGAT

>NM_004496

50 TTTTCATCCAGTGTTATGCACTTTCCACAGTTGGTGTTAGTATAGCCAGAGGGTTTCATTA
>AA846824
GGGAAGTAGGGATTATTCGTTAAATTCAATCGCGAGCACCAAGTCGGACTGGCCGGGGA

>BC017586
GGGACCAGGCCCTGGGACAGCCATGTGGCTCCAAATGACTAAATGTCAGCTCAAAAACCA

55 >AA456140
TCCGTTTATGGAGGCAATTCCATATCCTTTCTTGAACGCACATTCAGCTTACCCAGAGA
>NM_002283
AGAGTTAAGCCACTTCTGGGTCTCCTTCTTATGACTGTCTATGGGTGCATTGCCTTCTG

>AL023657

60 GTGGCCTGAGTAATGCATTATGGGTGGTTTACCATTTCTTGAGGTAAAAGCATCACATGA
>B001639
ACACATGCATGTGTCTGTGTATGTGTGAATGTGAGAGAGACACAGCCCTCCTTTCAGAAG

>BC015754
TCTGTAAGTGCACAACCCTGGGGTTTGCTGCAGAGCTATTTCTTCCATGTAAAGTAGTG

65 >AF332224
AAACACTCTTTCCGACTCCAGAGGAGAAGCTGGCAGCTCTCTGTAAGAAATATGCTGATC

>BC001270
 GCTTCCTCTATCGCCCAATGCAAAATCGATGAAATGGGGAGTTCTCTGGGCCAGGCCACA
 >AI147926
 GTAGAAATCCTCTGTTTATAATGAACAAGATGAACCAATGTGGATTAGAAAGAAGTCCGAG
 5 >AW298545
 CTGTTTTAAACTGAATGGCACGAAATTGTTTTCTCAACTCGGAGATTCCTGTATGGAG
 >AI802118
 AATAAATAGTAGCTCTGCTGATGATGACGTTGATAACCAAATGTTCTGTGGTCTTAAGT
 >AI683181
 10 CAAACAGCCCGGTCTTGATGCAGGAGAGTCTGAAAAGGAAGAAAATGGTTTCAGTTTCA
 >M95585
 AACATGGACCATCCAAATTTATGGCCGTATCAAATGGTAGCTGAAAAAATATATTTGAG
 >AK027147
 TTGTAATCATGCCAATTCAGATCAATAACTGCATGTCTGTTCTTTGGTAGAAATAGCTT
 15 >AW291189
 AAAGATTATTAACCCAAATCACCTTTCTTGCTTACTCCAGATGCCTCAGCCTCTGATATA
 >AI632869
 GACTTCCTTTAGGATCTCAGGCTTCTGCAGTTCTCATGACTCCTACTTTTCATCCTAGTC
 >BC006537
 20 CTGTATATTTTGCAATAGTTACCTCAAGGCCTACTGACCAAATGTTGTGTTGAGATGAT
 >R61469
 TGTTCAAACAGACTTTAACCTCTGCATCATACTTAACCCTGCGACATGCGTACAGTATGC
 >BC009084
 TGAGTCATATACATTTACTGACCACTGTTGCTTGTTGCTCACTGTGCTGCTTTTCCATGA
 25 >N64339
 CTGAAATGTGGATGTGATTGCCTCAATAAAGCTCGTCCCCATTGCTTAAGCCTTCAAAAA
 >AI200660
 ATCAAGAAAACCTAATCTTCTGACTCCCAGGCCAGGATGTTTTATTTCTCACATCATGTC
 >AK054605
 30 TTCATTTCCAAACATCATCTTTAAGACTCCAAGGATTTTTCCAGGCACAGTGGCTCATAC
 >NM_006115
 AGTTAGAAATAGAATCTGAATTTCTAAAGGGAGATTCTGGCTTGGGAAGTACATGTAGGA
 >X96757
 CAATTTTCTTTTTACTCCCCCTCTTAAGGGGGCCTTGAATCTATAGTATAGAATGAACT
 35 >AI804745
 GGGTGGAGTTTCAGTGAGAATAAACGTGTCTGCCTTTGTGTGTGTATATATACAGAGA
 >AJ000388
 CTCGCTCATTTTTTACCATGTTTTCCAGTCTGTTTAACTTCTGCAGTGCCTTCACTACAC
 >BC008764
 40 CTTTGGGCCGAGCACTGAATGTCTTGACTTTAAAAAATGTTTCTGAGACCTCTTTCTA
 >AI309080
 CTGGACCCTTGGAGCAGTGTGTGTGAACCTGCCTAGAACTCTGCCTTCTCCGTTGTCAA
 >AA865917
 CCACCTCCTTCGACCTCCACTGCGCCCCACCTCCCTGCCTGTGTGTGTTATTTCAAAGGA
 45 >AA946776
 TCTGGCTGGTGGCCTGCGCGAGGGTGCAGTCTTACTTAAAGACTTTTCAGTTAATTCTCA
 >AF104032
 AGATGCTGTGCGCACCATGTTTATTTATTTCCAGTGGTCATGCTCAGCCTTGCTGCTCTG
 >AW194680
 50 TCCTTCCTCTTCGGTGAATGCAGGTTATTTAACTTTGGGAAATGTACTTTTAGTCTGTC
 >BC001293
 GTCCTGTCCCTGTCTGGGAGTTGTGTTATTTAAAGATATTCTGTATGTTGTATCTTTTGC
 >BE962007
 ATTATATTTTCAGGTGTCCTGAACAGGTCACTAGACTCTACATTGGGCAGCCTTTAAATAT
 55 >BI493248
 AGGAATGGTACTACCGTTCCAGATTTTCTGTAATTGCTTCTGCAAAGTAATAGGCTTCTT
 >AF283645
 CTGTACCCAAAGGATGCCAGAATACTAGTATTTTTATTTATCGTAAACATCCACGAGTGC
 >AI669760
 60 ATTGCCCCCCTAACCAATCATGCAAATTTTCCCCCCTGGGGTAATTCACCAGTTAAAA
 >B001492
 CCCACAGTATTTAATGCCCTGTGAGTCCCTTCTAGTCTGACTCAATGGTAACTTGCTGTA
 >BC004453
 AAAACCAACTCTCTACTACACAGGCCTGATAACTCTGTACGAGGCTTCTCTAACCCTAG
 65 >BC010626
 CTCAGACTGGGCTCCACACTCTTGGGCTTCAGTCTGCCCATCTGCTGAATGGAGACAGCA

>B013117
 CCTAATGGGGATTCTCTGGTTGTTCACTGCCAAACTGTGGCATTTCATTACAGGAGA
 >BC011949
 CACTCACAATTGTTGACTAAAATGCTGCCTTTAAACATAGGAAAGTAGAATGGTTGAGT
 5 >AW083022
 CTTTGAAGGGCTGCTGCACATTGTTGAATCCATCGACCTTTAGCTGCAATGGGATCTCTA
 >R08175
 TGCCTCATCGATATTATAGGGGTCCATCACAACCCAACTGTGTGGCCGGATCCTGAGTCT
 >NM_000065
 10 AAAACAGACAAAAGCCTTTGCCTTCATGAAGCATAACATTCATTAGGGGTAGACACACAA
 >AK025181
 TAACAAACAAAGGCAGTAGCTCATCACTTGGGTAGCAGGTACCCATTTTAGGACCCTACA
 >NM_003914
 ATATCAGAAAGTCCAATAATCGTCATAGGCTTCTGCACGTTGGATCAACTAATGTTGTTT
 15 >AI123582
 ATCATAGCCCAACCATGTGAGAAGAAGGAGAAGGCCCCCTTTCTTCATTAATCTGAAAA
 >BC004331
 GCAGACCATTCTATCATACCTGGCAGGGCTTCTGTTTTATTTTGTAGGCTGGATGCTACC
 >AI341378
 20 ACTACAAGCCTCTTGTGTTTTACCAAAACCCTACATCTCAGGCTTACTAATTTTTGTGAT
 >NM_004063
 GCCATGCATACATGCTGCGCATGTTTTCTTCATTTCGTATGTTAGTAAAGTTTTGGTTATT
 >BC012926
 CACCTATTTATTTTACCTCTTTCCCAAACCTGGAGCATTTATGCCTAGGCTTGCAAGAA
 25 >AL110274
 GTGGACATAGCCACTAACCAACTAGTTACCTTTGGACTGCAACAAAAAATGTGAAAATGA
 >AW473119
 ACTTGTAACCTCTTTTGCACCTTGAAAAAGAATCCAGCGGGATGCTCGAGCACCTGTAA
 >AI685931
 30 AATTCTCTATAAACGGTTCACCAGCAAACCACCAATACATTCCATTGTTTGCCTAGAGAG
 >BF592799
 AATGGCCCATGCATGCTGTTTGCAGCAGTCAATTGAGTTGAATTAGAATTCCAACCATAC
 >BC000045
 GAGCTCAGTACTTGCCCTGTGAAAATCCCAGAAGCCCCCGCTGTCAATGTTCCCATCCA
 35 >BC015582
 ATGAAGCGGAATTAGGCTCCCGAGCTAAGGGACTCGCCTAGGGTCTCACAGTGAGTAGGA
 >M60502
 AGTGGCTATATCAACATCAGGGCTAGCACATCTTCTCTATTATCCTTCTATTGGAATTC

40 La invención también proporciona un segundo grupo de 90 secuencias genéticas de las que de 5 a 49 se pueden
 usar en la parte de la invención. Las 5 a 49 secuencias genéticas se pueden usar junto con la determinación de los
 niveles de expresión de secuencias adicionales siempre y cuando los niveles de expresión de las secuencias
 genéticas del conjunto de 90 se usan en la clasificación. Un ejemplo en la limitante de tales realizaciones de la
 invención es cuando la expresión de 5 a 49 de las 90 secuencias genéticas se mide junto con los niveles de
 45 expresión de una pluralidad de otras secuencias, tal como mediante el uso de una plataforma basada en micromatriz
 usada para realizar la invención. Cuando esos otros niveles de expresión no se usan en la clasificación, estos se
 pueden considerar los resultados de un "exceso" de secuencias transcritas y no es crítico para la práctica de la
 invención. Como alternativa, y cuando esos otros niveles de expresión se usan en la clasificación, éstos están dentro
 del alcance de la invención, en la que el uso de las secuencias que se han descrito anteriormente no excluye
 50 necesariamente el uso de niveles de expresión de secuencias adicionales.

38 miembros del conjunto de 90 se incluyen en el primer conjunto de 74 que se ha descrito anteriormente. Los
 números de referencia de estos miembros en común entre los dos conjuntos son AA456140, AA846824, AA946776,
 AF332224, AI620495, AI632869, AI802118, AI804745, AJ000388, AK025181, AK027147, AL157475, AW194680,
 55 AW291189, AW298545, AW473119, BC000045, BC001293, BC001504, BC004453, BC006537, BC008765,
 BC009084, BC011949, BC012926, BC013117, BC015754, BE962007, BF224381, BF437393, BI493248, M60502,
 NM_000065, NM_003914, NM_004063, NM_004496, NM_006115, y R61469. Las secuencias de ARNm que
 corresponden a miembros del conjunto de 90 quien están presentes en el conjunto de 74 secuencias genéticas
 también se proporcionan en el Apéndice 2 adjunto (Listado de Secuencias) junto con información de identificación
 60 adicional. El listado de la información de identificación para estos 52 miembros únicos mediante números de
 referencia, y como las secuencias de oligonucleótidos correspondientes que se pueden usar en la práctica de la
 invención, proporciona con lo siguiente.

>R15881
 65 ACTTCTGGTGATGATAAAAATGGTTTTATCACCCAGATGTGAAAGAAGCTGCCTGTTTAC
 >AI041545

GTGGTTCTGTAAAAACGCAGAGGAAAAGAGCCAGAAGGTTTCTGTTTAATGCATCTTGCC
 >NM_024423
 TTTATAAGGAAGCAGCTGTCTAAATGCAGTGGGGTTTGTGTTTGCAATGTTTAAACAGA
 >AB038160
 5 CTTATGAAGCTGGCCGGGCCACTCACGTTCAATGGTACATCTGGGTCTCTATGTGGTTCT
 >AK026790
 GTGAGCGAGCATTTCCCATAGCTAACCCCTATTCTCTTAGTCTTTCAAAATGTAGAATGGG
 >BC012727
 CTTTACACCTGATAAAATATTTTGCAGAGAGAGGTGTTCTTTTTCCTTACTGGTGCTGAA
 10 >BC016451
 GCATACATCTCATCCACAGGGGAAGATAAAGATGGTCACACAAACAGTTTCCATAAAGAT
 >H09748
 TGAGTTCAGCATGTGTCTGTCCATTTCAATTTGTACGCTTGTTCAAACCAAGTTTGTCT
 >NM_006142
 15 AAGACCGAGACTGAGGGAAAGCATGTCTGCTGGGTGTGACCATGTTTCTCTCAATAAAG
 >AF191770
 GGCATCTGGCCCTGGTAGCCAGCTCTCCAGAATTACTTGTAGGTAATTCCTCTCTTCAT
 >NM_006378
 TGGATGTTTGTGCGCGTGTGTGGACAGTCTTATCTTCCAGCATGATAGGATTTGACCATT
 20 >BC006819
 TCCTGGCAGAGCCATGGTCCCAGGCTTCCCAAAAGTGTGTTGTGGCAATTATCCCCTAGG
 >X79676
 TTTGATGATAGCAGACATTGTTACAAGGACATGGTGAGTCTATTTTAAATGCACCAATCT
 >BC006811
 25 TTCTTCCAGTTGCACTATTCTGAGGGAAAATCTGACACCTAAGAAATTTACTGTGAAAAA
 >NM_000198
 GAACAATTGTGGTCTCTCTTAACCTTGAGGTTCTCTTTGACTAATAGAGCTCCATTTCCT
 >AF301598
 GTTAAGTGTGGCCAAGCGCACGGCGGCAAGTTTTCAAGCACTGAGTTTCTATTCCAAGAT
 30 >NM_002847
 CGGCCTACTGAGCGGACAGAATGATGCCAAAATATTGCTTATGTCTCTACATGGTATTGT
 >NM_004062
 CAGGGTGTTTGCCCAATAATAAAGCCCCAGAGAACTGGGCTGGGCCCTATGGGATTGGTA
 >AW118445
 35 TGTACAGTTTGGTTGTTGCTGTAAATATGGTAGCGTTTTGTTGTTGTTGTTTTTCATGC
 >BC002551
 TACCAAACCTGGGACTCACAGCTTTATTGGGCTTTCTTTGTGTCTTGTGTGTTTCTTTTAT
 >AA765597
 CATTGAGGTTTGGATGGTGGCAGGTAAAACAGAAAGGCAAGATGTCATCTGACATTAGGC
 40 >AL137761
 AGTTCAGCACTGTGGTTATCATTGGTGATGCCAGAAAACATTAGTAGACTTAGACAATTG
 >X78202
 TAAAATTTCTTGATTGTGACTATGTGGTCATATGCCCGTGTTTGTCACTTACAAAAATGT
 >AK025615
 45 AGCCATCTGGTGTGAAGAACTCTATATTTGTATGTTGAGAGGGCATGGAATAATTGTATT
 >BC001665
 CTTATTGTCACTGGTTAAGAACTTGGCGAGATTGAAGGGCTTTTGTATTGTTGTTGGAT
 >AI985118
 CTTTCTAGTGAGCTAACCGTAACAGAGAGCCTACAGGATACACGTGAGATAATGTCACGT
 50 >AL039118
 TTGTCTTAAAATTTCTTGATTGTGATACTGTGGTCATATGCCCGTGTTTGTCACTTACAA
 >AA782845
 CCTGGGGGAAAGGGGCATTGATGACCTGAACTTTTGTAGCAAATTATTATTCTCAGTTTCC
 >BC016340
 55 TTCATTAACAGTACTAAGTGGAAGGGATCTGCAGATTCCAAATTGGAATAAGCTCTATCA
 >AA745593
 CCAATGCAGAAGAGTATTAAGAAAGATGCTCAAGTCCCATGGCACAGAGCAAGGCGGGCA
 >NM_004967
 CAAGGCTACGATGGCTATGATGGTCAGAATTACTACCACCACCAAGTGAAGCTCCAGCCTG
 60 >BF510316
 AGCTCACAGCTGGACAGGTGTTGTATATAGAGTGGAATCTCTTGGATGCAGCTTCAAGAA
 >AA993639
 TCCAAAGTAGAAAGGGTTCTTTTAGAAAACCTGAAGAATGTGCCTCCTCTTAGCATCTGT
 >AV656862
 65 GATGCATTTTTCAGTCCCTTTTCAGAGCAAATGCTTTTGCAATGGTAGTAATGTTTAGTT
 >X69699

CCTGTGGGGCTTCTCTCCTTGATGCTTCTTTCTTTTTTAAAGACAACCTGCCATTACCA
 >BC013282
 TTGACTAAGTCATGCTGTTTCCTCAAAGAAGCTTTGTTTTTGTAAACGTATTACTCAG
 >AI457360
 5 CTGGATCCCAGGCCCTGGCACCCCTCAGGAAATACAAGAAAAGAATATTCACATCTGTT
 >AW445220
 TTAGAGGGGCCACCTATCAACTCATCAGTGTTCAAAGAATATGCTGGGAGCATGGGTGAG
 >AF038191
 GGCCCATTTATGTCCCTCATGTCTCTAGATTTTCTCGTCACCCAGCCTCAAAAATATATG
 10 >X05615
 TCCCCAAAAACCTCACCCGAGGCTGCCCACTATGGTCATCTTTTTCTCTAAAATAGTTAC
 >BC005364
 GAAATTCCTCACACCTTGACCTTCCCTACTTTTCTGAATTGCTATGACTACTCCTTGTT
 >AK025701
 15 TGTCTGTCCACCACGAGATGGGAGGAGGAGAAAAAGCGGTACGATGCCTTCCTGACCTCA
 >BF446419
 GTCTTATCTCTCAGGGGGGGTTTAAGTGCCGTTTGCAATAATGTCGTCTTATTTATTTAG
 >AK025470
 CCGAGTAGTATGGGTCTCTGTGTGAGAAACCAGGAGATATTTTCATCTTGTTCCGAAATA
 20 >BE552004
 TTGTGCAAAAAGTCCCACAACCTTTCTGGATTGATAGTTTGTGGTGAAATAACAATTTTA
 >H05388
 TCCAGTATTCTGCAGGGCCAGTCAGTTGTACAGAAGTTGGAATATTCTGTTCCAGAATTA
 >NM_033229
 25 GTCTCGAACAGCGGTTGTTTTACTTTATTTATCTTAGGCCCTCAGCTCCCTGACGTCCT
 >BC010437
 AGTGAATCTTTTCTCTTGGTAGCATCAACACTGGGGATAAATCAGAACCATTCTGTGGA
 >AI952953
 TGAGAGCCCAGAACAAGAAGGAGCAGAAGGGCACTTTGACCTTCATTATTATGAAAATCA
 30 >R45389
 GGAAGAACTGATGCTTGCTGCTAACTAAAGTTTGGATGTATCGATTTAGAGAACCAATT
 >NM_001337
 GAATGAGAGAATAAGTCATGTTCTTCAAGATCATGTACCCCAATTTACTTGCCATTACT
 >AI499593
 35 TACGGAAAGGAAACAGGTTATACTCTTAGATTTAAAAAGTGAAAGAACTGCAGGCGCCT

En algunas realizaciones de la invención, los niveles de expresión de las secuencias genéticas se mide mediante la
 detección de las secuencias expresadas en una muestra que contiene células como que se intrigan a los
 oligonucleótidos mencionados anteriormente, que corresponden a secuencias en los Apéndices (Listado de
 40 Secuencias) como se indica con los números de referencia por.

En realizaciones adicionales, la invención proporciona el uso de cualquier número de las secuencias genéticas del
 conjunto de 74 poder conjunto de 90 en los métodos de la invención. Por lo tanto, en cualquier parte de 1 a todas las
 49 secuencias genéticas usadas en la invención puede ser de cualquiera o de ambos de los conjuntos mencionados
 45 anteriormente. De modo que para una, dos, tres, cuatro, o cinco, o más de las 5 a 49 secuencias pueden ser el
 conjunto de 74 o del conjunto de 90. De forma análoga, y cuando se usa de 10 a 49 secuencias, seis, siete, ocho,
 nueve, o diez de las secuencias pueden ser de uno de estos conjuntos.

Como se usa en el presente documento, una "muestra de tumor" o "muestra que contiene tumor" o "muestra que
 50 contiene célula tumoral" o variantes de las mismas, se refieren a muestras que contienen células de tejido o de fluido
 aisladas de un individuo del que se sospecha que está afectado con, o en riesgo de desarrollar, cáncer. Las
 muestras pueden contener células tumorales que se pueden aislar mediante métodos conocidos otros métodos
 apropiados según lo considere deseable el médico experto. Éstos incluyen, pero no se limitan a, microdissección,
 microdissección mediante captura por láser (LCM), o microdissección con láser (LMD) antes de su uso en la presente
 55 invención. Como alternativa, se pueden usar células sin diseccionar dentro de una "sección" de tejido. Los ejemplos
 no limitantes de tales muestras incluyen aislados primarios (a diferencia de células cultivadas) y se pueden recoger
 mediante cualquier medio no invasivo o mínimamente invasivo, que incluye, pero no se limita a, lavado ductal,
 aspiración con aguja fina, biopsia con aguja, los dispositivos y métodos que se describen en la Patente de Estados
 Unidos Nº 6.328.709, o cualquier otro medio adecuado reconocido en la técnica. Como alternativa, la muestra se
 60 puede recoger mediante un método invasivo, que incluye, pero no se limita a, biopsia quirúrgica.

La detección y medida de las secuencias transcritas se puede conseguir mediante una diversidad de medios
 conocidos en la técnica o según lo considere apropiado el médico experto. Básicamente, se puede usar cualquier
 método de ensayo siempre y cuando el ensayo refleje, de forma cuantitativa como cualitativa, la expresión de la
 65 secuencia transcrita que se está detectando.

La capacidad para clasificar muestras de tumor se proporciona mediante el reconocimiento de la relevancia del nivel de expresión de las secuencias genéticas (ya sean seleccionadas de forma aleatoria o específicas) y no mediante la forma del ensayo usado para determinar el nivel de expresión real. Un ensayo de la invención puede usar cualquier característica de identificación de una secuencia genética individual como se desvela en el presente documento siempre y cuando 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). Algunos ensayos adicionales incluyen los basados en la detección de fragmentos de polipéptido del miembro o miembros relevantes del proteoma. Los ejemplos no limitantes de este último incluye en la detección de fragmentos proteolíticos encontrados en un fluido biológico, tal como sangre o suero. Algunas características de identificación incluyen, pero no se limitan a, secuencias de ácidos nucleicos únicas usadas para codificar (ADN), o expresar (ARN), dicho gen o epítomos específicos para, o actividades de, una proteína codificada mediante una secuencia genética.

Los medios adicionales incluyen la detección de la amplificación de ácidos nucleicos como indicativo del aumento de los niveles de expresión e inactivación, supresión, o metilación de ácidos nucleicos, como indicativo de una disminución de los niveles de expresión. Indicado de forma diferente, la invención se puede poner en práctica sometiendo al ensayo uno o más aspectos del molde o moldes de ADN subyacen a la expresión de cada secuencia genética, del ARNm usado como un producto intermedio para expresar la secuencia, o del producto proteico expresado mediante la secuencia, así como fragmentos proteolíticos de tales productos. Como tal, la detección de la presencia de, cantidad de, estabilidad de, o degradación (incluyendo tasa) de, tal ADN, ARN y moléculas proteicas se puede usar en la práctica de la invención.

En algunas realizaciones, toda o parte de una secuencia genética se puede amplificar y detectar mediante métodos tales como la reacción en cadena de la polimerasa (PCR) y variaciones de la misma, tales como, pero no se limitan a, PCR cuantitativa (Q-PCR), PCR por transcripción inversa (RT-PCR), y PCR en tiempo real (que incluye un medio para medir las cantidades iniciales de copias de ARNm para cada secuencia en una muestra), opcionalmente RT-PCR en tiempo real o Q-PCR en tiempo real. Tales métodos usarían uno o dos cebadores que son complementarios con partes de una secuencia genética, en la que los celadores se usan para cebar la síntesis de ácidos nucleicos. Los ácidos nucleicos recién sintetizados se marcan opcionalmente y se pueden detectar directamente o mediante hibridación con un polinucleótido de la invención. Los ácidos nucleicos recién sintetizados se pueden poner en contacto con polinucleótidos (que contienen secuencias genéticas) de la invención en condiciones que permitan su hibridación. Algunos métodos adicionales para detectar la expresión de ácidos nucleicos expresados incluyen ensayos de protección de ARNsa, que incluyen hibridación es en fase líquida, e hibridación de células *in situ*.

Como alternativa, la expresión de secuencias genéticas en muestras de FFPE se puede detectar cómo se desvela en la patente de Estados Unidos N° US-7.504.214 presentada el 2 de diciembre de 2003, y el documento de Patente de Estados Unidos N° US-20050239083 A1, presentado el 6 de febrero de 2004. En resumen, la expresión de toda o parte de una secuencia genética expresada o transcrita se puede detectar mediante el uso de detección mediada por hibridación (tal como, pero no limitada a, tecnología basada en micromatriz, perla, o partícula) o detección mediada por PCR cuantitativa (tal como, pero no limitada a, PCR en tiempo real y PCR de transcriptasa inversa) como ejemplos no limitantes. La expresión de todo o parte de un polipéptido expresado se puede detectar mediante el uso de técnicas de inmunohistoquímica u otra detección mediada por anticuerpos (tal como, pero no se limita a, el uso de anticuerpos marcados que se unen de forma específica al menos parte del polipéptido con respecto a otros polipéptidos) como ejemplos no limitantes. Algunos medios adicionales para análisis de la expresión genéticas están disponibles, incluyendo detección de la expresión dentro de un ensayo para expresión genética global, o casi global, en una muestra (por ejemplo, como parte de un análisis de formación de perfiles de expresión genética tal como en una micromatriz). Algunos ejemplos no limitantes son la amplificación de ARN lineal y los que se describen en la Patente de Estados Unidos N° 6.794.141 (presentada el 25 de octubre de 2001).

En realizaciones que usan un ensayo basado en ácidos nucleicos para determinar la expresión se incluye la inmovilización de una o más secuencias genéticas en un soporte sólido, que incluye, pero no se limita a, un sustrato sólido como una matriz o a tecnología basada en perlas o perla como se conoce en la técnica. Como alternativa, también se pueden usar ensayos de expresión basados en solución conocidos en la técnica. La secuencia o secuencias genéticas inmovilizadas pueden estar en forma de polinucleótidos que son únicos o de otro modo específicos para el gen o genes de modo que los polinucleótido serían capaces de hibridarse con el ADN o ARN de dicho gen o genes. Estos polinucleótidos pueden ser la longitud completa del gen o genes o pueden ser secuencias cortas de los genes (hasta un nucleótido más corta que la secuencia de longitud completa conocida en la técnica mediante supresión del extremo en la posición 5' o 3' de la secuencia) que se interrumpen de forma mínima opcionalmente (tal como mediante pares de bases no complementarias desemparejadas o insertadas) de modo que la hibridación con un ADN o ARN correspondiente a los genes no se ve afectada. En algunas realizaciones, los polinucleótidos usados son del extremo en la posición 3' del gen, tal como dentro de aproximadamente 350, aproximadamente 300, aproximadamente 250, aproximadamente 200, aproximadamente 150, aproximadamente 100, o aproximadamente 50 nucleótidos a partir de la señal de poliadenilación o sitio de poliadenilación de un gen o secuencia expresada. También se pueden usar polinucleótidos que contienen mutaciones con respecto a las secuencias de los genes desvelados siempre y cuando la presencia de las mutaciones todavía permita la hibridación. Para producir una señal detectable. Por lo tanto, la práctica de la presente invención no se ve afectada por la

presencia de faltas de coincidencias menores entre las secuencias desveladas y las expresadas por células de una muestra de un sujeto. Un ejemplo no limitante de la existencia de tales faltas de coincidencias se observa en casos de polimorfismo de secuencias entre individuos de una especie, tal como pacientes humanos individuales dentro de *Homo sapiens*.

5 Como observarán los expertos en la materia, algunas secuencias genéticas incluyen tramos de poli A en la posición 3' (o poli T en la hebra complementaria) que no contribuyen a la singularidad de las secuencias desveladas. Por lo tanto, la invención se puede poner en práctica con secuencias genéticas que carecen de tramos de poli A en la posición 3' (o poli T). La singularidad de las secuencias desveladas se refiere a las partes o totalidades de las
10 secuencias que se encuentran solamente en ácidos nucleicos, incluyendo secuencias únicas encontradas en la parte sin traducir en la posición 3' de las mismas. Algunas secuencias únicas para la práctica de la invención son las que contribuyen a las secuencias consenso para los genes de modo que las secuencias públicas serán útiles en la detección de la expresión en una diversidad de individuos en lugar de ser específicas para un polimorfismo presente en algunos individuos. Como alternativa, se pueden usar secuencias únicas para un individuo o una subpoblación.
15 Las secuencias únicas pueden tener las longitudes de los polinucleótidos de la invención como se describe en el presente documento.

En realizaciones adicionales de la invención, se usan polinucleótidos que tienen secuencias presentes en las regiones sin traducir y/o no codificantes en la posición 3' de secuencias genéticas para detectar los niveles de
20 expresión en nuestras que contienen células de la invención. Tales polinucleótidos pueden contener opcionalmente secuencias encontradas en las partes en la posición 3' de las regiones de codificación de secuencias genéticas. Algunos polinucleótidos que contienen una combinación de secuencias de las regiones de codificación y no codificantes en la posición 3' tienen preferentemente las secuencias colocadas de forma contigua, sin intervención de secuencia o secuencias heterólogas.

25 Como alternativa, la invención se puede poner en práctica con polinucleótidos que tiene secuencias presentes en las regiones sin traducir y/o no codificantes en la posición 5' de las secuencias genéticas para detectar el nivel de expresión en células y muestras de la invención. Tales polinucleótidos pueden contener opcionalmente secuencias encontradas en las partes en la posición 5' de las regiones de codificación. Algunos polinucleótidos que contienen
30 una combinación de secuencias de las regiones de codificación y no codificación en la posición 5' pueden tener las secuencias colocadas de forma contigua, sin intervención de secuencia o secuencias heterólogas. La invención también se puede poner en práctica con secuencias presentes en las regiones de codificación de secuencias genéticas.

35 Los polinucleótidos de algunas realizaciones contienen secuencias de regiones sin traducir y/o no codificantes en la posición 3' o 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
40 aproximadamente 42, al menos aproximadamente 44, o al menos aproximadamente 46 nucleótidos consecutivos. El término "aproximadamente" como se ha usado en la frase anterior se refiere a un aumento o disminución de 1 a partir del valor numérico indicado. Otras realizaciones usan 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
45 aproximadamente 350, o al menos o aproximadamente 400 nucleótidos consecutivos. El término "aproximadamente" no se usa en la frase mencionada anteriormente se refiere a un aumento o disminución de un 10 % A partir del valor numérico indicado.

Las secuencias del extremo en la posición 3' o 5' de las regiones de codificación genética tal como se encuentran en los polinucleótidos de la invención tienen las mismas longitudes que las que se han descrito anteriormente, excepto en que se verían limitadas de forma natural por la longitud de la región de codificación. El extremo en la posición 3' de una región de codificación puede incluir secuencias de hasta la mitad de la posición 3' de la región de codificación. Por el contrario, el extremo en la posición 5' de una región de codificación puede incluir secuencias de hasta la mitad de la posición 5' de la región de codificación. Por supuesto, las secuencias que se han descrito
50 anteriormente, o las regiones de codificación y los polinucleótidos que contienen partes de las mismas, se pueden usar en su totalidad.

En otra realización de la invención, se pueden usar polinucleótidos que contienen supresiones de nucleótidos a partir del extremo en la posición 5' y/o 3' de las secuencias genéticas. Las supresiones son preferentemente 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, o 175-200 nucleótidos a partir del extremo en la posición 5' y/o 3', aunque el alcance de las supresiones estaría limitado de forma natural por la longitud de las secuencias y la necesidad para ser capaz de usar los polinucleótidos para la detección de niveles de expresión.

65 Otros polinucleótidos de la invención a partir del extremo en la posición 3' de las secuencias genéticas incluyen los de los cebadores y sondas opcionales para PCR cuantitativa. Preferentemente, los cebadores y sondas son los que

amplifican una región menor que aproximadamente 750, menor de aproximadamente 700, menor de aproximadamente 650, menor de aproximadamente 6000, menor de aproximadamente 550, menor de aproximadamente 500, menor de aproximadamente 450, menor de aproximadamente 400, menor de aproximadamente 350, menor de aproximadamente 300, menor de aproximadamente 250, menor de aproximadamente 200, menor de aproximadamente 150, menor de aproximadamente 100, o menor de aproximadamente 50 nucleótidos a partir de la señal de poliadenilación o sitio de poliadenilación de un gen o secuencia expresada. El tamaño de un amplicón de PCR de la invención puede tener cualquier tamaño, incluyendo 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, todos con la inclusión de la parte complementaria de los cebadores de PCR usados.

Otros polinucleótidos para uso en la práctica de la invención incluyen los que tienen una homología suficiente para las secuencias genéticas para detectar su expresión mediante el uso de técnicas de hibridación. Tales polinucleótidos tienen preferentemente aproximadamente o un 95 %, aproximadamente o un 96 %, aproximadamente o un 97 %, aproximadamente o un 98 %, o aproximadamente o un 99 % de identidad con las secuencias genéticas a usar. La identidad se determina usando el algoritmo BLAST, como se ha descrito anteriormente. Los otros polinucleótidos para uso en la práctica de la invención también se pueden describir sobre la base de la capacidad para hibridarse con los polinucleótidos de la invención en condiciones rigurosas de aproximadamente un 30 % en v/v a aproximadamente un 50 % de formamida y sal de aproximadamente 0,01 M a aproximadamente 0,15 M para hibridación y sal de aproximadamente 0,01 M a aproximadamente 0,15 M para condiciones de lavado de aproximadamente 55 °C a aproximadamente 65 °C o superior, o condiciones equivalentes a las mismas.

En una realización más de la invención, una población de moléculas de ácidos nucleicos monocatenarios que comprenden una o ambas hebras de una secuencia genética humana se proporciona como una sonda de modo que al menos una porción de dicha población se puede hibridar con una o ambas hebras de una molécula de ácido nucleico amplificada de forma cuantitativa a partir de ARN de una célula o muestra de la invención. La población puede ser solamente de la hebra antisentido de una secuencia genética humana de modo que una hebra sentido de una molécula de, o amplificada de, una célula se puede hibridar con una parte de dicha población. La población comprende preferentemente una cantidad lo suficientemente en exceso de dicha una o ambas hebras de una secuencia genética humana en comparación con la cantidad de moléculas de ácidos nucleicos expresadas (o amplificadas) que contienen una secuencia genética complementaria.

La invención también proporciona un método de clasificación de una muestra de tumor humano mediante la detección los niveles de expresión de aproximadamente 5 a 49 secuencias transcritas en una muestra que contiene ácido nucleico o célula obtenida de un sujeto humano, y clasificación de la muestra cómo que contiene una célula tumoral de un tipo de tumor encontrado en seres humanos hasta la exclusión de uno u otros tipos de tumor humano más. En algunas realizaciones, el método se puede usar para clasificar una muestra como que es, o tiene células de, uno de los 53 tipos de tumor que se han enumerado anteriormente para la exclusión de uno o más de los otros 52. En otras realizaciones, el método se usa para clasificar una muestra como que es, o tiene células de, uno de los 34 tipos de tumor que se han enumerado anteriormente para la exclusión de uno o más de los otros 33 tipos de tumor. En otras realizaciones, el método se usa para clasificar una muestra como que es, o tiene células de, uno de los 39 tipos de tumor que se han enumerado anteriormente para la exclusión de uno o más de los otros 38 tipos de tumor.

También se desvela un método para la clasificación de muestras de tumor como que son de un subconjunto de los posibles tipos de tumor que se describen en el presente documento mediante la detección de los niveles de expresión de 50 o más secuencias transcritas en una muestra de tumor que contiene ácidos nucleicos obtenida partir de un sujeto humano, y clasificación de la muestra como que es una de una serie de tipos de tumor encontrados en seres humanos para la función de uno otros tipos de tumor más. En algunas realizaciones de la invención, el número de los otros tipos de tumor es de 1 a aproximadamente 3, más preferentemente de 1 a aproximadamente 5, de 1 a aproximadamente 7, o de 1 a aproximadamente 9 o aproximadamente 10. En otras realizaciones, el número de tipos de tumor son todos del mismo origen de tejido humano tal como los que se han enumerado anteriormente. Este aspecto de la invención se relaciona con el análisis anterior de la Figura 10 y de la especificidad de la compensación a favor del aumento de confianza, y se puede aplicar de forma ventajosa a situaciones en las que la clasificación de una muestra de un solo tipo de tumor tiene un nivel de precisión o rendimiento que se puede mejorar mediante la clasificación de la muestra como una de un subconjunto de posibles tipos de tumor.

En realizaciones adicionales, la invención se puede poner en práctica mediante el análisis de la expresión genética de poblaciones de células individuales o células homogéneas que se han analizado minuciosamente a partir de, o de otro modo aislado o purificado a partir de, células contaminantes de la muestra como presentes en una biopsia sencilla. Una de las ventajas proporcionadas por estas realizaciones es que las células no tumorales, contaminantes (tales como linfocitos de infiltración u otras células del sistema inmune) se pueden retirar para que estén libres de influir en los genes identificados por el análisis posterior de los niveles de expresión genética como se proporciona

en el presente documento. Tal contaminación está presente cuando una biopsia se usa para generar perfiles de expresión genética.

En otras realizaciones de la invención que usan Q-PCR o Q-PCR de transcriptasa inversa como la plataforma de ensayo, los niveles de expresión de las secuencias genéticas de la invención se pueden comparar con los niveles de expresión de genes de referencia en la misma muestra o se puede usar una proporción de niveles de expresión. Esto proporciona un medio para "normalizar" los datos de expresión para comparación de datos en una pluralidad de tipos de tumor conocidos y una muestra que contiene células a someter a ensayo. Aunque se puede usar una diversidad de genes de referencia, la invención también se ponen en práctica con el uso de 8 secuencias genéticas de referencia en particular que se identificaron para su uso con el conjunto de 39 tipos de tumor. Además, la Q-PCR se puede realizar completamente buen parte con el uso de un formato multiplex.

Las secuencias de ARNm que corresponden a las 8 secuencias de referencia se proporcionan en los Apéndices adjuntos (Listado de Secuencias) junto con información de identificación adicional. El listado de la información de identificación, que incluye números de referencia y otra información, se proporciona mediante lo siguiente.

```
>Hs.77031_ARNm_1   gi|16741772|gb|BC016680.1|BC016680   clon   de   Homo sapiens   MGC:21349
IMAGE:4338754
poliA = 3
>Hs.775.41_ARNm_1   gi|12804364|gb|BC003043.1|BC003043   clon   de   Homo sapiens   MGC:4370
IMAGE:2822973
poliA = 3
>Hs.7001_ARNm_1     gi|6808256|emb|AL137727.1|HSM802274   ARNm   de   Homo sapiens;   ADNc
DKFZp434M0519
(del clon DKFZp434M0519); cds parcial poliA = 3
>Hs.302144_ARNm_1   gi|11493400|gb|AF130047.1|AF130047   clon   de   Homo sapiens   FLB3020   poliA = 0
>Hs.26510_ARNm_2    gi|1345385|gb|AF308803.1|AF308803   cromosoma 15 map 15q26 de Homo sapiens poliA =
3
>Hs.324709_ARNm_2   gi|12655026|gb|BC001361.1|BC001361   clon   de   Homo sapiens   MGC:2474
IMAGE:3050694
poliA = 2
>Hs.65756_ARNm_3    gi|3641494|gb|AF035154.1|AF035154   cromosoma 16 map 16p13.3 de Homo sapiens poliA
= 3
>Hs.165743_ARNm_2   gi|13543889|gb|BC006091.1|BC006091   clon   de   Homo sapiens   MGC:12673
IMAGE:3677524
poliA = 3
```

La detección de la expresión de cualquiera de las secuencias de referencia mencionadas anteriormente se puede realizar mediante la misma o diferente metodología que para las otras secuencias genéticas que se han descrito anteriormente. En algunas realizaciones de la invención, los niveles de expresión de las secuencias genéticas se miden mediante la detección de secuencias expresadas en una muestra que contiene células con que se hibridan con los oligonucleótidos siguientes, que corresponden a las secuencias mencionadas anteriormente, se indica mediante los números de referencia que se proporcionan.

```
>BC006091
TCATCTTCACCAAACAGTCCGAGGGGTGCGAAGCCAGACACGAGAGGAAGAGGGTCCTGG
>BC003043
CTCTGCTCCTGCTCCTGCCTGCATGTTCTCTCTGTTGTTGGAGCCTGGAGCCTTGCTCTC
>AF130047
TGCTCCCGGCTGTCTCCTCTCCTCTTCCCTAGTGAGTGGTTAATGAGTGTTAATGCCTA
>AF035154
CCCCATCTCTAAAACAGTAAATCAGCCAGCGAATACCCGGAAGCAAGATGCACAGGCGG
>BC001361
CCAGAAACAAGGAAGAGGAAAGACAAAGGGAAGGGACGGGAGCCCTGGAGAAGCCCGACC
>AF308803
AAGTACAACCCATGCTGCTAAGATGCGAGCAGGAAGAGGCATCCTTTGCTAAATCCTGTT
>BC016680
ACCTCACCCCTGCCCGGCCCAAGCTCTACTTGTGTACAGTGTATATTGTATAATAGACAA
>AL137727
TTCCCTTAATTCTCCTCCCGACCTTTTTTACCCCCCAGTTGCAGTATTTAACTGGGCT
```

En un aspecto adicional, los métodos proporcionados por la presente invención también se pueden automatizar total o parcialmente. Éstos incluyen la realización de la invención en software. Algunos ejemplos no limitantes incluyen instrucciones ejecutables con procesador en uno o más dispositivos de almacenamiento de lectura por ordenador en los que dichas instrucciones dirigen la clasificación de las muestras de tumor basándose en los niveles de expresión genética como se describe en el presente documento. Se contemplan instrucciones adicionales ejecutables con

procesador en uno o más sistemas dispositivos de almacenamiento de lectura por ordenador en los que dichas instrucciones provocan la representación y/o manipulación, a través de un dispositivo de salida de ordenador, del proceso o resultados de un método de clasificación.

- 5 La invención incluye realizaciones de software y hardware en las que los datos de expresión genética de un conjunto de secuencias genéticas en una pluralidad de tipos de tumor conocidos se realizan como un conjunto de datos. En algunas realizaciones, el conjunto de datos de expresión genética se usa para la práctica de un método de la invención. La invención también proporciona medios y sistemas relacionados con ordenadores para realizar los métodos que se desvelan en el presente documento. En algunas realizaciones, se proporciona un aparato para
10 clasificar una muestra que contiene células. Tal aparato puede comprender una entrada de búsqueda configurada para recibir un almacenamiento de la búsqueda configurado para almacenar un conjunto de datos de expresión genética, como se describe en el presente documento, recibidos a partir de una entrada de búsqueda; y un módulo para acceder a los datos de almacenamiento y usarlos en un algoritmo de clasificación como se describe en el presente documento. El aparato puede comprender adicionalmente un almacenamiento en serie para los resultados del algoritmo de clasificación, opcionalmente con un módulo para acceder a los datos y usarlos a partir del almacenamiento en serie en un algoritmo de salida como se describe en el presente documento.
15

Las etapas del método, proceso, o algoritmo que se describe en conexión con las realizaciones que se desvelan en el presente documento se pueden realizar directamente en hardware, en un módulo de software ejecutado con un procesador, o en una combinación de los dos. Las diversas etapas o actos en un método o proceso se pueden
20 realizar en el orden mostrado, o se pueden realizar en cualquier otro orden. Además, se puede omitir una o más etapas del proceso o método o se puede añadir una o más etapas del proceso a los métodos y procesos. Se puede añadir una etapa, lo que, o acción adicional al comienzo, final, o elementos existentes de intervención de los métodos y procesos.
25

Un aspecto adicional de la invención proporciona el uso de la presente invención con respecto a actividades clínicas. En algunas realizaciones, la determinación o medida de la expresión genética como se describe en el presente documento se realiza como parte de la provisión de cuidado médico a un paciente, que incluye la provisión de servicios de diagnóstico como soporte a la provisión de cuidado médico. Por lo tanto la invención incluye un método
30 en el cuidado médico de un paciente, método que comprende la determinación o medida de niveles de expresión de las secuencias genéticas en una muestra que contiene células obtenida a partir de un paciente como se describe en el presente documento. El método puede comprender adicionalmente la clasificación de la muestra, basándose en la determinación/medida, incluyendo una célula tumoral o un tipo de tumor u origen de tejido de una manera como se describe en el presente documento. La determinación y/o clasificación puede ser para uso con respecto a cualquier aspecto o realización de la invención como se describe en el presente documento.
35

La determinación o medida de niveles de expresión puede ir precedida por una diversidad de acciones relacionadas. En algunas realizaciones, la medida va precedida por una determinación o diagnóstico de uso de tu mano como con necesidad de dicha medida. La medida puede ir precedida por una determinación de una necesidad de la medida, tal
40 como la de un doctor en medicina, enfermera otro proveedor o profesional de cuidados para la salud, o las personas que trabajan bajo su instrucción, o personal que asegura u organización de mantenimiento de la salud en la aprobación de la realización de la medida como una base para realizar el reembolso o pago por la realización.

La medida también puede ir precedida por los actos necesarios para la medida real. Algunos ejemplos no limitantes incluyen la obtención real de una muestra que contiene células de un sujeto humano; o la recepción de la muestra que contiene células; o sección de una muestra que contiene células; o aislamiento de células de una muestra que contiene células; u obtención de ARN de células de una muestra que contiene células; o ARN de transcripción inversa de células de una muestra que contiene células. La muestra puede ser cualquiera de las que se describen en el presente documento para la práctica de la invención.
50

También se desvela un método para ordenar, o recibir una orden para, la realización de un método en el cuidado médico de un paciente u otro método de la invención. La orden la puede realizar un doctor en medicina, una enfermera, u otro proveedor de cuidados para la salud, o las personas que trabajan bajo sus instrucciones, a la vez que la recepción, directa o indirectamente, la puede realizar cualquier persona que realiza el método con métodos.
55 La orden se puede realizar mediante cualquier medio de comunicación, que incluye comunicación que está escrita, oral, electrónica, digital, analógica, telefónica, en persona, por fax, por correo electrónico, o de otro modo pasa a través de una jurisdicción dentro de los Estados Unidos.

También se desvelan métodos para el procesamiento de reembolso o pago por un ensayo, tal como el método mencionado anteriormente en el cuidado médico de un paciente u otro método de la invención. Un método para el procesamiento de reembolso o pago puede comprender la indicación de que 1) el pago se ha recibido, o 2) el pago lo realizará otro pagador, o 3) el pago permanece sin pagar en papel o en una base de datos después de la realización de un método de detección, determinación o medida del nivel de expresión de la invención. La base de datos puede estar en cualquier forma, con formularios electrónicos tales como bases de datos implementadas por ordenador incluidas dentro del alcance de la invención. La indicación puede estar en forma de un código (tal como un código CPT) en papel o en la base de datos. El "otro pagador" puede ser cualquier persona o entidad más allá de
60
65

los conocidos desde el momento en el que se realizó una solicitud previa de reembolso o pago.

Como alternativa, el método puede comprender la recepción del reembolso o pago por la realización técnica o real del método mencionado anteriormente en el cuidado médico de un paciente; para la interpretación de los resultados de dicho método; o para cualquier otro método de la invención. Por supuesto, la invención también incluye realizaciones que comprenden la instrucción de otra persona o equipo para recibir el reembolso o pago. La orden puede ser mediante cualquier medio de comunicación, que incluyen los que se han descrito anteriormente. La recepción puede ser de cualquier entidad, que incluye una compañía de seguros, organización de mantenimiento de la salud, agencia de salud gubernamental, o un paciente como ejemplos no limitantes. El pago se puede realizar totalmente o en parte. En el caso de un paciente, el pago se puede realizar en forma de pago parcial conocido como copago.

En otra realización más, el método puede comprender el reenvío o haber reenviado una solicitud de reembolso o pago a una compañía de seguros, organización para el mantenimiento de la salud, agencia de salud gubernamental, o a un paciente para la realización del método mencionó anteriormente el cuidado médico de un paciente u otro método de la invención. La solicitud puede ser mediante cualquier medio de comunicación, que incluye los que se han descrito anteriormente.

En una realización más, el método puede comprender la recepción de indicaciones de aprobación de pago, o de negación de pago, por la realización del método mencionado anteriormente en el cuidado médico de un paciente u otro método de la invención. Tal indicación puede venir de cualquier persona o equipo al que se realizó una solicitud de reembolso o pago. Algunos ejemplos no limitantes incluyen una compañía de seguros, organización de mantenimiento de la salud, o una agencia de salud gubernamental, tal como Medicare o Medicaid como ejemplos no limitantes. La indicación se puede realizar mediante cualquier medio de comunicación, incluyendo los que se han descrito anteriormente.

Una realización adicional es en la que el método comprende el envío de una solicitud de reembolso por la realización del método mencionado anteriormente en el cuidado médico de un paciente otro método de la invención. Tal solicitud se puede realizar mediante cualquier medio de comunicación, que incluye los que se han descrito anteriormente. La solicitud se puede haber realizado a una compañía de seguros, organización de mantenimiento de la salud, agencia de salud federal, o el paciente para el que se realizó el método.

Un método adicional comprende la indicación de la necesidad de reembolso o pago en un formulario o en una base de datos por la realización del método mencionado anteriormente en el cuidado médico de un paciente u otro método de la invención. Como alternativa, el método puede indicar simplemente la realización del método. La base de datos puede estar en cualquier forma, con formularios electrónicos tales como una base de datos implementada por ordenador incluida dentro del alcance de la invención. La indicación se puede realizar en forma de un código (tal como un código CPT) en papel o en la base de datos.

En los métodos mencionados anteriormente para el cuidado médico de un paciente u otro método de la invención, el método puede comprender el informe de los resultados del método, opcionalmente a una instalación de cuidados de la salud, un proveedor o profesional de cuidados de la salud, un doctor, una enfermera, o personal trabajando para los mismos. El informe también se puede proporcionar directa o indirectamente al paciente. El informe puede ser mediante cualquier medio de comunicación, que incluye los que se han descrito anteriormente.

También se describen kits para la determinación o medida de niveles de expresión genética en una muestra que contiene células como se describe en el presente documento. Un kit por lo general comprenderá uno o más reactivos para detectar la expresión genética como se describe en el presente documento para la práctica de la presente invención. Algunos ejemplos no limitantes incluyen sondas o cebadores de polinucleótidos para la detección de niveles de expresión, una o más enzimas usadas en los métodos de la invención, y uno o más tubos para uso en la práctica de la invención. En algunas realizaciones, el kit incluirá una matriz, o medio sólido capaz de su montaje en una matriz, para la detección de la expresión genética como se describe en el presente documento. En otras realizaciones, el kit puede comprender uno o más anticuerpos que son inmunorreactivos con epítopos presentes en un polipéptido que indica la expresión de una secuencia genética. En algunas realizaciones, el anticuerpo será un fragmento de anticuerpo.

Un kit también puede incluir materiales con instrucciones que desvelan o que describen el uso del kit o un cebador o sonda de la presente invención en un método de la invención como se proporciona en el presente documento. Un kit también puede incluir componentes adicionales para facilitar la aplicación en particular para la que está diseñado el kit. Por lo tanto, por ejemplo, un kit puede contener adicionalmente medios para detectar la marca (por ejemplo, sustratos enzimáticos para marcas enzimática es, conjuntos de filtros para detectar marcas fluorescentes, marcas secundarias apropiadas tales como una anti-ratón-HRP de oveja, o similares). Un kit puede incluir adicionalmente tampones y otros reactivos reconocidos para uso en un método de la invención.

Habiendo descrito ahora por lo general la invención, la misma se entenderá más fácilmente a través de la referencia a los siguientes ejemplos que se proporcionan a modo de ilustración, y no pretenden ser limitantes de la presente

invención, a menos que se especifique.

Ejemplos

5 Ejemplo 1: Capacidad de información de conjuntos de genes aleatorios

Se sometieron a ensayos subconjuntos de 100 secuencias genéticas expresadas seleccionadas de forma aleatoria usadas para clasificarse entre 39 tipos de tumor para su capacidad para clasificarse entre subconjuntos de los 39 tipos de tumor. Los niveles de expresión de combinaciones aleatorias de 5, 10, 15, 20, 25, 30, 35, 40, 45, y 49 (se toma muestra de cada combinación 10 veces) de las 100 secuencias expresadas se usaron con datos de tipos de tumor y a continuación se usaron para predecir conjuntos aleatorios de ensayo de muestras de tumor (se toma muestra de cada uno 10 veces) que varían de 2 a los 39 tipos. La Figura 1 muestra la capacidad de clasificación de diversos conjuntos de genes que se muestran con respecto al número de tipos de tumor clasificados. Como se esperaba, se necesita un número de secuencias genéticas más elevado para clasificar los tipos de tumor con precisiones más elevadas. La Figura 2 muestra el rendimiento de la clasificación para diversos números de tipos de tumores con respecto al número de secuencias genéticas usadas.

Los números de referencia en GenBank de las 100 secuencias genéticas son AF269223, BC006286, AK025501, AJ002367, AI469140, AW013883, NM_001238, AI476350, BC006546, AI041212, BF724944, AI376951, R56211, BC006393, X13274, BC001133, N62397, BC000885, AK001588, AK057901, AF146760, AI951287, AK025604, BC007581, BC015025, R43102, AW449550, AI922539, AI684144, AI277662, BC015999, AW444656, BC011612, BC015401, BF447279, BC009956, AL050163, BC001248, BE672684, AL137353, BC001340, U45975, BE856598, BC009060, AL137728, AA713797, AL583913, AK054617, AI028262, AI753041, BG939593, AL080179, AA814915, AF31798, AI961568, BC009849, AK021603, BC012561, AI570494, BC006973, AW294857, BC004952, AK026535, AI923614, AW082090, AI005513, AF339768, AK023167, AF169693, AF076249, BC007662, BC015520, AI814187, AI565381, AW271626, AK024120, AF139065, BC014075, AI887245, AF2S7081, AI767898, AF070634, AF155132, X69804, U65579, NM_004933, AI655104, AW131780, AI650407, AF131774, AA814057, AJ311123, BC009702, AF264036, AL161961, AJ010857, AF106912, AK023542, AF073518, y D83032. Se hizo un índice de 1 a 100, y los conjuntos aleatorios representativos usados en la invención son los que siguen a continuación:

Para 2 genes, los genes 33 y 63, los genes 17 y 72, los genes 64 y 21, los genes 48 y 25, los genes 88 y 54, los genes 80 y 32, los genes 24 y 99, los genes 14 y 31, los genes 80 y 23, y genes 18 y 34 se usaron como los 10 conjuntos aleatorios.

Para 5 genes, en el conjunto 1, se usaron los genes 27, 97, 56, 88, y 50. En el conjunto 2, se usaron los genes 24, 26, 35, 48, y 83. En el conjunto 3, se usaron los genes 46, 62, 75, 91, y 2. En el conjunto 4, se usaron los genes 19, 61, 34, 87, y 13. En el conjunto 5, se usaron los genes 56, 32, 66, 20, y 55. En el conjunto 6, se usaron los genes 90, 21, 6, 78, y 66. En el conjunto 7, se usaron los genes 73, 47, 3, 82, y 86. En el conjunto 8, se usaron los genes 74, 39, 13, 7, y 67. En el conjunto 9, se usaron los genes 34, 1, 24, 85, y 62. En el conjunto 10, se usaron los genes 23, 89, 15, 54, y 98.

Para 10 genes, en el conjunto 1, se usaron los genes 11, 58, 90, 40, 20, 44, 10, 78, 72, y 74. En el conjunto 2, se usaron los genes 79, 71, 42, 48, 93, 56, 55, 14, 92, y 52. En el conjunto 3, se usaron los genes 62, 53, 52, 19, 98, 26, 76, 65, 33, y 40. En el conjunto 4, se usaron los genes 94, 8, 16, 99, 58, 19, 97, 92, 76, y 86. En el conjunto 5, se usaron los genes 18, 97, 16, 94, 84, 52, 11, 24, 89, y 92. En el conjunto 6, se usaron los genes 12, 42, 45, 51, 2, 75, 63, 28, 13, y 58. En el conjunto 7, se usaron los genes 67, 98, 55, 32, 82, 42, 2, 45, 37, y 23. En el conjunto 8, se usaron los genes 40, 43, 69, 68, 13, 97, 35, 3, 44, y 42. En el conjunto 9, se usaron los genes 69, 47, 96, 80, 100, 50, 42, 26, 65, y 17. En el conjunto 10, se usaron los genes 83, 84, 69, 67, 19, 85, 35, 11, 70, y 64.

Para 15 genes, en el conjunto 1, se usaron los genes 98, 81, 43, 63, 18, 56, 19, 97, 47, 13, 48, 99, 75, 45, y 83. En el conjunto 2, se usaron los genes 5, 72, 31, 59, 81, 40, 92, 3, 23, 50, 57, 74, 62, 21, y 93. En el conjunto 3, se usaron los genes 11, 69, 91, 100, 38, 1, 73, 64, 90, 26, 62, 2, 37, 23, y 18. En el conjunto 4, se usaron los genes 76, 9, 53, 4, 11, 41, 77, 44, 87, 51, 54, 49, 43, 56, y 67. En el conjunto 5, se usaron los genes 55, 34, 13, 89, 52, 74, 96, 80, 48, 22, 31, 39, 43, 91, y 54. En el conjunto 6, se usaron los genes 59, 88, 15, 90, 4, 73, 93, 7, 10, 18, 98, 83, 43, 3, y 5. En el conjunto 7, se usaron los genes 68, 91, 77, 33, 88, 94, 95, 41, 46, 27, 36, 51, 97, 7, y 2. En el conjunto 8, se usaron los genes 7, 10, 78, 40, 70, 84, 55, 1, 98, 22, 99, 91, 8, 17, y 89. En el conjunto 9, se usaron los genes 65, 10, 38, 8, 77, 98, 37, 43, 93, 99, 86, 16, 82, 27, y 9. En el conjunto 10, se usaron los genes 97, 27, 78, 38, 24, 19, 55, 47, 77, 13, 45, 25, 43, 70, y 68.

Para 20 genes, en el conjunto 1, se usaron los genes 41, 94, 38, 76, 35, 65, 92, 26, 49, 7, 85, 54, 77, 66, 98, 15, 86, 69, 70, y 67. En el conjunto 2, se usaron los genes 43, 87, 1, 81, 7, 14, 94, 28, 25, 55, 100, 41, 18, 47, 96, 89, 26, 53, 29, y 32. En el conjunto 3, se usaron los genes 48, 80, 90, 99, 50, 98, 36, 91, 6, 41, 61, 96, 74, 66, 9, 5, 16, 18, 20, y 1. En el conjunto 4, se usaron los genes 49, 58, 73, 24, 94, 22, 41, 52, 18, 19, 63, 91, 74, 37, 59, 95, 53, 87, 72, y 13. En el conjunto 5, se usaron los genes 67, 74, 2, 98, 46, 69, 5, 42, 22, 66, 60, 20, 100, 80, 24, 76, 63, 9, 39, y 15. En el conjunto 6, se usaron los genes 10, 74, 50, 92, 69, 68, 52, 56, 63, 71, 11, 17, 29, 64, 88, 59, 25, 94, 35, y 57. En el conjunto 7, se usaron los genes 97, 72, 16, 19, 14, 42, 70, 31, 29, 13, 22, 37, 95, 69, 87, 39, 18, 81, 58, y 100. En el conjunto 8, se usaron los genes 5, 3, 18, 91, 77, 19, 82, 31, 92, 22, 93, 45, 76, 84, 46, 100, 53, 99, 89, y 42. En el conjunto 9, se usaron los genes 62, 3, 85, 37, 34, 93, 52, 40, 74, 25, 86, 57, 33, 60, 20, 77, 78, 17, 28, y 13. En el conjunto 10, se usaron los genes 22, 26, 23, 39, 35, 10, 43, 32, 65, 38, 54, 45, 8, 17, 90, 20, 83, 60, 6, y 58.

Para 25 genes, en el conjunto 1, se usaron los genes 21, 28, 50, 27, 8, 48, 74, 80, 38, 96, 71, 15, 89, 84, 32, 26, 55, 36, 29, 68, 13, 7, 18, 63, y 72. En el conjunto 2, se usaron los genes 61, 38, 59, 92, 3, 80, 33, 68, 79, 70, 44, 26, 95, 63, 85, 27, 60, 43, 75, 96, 42, 99, 58, 48, y 91. En el conjunto 3, se usaron los genes 75, 83, 78, 5, 99, 56, 26, 36, 57, 23, 37, 28, 88, 16, 63, 2, 72, 59, 9, 80, 52, 91, 62, 3, y 27. En el conjunto 4, se usaron los genes 48, 75, 84, 83, 88, 29, 13, 9, 98, 6, 31, 63, 45, 5, 51, 52, 39, 22, 100, 91, 74, 12, 94, 21, y 8. En el conjunto 5, se usaron los genes 79, 84, 47, 43, 26, 37, 46, 19, 85, 91, 2, 10, 81, 89, 38, 71, 17, 57, 7, 93, 31, 87, 29, 78, y 73. En el conjunto 6, se usaron los genes 62, 93, 83, 42, 97, 96, 78, 98, 47, 22, 67, 48, 89, 95, 24, 81, 16, 45, 8, 90, 66, 64, 2, 3, y 58. En el conjunto 7, se usaron los genes 100, 34, 58, 28, 104, 35, 88, 76, 6, 30, 83, 81, 67, 36, 39, 87, 66, 45, 20, 15, 86, 56, 55, y 95. En el conjunto 8, se usaron los genes 17, 43, 50, 63, 47, 58, 95, 32, 79, 60, 16, 91, 86, 22, 97, 21, 9, 55, 72, 78, 77, 45, 100, 14, y 30. En el conjunto 9, se usaron los genes 24, 67, 60, 94, 59, 14, 70, 84, 8, 89, 63, 23, 39, 11, 81, 42, 33, 3, 12, 93, 54, 35, 78, 73, y 90. En el conjunto 10, se usaron los genes 11, 2, 19, 62, 13, 51, 30, 80, 81, 82, 52, 34, 67, 57, 25, 95, 93, 39, 26, 48, 44, 89, 61, 17, y 18.

Para 30 genes, en el conjunto 1, se usaron los genes 30, 97, 54, 21, 34, 9, 56, 71, 62, 14, 24, 23, 89, 61, 76, 41, 29, 67, 94, 22, 88, 4, 40, 33, 38, 78, 82, 66, 84, y 100. En el conjunto 2, se usaron los genes 89, 41, 56, 43, 98, 44, 35, 26, 19, 86, 15, 67, 8, 69, 3, 76, 48, 17, 55, 31, 25, 91, 72, 36, 18, 82, 37, 50, 9, y 75. En el conjunto 3, se usaron los genes 28, 39, 78, 15, 65, 93, 66, 29, 88, 35, 49, 69, 50, 9, 53, 80, 81, 95, 76, 44, 48, 64, 83, 11, 70, 33, 73, 96, 56, y 92. En el conjunto 4, se usaron los genes 4, 2, 19, 6, 11, 84, 94, 44, 60, 37, 29, 97, 53, 83, 98, 45, 65, 9, 85, 35, 20, 89, 10, 17, 23, 74, 70, 41, 18, y 76. En el conjunto 5, se usaron los genes 27, 4, 43, 1, 10, 95, 88, 74, 77, 47, 63, 81, 31, 9, 41, 100, 87, 57, 8, 79, 24, 6, 26, 20, 55, 61, 34, 42, 25, y 39. En el conjunto 6, se usaron los genes 47, 67, 98, 56, 37, 44, 5, 70, 48, 12, 20, 86, 83, 89, 27, 59, 19, 54, 69, 97, 43, 71, 58, 82, 8, 50, 51, 10, 25, y 72. En el conjunto 7, se usaron los genes 100, 99, 37, 58, 44, 60, 39, 3, 59, 96, 50, 68, 94, 69, 83, 90, 17, 4, 5, 67, 88, 56, 29, 79, 23, 1, 38, 25, 49, y 74. En el conjunto 8, se usaron los genes 26, 23, 58, 47, 6, 68, 41, 31, 16, 64, 19, 75, 36, 32, 87, 2, 12, 97, 73, 21, 53, 78, 15, 94, 1, 20, 79, 81, 70, y 7. En el conjunto 9, se usaron los genes 61, 48, 78, 75, 12, 36, 37, 66, 91, 2, 92, 32, 8, 26, 6, 82, 14, 68, 4, 88, 39, 89, 43, 41, 40, 87, 69, 74, 42, y 9. En el conjunto 10, se usaron los genes 58, 99, 60, 39, 50, 25, 22, 57, 48, 85, 24, 10, 97, 68, 36, 38, 93, 62, 52, 56, 34, 18, 32, 64, 95, 81, 74, 88, 61, y 96.

Para 35 genes, en el conjunto 1, se usaron los genes 52, 68, 22, 92, 43, 75, 20, 62, 15, 76, 99, 61, 64, 36, 12, 66, 24, 21, 31, 88, 25, 6, 93, 91, 55, 74, 69, 90, 23, 4, 80, 72, 97, 58, y 1. En el conjunto 2, se usaron los genes 48, 21, 68, 16, 96, 10, 1, 69, 36, 20, 3, 14, 59, 53, 12, 84, 90, 17, 9, 65, 4, 32, 75, 81, 88, 37, 38, 5, 94, 60, 64, 45, 7, 43, y 55. En el conjunto 3, se usaron los genes 33, 95, 59, 86, 83, 76, 36, 55, 90, 22, 62, 98, 34, 46, 4, 87, 5, 66, 38, 78, 97, 100, 71, 25, 30, 2, 21, 99, 12, 54, 9, 81, 32, y 52. En el conjunto 4, se usaron los genes 27, 64, 40, 59, 63, 100, 50, 19, 1, 10, 96, 2, 34, 28, 67, 26, 87, 41, 57, 33, 11, 94, 66, 82, 6, 52, 55, 84, 47, 97, 83, 80, 62, y 5. En el conjunto 5, se usaron los genes 99, 86, 92, 72, 83, 48, 79, 46, 91, 2, 90, 9, 23, 44, 85, 31, 38, 81, 76, 54, 71, 14, 3, 13, 62, 11, 39, 4, 95, 36, 20, 30, 75, 63, y 51. En el conjunto 6, se usaron los genes 41, 89, 81, 29, 86, 95, 34, 42, 50, 9, 45, 21, 64, 84, 74, 91, 69, 98, 57, 79, 39, 87, 93, 63, 26, 82, 2, 59, 30, 71, 83, 38, 77, 24, y 73. En el conjunto 7, se usaron los genes 87, 60, 59, 98, 43, 38, 28, 64, 29, 92, 22, 27, 40, 33, 69, 71, 73, 79, 15, 70, 32, 90, 76, 93, 6, 50, 55, 9, 49, 54, 36, 5, 48, 19, y 10. En el conjunto 8, se usaron los genes 100, 70, 98, 79, 91, 23, 37, 29, 73, 65, 78, 31, 3, 11, 30, 51, 16, 40, 95, 94, 62, 38, 67, 39, 82, 72, 22, 5, 87, 57, 6, 75, 35, 99, y 46. En el conjunto 9, se usaron los genes 46, 61, 59, 86, 29, 74, 56, 89, 52, 26, 54, 20, 84, 97, 33, 71, 14, 36, 38, 49, 28, 60, 19, 90, 11, 42, 87, 92, 82, 21, 94, 3, 22, 2, y 39. En el conjunto 10, se usaron los genes 31, 76, 77, 27, 72, 38, 42, 36, 53, 82, 61, 39, 98, 81, 34, 80, 22, 100, 8, 32, 17, 21, 28, 56, 59, 29, 55, 5, 62, 40, 90, 87, 24, 68, y 37.

Para 40 genes, en el conjunto 1, se usaron los genes 64, 50, 46, 22, 51, 6, 47, 12, 2, 30, 45, 7, 63, 55, 91, 90, 80, 49, 71, 8, 79, 82, 77, 76, 97, 5, 95, 11, 32, 70, 20, 62, 38, 26, 41, 58, 44, 87, 35, y 23. En el conjunto 2, se usaron los genes 44, 26, 16, 12, 30, 45, 71, 90, 37, 68, 32, 70, 58, 43, 51, 6, 62, 92, 87, 20, 56, 5, 47, 48, 86, 29, 98, 22, 59, 76, 8, 79, 64, 14, 50, 3, 54, 83, 96 y 80. En el conjunto 3, se usaron los genes 20, 34, 57, 70, 39, 15, 25, 33, 78, 51, 87, 46, 67, 80, 28, 52, 66, 72, 22, 88, 96, 3, 90, 6, 82, 42, 41, 94, 85, 61, 54, 84, 14, 9, 81, 19, 7, 91, 23, y 40. En el conjunto 4, se usaron los genes 61, 46, 64, 71, 35, 58, 100, 23, 95, 17, 87, 68, 54, 8, 50, 4, 27, 49, 47, 52, 53, 28, 24, 34, 45, 2, 89, 48, 3, 65, 42, 9, 92, 36, 6, 84, 51, 60, 77, y 94. En el conjunto 5, se usaron los genes 28, 97, 21, 43, 22, 89, 94, 87, 99, 5, 4, 20, 13, 61, 37, 42, 72, 62, 7, 12, 31, 23, 60, 98, 48, 38, 53, 56, 29, 69, 26, 82, 24, 74, 86, 10, 67, 2, 47, y 46. En el conjunto 6, se usaron los genes 12, 74, 96, 77, 78, 72, 53, 87, 47, 29, 40, 98, 52, 22, 69, 3, 58, 97, 60, 48, 55, 80, 57, 39, 50, 89, 71, 9, 63, 51, 21, 23, 73, 32, 20, 19, 25, 5, 38, y 46. En el conjunto 7, se usaron los genes 88, 79, 54, 44, 37, 36, 32, 91, 47, 50, 60, 92, 82, 80, 46, 19, 98, 20, 76, 29, 9, 95, 2, 77, 97, 74, 90, 73, 100, 1, 34, 85, 24, 71, 57, 99, 68, 13, 43, y 53. En el conjunto 8, se usaron los genes 23, 39, 7, 64, 20, 27, 69, 43, 38, 89, 50, 3, 16, 79, 83, 72, 65, 66, 32, 30, 100, 82, 28, 22, 54, 84, 53, 75, 59, 37, 34, 49, 12, 86, 71, 97, 26, 88, 70, y 57. En el conjunto 9, se usaron los genes 74, 96, 80, 39, 40, 82, 38, 56, 35, 93, 55, 73, 44, 17, 81, 27, 2, 83, 65, 89, 76, 8, 18, 45, 58, 77, 14, 49, 21, 6, 4, 92, 33, 13, 12, 88, 98, 24, 84, y 36. En el conjunto 10, se usaron los genes 35, 77, 48, 62, 26, 12, 41, 68, 81, 5, 37, 70, 28, 72, 50, 83, 64, 99, 74, 57, 84, 76, 52, 14, 87, 97, 3, 31, 73, 58, 44, 24, 15, 66, 45, 91, 4, 32, 46, y 49.

Para 45 genes, en el conjunto 1, se usaron los genes 52, 97, 84, 72, 96, 34, 18, 38, 88, 80, 91, 49, 71, 64, 93, 26, 62, 40, 68, 29, 67, 39, 60, 9, 13, 74, 95, 99, 27, 47, 25, 45, 31, 8, 69, 17, 75, 53, 51, 12, 23, 1, 6, 30, y 50. En el conjunto 2, se usaron los genes 97, 80, 55, 32, 94, 84, 28, 3, 6, 48, 17, 41, 65, 37, 79, 34, 61, 83, 35, 49, 27, 38, 43, 2, 24, 77, 25, 71, 58, 14, 8, 30, 46, 98, 82, 75, 22, 72, 26, 74, 93, 66, 73, 1, y 53. En el conjunto 3, se usaron los genes 64, 45, 38, 92, 23, 74, 66, 60, 100, 3, 82, 20, 54, 11, 19, 16, 80, 86, 14, 75, 62, 10, 52, 47, 13, 31, 35, 53, 41, 9, 79, 39, 17, 22, 99, 58, 46, 83, 43, 40, 44, 90, 95, 12, y 81. En el conjunto 4, se usaron los genes 20, 66, 9, 24, 16, 76, 99, 42, 86, 58, 15, 93, 48, 28, 26, 50, 68, 12, 2, 37, 82, 36, 27, 57, 45, 41, 32, 1, 52, 54, 30, 39, 7, 100, 59, 23, 94, 75, 8, 60, 55, 34, 38, 29, y 87. En el conjunto 5, se usaron los genes 66, 88, 73, 53, 51, 69, 36,

87, 78, 40, 58, 76, 31, 65, 56, 42, 100, 68, 5, 18, 17, 91, 45, 22, 74, 82, 1, 44, 67, 43, 10, 63, 79, 92, 6, 72, 80, 75, 9, 30, 19, 61, 99, 3, y 38. En el conjunto 6, se usaron los genes 75, 66, 84, 59, 9, 70, 100, 27, 79, 41, 73, 67, 23, 39, 28, 68, 21, 69, 38, 72, 86, 82, 36, 46, 77, 34, 47, 54, 13, 16, 7, 88, 22, 26, 4, 89, 55, 24, 61, 12, 35, 50, 95, 92, y 80. En el conjunto 7, se usaron los genes 59, 86, 10, 29, 53, 88, 43, 64, 11, 13, 19, 17, 36, 65, 73, 94, 20, 51, 80, 24, 66, 83, 44, 47, 21, 6, 52, 82, 69, 54, 100, 28, 18, 34, 35, 30, 74, 91, 49, 46, 60, 5, 38, 71, y 2. En el conjunto 8, se usaron los genes 77, 32, 55, 44, 6, 98, 94, 19, 10, 71, 72, 85, 67, 75, 78, 88, 90, 58, 89, 27, 69, 42, 31, 47, 1, 37, 52, 7, 57, 45, 11, 83, 49, 46, 34, 64, 14, 24, 87, 9, 56, 8, 20, 36, y 15. En el conjunto 9, se usaron los genes 4, 27, 83, 61, 46, 15, 35, 26, 51, 54, 23, 38, 100, 7, 42, 58, 44, 8, 22, 37, 20, 89, 56, 91, 70, 29, 11, 19, 87, 99, 21, 65, 72, 75, 49, 40, 45, 30, 43, 48, 63, 3, 18, 74, y 1. En el conjunto 10, se usaron los genes 68, 19, 90, 52, 55, 23, 17, 53, 3, 2, 74, 82, 26, 88, 48, 6, 8, 43, 15, 73, 57, 67, 85, 91, 13, 44, 81, 1, 75, 33, 51, 21, 4, 41, 77, 86, 40, 18, 31, 78, 92, 10, 64, 99, y 69.

Se realizó la clasificación de subconjuntos de los 39 tipos de tumores con el uso de selecciones aleatorias de tipos de tumor del grupo de 39. Los niveles de expresión de los conjuntos de secuencias genéticas como se describe en el presente documento se usaron para clasificar combinaciones aleatorias de tipos de tumor. Se usaron diferentes conjuntos aleatorios de tipos de tumor con cada uno de los conjuntos de 100, 74, y 90 secuencias genéticas como se describe en estos ejemplos. Algunos ejemplos representativos, y no limitantes de conjuntos aleatorios de 2 a 20 tipos de tumor usa dos como sigue a continuación, en los que se hizo un índice del conjunto de 1 a 39 de los 39 tipos de tumor.

Para 2 tipos de tumor, el conjunto 1 usó los tipos 26 y 16. El conjunto 2 usó los tipos 8 y 5. El conjunto 3 usó los tipos 39 y 8. El conjunto 4 usó los tipos 27 y 23. El conjunto 5 usó los tipos 8 y 19. El conjunto 6 usó los tipos 12 y 21. El conjunto 7 usó los tipos 30 y 15. El conjunto 8 usó los tipos 30 y 5. El conjunto 9 usó los tipos 18 y 22. El conjunto 10 usó los tipos 27 y 26.

Para 4 tipos de tumor, el conjunto 1 usó los tipos 20, 35, 15 y 7. El conjunto 2 usó los tipos 36, 1, 28 y 19. El conjunto 3 usó los tipos 13, 4, 12 y 21. El conjunto 4 usó los tipos 12, 33, 14 y 28. El conjunto 5 usó los tipos 6, 28, 5 y 37. El conjunto 6 usó los tipos 5, 25, 36 y 15. El conjunto 7 usó los tipos 12, 26, 21 y 19. El conjunto 8 usó los tipos 19, 3, 20 y 17. El conjunto 9 usó los tipos 18, 10, 8 y 9. El conjunto 10 usó los tipos 28, 20, 2 y 22.

Para 6 tipos de tumor, el conjunto 1 usó los tipos 27, 3, 10, 39, 11 y 20. El conjunto 2 usó los tipos 33, 10, 20, 32, 13 y 19. El conjunto 3 usó los tipos 31, 27, 18, 39, 8 y 16. El conjunto 4 usó los tipos 25, 28, 10, 12, 7 y 39. El conjunto 5 usó los tipos 14, 13, 28, 24, 30 y 36. El conjunto 6 usó los tipos 9, 24, 8, 17, 36 y 26. El conjunto 7 usó los tipos 20, 1, 34, 26, 6 y 19. El conjunto 8 usó los tipos 12, 13, 3, 17, 34 y 22. El conjunto 9 usó los tipos 7, 1, 17, 13, 20 y 34. El conjunto 10 usó los tipos 5, 11, 25, 29, 28 y 35.

Para 8 tipos de tumor, el conjunto 1 usó los tipos 34, 33, 28, 3, 23, 25, 9 y 29. El conjunto 2 usó los tipos 27, 8, 38, 28, 20, 14, 12 y 9. El conjunto 3 usó los tipos 29, 21, 19, 1, 13, 26, 11 y 31. El conjunto 4 usó los tipos 25, 17, 7, 20, 34, 8, 28 y 10. El conjunto 5 usó los tipos 36, 28, 35, 26, 2, 8, 29 y 7. El conjunto 6 usó los tipos 10, 23, 2, 27, 33, 21, 25 y 35. El conjunto 7 usó los tipos 10, 18, 38, 2, 6, 7, 19 y 32. El conjunto 8 usó los tipos 11, 37, 6, 28, 3, 9, 2 y 16. El conjunto 9 usó los tipos 22, 2, 10, 8, 17, 19 y 33. El conjunto 10 usó los tipos 35, 39, 8, 10, 37, 4, 36 y 6.

Para 10 tipos de tumor, el conjunto 1 usó los tipos 25, 10, 26, 2, 32, 31, 39, 23, 22 y 18. El conjunto 2 usó los tipos 12, 35, 6, 16, 20, 3, 39, 36, 11 y 2. El conjunto 3 usó los tipos 34, 1, 15, 29, 5, 39, 2, 12, 25 y 18. El conjunto 4 usó los tipos 10, 8, 14, 18, 31, 19, 23, 20, 32 y 33. El conjunto 5 usó los tipos 10, 18, 37, 15, 4, 35, 33, 24, 39 y 20. El conjunto 6 usó los tipos 22, 16, 4, 3, 18, 21, 1, 25, 37 y 13. El conjunto 7 usó los tipos 14, 6, 28, 18, 11, 13, 2, 32, 33 y 19. El conjunto 8 usó los tipos 39, 2, 38, 4, 34, 8, 25, 6, 32 y 35. El conjunto 9 usó los tipos 3, 10, 11, 16, 6, 15, 18, 14, 12 y 26. El conjunto 10 usó los tipos 24, 25, 21, 9, 36, 29, 20, 39, 10 y 37.

Para 12 tipos de tumor, el conjunto 1 usó los tipos 26, 20, 4, 12, 2, 31, 38, 18, 16, 39, 3 y 33. El conjunto 2 usó los tipos 25, 16, 4, 9, 29, 27, 14, 24, 21, 7, 23 y 2. El conjunto 3 usó los tipos 31, 18, 23, 13, 25, 1, 29, 21, 35, 10, 32 y 39. El conjunto 4 usó los tipos 8, 34, 23, 9, 35, 14, 25, 21, 2, 33, 18 y 28. El conjunto 5 usó los tipos 6, 11, 21, 8, 5, 7, 19, 32, 3, 13, 36 y 9. El conjunto 6 usó los tipos 12, 33, 14, 26, 27, 15, 2, 21, 36, 35, 9 y 39. El conjunto 7 usó los tipos 26, 29, 32, 17, 31, 19, 6, 5, 20, 34, 2 y 24. El conjunto 8 usó los tipos 17, 12, 8, 22, 28, 9, 27, 29, 14, 35, 4 y 32. El conjunto 9 usó los tipos 29, 9, 36, 23, 33, 18, 21, 35, 3, 6, 2 y 1. El conjunto 10 usó los tipos 1, 3, 35, 29, 22, 27, 8, 23, 2, 36, 14 y 19.

Para 14 tipos de tumor, el conjunto 1 usó los tipos 9, 26, 38, 25, 31, 3, 15, 14, 17, 33, 12, 35, 39 y 16. El conjunto 2 usó los tipos 1, 26, 16, 25, 20, 12, 14, 37, 38, 24, 23, 33, 27 y 35. El conjunto 3 usó los tipos 11, 21, 35, 38, 32, 34, 27, 39, 16, 15, 4, 5, 13 y 18. El conjunto 4 usó los tipos 27, 5, 13, 28, 18, 17, 15, 20, 29, 37, 21, 36, 25 y 14. El conjunto 5 usó los tipos 5, 12, 17, 9, 25, 21, 33, 37, 8, 15, 24, 3, 34 y 28. El conjunto 6 usó los tipos 11, 19, 34, 26, 9, 6, 32, 14, 27, 29, 30, 16, 24 y 17. El conjunto 7 usó los tipos 31, 26, 11, 18, 19, 20, 9, 8, 5, 36, 12, 6, 27 y 38. El conjunto 8 usó los tipos 20, 17, 11, 5, 15, 9, 2, 39, 34, 24, 27, 26, 35 y 10. El conjunto 9 usó los tipos 1, 14, 39, 30, 17, 6, 10, 35, 31, 33, 15, 29, 32 y 7. El conjunto 10 usó los tipos 1, 19, 24, 28, 34, 12, 13, 18, 32, 11, 14, 21, 22 y 25.

Para 16 tipos de tumor, el conjunto 1 usó los tipos 27, 15, 8, 12, 6, 20, 26, 19, 25, 2, 37, 38, 7, 39, 4 y 33. El conjunto

2 usó los tipos 17, 18, 28, 5, 6, 31, 25, 13, 8, 20, 37, 36, 35, 9, 23 y 27. El conjunto 3 usó los tipos 23, 37, 34, 14, 16, 27, 32, 33, 21, 38, 4, 30, 24, 22, 17 y 25. El conjunto 4 usó los tipos 7, 37, 38, 21, 34, 31, 32, 25, 10, 36, 19, 11, 6, 26, 18 y 35. El conjunto 5 usó los tipos 9, 32, 12, 24, 20, 13, 38, 21, 39, 23, 36, 18, 37, 22, 5 y 3. El conjunto 6 usó los tipos 14, 21, 5, 17, 6, 20, 18, 35, 22, 10, 3, 23, 13, 2, 34 y 26. El conjunto 7 usó los tipos 1, 8, 19, 6, 9, 39, 28, 18, 13, 31, 14, 16, 37, 12, 3 y 25. El conjunto 8 usó los tipos 32, 36, 28, 38, 9, 33, 2, 5, 4, 11, 19, 18, 13, 8, 12 y 3. El conjunto 9 usó los tipos 9, 14, 10, 5, 28, 32, 23, 6, 39, 3, 17, 8, 19, 1, 31 and 12. El conjunto 10 usó los tipos 4, 34, 11, 6, 38, 19, 7, 20, 23, 3, 25, 37, 26, 1, 15 y 12.

Para 18 tipos de tumor, el conjunto 1 usó los tipos 15, 24, 39, 35, 7, 30, 16, 13, 20, 3, 26, 4, 12, 10, 34, 25, 21 y 28. El conjunto 2 usó los tipos 21, 23, 29, 11, 10, 19, 13, 28, 4, 20, 17, 24, 30, 12, 39, 34, 31 y 9. El conjunto 3 usó los tipos 7, 17, 27, 6, 30, 8, 22, 2, 32, 26, 21, 14, 4, 38, 1, 35, 16 y 28. El conjunto 4 usó los tipos 17, 13, 20, 33, 10, 3, 16, 22, 1, 38, 2, 9, 28, 5, 6, 19, 12 and 11. El conjunto 5 usó los tipos 4, 35, 21, 25, 18, 17, 8, 14, 31, 30, 9, 1, 2, 23, 36, 29, 32 y 37. El conjunto 6 usó los tipos 17, 34, 2, 18, 19, 15, 16, 13, 4, 24, 5, 35, 6, 22, 28, 37, 38 y 1. El conjunto 7 usó los tipos 34, 26, 12, 25, 27, 3, 17, 7, 2, 32, 9, 36, 21, 18, 22, 8, 20 y 29. El conjunto 8 usó los tipos 12, 34, 38, 25, 17, 22, 14, 39, 10, 7, 31, 2, 3, 11, 29, 30, 16 y 24. El conjunto 9 usó los tipos 13, 26, 27, 14, 5, 10, 8, 7, 16, 30, 37, 4, 6, 35, 28, 1, 36 y 20. El conjunto 10 usó los tipos 15, 2, 17, 23, 26, 28, 36, 38, 12, 6, 19, 37, 20, 14, 9, 39, 11 y 21.

Para 20 tipos de tumor, el conjunto 1 usó los tipos 25, 13, 21, 15, 37, 20, 12, 28, 9, 10, 26, 22, 14, 24, 16, 7, 39, 34, 33 y 4. El conjunto 2 usó los tipos 20, 17, 10, 27, 19, 28, 5, 1, 23, 21, 38, 7, 13, 22, 32, 31, 9, 4, 3 y 24. El conjunto 3 usó los tipos 17, 13, 7, 20, 11, 38, 34, 3, 15, 12, 5, 39, 9, 10, 4, 35, 27, 6, 21 y 33. El conjunto 4 usó los tipos 6, 13, 17, 26, 1, 7, 33, 5, 10, 32, 3, 23, 35, 4, 14, 28, 12, 38, 8 y 27. El conjunto 5 usó los tipos 10, 23, 9, 38, 5, 29, 12, 27, 25, 6, 7, 26, 37, 31, 24, 36, 19, 15, 16 y 11. El conjunto 6 usó los tipos 30, 24, 21, 11, 23, 25, 8, 9, 7, 31, 27, 5, 14, 29, 1, 19, 16, 12, 22 y 17. El conjunto 7 usó los tipos 26, 13, 23, 19, 22, 11, 25, 21, 33, 20, 6, 17, 2, 10, 31, 34, 27, 37, 7 y 9. El conjunto 8 usó los tipos 30, 1, 38, 7, 31, 37, 11, 25, 6, 19, 28, 33, 7, 29, 10, 27, 16, 3, 14 y 15. El conjunto 9 usó los tipos 15, 19, 26, 24, 5, 33, 11, 2, 13, 18, 31, 22, 32, 20, 23, 6, 10, 25, 36 y 3. El conjunto 10 usó los tipos 24, 25, 21, 29, 14, 18, 31, 2, 20, 39, 23, 9, 38, 12, 6, 32, 22, 26, 33 y 7.

Ejemplo 4: Conjuntos de genes especificados

Un primer conjunto de 74 genes y un segundo conjunto de 90 genes, en los que los dos conjuntos tienen 38 miembros en común, se usaron en la práctica de la invención.

Subconjuntos aleatorios de aproximadamente 5 a 49 miembros del conjunto de 74 secuencias genéticas expresadas se evaluaron de manera análoga a la que se ha descrito en el Ejemplo 3. De nuevo, los niveles de expresión de combinaciones aleatorias de 5, 10, 15, 20, 25, 30, 35, 40, 45, y 49 (se tomara muestra de cada combinación 10 veces) de las 74 secuencias expresadas se usaron con datos de los tipos de tumor y a continuación se usaron para predecir los conjuntos de muestras de tumor aleatorios del ensayo (se toma muestra de cada una 10 veces) que varía de 2 a los 39 tipos. Los datos resultantes se muestran en las Figuras 4-6.

Se hizo un índice de los miembros de las 74 secuencias genéticas de 1 a 74, y los conjuntos aleatorios representativos usados en la invención son los que siguen a continuación:

Para 2 genes, en el conjunto 1, se usaron los genes 64 y 6. Para el conjunto 2, se usaron los genes 64 y 13. Para el conjunto 3, se usaron los genes 67 y 51. Para el conjunto 4, se usaron los genes 51 y 29. Para el conjunto 5, se usaron los genes 46 y 12. Para el conjunto 6, se usaron los genes 68 y 65. Para el conjunto 7, se usaron los genes 6 y 28. Para el conjunto 8, se usaron los genes 9 y 55. Para el conjunto 9, se usaron los genes 55 y 71. Para el conjunto 10, se usaron los genes 63 y 39.

Para 5 genes, en el conjunto 1, se usaron los genes 8, 64, 50, 54, y 4. En el conjunto 2, se usaron los genes 39, 17, 45, 34, y 15. En el conjunto 3, se usaron los genes 10, 4, 61, 21, y 55. En el conjunto 4, se usaron los genes 59, 37, 21, 23, y 64. En el conjunto 5, se usaron los genes 69, 8, 25, 59, y 63. En el conjunto 6, se usaron los genes 45, 71, 19, 59, y 38. En el conjunto 7, se usaron los genes 21, 43, 14, 48, y 30. En el conjunto 8, se usaron los genes 73, 35, 36, 10, y 9. En el conjunto 9, se usaron los genes 62, 28, 11, 70, y 64. En el conjunto 10, se usaron los genes 8, 16, 70, 18, y 59.

Para 10 genes, en el conjunto 1, se usaron los genes 49, 72, 38, 68, 52, 21, 1, 10, 2, y 40. En el conjunto 2, se usaron los genes 54, 70, 28, 64, 68, 41, 44, 20, 7, y 2. En el conjunto 3, se usaron los genes 71, 49, 51, 11, 18, 53, 8, 42, 36, y 58. En el conjunto 4, se usaron los genes 72, 15, 35, 3, 23, 8, 2, 48, 22, y 65. En el conjunto 5, se usaron los genes 44, 19, 6, 22, 38, 5, 37, 9, 30, y 14. En el conjunto 6, se usaron los genes 15, 27, 3, 10, 31, 19, 44, 39, 48, y 46. En el conjunto 7, se usaron los genes 70, 30, 9, 33, 63, 71, 32, 34, 20, y 7. En el conjunto 8, se usaron los genes 45, 29, 54, 58, 15, 21, 68, 5, 42, y 62. En el conjunto 9, se usaron los genes 74, 17, 66, 46, 10, 8, 63, 5, 24, y 2. En el conjunto 10, se usaron los genes 33, 2, 34, 19, 60, 71, 42, 51, 70, y 66.

Para 15 genes, en el conjunto 1, se usaron los genes 13, 22, 26, 67, 64, 40, 68, 71, 4, 28, 24, 33, 46, 69, y 41. En el conjunto 2, se usaron los genes 10, 1, 14, 70, 71, 64, 46, 67, 45, 48, 65, 74, 34, 49, y 37. En el conjunto 3, se usaron los genes 58, 30, 44, 40, 51, 36, 33, 60, 39, 21, 54, 64, 25, 13, y 35. En el conjunto 4, se usaron los genes 63, 70, 60, 32, 31, 16, 49, 65, 38, 5, 72, 47, 40, 2, y 46. En el conjunto 5, se usaron los genes 43, 6, 40, 13, 39, 72, 68, 41, 27, 73, 36, 25, 33, 34, y 1. En el conjunto 6, se usaron los genes 68, 67, 71, 59, 73, 62, 31, 43, 7, 44,

21, 48, 54, 58, y 6. En el conjunto 7, se usaron los genes 16, 50, 61, 62, 27, 2, 21, 1, 41, 28, 68, 35, 17, 47, y 46. En el conjunto 8, se usaron los genes 27, 18, 44, 66, 2, 20, 53, 64, 46, 70, 57, 7, 51, 10, y 45. En el conjunto 9, se usaron los genes 65, 8, 43, 23, 50, 46, 21, 41, 44, 3, 31, 17, 7, 66, y 70. En el conjunto 10, se usaron los genes 16, 14, 61, 51, 39, 33, 43, 31, 53, 65, 74, 42, 29, 9, y 11.

5 Para 20 genes, en el conjunto 1, se usaron los genes 14, 60, 6, 71, 74, 16, 62, 39, 56, 44, 32, 72, 18, 42, 66, 49, 1, 9, 69, y 21. En el conjunto 2, se usaron los genes 23, 1, 7, 27, 26, 71, 12, 4, 22, 69, 62, 44, 6, 25, 57, 28, 33, 9, 21, y 51. En el conjunto 3, se usaron los genes 46, 48, 29, 54, 55, 69, 73, 47, 6, 27, 24, 21, 15, 43, 45, 7, 62, 25, 22, y 74. En el conjunto 4, se usaron los genes 12, 65, 24, 73, 45, 57, 49, 63, 61, 1, 58, 10, 2, 18, 8, 51, 67, 69, 59, y 13. En el conjunto 5, se usaron los genes 33, 43, 9, 52, 54, 38, 8, 16, 48, 1, 39, 60, 17, 6, 15, 66, 68, 63, 37, y 42. En el conjunto 6, se usaron los genes 43, 19, 44, 28, 56, 34, 66, 42, 73, 40, 65, 38, 54, 20, 51, 37, 30, 35, 53, y 61. En el conjunto 7, se usaron los genes 61, 6, 20, 4, 34, 53, 70, 38, 35, 46, 36, 16, 1, 23, 68, 12, 59, 71, 65, y 14. En el conjunto 8, se usaron los genes 25, 68, 69, 3, 33, 49, 19, 56, 54, 4, 32, 6, 45, 16, 67, 52, 65, 14, 12, y 40. En el conjunto 9, se usaron los genes 47, 7, 36, 32, 61, 74, 14, 45, 26, 51, 69, 12, 41, 42, 64, 25, 27, 57, 23, y 58. En el conjunto 10, se usaron los genes 27, 13, 3, 17, 51, 7, 37, 43, 20, 12, 52, 21, 25, 2, 5, 32, 62, 47, 4, y 26.

15 Para 25 genes, en el conjunto 1, se usaron los genes 57, 61, 31, 38, 3, 7, 72, 43, 32, 23, 28, 71, 48, 17, 2, 49, 10, 30, 66, 12, 69, 41, 20, 63, y 68. En el conjunto 2, se usaron los genes 18, 54, 47, 57, 24, 42, 66, 46, 16, 58, 37, 60, 62, 9, 2, 27, 36, 52, 13, 32, 45, 6, 43, 21, y 56. En el conjunto 3, se usaron los genes 47, 48, 52, 16, 56, 54, 42, 37, 17, 41, 35, 21, 6, 9, 63, 10, 49, 68, 23, 25, 70, 3, 58, 2, y 31. En el conjunto 4, se usaron los genes 50, 10, 25, 16, 68, 15, 29, 73, 27, 63, 3, 17, 28, 66, 19, 13, 4, 9, 36, 48, 23, 57, 59, 26, y 14. En el conjunto 5, se usaron los genes 40, 39, 43, 49, 66, 15, 14, 29, 36, 21, 19, 44, 72, 58, 69, 12, 11, 9, 37, 46, 32, 51, 3, 24, y 6. En el conjunto 6, se usaron los genes 42, 49, 44, 32, 46, 35, 70, 40, 3, 21, 11, 67, 25, 56, 37, 43, 60, 55, 16, 27, 30, 53, 63, 23, y 33. En el conjunto 7, se usaron los genes 70, 27, 68, 17, 64, 65, 18, 69, 10, 67, 42, 23, 48, 14, 31, 11, 55, 25, 52, 34, 13, 45, 12, 29, y 47. En el conjunto 8, se usaron los genes 48, 10, 17, 27, 25, 55, 12, 62, 30, 65, 15, 49, 70, 14, 54, 24, 33, 26, 50, 60, 6, 40, 67, 11, y 2. En el conjunto 9, se usaron los genes 41, 47, 24, 59, 7, 44, 2, 67, 12, 19, 13, 17, 35, 56, 28, 14, 61, 15, 60, 58, 1, 64, 31, 45, y 23. En el conjunto 10, se usaron los genes 42, 72, 41, 38, 57, 27, 4, 13, 9, 43, 34, 28, 8, 62, 64, 46, 12, 70, 21, 66, 16, 7, 48, 3, y 54.

20 Para 30 genes, en el conjunto 1, se usaron los genes 16, 47, 67, 9, 22, 10, 64, 72, 46, 6, 60, 74, 3, 68, 57, 63, 14, 54, 58, 30, 28, 18, 70, 73, 52, 39, 34, 61, 12, 21. En el conjunto 2, se usaron los genes 18, 1, 44, 24, 68, 26, 62, 10, 47, 67, 37, 55, 32, 35, 34, 14, 49, 30, 17, 16, 51, 45, 74, 31, 9, 57, 66, 39, 53, y 8. En el conjunto 3, se usaron los genes 58, 45, 55, 39, 22, 32, 9, 49, 31, 13, 51, 56, 28, 12, 3, 59, 74, 35, 42, 67, 69, 47, 66, 18, 52, 57, 43, 5, 26, y 4. En el conjunto 4, se usaron los genes 45, 1, 74, 12, 17, 23, 59, 27, 38, 40, 72, 56, 50, 20, 52, 32, 5, 16, 9, 21, 60, 64, 49, 70, 30, 61, 6, 10, 31, y 24. En el conjunto 5, se usaron los genes 60, 53, 7, 32, 73, 25, 69, 48, 17, 45, 16, 3, 14, 9, 37, 41, 72, 43, 68, 39, 20, 51, 59, 23, 6, 15, 74, 19, 31, y 66. En el conjunto 6, se usaron los genes 47, 54, 9, 38, 60, 33, 40, 12, 57, 45, 26, 56, 11, 27, 67, 25, 69, 59, 68, 7, 61, 72, 23, 21, 28, 48, 29, 65, 37, y 15. En el conjunto 7, se usaron los genes 21, 42, 30, 57, 65, 59, 53, 74, 45, 66, 68, 41, 19, 24, 8, 10, 61, 43, 38, 67, 37, 47, 40, 22, 63, 35, 70, 72, 5, y 6. En el conjunto 8, se usaron los genes 58, 11, 28, 36, 24, 34, 53, 9, 44, 23, 51, 70, 22, 17, 15, 59, 5, 60, 1, 64, 21, 50, 35, 52, 31, 43, 38, 39, 32, y 62. En el conjunto 9, se usaron los genes 43, 30, 63, 7, 60, 40, 39, 1, 48, 17, 69, 57, 6, 62, 19, 38, 36, 13, 66, 64, 25, 31, 65, 47, 27, 16, 53, 68, 37, y 41. En el conjunto 10, se usaron los genes 22, 17, 4, 2, 37, 16, 49, 7, 63, 64, 14, 15, 74, 43, 25, 54, 46, 50, 53, 67, 39, 62, 59, 10, 55, 72, 65, 52, 58, y 19.

30 Para 35 genes, en el conjunto 1, se usaron los genes 4, 43, 55, 49, 13, 26, 32, 21, 18, 50, 14, 20, 65, 7, 24, 52, 58, 8, 30, 37, 54, 71, 2, 31, 44, 61, 66, 67, 28, 39, 10, 70, 17, 19, y 45. En el conjunto 2, se usaron los genes 14, 13, 67, 21, 48, 28, 69, 47, 50, 3, 68, 63, 22, 41, 60, 61, 5, 44, 56, 65, 7, 66, 15, 6, 45, 2, 36, 5, 30, 72, 34, 46, 24, 29, y 12. En el conjunto 3, se usaron los genes 67, 25, 58, 11, 17, 16, 3, 69, 21, 1, 59, 26, 72, 41, 47, 2, 34, 24, 10, 19, 33, 5, 50, 9, 71, 20, 62, 8, 68, 61, 23, 37, 35, 60, y 32. En el conjunto 4, se usaron los genes 5, 30, 14, 1, 59, 27, 28, 51, 55, 61, 18, 37, 17, 73, 6, 44, 67, 12, 35, 11, 53, 72, 70, 25, 21, 7, 34, 13, 74, 43, 52, 39, 54, 2, y 19. En el conjunto 5, se usaron los genes 56, 64, 58, 35, 1, 23, 43, 4, 73, 28, 54, 6, 51, 68, 49, 37, 16, 71, 3, 21, 48, 69, 70, 10, 26, 22, 50, 44, 2, 60, 38, 40, 66, 63, y 65. En el conjunto 6, se usaron los genes 72, 49, 51, 44, 19, 28, 1, 11, 3, 40, 33, 41, 70, 29, 48, 62, 50, 4, 47, 60, 68, 10, 61, 32, 20, 13, 22, 59, 65, 64, 67, 21, 35, 39, y 24. En el conjunto 7, se usaron los genes 14, 35, 31, 20, 8, 59, 50, 15, 52, 62, 19, 30, 71, 68, 72, 47, 38, 74, 36, 49, 73, 22, 41, 25, 69, 16, 32, 24, 51, 43, 65, 3, 6, 53, y 29. En el conjunto 8, se usaron los genes 22, 44, 23, 9, 26, 56, 72, 59, 35, 61, 51, 69, 64, 30, 53, 27, 11, 55, 39, 67, 48, 28, 14, 10, 8, 12, 40, 24, 57, 34, 50, 32, 42, 41, y 38. En el conjunto 9, se usaron los genes 15, 7, 27, 6, 67, 9, 26, 57, 30, 37, 58, 23, 42, 11, 36, 52, 32, 29, 62, 21, 41, 61, 64, 18, 40, 35, 66, 1, 2, 56, 16, 3, 55, 10, y 51. En el conjunto 10, se usaron los genes 9, 14, 71, 25, 44, 37, 49, 46, 66, 53, 7, 33, 22, 12, 73, 50, 27, 24, 13, 5, 41, 51, 61, 16, 28, 56, 23, 20, 10, 8, 70, 48, 42, 52, y 34.

55 Para 40 genes, en el conjunto 1, se usaron los genes 26, 36, 43, 30, 62, 19, 20, 51, 41, 71, 1, 63, 10, 56, 65, 17, 15, 50, 5, 35, 4, 54, 12, 70, 48, 31, 47, 37, 34, 8, 3, 69, 40, 44, 46, 59, 61, 74, 23, 27. En el conjunto 2, se usaron los genes 1, 4, 38, 24, 37, 69, 21, 52, 13, 2, 63, 51, 30, 16, 27, 58, 74, 20, 32, 53, 59, 31, 50, 10, 42, 8, 54, 36, 5, 47, 70, 41, 12, 46, 28, 19, 35, 9, 61, y 48. En el conjunto 3, se usaron los genes 35, 48, 40, 47, 20, 67, 57, 72, 15, 17, 46, 37, 9, 2, 60, 30, 65, 49, 29, 64, 16, 21, 7, 74, 61, 11, 58, 71, 62, 23, 24, 55, 3, 53, 52, 27, 18, 50, 25, y 66. En el conjunto 4, se usaron los genes 35, 10, 59, 19, 27, 40, 30, 4, 9, 52, 2, 29, 26, 41, 55, 17, 13, 53, 71, 63, 58, 44, 45, 62, 70, 16, 64, 48, 43, 8, 38, 72, 49, 37, 18, 36, 74, 42, 46, y 54. En el conjunto 5, se usaron los genes 16, 61, 1, 10, 20, 51, 22, 6, 43, 65, 66, 24, 30, 9, 14, 40, 32, 74, 18, 71, 15, 28, 52, 31, 56, 55, 23, 4, 58, 36, 60, 54, 25, 63, 27, 64, 50, 29, 44, y 45. En el conjunto 6, se usaron los genes 15, 30, 3, 50, 61, 47, 13, 48, 45, 17, 46, 10, 28, 37, 8, 54, 9, 5, 63, 18, 39, 49, 34, 68, 14, 23, 43, 11, 1, 51, 56, 67, 20, 57, 6, 19, 25, 31, 21, y 12. En el

conjunto 7, se usaron los genes 45, 73, 53, 29, 35, 56, 70, 51, 30, 59, 49, 22, 6, 43, 28, 31, 40, 4, 66, 25, 37, 19, 12, 65, 26, 74, 46, 50, 23, 62, 17, 69, 36, 41, 34, 27, 67, 7, 24, y 13. En el conjunto 8, se usaron los genes 62, 30, 38, 41, 18, 13, 49, 71, 68, 47, 50, 70, 66, 5, 23, 33, 27, 56, 6, 7, 34, 28, 26, 58, 53, 46, 16, 52, 72, 42, 10, 54, 67, 64, 12, 8, 19, 57, 73, y 17. En el conjunto 9, se usaron los genes 11, 32, 48, 54, 42, 67, 13, 53, 21, 44, 57, 22, 40, 12, 5, 29, 69, 37, 17, 39, 45, 73, 60, 26, 14, 72, 4, 59, 24, 46, 18, 51, 36, 61, 35, 9, 19, 16, 38, y 28. En el conjunto 10, se usaron los genes 58, 1, 55, 59, 11, 63, 3, 26, 49, 69, 34, 47, 65, 46, 14, 39, 5, 67, 16, 66, 64, 38, 44, 32, 15, 22, 19, 71, 23, 52, 45, 53, 48, 8, 60, 73, 9, 30, 25, y 37.

Para 45 genes, en el conjunto 1, se usaron los genes 26, 21, 17, 34, 19, 27, 6, 61, 24, 42, 3, 60, 70, 43, 54, 13, 9, 20, 28, 58, 12, 23, 33, 4, 63, 56, 67, 1, 11, 68, 41, 59, 45, 5, 48, 32, 10, 44, 16, 65, 51, 62, 22, 38, y 74. En el conjunto 2, se usaron los genes 21, 41, 67, 5, 51, 53, 28, 25, 31, 60, 52, 17, 50, 11, 29, 45, 2, 32, 71, 13, 68, 22, 74, 33, 48, 56, 62, 42, 26, 14, 61, 23, 9, 46, 66, 10, 64, 59, 54, 69, 27, 47, 44, 34, y 40. En el conjunto 3, se usaron los genes 68, 48, 43, 74, 17, 4, 49, 34, 38, 60, 12, 42, 18, 5, 51, 32, 1, 57, 9, 11, 30, 13, 37, 15, 29, 33, 44, 20, 55, 70, 45, 41, 24, 56, 35, 52, 59, 7, 25, 2, 31, 64, 71, 22, y 39. En el conjunto 4, se usaron los genes 44, 61, 51, 69, 65, 72, 29, 57, 40, 62, 66, 63, 67, 55, 74, 14, 56, 11, 16, 58, 1, 15, 3, 48, 42, 7, 8, 30, 18, 19, 23, 60, 4, 10, 21, 43, 12, 37, 32, 25, 22, 50, 34, 59, y 2. En el conjunto 5, se usaron los genes 67, 54, 33, 41, 5, 61, 3, 10, 2, 71, 73, 53, 25, 42, 44, 23, 9, 38, 45, 62, 32, 46, 40, 8, 66, 49, 16, 24, 68, 69, 21, 52, 20, 6, 48, 11, 57, 39, 22, 31, 63, 36, 34, 35, y 17. En el conjunto 6, se usaron los genes 43, 45, 19, 17, 4, 58, 37, 7, 42, 52, 2, 62, 25, 66, 24, 15, 22, 74, 68, 67, 8, 1, 33, 70, 31, 50, 64, 14, 61, 51, 6, 38, 35, 39, 72, 5, 27, 36, 11, 18, 12, 48, 46, 54, y 71. En el conjunto 7, se usaron los genes 41, 45, 58, 11, 66, 26, 53, 13, 60, 4, 65, 18, 67, 73, 28, 55, 56, 57, 29, 68, 23, 19, 42, 17, 22, 62, 61, 10, 43, 64, 38, 71, 7, 40, 16, 34, 74, 12, 37, 8, 63, 44, 49, 47, y 3. En el conjunto 8, se usaron los genes 47, 40, 59, 14, 50, 71, 1, 57, 19, 28, 6, 34, 68, 4, 30, 20, 31, 33, 38, 39, 17, 41, 24, 65, 70, 61, 3, 35, 45, 11, 9, 8, 73, 42, 26, 23, 46, 72, 25, 64, 16, 53, 62, 18, y 7. En el conjunto 9, se usaron los genes 61, 5, 69, 22, 7, 17, 26, 13, 2, 30, 55, 33, 47, 14, 59, 32, 9, 44, 23, 45, 42, 25, 15, 57, 48, 50, 1, 68, 18, 72, 46, 73, 67, 36, 63, 60, 28, 21, 20, 8, 29, 35, 37, 38, y 71. En el conjunto 10, se usaron los genes 22, 31, 58, 50, 64, 11, 17, 67, 41, 2, 21, 4, 61, 70, 54, 3, 71, 25, 40, 43, 69, 38, 9, 73, 45, 16, 34, 10, 7, 52, 35, 19, 66, 24, 5, 60, 18, 14, 59, 32, 68, 15, 56, 63, y 65.

Un experimento similar se realizó con subconjuntos aleatorios de aproximadamente 5 a 49 miembros del conjunto de 90 secuencias genéticas expresadas. De nuevo, los niveles de expresión de combinaciones aleatorias de 5, 10, 15, 20, 25, 30, 35, 40, 45, y 49 (se toma muestra de cada combinación 10 veces) de las 90 secuencias expresadas se usaron con datos de los tipos de tumor y se usaron a continuación para predecir los conjuntos aleatorios del ensayo de muestras de tumor (se toma muestra de cada una 10 veces) quedarían de 2 a los 39 tipos. Los datos resultantes se muestran en las Figuras 7-9.

Se hizo un índice de los miembros de las 90 secuencias genéticas de 1 a 90, y los conjuntos aleatorios representativos usados en la invención son los que siguen a continuación:

Para 2 genes, en el conjunto 1, se usaron los genes 30 y 72. Para el conjunto 2, se usaron los genes 65 y 88. Para el conjunto 3, se usaron los genes 76 y 88. Para el conjunto 4, se usaron los genes 5 y 86. Para el conjunto 5, se usaron los genes 30 y 32. Para el conjunto 6, se usaron los genes 6 y 59. Para el conjunto 7, se usaron los genes 57 y 2. Para el conjunto 8, se usaron los genes 49 y 28. Para el conjunto 9, se usaron los genes 37 y 35. Para el conjunto 10, se usaron los genes 34 y 18.

Para 5 genes, en el conjunto 1, se usaron los genes 1, 83, 59, 36, 66, y 88. En el conjunto 2, se usaron los genes 58, 13, 59, 22, y 64. En el conjunto 3, se usaron los genes 46, 72, 51, 88, y 14. En el conjunto 4, se usaron los genes 23, 74, 22, 27, y 20. En el conjunto 5, se usaron los genes 58, 54, 78, 87, y 50. En el conjunto 6, se usaron los genes 59, 6, 56, 78, y 9. En el conjunto 7, se usaron los genes 30, 78, 69, 83, y 21. En el conjunto 8, se usaron los genes 5, 39, 54, 56, y 55. En el conjunto 9, se usaron los genes 9, 70, 54, 67, y 43. En el conjunto 10, se usaron los genes 80, 81, 63, 90, y 53.

Para 10 genes, en el conjunto 1, se usaron los genes 70, 17, 45, 5, 2, 37, 6, 76, 39, y 14. En el conjunto 2, se usaron los genes 54, 16, 80, 26, 15, 45, 50, 8, 73, y 48. En el conjunto 3, se usaron los genes 66, 87, 31, 74, 37, 45, 19, 1, 70, y 7. En el conjunto 4, se usaron los genes 85, 17, 78, 61, 23, 59, 27, 18, 58, y 24. En el conjunto 5, se usaron los genes 44, 89, 36, 76, 49, 3, 21, 24, 38, y 69. En el conjunto 6, se usaron los genes 32, 72, 55, 2, 86, 81, 53, 45, 17, y 74. En el conjunto 7, se usaron los genes 27, 55, 62, 33, 32, 84, 21, 45, 23, y 7. En el conjunto 8, se usaron los genes 62, 45, 68, 31, 69, 39, 33, 63, 19, y 22. En el conjunto 9, se usaron los genes 71, 39, 11, 56, 88, 80, 37, 77, 62, y 35. En el conjunto 10, se usaron los genes 38, 83, 41, 47, 66, 87, 10, 4, 88, y 22. Para 15 genes, en el conjunto 1, se usaron los genes 61, 17, 64, 14, 1, 41, 72, 47, 69, 48, 49, 70, 12, 20, y 35. En el conjunto 2, se usaron los genes 26, 49, 69, 31, 84, 42, 24, 56, 82, 12, 29, 2, 21, 15, y 71. En el conjunto 3, se usaron los genes 54, 62, 8, 32, 58, 65, 39, 44, 35, 22, 34, 77, 43, 83, y 75. En el conjunto 4, se usaron los genes 62, 50, 57, 80, 28, 83, 32, 56, 14, 2, 3, 48, 67, 79, y 72. En el conjunto 5, se usaron los genes 55, 58, 77, 68, 90, 76, 17, 72, 85, 34, 43, 33, 62, 6, y 64. En el conjunto 6, se usaron los genes 41, 63, 90, 9, 25, 35, 2, 14, 65, 87, 11, 36, 10, 79, y 17. En el conjunto 7, se usaron los genes 69, 89, 77, 33, 71, 4, 6, 46, 72, 13, 68, 81, 31, 50, y 32. En el conjunto 8, se usaron los genes 29, 69, 34, 47, 32, 52, 63, 73, 23, 25, 33, 10, 37, 17, y 55. En el conjunto 9, se usaron los genes 24, 13, 45, 17, 51, 48, 20, 30, 29, 40, 53, 19, 88, 76, y 28. En el conjunto 10, se usaron los genes 86, 33, 19, 4, 84, 25, 78, 29, 88, 10, 7, 67, 85, 45, y 8.

Para 20 genes, en el conjunto 1, se usaron los genes 57, 78, 43, 50, 14, 71, 56, 25, 80, 31, 88, 4, 49, 13, 3, 38, 32, 8, 52, y 75. En el conjunto 2, se usaron los genes 84, 46, 23, 85, 55, 82, 56, 83, 48, 89, 8, 60, 21, 40, 20, 17,

- 87, 24, 34, y 39. En el conjunto 3, se usaron los genes 72, 88, 53, 46, 82, 9, 34, 21, 76, 24, 14, 35, 90, 31, 58, 30, 15, 41, 7, y 28. En el conjunto 4, se usaron los genes 22, 62, 21, 3, 45, 50, 58, 72, 69, 82, 49, 42, 47, 9, 15, 59, 17, 24, 40, y 52. En el conjunto 5, se usaron los genes 71, 18, 74, 53, 43, 75, 76, 54, 63, 64, 10, 5, 90, 51, 31, 58, 28, 35, 70, y 23. En el conjunto 6, se usaron los genes 7, 30, 77, 25, 17, 16, 35, 68, 56, 37, 78, 87, 45, 8, 42, 82, 72, 23, 58, y 54. En el conjunto 7, se usaron los genes 3, 58, 67, 5, 87, 62, 56, 88, 73, 50, 22, 52, 10, 60, 57, 42, 46, 26, 7, y 82. En el conjunto 8, se usaron los genes 63, 19, 22, 13, 82, 12, 44, 52, 8, 90, 35, 81, 79, 15, 83, 76, 51, 27, 45, y 56. En el conjunto 9, se usaron los genes 65, 34, 76, 81, 58, 86, 83, 46, 40, 55, 48, 42, 57, 70, 21, 72, 71, 17, 22, y 24. En el conjunto 10, se usaron los genes 34, 74, 2, 53, 76, 73, 19, 72, 88, 87, 44, 70, 40, 39, 22, 45, 83, 77, 30, y 46.
- 10 Para 25 genes, en el conjunto 1, se usaron los genes 13, 77, 22, 85, 58, 8, 23, 2, 40, 81, 50, 31, 14, 41, 21, 52, 6, 74, 11, 17, 83, 7, 9, 19, 18. En el conjunto 2, se usaron los genes 3, 12, 8, 87, 34, 75, 31, 88, 77, 39, 40, 60, 54, 9, 37, 5, 51, 53, 32, 35, 66, 4, 26, 59, y 29. En el conjunto 3, se usaron los genes 29, 41, 44, 56, 88, 72, 90, 6, 19, 63, 42, 24, 49, 70, 39, 17, 82, 13, 9, 4, 51, 40, 22, 71, y 25. En el conjunto 4, se usaron los genes 70, 82, 55, 43, 40, 32, 16, 13, 22, 41, 7, 85, 46, 42, 73, 76, 14, 60, 50, 72, 5, 81, 67, 57, y 83. En el conjunto 5, se usaron los genes 88, 83, 53, 26, 29, 4, 38, 71, 11, 66, 14, 89, 39, 34, 84, 41, 7, 64, 87, 3, 67, 43, 50, 79, y 6. En el conjunto 6, se usaron los genes 88, 16, 83, 4, 7, 39, 56, 82, 10, 20, 87, 79, 3, 35, 76, 49, 43, 11, 74, 13, 48, 22, 64, 34, y 89. En el conjunto 7, se usaron los genes 6, 64, 39, 50, 44, 46, 61, 28, 79, 43, 35, 85, 48, 9, 59, 47, 57, 5, 24, 33, 80, 11, 42, 20, y 26. En el conjunto 8, se usaron los genes 59, 24, 46, 33, 50, 71, 53, 21, 86, 10, 75, 23, 74, 60, 43, 22, 16, 62, 85, 79, 81, 34, 73, 2, y 1. En el conjunto 9, se usaron los genes 68, 11, 64, 54, 37, 28, 44, 73, 83, 89, 2, 41, 59, 75, 21, 23, 88, 71, 34, 29, 1, 47, 84, 60, y 72. En el conjunto 10, se usaron los genes 5, 12, 60, 84, 32, 58, 70, 2, 38, 41, 24, 13, 85, 10, 49, 90, 55, 81, 39, 27, 65, 56, 31, 34, y 57.
- 15 Para 30 genes, en el conjunto 1, se usaron los genes 24, 88, 10, 69, 64, 8, 19, 54, 80, 70, 11, 9, 29, 56, 36, 79, 30, 65, 2, 58, 23, 74, 41, 16, 77, 4, 78, 14, 85, y 32. En el conjunto 2, se usaron los genes 73, 27, 19, 52, 87, 51, 63, 4, 76, 64, 90, 81, 42, 47, 9, 62, 40, 65, 83, 30, 39, 59, 10, 11, 54, 44, 43, 6, 86, y 41. En el conjunto 3, se usaron los genes 28, 47, 41, 8, 24, 54, 26, 49, 61, 17, 46, 64, 20, 16, 1, 33, 82, 79, 85, 5, 86, 69, 31, 65, 83, 7, 67, 35, 48, y 57. En el conjunto 4, se usaron los genes 13, 21, 83, 35, 47, 57, 8, 66, 75; 17, 38, 70, 39, 23, 9, 1, 2, 28, 68, 81, 36, 80, 52, 22, 44, 37, 85, 15, 72, y 86. En el conjunto 5, se usaron los genes 81, 20, 36, 89, 13, 14, 46, 58, 59, 62, 28, 7, 1, 25, 35; 83, 26, 50, 51, 15, 16, 56, 71, 5, 47, 6, 78, 80, 85, y 84. En el conjunto 6, se usaron los genes 68, 74, 73, 89, 38, 72, 33, 35, 15, 79, 3, 37, 23, 67, 10, 62, 64, 77, 44, 60, 75, 7, 51, 12, 46, 76, 81, 26, 42, y 6. En el conjunto 7, se usaron los genes 34, 55, 62, 40, 78, 35, 76, 30, 21, 77, 46, 71, 66, 69, 63, 81, 51, 38, 84, 53, 82, 89, 29, 14, 36, 45, 60, 7, 52, y 27. En el conjunto 8, se usaron los genes 56, 12, 35, 79, 57, 4, 16, 9, 24, 58, 40, 72, 80, 67, 23, 76, 88, 69, 52, 78, 32, 47, 14, 46, 64, 83, 17, 59, 81, y 20. En el conjunto 9, se usaron los genes 73, 27, 12, 58, 54, 62, 48, 43, 16, 41, 49, 84, 9, 75, 13, 50, 19, 3, 76, 78, 56, 68, 71, 25, 24, 60, 18, 35, 45, y 51. En el conjunto 10, se usaron los genes 82, 21, 24, 85, 51, 18, 72, 28, 89, 22, 34, 4, 53, 75, 83, 23, 50, 5, 42, 13, 88, 63, 40, 64, 38, 35, 39, 44, 59, y 70.
- 25 Para 35 genes, en el conjunto 1, se usaron los genes 2, 69, 70, 89, 9, 11, 5, 17, 63, 18, 12, 59, 58, 85, 26, 71, 61, 10, 3, 1, 22, 79, 84, 30, 48, 82, 38, 44, 56, 42, 88, 6, 60, 14, y 28. En el conjunto 2, se usaron los genes 84, 81, 88, 46, 12, 50, 38, 78, 62, 48, 19, 43, 26, 66, 4, 20, 40, 58, 9, 52, 87, 47, 6, 55, 21, 75, 31, 77, 57, 53, 45, 34, 30, 32, y 39. En el conjunto 3, se usaron los genes 6, 3, 22, 89, 8, 78, 87, 71, 42, 63, 18, 40, 68, 77, 64, 88, 5, 58, 43, 72, 80, 10, 21, 56, 11, 59, 61, 2, 19, 76, 30, 20, 14, 69, y 35. En el conjunto 4, se usaron los genes 55, 42, 89, 41, 56, 33, 24, 28, 15, 61, 63, 18, 90, 60, 35, 76, 70, 52, 8, 1, 64, 23, 13, 39, 71, 31, 3, 81, 10, 34, 66, 44, 16, 7, y 78. En el conjunto 5, se usaron los genes 59, 58, 12, 50, 47, 42, 28, 22, 76, 54, 1, 18, 7, 53, 68, 73, 20, 67, 14, 72, 23, 13, 39, 10, 70, 55, 45, 17, 31, 51, 80, 3, 24, 30, y 46. En el conjunto 6, se usaron los genes 53, 66, 26, 31, 73, 47, 61, 63, 51, 41, 29, 5, 19, 10, 57, 22, 64, 11, 34, 89, 43, 24, 31, 60, 27, 76, 17, 86, 70, 81, 50, 46, 36, 14, y 45.
- 30 En el conjunto 7, se usaron los genes 18, 88, 90, 13, 73, 81, 64, 56, 84, 2, 4, 22, 3, 25, 35, 54, 89, 86, 27, 41, 6, 34, 38, 14, 74, 36, 59, 8, 40, 55, 42, 83, 39, 44, y 60. En el conjunto 8, se usaron los genes 46, 32, 22, 15, 67, 89, 14, 5, 70, 39, 49, 9, 84, 71, 12, 78, 27, 86, 26, 57, 20, 43, 58, 87, 42, 8, 31, 1, 54, 62, 69, 40, 29, 52, y 64. En el conjunto 9, se usaron los genes 3, 39, 55, 25, 90, 10, 9, 77, 62, 78, 18, 12, 58, 51, 22, 67, 7, 61, 59, 35, 52, 4, 65, 38, 32, 71, 87, 88, 63, 50, 73, 70, 44, 45, y 84. En el conjunto 10, se usaron los genes 65, 54, 51, 38, 40, 5, 43, 71, 34, 30, 22, 6, 36, 64, 63, 13, 70, 85, 21, 88, 77, 86, 79, 66, 25, 18, 26, 19, 76, 56, 23, 60, 75, 2, y 49.
- 35 Para 40 genes, en el conjunto 1, se usaron los genes 81, 80, 68, 77, 17, 71, 34, 33, 48, 88, 90, 32, 23, 2, 38, 59, 75, 82, 50, 56, 12, 36, 6, 87, 72, 37, 26, 15, 35, 66, 13, 76, 55, 3, 78, 18, 52, 47, 73, y 20. En el conjunto 2, se usaron los genes 11, 65, 27, 44, 88, 49, 55, 57, 1, 72, 9, 28, 56, 67, 13, 58, 42, 36, 8, 31, 40, 14, 26, 35, 62, 22, 19, 84, 78, 21, 2, 41, 74, 71, 52, 30, 25, 76, 85, y 63. En el conjunto 3, se usaron los genes 50, 22, 10, 54, 9, 51, 15, 34, 29, 35, 76, 89, 33, 6, 88, 56, 36, 70, 87, 40, 83, 62, 1, 42, 25, 78, 30, 26, 44, 60, 69, 47, 49, 31, 18, 59, 37, 52, 61, y 17. En el conjunto 4, se usaron los genes 27, 33, 7, 89, 36, 59, 48, 42, 66, 39, 90, 52, 2, 14, 30, 80, 9, 56, 21, 87, 65, 67, 41, 73, 82, 20, 4, 46, 5, 84, 88, 15, 44, 58, 78, 85, 3, 64, 6, y 8. En el conjunto 5, se usaron los genes 43, 24, 86, 29, 46, 90, 40, 1, 71, 57, 12, 84, 69, 19, 42, 62, 28, 35, 5, 63, 52, 17, 39, 4, 67, 81, 50, 47, 61, 54, 87, 70, 77, 6, 10, 38, 37, 79, 31, y 36. En el conjunto 6, se usaron los genes 28, 5, 78, 85, 16, 20, 36, 52, 43, 29, 67, 83, 12, 79, 84, 8, 81, 46, 11, 3, 54, 86, 10, 60, 71, 51, 39, 53, 59, 69, 44, 61, 7, 56, 27, 50, 66, 70, 1, y 25.
- 40 En el conjunto 7, se usaron los genes 39, 47, 48, 24, 25, 3, 41, 16, 65, 73, 63, 14, 70, 57, 12, 64, 90, 23, 27, 38, 66, 71, 54, 21, 83, 28, 72, 53, 11, 30, 80, 15, 6, 88, 89, 85, 81, 61, 78, y 34. En el conjunto 8, se usaron los genes 61, 8, 57, 16, 24, 64, 48, 36, 58, 28, 27, 40, 70, 77, 25, 76, 52, 35, 62, 4, 60, 7, 54, 37, 11, 20, 72, 34, 56, 78, 10, 86, 51, 29, 84, 47, 30, 21, 59, y 67. En el conjunto 9, se usaron los genes 67, 3, 83, 33, 35, 26, 25, 79, 68, 19, 18, 84, 14, 58, 66, 57, 1, 2, 27, 64, 23, 24, 76, 81, 17, 37, 38, 30, 45, 75, 49, 39, 5, 53, 43, 15, 51, 40, 69, y 12. En el conjunto 10, se usaron los genes 39, 77, 29, 70, 85, 45, 54, 79, 31, 43, 15, 11, 47, 83, 76, 21, 67, 14, 4, 19, 49,
- 45
- 50
- 55
- 60
- 65

42, 18, 13, 12, 7, 88, 8, 3, 35, 81, 55, 71, 60, 72, 57, 46, 40, 56, y 32.

Para 45 genes, en el conjunto 1, se usaron los genes 7, 63, 45, 87, 19, 55, 36, 42, 9, 4, 79, 68, 46, 35, 40, 80, 59, 58, 38, 17, 50, 30, 13, 39, 33, 84, 34, 64, 2, 57, 24, 88, 65, 16, 53, 18, 28, 8, 60, 15, 43, 73, 77, 20, y 78. En el conjunto 2, se usaron los genes 70, 19, 81, 68, 38, 35, 48, 9, 53, 11, 73, 42, 54, 28, 32, 40, 60, 88, 25, 7, 67, 17, 36, 51, 44, 46, 10, 89, 14, 80, 39, 41, 27, 8, 75, 47, 61, 57, 59, 76, 86, 65, 63, 74, y 77. En el conjunto 3, se usaron los genes 55, 24, 63, 17, 32, 81, 2, 67, 51, 85, 27, 46, 60, 90, 25, 35, 58, 11, 47, 33, 73, 3, 74, 52, 15, 86, 6, 78, 36, 66, 57, 13, 49, 28, 75, 70, 4, 77, 43, 26, 61, 64, 20, 1, y 23. En el conjunto 4, se usaron los genes 49, 72, 13, 51, 55, 11, 29, 5, 43, 44, 40, 6, 38, 67, 47, 35, 36, 28, 81, 24, 80, 32, 16, 88, 63, 87, 86, 79, 21, 1, 30, 10, 62, 58, 23, 12, 78, 26, 69, 56, 85, 42, 17, 84, y 39. En el conjunto 5, se usaron los genes 53, 33, 18, 65, 22, 83, 50, 88, 76, 40, 82, 68, 85, 5, 63, 45, 78, 16, 42, 54, 27, 66, 70, 74, 7, 51, 89, 64, 49, 37, 84, 86, 34, 39, 80, 31, 61, 87, 69, 4, 81, 30, 14, 41, y 29. En el conjunto 6, se usaron los genes 7, 60, 38, 14, 73, 9, 79, 81, 22, 10, 85, 51, 40, 87, 3, 26, 57, 56, 12, 72, 39, 59, 63, 28, 64, 71, 69, 21, 67, 48, 50, 66, 46, 88, 11, 13, 24, 8, 58, 75, 2, 41, 5, 44, y 55. En el conjunto 7, se usaron los genes 15, 65, 31, 19, 11, 38, 2, 9, 64, 66, 22, 35, 49, 3, 77, 43, 32, 56, 39, 54, 80, 21, 6, 40, 27, 86, 10, 16, 70, 30, 85, 23, 26, 4, 55, 73, 42, 13, 41, 68, 29, 57, 28, 72, y 58. En el conjunto 8, se usaron los genes 83, 27, 9, 62, 84, 78, 13, 5, 74, 55, 12, 34, 58, 3, 67, 57, 24, 45, 42, 47, 75, 25, 29, 44, 46, 61, 56, 70, 86, 37, 14, 49, 60, 89, 28, 72, 59, 38, 2, 81, 50, 7, 6, 21, y 82. En el conjunto 9, se usaron los genes 7, 10, 35, 14, 79, 66, 33, 52, 16, 55, 68, 59, 57, 19, 11, 47, 22, 38, 61, 30, 71, 50, 63, 88, 53, 80, 6, 54, 77, 21, 37, 84, 9, 65, 12, 49, 40, 73, 76, 2, 28, 29, 3, 72, y 18. En el conjunto 10, se usaron los genes 12, 19, 9, 80, 84, 15, 7, 2, 39, 21, 48, 40, 51, 69, 74, 83, 5, 66, 27, 26, 89, 60, 4, 86, 41, 44, 35, 10, 76, 53, 63, 16, 37, 79, 11, 42, 68, 3, 59, 82, 77, 73, 85, 67, y 14.

Para 49 genes, en el conjunto 1, se usaron los genes 84, 47, 56, 1, 18, 21, 57, 54, 27, 89, 44, 85, 64, 10, 77, 34, 65, 66, 80, 70, 46, 23, 53, 61, 24, 81, 43, 35, 30, 74, 83, 51, 20, 17, 72, 4, 49, 68, 60, 28, 67, 19, 42, 55, 73, 36, 7, 39, y 33. En el conjunto 2, se usaron los genes 47, 29, 58, 36, 21, 53, 40, 7, 83, 77, 24, 89, 71, 64, 60, 4, 37, 86, 27, 57, 62, 63, 72, 1, 88, 78, 68, 17, 51, 16, 82, 42, 81, 18, 32, 49, 55, 10, 11, 66, 35, 23, 70, 20, 61, 25, 48, 43, y 54. En el conjunto 3, se usaron los genes 54, 2, 62, 67, 44, 25, 8, 53, 86, 33, 75, 32, 45, 76, 43, 65, 59, 58, 42, 64, 47, 78, 3, 57, 71, 88, 14, 23, 51, 83, 1, 41, 7, 56, 40, 20, 39, 72, 70, 19, 5, 35, 50, 82, 37, 48, 15, 31, y 16. En el conjunto 4, se usaron los genes 35, 65, 48, 43, 69, 62, 64, 74, 82, 39, 37, 1, 88, 45, 66, 12, 79, 55, 38, 84, 17, 30, 25, 26, 89, 56, 28, 57, 59, 34, 85, 14, 47, 44, 41, 19, 60, 20, 73, 2, 63, 75, 49, 80, 58, 77, 27, 54, y 29. En el conjunto 5, se usaron los genes 64, 51, 36, 12, 84, 24, 65, 47, 88, 26, 10, 19, 73, 90, 35, 53, 18, 55, 80, 70, 79, 82, 87, 77, 15, 85, 83, 7, 72, 1, 6, 57, 38, 45, 74, 33, 62, 86, 31, 69, 27, 14, 4, 29, 54, 44, 63, 78, y 42. En el conjunto 6, se usaron los genes 24, 39, 85, 42, 88, 32, 65, 23, 6, 75, 53, 77, 64, 90, 13, 82, 47, 31, 48, 8, 78, 67, 63, 44, 26, 40, 14, 34, 18, 59, 2, 17, 20, 56, 83, 68, 86, 9, 38, 73, 89, 55, 29, 69, 72, 16, 28, 51, y 81. En el conjunto 7, se usaron los genes 32, 70, 57, 67, 1, 73, 52, 38, 65, 83, 5, 40, 49, 31, 66, 85, 6, 82, 12, 48, 89, 3, 19, 41, 62, 16, 46, 61, 24, 18, 55, 30, 33, 56, 68, 20, 81, 10, 86, 9, 15, 63, 78, 22, 75, 14, 13, 43, y 77. En el conjunto 8, se usaron los genes 17, 30, 47, 85, 7, 3, 6, 35, 76, 77, 25, 86, 36, 75, 44, 29, 69, 60, 63, 64, 82, 51, 19, 68, 41, 28, 73, 18, 10, 26, 42, 78, 67, 12, 80, 33, 13, 57, 38, 87, 49, 59, 74, 50, 90, 46, 8, 81, y 4. En el conjunto 9, se usaron los genes 19, 76, 42, 36, 66, 21, 8, 28, 22, 15, 56, 5, 2, 86, 17, 62, 23, 1, 80, 73, 52, 83, 32, 65, 44, 82, 35, 60, 47, 90, 74, 9, 84, 50, 4, 77, 55, 57, 19, 71, 25, 48, 81, 53, 34, 38, 3, 37, y 16. En el conjunto 10, se usaron los genes 84, 87, 3, 41, 36, 71, 33, 57, 85, 26, 53, 22, 82, 31, 2, 45, 24, 18, 37, 35, 77, 20, 63, 25, 6, 17, 58, 7, 9, 49, 28, 76, 79, 67, 13, 80, 66, 5, 43, 4, 74, 75, 21, 86, 23, 39, 42, 27, y 54.

Ejemplo 5: Detección basada en PCR

Como se ha indicado anteriormente, la determinación o medida de la expresión genética se puede realizar mediante PCR, tal como el uso de PCR cuantitativa. La detección de la expresión de aproximadamente 5 a 49 secuencias expresadas en el genoma humano se puede usar en tales realizaciones de la invención. Además, también se pueden usar los niveles de expresión de aproximadamente 5 a 49 secuencias genéticas en el conjunto de 74, el conjunto de 90, o una combinación del conjunto de los dos (con un total de 126 secuencias genéticas dada la presencia de 38 secuencias genéticas en común entre los dos conjuntos). La invención contempla el uso de PCR cuantitativa para medir los niveles de expresión, como se ha descrito anteriormente, de aproximadamente 5 a 49 de 87 secuencias genéticas, todas las cuales están presentes en cualquiera del conjunto de 74 o el conjunto de 90. De las 87 secuencias genéticas, 60 están presentes en el conjunto de 74, y 63 están presentes en el conjunto de 90. Los identificadores/números de referencia de las 87 secuencias genéticas son AA456140, AA745593, AA765597, AA782845, AA865917, AA946776, AA993639, AB038160, AF104032, AF133587, AF301598, AF332224, AI041545, AI147926, AI309080, AI341378, AI457360, AI620495, AI632869, AI683181, AI685931, AI802118, AI804745, AI952953, AI985118, AJ000388, AK025181, AK027147, AK054605, AL023657, AL039118, AL110274, AL157475, AW118445, AW194680, AW291189, AW298545, AW445220, AW473119, AY033998, BC000045, BC001293, BC001504, BC001639, BC002551, BC004331, BC004453, BC005364, BC006537, BC006811, BC006819, BC008764, BC008765, BC009084, BC009237, BC010626, BC011949, BC012926, BC013117, BC015754, BC017586, BE552004, BE962007, BF224381, BF437393, BF446419, BF592799, BI493248, H05388, H07885, H09748, M95585, N64339, NM_000065, NM_001337, NM_003914, NM_004062, NM_004063, NM_004496, NM_006115, NM_019894, NM_033229, R15881, R45389, R61469, X69699, y X96757.

El uso de aproximadamente 5 a 49 de estas secuencias en la práctica de la invención puede incluir el uso de los niveles de expresión medidos para secuencias genéticas de referencia como se describe en el presente documento. En algunas realizaciones, las secuencias genéticas de referencia son una o más de las 8 que se desvelan en el

presente documento. La invención contempla el uso de una o más de las secuencias de referencia identificadas como AF308803, AL137727, BC003043, BC006091, y BC016680 en realizaciones de la invención basadas en PCR o QPCR. Por supuesto, las 5 de estas secuencias de referencia también se pueden usar en combinación.

- 5 Habiendo descrito ahora totalmente la presente invención, los expertos en la materia observarán que la misma se puede realizar dentro de un amplio intervalo de parámetros, concentraciones, y condiciones equivalentes sin apartarse del alcance de la invención y sin experimentación indebida.

- 10 Aunque la presente invención se ha descrito en conexión con realizaciones específicas de la misma, se entenderá que se pueden realizar modificaciones adicionales. La presente solicitud pretende cubrir cualquier variación, uso, o adaptación de la invención siguiendo, en general, los principios de la invención e incluyendo tales desviaciones de la presente divulgación ya que forman parte del conocimiento o la práctica habitual dentro de la técnica a la que pertenece la invención y como tal se puede aplicar a las características esenciales que se han expuesto anteriormente en el presente documento.

15

APÉNDICE 1

Apéndice de secuencias de ARNRNA (Listado de Secuencias)

- 20 >Hs.73995_ARNm_1 gi|190403|gb|M60502.1|PERFILHUM ARNm de profilagrina humana, extremo en la posición 3' poli = 1

GGCCACTCTGCAGACAGCTCCAGACAATCAGGCACTCGTCCACACAGAGTCTTCCTCTCGT
 GGACAGGCTGCGTCATCCCATGAACAGGCAAGATCAAGTGCAGGAGAAAGACATGGATCC
 CACCACCAGCAGTCAGCAGACAGCTCCAGACACGCAGGCATTGGGCACGGACAAGCTTCA
 TCTGCAGTCAGAGACAGTGGACACCGAGGGTACAGAGGTAGTCAGGCCACTGCAGTGAG
 GGACATTCAGAAGACTCAGACACACAGTCAGTGTTCAGCACAGGGACAAGCTGGGCCCCAT
 CAGCAGAGCCACCAAGAGTCCGCACGTGGCCAGTCAGGGGAAAGCTCTGGACGTTTCAGGG
 TCTTTCTCTTACCAGGTGAGCACTCATGAACAGTCTGAGTCCACCCATGGACAGTCTGTG
 CCCAGCACTGGAGGAAGACAAGGATCCCACCATGATCAGGCACAAGACAGCTCCAGGCAC
 TCAGCATCCCAAGAGGGTTCAGGACACCATTTCGTGGACACCCGGGGCCAAGCAGAGGAGGA
 AGACAGGGGTCCCACCACGAGCAATCGGTAGATAGGTCTGGACACTCAGGGTCCCATCAC
 AGCCACACCACATCCAGGGAAGGTCTGATGCCTCCCGTGGGCAGTCAGGATCCAGAAGT
 GCAAGCAGACAAACACATGACCAGGAACAATCAGGAGACGGCTCTAGGCACTCAGGGTCCG
 CGTCATCAGGAAGCTTCTCTTGGGCCGACAGCTCTAGACACTCACAGGCAGTCCAGGGA
 CAATCAGAGGGGTCCAGGACAAGCAGGCGCCAGGGATCCAGTGTTCAGGACAGTGCAG
 AGTCAGGGACACTCAGAAGACTCTGAGAGGCGGTCTGGGTCTGCTTCCAGAAACCATCGT
 GGATCTGCTCAGGAGCAGTCAAGAGATGGCTCCAGACACCCAGGTCCCATCACGAAGAC
 AGAGCCGGTCCAGGGGACTCTGCAGAGAGCTCCAGACAATCAGGCACTCATCATGCAGAG
 AATTCCTCTGGTGGACAGGCTGCATCATCCCATGAACAGGCAAGATCAAGTGCAGGAGAG
 AGACATGGATCCCCTACCAGCAGTCAGCAGACAGCTCCAGACACTCAGGCATTGGGCAC
 GGACAAGCTTCTATCTGCAGTCAGAGACAGTGGACACCGAGGGTCCAGTGGTAGTCAGGCC
 AGTGACAATGAGGGACATTCAGAAGACTCAGACACACAGTCAGTGTTCAGCCACCCGACAG
 GCTGGGCGCCATCACGAGAGCCACCAAGAGTCCACACGTGGCCGGTCCAGGGAAGGTCT
 GGACGTTTCAGGGTCTTCTCTTACCAGGTGAGCACTCATGAACAGTCTGAGTCTGCCCAT
 GGACGGGCTGGGCCAGTACTGGAGGAAGACAAGGATCCCGCCACGAGCAGGCACGAGAC
 AGCTCCAGGCAGTCAGCGTCCCAAGAGGGTTCAGGACACCATTTCGTGGACACCCGGGGTCA
 AGGAGAGGAGGAAGACAGGGATCCTACCACGAGCAATCGGTAGATAGGTCTGGACACTCA
 GGGTCCCATCACAGCCACACCACATCCAGGGAAGGTCTGATGCCTCCCATGGGCAGTCA
 GGATCCAGAAGTGCAAGCAGAGAAACACGTAATGAGGAACAGTCAGGAGACGGCTCCAGG
 CACTCAGGGTCCGCTCACCATGAAGCTTCCACTCAGGCTGACAGCTCTAGACACTCACAG
 TCCGGCCAGGGTGAATCAGCGGGGTCCAGGAGAAGCAGGCGCCAGGGATCCAGTGTTCAG
 CAGGACAGTGACAGTGAGGCATACCCAGAGGACTCTGAGAGGCGATCTGAGTCTGCTTCC
 AGAAACCATCATGGATCTTCTCGGGAGCAGTCAAGAGATGGCTCCAGACACCCCGGATCC
 TCTCACCGGATACAGCCAGTCATGTACAGTCTTCACCTGTACAGTCAGACTCTAGTACC
 GCTAAGGAACATGGTCACTTTAGTAGTCTTTCACAAGATTCTGCGTATCACTCAGGAATA
 CAGTACCGTGGCAGTCTTCACAGTTCTAGTTCTTATCATTATCAATCTGAGGGCACTGAA
 AGGCAAAAAGGTCAATCAGGTTTAGTTTGGAGACATGGCAGCTATGGTATGTCAGATTAT
 GATTATGGTGAATCCGGGTTTAGACACTCTCAGCACGGAAGTGTAGTTACAATTCGAAT
 CCTGTTCTTTTCAAGGAAAGATCTGATATCTGTAAAGCAAGTGCGTTTGGTAAAGATCAT
 CCAAGGTATTATGCAACGTATATTAATAAGGACCCAGGTTTATGTGGCCATTCTAGTGAT
 ATATCGAAACAACCTGGGATTTAGTCAGTCACAGAGATACTATTACTATGAGTAAGAAATT
 AATGGCAAAGGAATTAATCCAAGAATAGAAGAATGAAGCAAGTTCACTTTCAATCAAGAA
 ACTTCATAATACTTTCAGGGAAGTTATCTTTTCTGTCAATCTGTTTAAATATGCTATA
 GTATTTTCAATTAGTTTGGTGGTAACTTATTTTATTTGTGTAATGATCTTTAAACGCTATAT
 TTCAGAAATATTAAATGGAAGAAATCAATATCATGGAGAGCTAACTTTAGAAAACCTAGCT
 GGAGTATTTTAGGAGATTCTGGGTCAAGTAATGTTTATGTTTGTGAAAGTTTAAAGTTT
 AGACACTCCCCAAATTTCTAAATTAATCTTTTCAGAAATATCGAAGGAGCCAAAATAT
 AAAACAGTTCTGATATCCAAAGTGGCTATATCAACATCAGGGCTAGCACATCTTTCTCTA
 TTATCCTTCTATTGGAATTCTAGTATTCTGTATTCAAAAAATCATCTTGGACATAATTAA
 TATTTTAGTAAGCTGCATCTAAATTAAAAATAAACTATTCATCATATAAT

>Hs.75236_ARNm_4 gi|14280328|gb|AY033998.1| Homo sapiens poliA = 3

TAGAATCGGGGGTTTCAGCTCACTGCTCCTTTTCTTTTTTTTTCTTTCTCTCCCCCGCCCA
 CCCCCCAAAAATAATTGATTTGCTTTACAATCATCCACACTGTGTTTTGTGGATCTTTTA
 ATTATATATAACAATAGTAGTCATTTTAAATATATATTCTGAAATCTTTGCAAAATTTAA
 CAGAAGAGTCGAAGCTCTGCGAGACCCAAATTTGCCAATAAGAATGGTTATGATAATTA
 GCACCATGGAGCCTCAGGTGTCAAATGGTCCGACATCCAATACAAGCAATGGACCCCTCA

GCAACAACAGAACTGTCCTTCTCCCATGCAAAACAGGGGCAACCACAGATGACAGCAAAA
 CCAACCTCATCGTCAACTATTTACCCAGAAATATGACCCAAGAAGAATTCAGGAGTCTCT
 TCGGGAGCATTTGGTGAAATAGAATCCTGCAAACCTGTGAGAGACAAAATTACAGGACAGA
 GTTTAGGGTATGGATTTGTAACTATATTGATCCAAAGGATGCAGAGAAAGCCATCAACA
 CTTTAAATGGACTCAGACTCCAGACCAAAACCATAAAGGTCTCATATGCCCGTCCGAGCT
 CTGCCTCAATCAGGGATGCTAACCTCTATGTTAGCGGCCCTTCCCAAAACCATGACCCAGA
 AGGAACTGGAGCAACTTTTCTCGCAATACGGCCGTATCATCACCTCACGAATCCTGGTTG
 ATCAAGTCACAGGAGTGTCCAGAGGGGTGGGATTTCATCCGCTTTGATAAGAGGATTGAGG
 CAGAAGAAGCCATCAAAGGGCTGAATGGCCAGAAGCCCAGCGGTGCTACGGAACCGATTA
 CTGTGAAGTTTGCCAACAACCCAGCCAGAAGTCCAGCCAGGCCCTGCTCTCCAGCTCT
 ACCAGTCCCCTAACCGGCGCTACCCAGGTCCACTTCACCACCAGGCTCAGAGGTTTCAGGC
 TGGACAATTTGCTTAATATGGCCTATGGCGTAAAGAGACTGATGTCTGGACCAGTCCCCC
 CTTCTGCTTTGTTCCCCCAGGTTCTCCCCAATTACCATTGATGGAATGACAAGCCTTGTTGG
 GAATGAACATCCCTGGTCACACAGGAACTGGGTGGTGCATCTTTGTCTACAACCTGTCCC
 CCGATTCCGATGAGAGTGTCTCTGGCAGCTCTTTGGCCCCCTTTGGAGCAGTGAACAACG
 TAAAGGTGATTTCGTGACTTCAACACCAACAAGTGCAAGGGATTTCGGCTTTGTACCATGA
 CCAACTATGATGAGGCGGCCATGGCCATCGCCAGCCTCAACGGGTACCGCCTGGGAGACA
 GAGTGTGCAAGTTTCCTTTAAAACCAACAAAGCCCAAGTCCTGAATTTCCCATTCTT
 ACTTACTAAAATATATATAGAAATATATACGAACAAAACACACGCGCGCACACACACA
 TACACGAAAGAGAGAGAAACAACTTTTCAAGGCTTATATTCAACCATGGACTTTATAAG
 CCAGTGTTCCTAAGTATTAAACATTGGATTATCCTGAGGTGTACCAGGAAAGGATTTT
 ATAATGCTTAGAAAAA

>Hs.299867_ARNm_1 gi|4758533|ref|NM_004496.1| factor nuclear 3 de hepatocitos de Homo sapiens, alfa (HNF3A), ARNm poliA = 3

TCCAGGAATCGATAGTGCATTCGTGCGCGCGGCCCGCCGTCGCTTCGCACAGGGCTGGAT
GGTTGTATTGGGCAGGGTGGCTCCAGGATGTTAGGAACTGTGAAGATGGAAGGGCATGAA
ACCAGCGACTGGAACAGCTACTACGCAGACACGCAGGAGGCCCTACTCCTCGGTCCCGGTC
AGCAACATGAACTCAGGCCTGGGCTCCATGAACTCCATGAACACCTACATGACCATGAAC
ACCATGACTACGAGCGGCAACATGACCCCGCGCTCCTTCAACATGTCTATGCCAACCCG
GCCTTAGGGGCGCGCTGAGTCCCGCGCAGTAGCCGGCATGCCGGGGGGCTCGGCGGGC
GCCATGAACAGCATGACTGCGGCCGGCGTGACGGCCATGGGTACGGCGCTGAGCCCCGAGC
GGCATGGGCGCCATGGGTGCGCAGCAGGCGGCCCTCCATGATGAATGGCCTGGGCCCCCTAC
GCGGCCGCCATGAACCCGTGCATGAGCCCCATGGCGTACGCGCCGTCCAACCTGGGCCGC
AGCCGCGCGGGCGGGCGGGCGGCGACGCCAAGACGTTCAAGCGCAGTTACCCGCACGCCAAG
CCGCCCTACTCGTACATCTCGCTCATCACCATGGCCATCCAGCGGGCGCCCAGCAAGATG
CTCACGCTGAGCGAGATCTACCAGTGGATCATGGACCTCTTCCCCCTATTACCGGCAGAAC
CAGCAGCGCTGGCAGAACTCCATCCGCCACTCGCTGTCTTCAATGACTGCTTCGTCAAG
GTGGCACGCTCCCCGGACAAGCCGGGCAAGGGCTCCTACTGGACGCTGCACCCGGACTCC
GGCAACATGTTTCGAGAACGGCTGCTACTTTCGCGCCGCCAGAAGCGCTTCAAGTGCAGAAAG
CAGCCGGGGGCGCGCGGGGGCGGGGAGCGGAAGCGGGGCGAGCGGGCGCAAGGGCGGC
CCTGAGAGCCGCAAGGACCCCTCTGGCGCCTCTAACCCAGCGCCGACTCGCCCCCTCCAT
CGGGGTGTGCACGGGAAGACCGGCCAGCTAGAGGGCGCGCCGGCCCCGGGCGGCCGCC
AGCCCCCAGACTCTGGACCACAGTGGGGCGACGGCGACAGGGGGCGCCTCGGAGTTGAAG
ACTCCAGCTCCTCAACTGCGCCCCCATAGCTCCGGGGCGGGGCGCTGGCCTCTGTG
CCCGCCTCTCACCCGGCACACGGCTTGGCACCCACAGAGTCCAGCTGCACCTGAAAGGG
GACCCCCACTACTCCTTCAACCACCCGTTCTCCATCAACAACCTCATGTCTCTCTCGGAG
CAGCAGCATAAGCTGGACTTCAAGGCATACGAACAGGCACTGCAATACTCGCCTTACGGC
TCTACGTTGCCCGCCAGCCTGCCTCTAGGCAGCGCCTCGGTGACCACCAGGAGCCCCATC
GAGCCCTCAGCCCTGGAGCGGGCTACTACCAAGGTGTGTATTCCAGACCCGTCCTAAAC
ACTTCTTAGCTCCCGGACTGGGGGGTTTGTCTGGCATAGCCATGCTGGTAGCAAGAGAG
AAAAATCAACAGCAAACAAAACACAAACCAACCGTCAACAGCATAATAAAATCCA
ACAACTATTTTATTTTATTTTATTTTATGTCACAACCTTGGCCCCAGTGCAAAAGACTGTTAC
TTTATTATTGTATTCAAATTCATTGTGTATATTACTACAAAGACGGCCCCAAACCAATT
TTTTTCTGCGAAGTTTAATGATCCACAAGTGTATATATGAAATTCTCCTCCTTCTTGC
CCCCCTCTCTTCTTCTTCTTGGCCCTCCAGACATTCTAGTTTGTGGAGGGTTATTTAA
AAAACAAAAGGAAGATGGTCAAGTTTGTAAATATTTGTTTGTGCTTTTCCCCCTCCT
TACCTGACCCCTACGAGTTTACAGGCTTGTGGCAATACTCTTAACCATAAGAATTGAAA
TGGTGAAGAAACAAGTATACACTAGAGGCTCTTAAAGTATTGAAAAGACAATACTGCTG
TTATATAGCAAGACATAAACAGATTATAAACATCAGAGCCATTTGCTTCTCAGTTTACAT
TTCTGATACATGCAGATAGCAGATGTCTTAAATGAAATACATGTATATTGTGTATGGAC
TTAATTATGCACATGCTCAGATGTGTAGACATCCTCCGTATATTTACATAACATATAGAG

GTAATAGATAGGTGATATACGTGATACGTTCTCAAGAGTTGCTTGACCGAAAGTTACAAG
GACCCCAACCCCTTTGCTCTTACCCACAGATGGCCCTGGGAACAATCCTCAGGAATTGC
CCTCAAGAACTCGCTTCTTTGCTTTGAGAGTGCCATGGTCATGTCATTCTGAGGTACATA
ACACATAAATAGTTTCTATGAGTGTATACCATTAAAGATTTTTCAGTAAAGGGAATA
TTACATGTTGGGAGGAGGAGATAAGTTATAGGGAGCTGGATTCAAACGGTGGTCCAAGA
TTCAAAAATCCTATTGATAGTGGCCATTTAATCATTGCCATCGTGTGCTTGTTCATCC
AGTGTTATGCACTTCCACAGTTGGTGTGTAGTATAGCCAGAGGGTTTCATTATTATTCT
CTTTGCTTTCTCAATGTTAATTTATTGCATGGTTTATTCTTTTCTTTTACAGCTGAAATT
GCTTTAAATGATGGTTAAATTTACAAATTAATTTGGGAATTTTATCAATGTGATTGTAA
TTAAAAATATTTTGATTAAATAACAAAATAATACCAGATTTTAAAGCCGCGGAAATGT
TCTTGATCATTTGCAGTTAAGGACTTTAAATAAATCAAATGTTAACAAAAA

>Hs.285401_contig1

AI147926|AI880620|AA768316|AA761543|AA279147|AI216016|AI738663|N79248|AI
684489|AA960845|AI718599|AI379138|N29366|BF002507|AW044269|R34339|R66326
|H04648|R67467|AI523112|BF941500 poliA = 2 poliA = 3

5

TGTTTTTCTAGTTCATTTTGTGTTTCCAACCTTTTCATGTAAAAATTTAATTATTTTTGAA
 TGTGTGGATGTGAGACTGAGGTGCCTTTTGGTACTGAAATTCCTTTTCCATGTACCTGAA
 GTGTTACTTTTGTGATATAGGAAATCCTTGTATATATACTTTATTTGGTCCCTAGGCTTCC
 TATTTTGTACCTTGCTTTCTCTATGGCATCCACCATTTGATTGTTCTACTTTTATGAT
 ATGTTTTTCATAAGTGGTTAAGCAAGTATTCTCGTTACTTTTGCTCTTAAATCCCTATTCA
 TTACAGCAATGTTGGTGGTCAAAGAAAATGATAAAACAACTTGAATGTTCAATGGTCCTGA
 AATACATAACAACATTTTAGTACATTGTAAAGTAGAATCCTCTGTTTCATAATGAACAAGA
 TGAACCAATGTGGATTAGAAAAGAAGTCCGAGATATTAATTCCAAATATCCAGACATTGT
 TAAAGGGAAAAAATTTGCAATAAAATATTTGTAACATAAAAAAAAAAAAAAAAAAAAAA

>Hs.182507_ARNm_1 gi|15431324|ref|NM_002283.2|queratina de Homo sapiens, pelo, básico, 5 (KRTHB5),
 ARNm poliA = 3

AGCTCTCCCCACCAATAAAAGGACCAGSGAGGATCAGAGAGAGCAGAAGGATCCTGAGCC
 TCGCACTCTGCCGCCCGCACCACCTTCCGCTGCCTCTCAGACTCTGCTCAGCCTCACACG
 ATGTCGTGCCGCTCCTACAGGATCAGCTCAGGATGCCGGGTCAACAGGAACCTCAGCTCC
 TGCTCAGCTGTGGCCCCCAAACCTGGCAACCGCTGCTGCATCAGCGCCGCCCTTACCGA
 GGGGTGTCTCTACCGAGGGCTGACGGGCTTCGGCAGCCGACGCTCTGCAACCTGGGC
 TCCTGCGGGCCCCGGATAGCTGTAGGTGGCTTCCGAGCCGGCTCCTGCGGACGACGCTTC
 GGCTACCGCTCCGGGGGCTGTGCGGACCCAGCCCCCATGCATCACTACCGTGTGCGTTC
 AACGAGAGCCTCCTCACGCCCTCAACCTGGAGATCGACCCCAACGCACAGTGCCTGAAG
 CAGGAGGAGAAGGAGCAGATCAAGTCCCTCAACAGCAGGTTCCGCGCCTTCATCGACAAG
 GTGCGCTTCTTGAGCAGCAGAACAAGCTGCTGGAGACCAAGTGGCAGTTCTACCAGAAC
 CAGCGCTGCTGCGAGAGCAACCTGGAGCCACTGTTCAAGTGGCTACATCGAGACTCTGCGG
 CGGGAGGCCGAGTGCCTGGAGGCCGACAGCGGGAGGCTGGCCTCAGAGCTCAACCATGTG
 CAGGAGGTGCTGGAGGGCTACAAGAAGAAGTATGAAGAGGAGGTGGCCCTGAGAGCCACA
 GCAGAGAATGAGTTTGTCTTCTAAAGAAGGACGTGGACTGTGCCTACCTGCGGAAATCA
 GACCTGGAGGCCAATGTGGAGGCCCTGGTGGAGGAGTCTAGCTTCCTGAGGCGCCTCTAT
 GAAGAGGAGATCCGCGTTCTCCAAGCCACATCTCAGACACCTCGGTTCATAGTCAAGATG
 GACAACAGCCGAGACCTGAACATGGACTGCATCATCGCTGAGATCAAGGCTCAGTATGAC
 GATGTTGCCAGCCGACGCCGGGCCGAGGCTGAGTCTCGGTACCGTAGCAAGTGTGAGGAG
 ATGAAGGCCACGGTGATCAGGCATGGGGAGACCCTGCGCCGCACCAAGGAGGAGATCAAC
 GAGCTGAACCGCATGATCCAGAGGCTGACGGCCGAGATTGAGAATGCCAAGTGCCAGCGT
 GCCAAGCTGGAGGCTGCTGTGGCTGAGGCAGAGCAGCAGGGTGAGGCGGCCCTCAGCGAT
 GCCCGCTGCAAGCTGGCTGAGCTGGAGGGCGCCCTGCAGAAGGCCAAGCAGGACATGGCC
 TGCCCTCAAGGAGTACCAGGAGGTGATGAACCTCAAGCTGGCCCTGGACATCGAGATC
 GCCACCTACAGGCGCCTGCTGGAGGGCGAGGAACACAGGCTGTGTGAAGGTGTGGGCTCT
 GTGAATGTCTGTGTCAGCAGCTCCCGTGGTGGAGTCTCCTGTGGGGGCCCTCTCTACAGC
 ACCACCCAGGGCGCCAGATCACTTCTGGCCCTCAGCCATAGGCGGCAGCATCACGGTG
 GTGGCCCTGACTCCTGTGCCCTTGCAGCCTCGTTCTCCAGCTTCAGCTGCGGGAGT
 AGCCGGTCCGTCCGCTTTGCCCTAGTAGAGTCACTGGAGCCAGGGCTCCTGCCAAGCACCT
 GCCTGCCTGCATCACTGCCTGAATGGCATGTGAATGGAAAATGTGTGCTTGCCTCCAGA
 ATCTTCTGGATGTTCTTACAGAGGGAAAGACCTACAGAGGGAAAGACCCTCGGGCCGCTC
 CCCTGCGCCTTTTCATGCTAGGGAGATGCATCCTAGTTGTCTCCTGCGCAGCTGTTTTCA
 GAGGCATTCCAGCCCTTCACTTAACCTCTACTTAGCTCCAAAATACCTGTATCCAATTT
 GTATTATTCCCCAGCTCTCAGGGACAAGACCAGTCCCCAGCGTGGTGGTGCAGCACGGA
 AGCTCCACCTTCTGGGTGGAGGCGCCATCCTAACCATCCAGCCAGGCCACCCACAACCCG
 AGAATCAGGGAGAAAGTCCCTCCCGCAGCAGCCCTCCTCCTGGCTGGGAAGAATGGTCC

CCCAGCAAGCACTTGCTGTTTCAATCCCGTTTCATGTTTGTCTCTCTCAGACTGCCTT
 CCTGCTTCTGGGCTAACCTGTTCCAGCCAGGCTCCTCATGTGACCTCGCAGTTGAGAAGC
 CCATTATCGTGGGGCATCCTTTTGCCTACAGCCCTGGTTAGGGCACTTTGGACAGGTCT
 TGCTATTCAAGTGAACCTTTGTACATTTCAAAGAAGACTCCATGGCTGCTCCAGATGCCCC
 CTTGCTGGGTGCAGGTGGGGACTGTCCAATGCAGAGCTGGCGGGACAGAGAGTTAAGCCA
 CTTCTGGGTCTCCTTCTTATGACTGTCTATGGGTGCATTGCCCTTCTGGGTGTCTCGAT
 CTGTGTTTCAATAAATGCCGCTGCAATGCAAAAAAAAAAAAAAAAAAAAAA

>Hs.292653_contig1 AI200660|AW014007|AI341199|AI692279|AI393765|AI378686|AI695373|AW292108|

T10352|R44346|AW470408|AI380925|BF938983|AW003704|H08077|F03856|H08075|F08895|AW468398|AI865976|H22568|AI858374|AI216499 poliA = 2 poliA = 3

CAATCAGTGAAAATTCTATATTCCTTTGGCATTTTTGTGACATATTCAATTCAGTTNTAT
 GTTCCAGCAGAGATCATTATCCCTGGGATCACATCCAAATTTTCACTAAATGGAAGCAA
 ATCTGTGAATTTGGGATAAGATCCCTTCTTGGTTAGTATTACTTGCGCCGGAGCAATGTCT
 TATTCCTCGTTTTAGACATTGTGATTTCCTTCGTTGGAGCTGTGAGCAGCAGCACATTGGC
 CCTAATCCTGCCACCTTTGGTTGAAATTCCTTACATTTTTCGAAGGAACATTATAATATATG
 GATGGTCCTGAAAAATATTTCTATAGCATTCACTGGAGTTGTTGGCTTCTTATTAGGTAC
 ATATATAACTGTTGAAGAAATATTTATCTTACTCCCAAAGTTGTAGCTGGCACTCCACA
 GAGTCCTTTTCTAAATTTGAATTCACATGCTTAACATCTGGTTTGAAATAGTAAAAGCA
 GAATCATGAGTCTTCTATTTTGTCCCATTTCTGAAAATTATCAAGATAACTAGTAAAT
 ACATTGCTATATACATAAAAAATGGTAACAACTCTGTTTTCTTTGGCAGCATATTAATAT
 TTTGGAAGTAATCATAACTCTTACCAGTAGTGGTAAACCTATGAAAAATCCTTGCTTTT
 AAGTGTAGCAATAGTTCAAAAAATTAAGTCTGAAAATTGAAAAATTAATAATGTAAAA
 AATTAAAGAATAAAAAATACTTCTATTTATCTTTTATCTCAGTAAGAAATACCTTAACCA
 AGATATCTCTCTTTTATGCTACTCTTTTGGCACTCACTTGAGAACAGAAATAGGATTTCAA
 CAATAAGAGAATAAAAAATAAGAACATGTATAACAAAAAGCTCTCTCCAGATCATCCCTGTG
 AATGCCAAAGTAACTTTATGTACAGTGTAAAAAATACTCTCAGTTATGTTTAT
 TAGCCAAATCTAATGATTGGCTCCTGGAAGTATAGAAAATCCCATTAAACATAATATAA
 GCATCAGAAAATTGCAACACTAGAATTAATTTTACACTCTAATGGTAGTTGATCTTCAT
 AGTCAAGAGGCAGTGTCAAGATCATGACTTAGTGTTCATGAAATTTGAAAAGGGACT
 TTAAACTTATCCAGTGCAACTCCCTTGTTTTTCGTGAGAGGAAAGGAGCCCTAGAAAG
 GTTAAGTAACTTGGTCGAGACCACTCAGCCTTGAGATCAAGAAAACCTAATCTTCTGACT
 CCCAGGCCAGGATGTTTTATTTCTCACATCATGTCCAAGAAAAGAATAAATTATGTTCA
 GCTTAAAAA

5

>Hs.97616_ARNm_3 gi|12654852|gb|BC001270.1|BC001270 clon de Homo sapiens MGC:5069 IMAGE:3458016
 poliA = 3

CGGAGGCGGCGCCGACGGGGACTGCTGAGGCGCGCAGAGGGTCGGCGGCGCCCGGGAGCC
 TGTCGCTGGCGCGGTCCGGGCGGGAGGCTCGGCGGCGGGCGGCAGCATGTCGGTGGCGGG
 GCTGAAGAAGCAGTTCTACAAGGCGAGCCAGCTGGTCACTGAGAAGGTCCGAGGGGCGCA
 GGGACCAAGCTGGATGATGACTTCAAAGAGATGGAGAAGAAGGTGGATGTCACAGCA
 GCGGCTGACAGAGTGCTGGCCAGGACCATCGAGTACCTGCAGCCCAACCCAGCCTCGCG
 GGCTAAGCTGACCATGCTCAACACGGTGTCCAAGATCCGGGGCCAGGTGAAGAACCCCGG
 CTACCCGCGAGTCGGAGGGGCTTCTGGGCGAGTGCATGATCCGCCACGGGAAGGAGCTGGG
 CGGCGAGTCCAACCTTTGGTGACGCATTGCTGGATGCCGGCGAGTCCATGAAGCGCCTGGC
 AGAGGTGAAGGACTCCCTGGACATCGAGGTCAAGCAGAACTTCATTGACCCCCCTCCAGAA
 CCTGTGCGAGAAAGACCTGAAGGAGATCCAGCACCACCTGAAGAACTGGAGGGCCGCG
 CCTGGACTTTGACTACAAGAAGAAGCGGCAGGGCAAGATCCCCGATGAGGAGCTACGCCA
 GGCGCTGGAGAAGTTCGAGGAGTCCAAGGAGGTGGCAGAAACCAGCATGCACAACCTCCT
 GGAGACTGACATCGAGCAGGTGAGTCAGCTCTCGGCCCTGGTGGATGCACAGCTGGACTA
 CCACCGGCAGGCCGTGACAGATCCTGGACGAGCTGGCGGAGAAGCTCAAGCGCAGGATGCG
 GGAAGCTTCTCAGCCCTAAGCGGGAGTATAAGCCGAAGCCCCGGGAGCCCTTTGACCT
 TGGAGAGCCTGAGCAGTCCAACGGGGGCTTCCCTTGACACAGCCCCCAAGATCGCAGC
 TTCATCGTCTTTCCGATCTTCCGACAAGCCCATCCGACCCCTAGCCGGAGCATGCCGCC
 CCTGGACAGCCGAGCTGCAAGGCGCTGTACGACTTCGAGCCCGAGAACGACGGGGAGCT
 GGGCTTCCATGAGGGCGACGTATCACGCTGACCAACCAGATCGATGAGAACTGGTACGA
 GGGCATGCTGGACGGCCAGTCGGGCTTCTTCCCGCTCAGCTACGTGGAGGTGCTTGTGCC
 CCTGCCGAGTGACTACCCGTGTCCCGCCCCCGCCCTCCGTCCACACTGGCCGGCACC
 CCCTGTGGTCTCTCTGCAATTCACGGAGCCCTGCTGCCAGGGCGGTGTCTGAGCCTGC
 CGGCGCCACTTGGGCCCCGGCCCTTGAGGTACTCCCTGAGCAGGACCCACACTTGGGTG
 GGGGGCTTATCTGGGTGGGTGGGATGCTTGTACACTAGCGCTGACTCCCAACGGTG
 ACGGCTCCCTTCCCCACTCCATGGCGCCAGCCTCCTCCCCGCTCCCCAATCTCTCGCC

AGCTGGCCGAGGCGGGGCAACACTAAGGTGCTCTTAGAAACACTAATGTTCTCTGGGGC
AGCCCCACCTCCGTCCTGACCCGACGGGGGCCCCGGCCACTGCCTACCCTCGAGTCCCG
CAGCCTTAACAGGATGGGATCGAGGGTCCCCATGGGGTGGCTCAGAGATAGGACCCCTGGT
TTTAAATCCCTCCAGCCTGGTGCTGGTGATGGGCCCTGGCCCTACTCCAGGGCCAATGC
ACCCCCGCTCACACACGCACTCCTTCTCCTCAAGGCCAGGGCAGAGGGCCTCACC3CCT
CCCCGGCCTGCTGTCAGCTTGACGCCGGGGACAGAGGCCAGCTGGGATCTGCCTGAGGA
CAGAGAACATGGTCTCCTGCAGGGCCCTGCCTCCCAAGCCCCGCCCTCAGAAAGCCAAGT
ACCTTTTTCAGCTTTTAACTGCCCCCATCCCAACCCAGGGAGGCCTGTGTCACTCTGGCA
CAAGCTGCCACCACCAGCCACCCACACCCACCCAGCACACCTCACACGGGACCACAGCC
GCGCTGCCGAGGGCCAAGCACAAAGGTTCCAGTGAGCGCATGTCCAGCCCCCTGGTGGCC
AGGCTCCCCTTGCTGAGCCGCTGCCACTTCACCTGTGGGAAGTGGCCCCAGCCATCTCC
TCTAGACCAAGGCAGGCAGCCCCGACATCTGCTTCCTCTATCGCCCAATGCAAAATCGAT
GAAATGGGGAGTTCTCTGGGCCAGGCCACATTACATTCCCCTCCCCCTGTGGTCCAGTG
AAGCCTCCGGACCCAGGCTCTGCTCTGCCCTGCCCTGCACCCCCCTCGTCAGAAGTACA
TGAGGGGCGCAGAGATGAGCACACAGCTTTGGGCGACGGTCCAGGGCAAACCTGAAATGTAC
GCCTGAATTTTGTAAACAGAAGTATTAAATGTCTTTTCTACAAAAAAAAAAAAAAAAAA

>Hs.123078_ARNm_3 gj|14328043|gb|BC009237.1|BC009237 clon de Homo sapiens MGC:2216
IMAGE:2989823 poliA = 3

GGCACGAGGGAGGTGCAGAGCTGAGAATGAGGCGATTTCCGAGGATGGAGAAATAGCCCC
GAGTCCCGTGAGAAATGAGGCCGGCGGACTTGCTGCAGCTGGTGCTGCTGCTCGACCTGC
CCAGGGACCTGGGCGGAATGGGGTGTTCGTCTCCACCCTGCGAGTGCCATCAGGAGGAGG
ACTTCAGAGTCACCTGCAAGGATATTCAACGCATCCCCAGCTTACCGCCAGTACGCAGA
CTCTGAAGCTTATTGAGACTCACCTGAGAACTATTCCAAGTCATGCATTTTCTAATCTGC
CCAATATTTCCAGAATCTACGTATCTATAGATGTGACTCTGCAGCAGCTGGAATCACACT
CCTTCTACAATTTGAGTAAAGTGACTCACATAGAAATTCGGAATACCAGGAACCTAACTT
ACATAGACCCTGATGCCCTCAAAGAGCTCCCCCTCCTAAAGTTCCTTGGCATTTCACA
CTGGACTTAAATGTTCCCTGACCTGACCAAAGTTTATTCACCTGATATATTCTTTATAC
TTGAAATTACAGACAACCCTTACATGACGTCAATCCCTGTGAATGCTTTTCAGGGACTAT
GCAATGAAACCTTGACACTGAAGCTGTACAACAATGGCTTTACTTCAGTCCAAGGATATG
CTTTCAATGGGACAAAGCTGGATGCTGTTTACCTAAACAAGAATAAATACCTGACAGTTA
TTGACAAAGATGCATTTGGAGGAGTATACAGTGGAACCAAGCTTGCTGCTGCCTCTTGGAA
GAAAGTCTTGTCTTTGAGACTCAGAAGGCCCAAGCTCCAGTATGCCATCATGATGCC
TGCTAAGGCAGCCACCTTGGTGATCATGCTCACAGAGGCTCTGTTTCATGGAGCAGCTGCT
GTTTGAAGAAATTTTGAATGCAAGATCCACAACCTAGATGGAAGGCACTCTAGTCTTTGCA
GAAAAAATGTACCTGAATGTACATTGCACAATGCCTGGCACAAGAAGGAAGAATATAA
ATGATAGTTCGACTCGTCTGTGGAAGAACTTACAATCATGGGGAAAGATGGAATAAAAC
ATTTTTTAAACAGCAAAAAAAAAAAAAAAAAAA

5

>Hs.285508_contig1 AW194680|BF939744|BF516467 poliA = 1 poliA = 1

CCCCAGCCCCACTCACCCACCCTCCTTCCACCAGCCTGCTCTCCGAGGCCCCACTGTCT
TTGGGTTTAAATGACGTCTCTTCTCTGTGGAACCTTACGATTCTTCCCACGGTCAACTCG
GGACCTCCCAGCGACCACTGCAGCCTGCGGACGAGGCCGGGACTTGCGCGAGCGGATCCT
AATAAGGGGAAAAATGGTAAATGCAAACGTCCCGTTACAATTTTACCGCCAGTGTGCTGTC
GTTCCCCCTCCCCCTCTCCGAGTCTCTGTTGGGGACACGGCGGGGTCTGTAGGAAGTTGGG
CCGGGTTGGGGGTGCTAGAAGGCGCTGGTGTTTTGTCTCTGAGTTTTAAGAGATCCCTTC
CTTCTCTTCGGTGAATGCAGGTTATTTAAACTTTGGGAAATGTACTTTTAGTCTGTTCAT
ATCAAGGCATGAGTCACTGTCTTTTTTTGTGTGAATAAATGGTTTCTAGTACAATGGA

10

>Hs.183274_contig1
BF437393|BF064008|BF509951|AW134603|AI277015|AI803254|AA887915|BF054958|
AI004413|AI393911|AI278517|AW612644|AI492162|AI309226|AI863671|AA448864|
AI640165|AA479926|AA461188|AA780161|BF591180|AI918020|AI758226|AI291375|

15

BF001845|BF003064|AI337393|AI522206|BE856784|BF001760|AI280300 FLAG = 1 poliA = 2 WARN poliA = 3

GCGGCCGCCCCGACGTCCGCGGGTCCCCGGCCGCGCCGCGCCGCGCCCCCTGCCCCGAGA
 GAGCTCTGGCCCCGCTAGCGGGGCCAGGAGCCGGGCCTCCCACCGCAGCGTCCCCCGCCG
 CGCCAGTCCCCGCTAGTGGTAGTATCTCGTAATAGCTTCTGTGTGTGAGCTACCGTGGAT
 CTCCTTCCCTTCTCTTGGGGGCCGGGGGAAAGAAAAGGATTTAAGCAAAGGCTCCCTCG
 CCCTGTGAGGGCGAGCGGCAAAGGCCCGGCTGAGCCCCCATGCCCCCTCCCTCCCGTG
 TAAAAAGCCTCCTTGTGCAATTGTCTTTTTTTTCTTTTGAACGTGCTTCTTTGTAATGAC
 CAAGGTACCGATTTCTGCTAAGTTCTCCCAACAACATGAAACTGCCTATTCACGCCGTAA
 TTCTTTCTGTCTCCCTTCTCTCTCTCTCTCTCGCTCGCTCGCTCTCGCTCTCGCTCTCTC
 TCGCTGCGTCTCATTTCCCTCCCAATCTCTCTCCCTCTGCAACCCCCCAGCTCGCT

GGCTTTCTCTCTGGCTTCTCTCTTTTCTCTCTCCACCCACCCCTTTGGTTTGACAATTT
 TGTCTTAAGTGTCTTCTCAAAGAGGTTACTTTAGTTAGCATGCGCGCTGTGGGCAATTGT
 TACAAGTGTCTTAGGTTTACTGTGAAGAGAATGTATTCTGTATCCGTGAATTGCTTTAT
 GGGGGGAGGGAGGGCTAATTATATATTTTGTGTTTCTCTATACTTTGTTCTGTTGTCT
 GCGCCTGAAAAGGGCGGAAGAGTTACAATAAAGTTTACAAGCGAGAACCAGAAAAAAA
 AAAAA

>Hs.334841_ARNm_3 gij14290606|gb|BC009084.1|BC009084 clon de Homo sapiens MGC:9270
 IMAGE:3853674 poliA = 3

5

CACCAGCACAGCAAACCCGCCGGGATCAAAGTGTACCAGTCGGCAGCATGGCTACGAAAT
 GTGGGAATTGTGGACCCGGCTACTCCACCCCTCTGGAGGCCATGAAAGGACCCAGGGAAG
 AGATCGTCTACCTGCCCTGCATTTACCGAAACACAGGCACTGAGGCCCCAGATTATCTGG
 CCACTGTGGATGTTGACCCCAAGTCTCCCCAGTATTGCCAGGTCATCCACCGGCTGCCCA
 TGCCCAACCTGAAGGACGAGCTGCATCACTCAGGATGGAACACCTGCAGCAGCTGCTTCG
 GTGATAGCACCAAGTCGCGCACCAAGCTGGTGTCTGCCAGTCTCATCTCCTCTCGCATCT
 ATGTGGTGGACGTGGGCTCTGAGCCCCGGGCCCCAAAGCTGCACAAGGTCAATTGAGCCCA
 AGGACATCCATGCCAAGTGCGAAGTGGCTTTCTCCACACCAGCCACTGCCTGGCCAGCG
 GGGAAGTGATGATCAGCTCCCTGGGAGACGTCAAGGGCAATGGCAAAGGGGGTTTTGTGC
 TGCTGGATGGGGAGACGTTTCGAGGTGAAGGGGACATGGGAGAGACCTGGGGGTGCTGCAC
 CGTTGGGCTATGACTTCTGGTACCAGCCTCGACACAATGTCTATGATCAGCACTGAGTGGG
 CAGCTCCCAATGTCTTACGAGATGGCTTCAACCCCGCTGATGTGGAGGCTGGACTGTACG
 GGAGCCACTTATATGTATGGGACTGGCAGCGCCATGAGATTGTGCAGACCCTGTCTCTAA
 AAGATGGGCTTATTCCCTTGGAGATCCGCTTCTGCAACAACCCAGACGCTGCCCAAGGCT
 TTGTGGGCTGCGCACTCAGCTCCACCATCCAGCGCTTCTACAAGAACGAGGGAGGTACAT
 GGTCAGTGGAGAAGGTGATCCAGGTGCCCCCAAGAAAGTGAAGGGCTGGCTGCTGCCCG
 AAATGCCAGGCTGATCACCAGACATCTGCTCTCCCTGGACGACCGCTTCTCTACTTCA
 GCAACTGGCTGTCATGGGGACCTGAGGCAGTATGACATCTCTGACCCACAGAGACCCCGCC
 TCACAGGACAGCTCTTCTCGGAGGCAGCATTTGTTAAGGGAGGCCCTGTGCAAGTGCTGG
 AGGACGAGGAACTAAAGTCCCAGCCAGAGCCCTAGTGGTCAAGGGAAAACGGGTGGCTG
 GAGGCCCTCAGATGATCCAGCTCAGCCTGGATGGGAAGCGCCTCTACATCACCAGTCTGC
 TGTACAGTGCCTGGGACAAGCAGTTTACCCTGATCTCATCAGGGAAGGCTCTGTGATGC
 TGCAGGTTGATGTAGACACAGTAAAAGGAGGGCTGAAGTTGAACCCCAACTTCTTGGTGG
 ACTTCGGGAAGGAGCCCTTGGCCAGCCCTTGGCCATGAGCTCCGCTACCTTGGGGGCG
 ATTGTAGCTCTGACATCTGGATTTGAACTCCACCCTCATCACCACACTCCCTATTTTGG
 GCCCTCACTTCTTGGGGACCTGGCTTCATTCTGCTCTCTCTTGGCACCCGACCCTTGGC
 AGCATGTACCACACAGCCAAGCTGAGACTGTGGCAATGTGTTGAGTCATATACATTTACT
 GACCACTGTTTGCTTGTGCTCACTGTGCTGCTTTTCCATGAGCTCTTGGAGGCACCAAGA
 AATAAAGTCTGTAACCCCTGTCTTCAAAAAAAAAAAAAAAAAA

>Hs.3321_contig1
 A1804745|A1492375|AA594799|BE672611|AA814147|AA722404|AW170088|D11718|BG
 153444|A1680648|AA063561|BE219054|A1590287|R55185|A1479167|A1796872|A101
 8324|A1701122|BE218203|AA905336|A1681917|B1084742|A1480008|A1217994|A140 1468 poliA = 2 poliA = 3

10

CCGGAGATAA CTTGAGGGCTATAGAGGACCGGCTAATACTGGTCCTGAATTTGGCTTCAG
GCCTCACCAACCAAGTGGCCGTGGCCTTTGCCGTCTTGCCCGTCGGCCCCCGGTGAGGCCT
GGACCCCTGGGGTCCCCGGCACCAGGCCCCGGCTTCCGACCCTGGCAGAAAGCCCAAGATCT
GGTCCCTCGCGGAGACTGCCACAAGCCCCGGACACCCGCGCCGGCTCGCCTCCCGGCGCG
GGGGGGTCTCCACCGGGGGGCAACGGTCGCGCCTTTCCGCCCTGCAGCTCTCTCCGGGGCC
GCCGCCGCCGCCGCCGCTCACAGACTGGTCTCAGCGCCGCTGGGCAAGTTCCCGGCTTGG
ACCAACCGGCCGTTTCCAGGCCACCGCCCGGCCCGGCCCGCACCCGCTCTCCCTGCTG
GGCTCTGCCCCCTCCGCACCTGCTGGGACTTCCCGGAGCCGCGGGCCACCCGGCTGCCGCC
GCCGCTTCGCTCGGCCAGCGGAGCCCGAAGGCGGAACAGATCGCTGTAGTGCCCTTGGAA
GTGGAGAAAAAGTTACTCAAGACAGCTTTCCATCCCGTGCCAGGCGGCCCGAACCAT
CTGGACGCCGCCCTGGTCTTATCGGCTCTCTCCTCATCCTAGTTCTTTAAAAAAAACAA
AAAAACAAAAAACTTTTAAATCGTTGTAATAATTGTATAAAAAAATCGCTCTGTA
TAGTTACAACTTGTAAGCATGTCCGTGTATAAATACCTAAAAGCAAACTAAACAAAGAA
AGTAAGAAAAAGAAATAAAACAGTCCTCTCAGCCCTCCCCAAGTCGCTTCTGTGGCAC
CCCGCATTGCTGTGAGGTTTGTGTGTCGGTGTGATTTTGGGGGGTGGAGTTTCAGTGAG
AATAACGTGTCTGCCCTTGTGTGTGTGTATATATACAGAGAAATGTACATATGTGTGAA
CCAAATTGTACGAGAAAGTATCTATTTTGGCTAAATAAATGAGCTGCTGCCACTTTGAC
TATAAAAAA

>Hs.306216_single1 AW083022 poliA = 1 poliA = 2

TATGAGCACCTTCACATGGATCCACTTGAGGAAAGAAGGTGGACCGAATTTGTAAACGGT
GTGCAGCAATATATATCAATTCGTTCTGAGATAATCGCCACTTACGCTCTCTGTGGTTTT
GCCAATATCGGGTCCCTAGGAATCGTGATCGGCGGACTCACATCCATGGCTCCTTCCAGA
AAGCGTGATATCGCCTCGGGGGCAGTGAGAGCTCTGATTGCGGGGACCGTGGCCTGCTTC
ATGACAGCCTGCATCGCAGGCATACTCTCCAGCACTCCTGTGGACATCAACTGCCATCAC
GTTTTAGAGAATGCCTTCAACTCCACTTTCCCTGGAACCCCAACCAAGGGTGATAGCTTG
TTGCCAAAGTCTGTTGAGCAGCCCTGTTGCCAGGGTCTGGTGAAGTCATCCAGGAGG
AAACCCAGTCTGTATTCTTTGAAGGGCTGCTGCACATTGTTGAATCCATCGACCTTTAG
CTGCAATGGGATCTCTAATACATTTTGAAGGTGAGCCACTTCTCCAGTGGAACTCTGAAGT
ACAGATGCTGAATTTCTGCTTTGGAAGAAAAA

5

>Hs.99235_contig1 AA456140|AI167259|AA450056 poliA = 2 poliA = 3

ACTCGGCATGTGATGAACACCCATAGTTAAGAAACCATGGAGCAAGAAAGCTTGTGGAAA
GTCTCTCTCCTTCCCTCATAAGACATGCACACTAATACACATACACACCAAAAAATTACAC
ATTTTAAACTGCTAAGCTTGGATTAACTGAATCATATATCTTTTATCATGTTATCCTA
AAAGTGAGAAAGACATAACCAAGACATGGAAATAAATGTGAAAGCTGGAGCCGAAGAGTCA
AAGAGCTAAAAAATTAAGTCTAGAACATTCTATGAGGATAGTATAAATAAAGAAATAC
AGTCTAGACATGCTGCAAGGAAAGAAGATTCTAAAGTCCGTTTATGGAGGCAATTCCATA
TCCTTTCTTGAACGCACATTCAGCTTACCCAGAGAGCAAGTGAGGCAATCTGGCAAAAG
ATTAATAAAGATGTAAACCCCTGGAAAAA

10

>Hs.169172_ARNm_2 gi|2274961|emb|AJ000388.1|HSCANPX ARNm de Homo sapiens para proteasa de tipo calpaína CANPX poliA = 3

GAATTCGGCACGAGATAGTTTTTCAGGTTAAGAAAGCCAGAATCTTTGTTTCAGCCACACTG
 ACTGAACAGACTTTTAGTGGGGTTACCTGGCTAACAGCAGCAGCGGCAACGGCAGCAGCA
 GCAGCAGCAGCAGCAGCAGCAGCAGCGGGCTCCTGGGATAACTCAGGCATAGTTCAACA
 CTATGGGTCCCTCCTCTGAAGCTCTTCAAAAACCAGAAATACCAGGAACTGAAGCAGGAAT
 GCATCAAAGACAGCAGACTTTTCTGTGATCCAACATTTCTGCCTGAGAATGATTCTCTTT
 TCTACTTCCGACTGCTTCCTGGAAAGGTGGTGTGGAAACGTCCCCAGGACATCTGTGATG
 ACCCCCATCTGATTGTGGGCAACATTAGCAACCACCAGCTGACCCAAGGGAGACTGGGGC
 ACAAGCCAATGGTTTCTGCATTTTCTGTGTTGGCTGTTTCAGGAGTCTCATTGGACAAAGA
 CAATTCCTCAACCATAAGGAACAGGAATGGGACCTCAAAAACAGAAAAATACGCTGGGA
 TATTTCACTTTTCGTTTCTGGCATTTTGGAGAATGGACTGAAGTGGTGATTGATGACTTGT
 TGCCCAACCATTAACGGAGATCTGGTCTTCTCTTCTCCACTTCCATGAATGAGTTTGGGA
 ATGCTCTGCTGGAAAAAGCTTATGCAAAGCTGCTAGGCTGTTATGAGGCCCTGGATGGTT
 TGACCATCACTGATATTATTGTGGACTTCACGGGCACATTGGCTGAAACTGTTGACATGC
 AGAAAGGAAGATACACTGAGCTTGTGAGGAGAAAGTACAAGCTATTTCGAGAACTGTACA
 AAACATTTACCAAAGGTGGTCTGATCTGCTGTTCCATTGAGTCTCCAATCAGGAGGAGC
 AAGAAGTTGAACTGATTGGGGTCTGCTGAAGGGCCATACCTATACCATGACTGATATTC
 GCAAAATTCGCTCTTGAGAGAGACTTGTGGAAGTCTTCAGTGTGAGAAGGTGTATATGG
 TTCGCTGAGAAACCCCTTGGAAGACAGGAATGGAGTGGCCCTGGAGTGAAATTTCTG
 AAGAGTGGCAGCAACTGACTGCATCAGATCGCAAGAACCTGGGGCTTGTATGTCTGATG
 ATGGAGAGTTTTGGATGAGCTTGGAGGACTTTTGCCGCAACTTTCACAACTGAATGTCT
 GCCGCAATGTGAACAACCTATTTTTGGCCGAAAGGAGCTGGAATCGGTGTTGGGATGCT
 GGACTGTGGATGATGATCCCCTGATGAACCGCTCAGGAGGCTGCTATAACAACCGTGATA
 CCTTCTGTCAGAAATCCCAGTACATCTTCACTGTGCTGAGGATGGGCACAAGGTCATTA
 TGTCACTGCAGCAGAAGGACCTGCGCACTTACCGCCGAATGGGAAGACCTGACAATTACA
 TCATTGGCTTTGAGCTCTTCAAGGTGGAGATGAACCGCAAATTCGCTCCACCACCTCT
 ACATCCAGGAGCGTGTGCTGGGACTTCCACCTATATTGACACCCGCACAGTGTCTCTGAGCA
 AGTACCTGAAGAAGGGCAACTATGTGCTTGTCCCAACCATGTTCCAGCATGGTCCGACCA
 GCGAGTTTCTCCTGAGAATCTTCTCTGAAGTGCCTGTCCAGCTCAGGGAACCTGACTCTGG
 ACATGCCCCAAAATGTCTGCTGGAACCTGGCTCGTGGCTACCCGAAAGTAGTTACTCAGA
 TCACTGTTTACAGTGTGCTGAGGACCTGGAGAAGAAGTATGCCAATGAACTGTAAACCCAT
 ATTTGGTCATCAAATGTGGAAAGGAGGAAGTCCGTTCTCCTGTCCAGAAGAATACAGTTC
 ATGCCATTTTTGACACCCAGGCCATTTTCTACAGAAGGACCACTGACATTCTATTATAG
 TACAGGTCTGGAACAGCCGAAAATTCTGTGATCAGTTCTTGGGGCAGGTTACTCTGGATG
 CTGACCCAGCGACTGCCGTGATCTGAAGTCTCTGTACCTGCGTAAGAAGGGTGGTCCAA
 CTGCCAAAGTCAAGCAAGGCCACATCAGCTTCAAGGTTATTTCCAGCGATGATCTCACTG
 AGCTCTAAATCTGCAATCCCAGAGAATCCTGACAAAGCGTGCCACCCTTTTATTTTCCGT
 CAGGTGCCAGGTCTTAGTTAAGATTACAATCTTTAGAAAGAATGAGATTACAATAATT
 AACTCTTCTCTCTTCTGATAAATCCCCATACCTCCCAATCCAAGTAGCATCTGTAGCT
 ACATAACCTATATACCTCCAGCAGCTGGACATGGGGAGCGACAGTCTATCTAGACATCA
 TACACATTTGCCAAGAAAGGATCTCTGGGGCTTCCGGGGTGAGATTCAAGCAGGACAAT

AACAAGAGGCTGGACACCCCTACAGATGTCTTTGATGTTTTAGTTGTTTGATATATCTCC
CCTGTAGGGCATGTTGAGGAAGGAGGAGGGCTGATCAAGGCCAAGCTGGTCTAGCCTGAC
ATCCTAGCTCCTGACTGAACACTATAGACTTCCCAGCAGCATTTCACCCAGCAGCCAGA
GCCGGCTTTAAGTCCCCAACCCCTTACAGACACCACTGCCACCACCACCAACCACGACCAC
CACCACCACCACCACTCACCACCATCATCACCTCCGGAAGTGTAGTCCCTGCCCTAACCC
TAACCCCAAGTCACCCCCACAGTAAATTTTACCTTCATGTTGAGAAAGCTTCCTGGTGC
TTAATCAAGAGCTGGAGTTCAATGAGTCCCTAGACAGTGAGAGGGGCTGAGCTTCAGCTC
AATGGAAGCCTGCTGTGTGCTCACAAGACGGAAGTGGGAAGAAGCTGCAGTGGGAGACA
AAGCCTCGGTCCCCCACCATCCACACACACCTACACTCACACACGCGCACATGGGCGCG
CAACGGAATACCATTTTCAGGCAGTCAGTGGGCAAGAGGAAAGATAAGTAAGTACCATAC
ACACCTTAAAAGATGAGGAGAATTCATCCAGACATATTACAGCCAGTTTGGGGCCCTGA
CTTGCAATGTGAAACCTCTTCGCTTGCTGCTAGGTTTACAAACAAGCCCATTTGTTCTGT
GCCTCCTAATATTTCATTTGTTACTGAAGGACCCCATCTGGGGACTTGAGACTTTGGTCCC
AGCCCAGACGCCTCAGACTGGTCTCAAAGTCAAGCAAGGCTTCACATCAGCTGCAAGTGT
TAGTTTGCCAGCGCATGATCTCACTGAGCTTCTACAGAATCTGCAATCCAGAGTCAATC
ATGACGAAATGTACGTCCCAACATCTTAACCTATCAACTTTCTGCCCCCTCCTTCAAGGCC
CAGTATAAATGCCACCTCCTCCATGAAGCCTTCCCTAATTCACCCCCAACCCCCACCTT
CAACAATATTTCAACGCTTCTGCAATGATGAAAAAGAAACATAGTTGTAGTACTTAGCCT
ACCTAGACCAGCAAGCATTCATTTTCTAGCTCGCTCATTTTTTACCATGTTTTCCAGTCTG
TTTAACTTCTGCACTGCCTTCACTACACTGCCTTACATAAACCAATCACAATAAAGTTC
ATATTTCAGTACAATTAATAAAAAAAAAA

>Hs.351486_ARNm_1_gi|16549178|dbj|AK054605.1|AK054605 ADNc de Homo sapiens FLJ30043 fis, clon
3NB692001548 poliA = 0

5

TATGCAAGTGTTTAAACAGATGCTTCACTATTAAATATTTTTCCCCCAAGTCTCAAATAT
TGAAGAATCTCTAACAGGGACACCAGTCCCTACGAAGACCTTGGGCGATTTTGAAGTGC
GGGCACCTCGATTCCCCGAATCTGTAGTGTGGCTGGTATCGGTGTTCCCTGGTTTAACT
AGCCTGTTTGAAGGCACAGATCATTCATGGGGAAGTATAACCGAATCCAGTCTCTCCAC
CGCCTGGGGATCTTCACTTTCGAGTCTACGACTGCCTGTGACTCCAGAAAGACAACTG
CAGATTGGCCAAGATGGGGAATTGAGGCAGAGAAGCCAAGACATGTGCTAAAGTTCATG
CAGGCTATGAATGGAGCTGGAATGTGAACGCAGGCCATATGACCCAGAGCCCATGTTCT
TGAACCCTTAGAAAGACAGCAGCAACACACCTGGTGCAGCAGCTGCTTAGTTGGAGTGGC
TGACAAGGAGAGAATGATTTCCAGGAAGAGCGGAACACATATGGAAGGCCCTTAGCTTATC
TTTAGCGCCTCATACACCCGTTCTGGACTTCAGAAAGGCCAGTGAGTGGGATTAGGCCTC
AGAGATAGGATGTGAGTCCCAGTGAGGGATGGCCTAGAGCATTCCTTAATTCCTTCCCTT
GGGTACACATAAGAAACAATTTCCAGCACTGATGAGTGTATTAACAATGAGATGGGA
TAGAATTTAGTTTCCCTATGGCTGTGCTTCAAAAATAGAAAAGCTGTCTTTCTCTGGA
ATGATTGAATGAAGCTCTGGGAGGAAAAGGTGGATTGGCAGATCTCTTAAAGGAAGCTT
CTCCTTCTAGGCATATTCTAAGGCTTAATATTTTAACTCCCTATATTAACCTAGTTCAA
CTAAACAGTGATCTGAGTAATTTTATTTTATTAAAGCTCAGATCAAAATGCCATTAAACA
TTGATTGAGAAAATCAAAGGAATCTTTGATGTGAGTGGTTAAATTGGAATTTTTCAG
TCCCATACCTCACAGCATGAGTACCTGATCTGATAGACTTCTTTGGAATTCCTTTTTTG
TTTGAGACAGAGTCTTGCTCTGTGCGCCAGGCTGGAGTGCAGCGGTGTGATCTCAACCAT
TGCAACCTCCACCTCCAGGTTCAAGTGATTCTCATGCCTCAGCCTCCTGAGTAGCTGGG
ATTACAGATGTGCACCACCATGCCCGGCTAATTTATTTTGTATCTTTAGTAGAGATGAAGT
TTTGCCATGTGGGCCAGGCTGTTCTCAAACTACTGGCCTCAAGTGATCTGCCCGCCTCGG
CCTCCCAGACTGCTGGGATTACAGGCGTGAGGCACCGTGCCTGGCTGGGATTCCATAATA
AATCCCTCTGTGTCTATTTCTTTTCAAATATAATTTTCTTCATTTCCAAACATCATCT
TTAAGACTCCAAGGATTTTCCAGGCACAGTGGCTCATACCTGTAATCCCATGTCTGGA
GAGGCCAAGGTGGAAGTTCATTTGAGGCCAGGAGTTCGAGACCAGGTGGGCAACATAGTG
AAACCTTGTCTCTACAACAT

>Hs.153504_contig2
BE962007|AW016349|AW016358|AW139144|AA932969|AI025620|AI688744|AI865632|
AA854291|AA932970|AU156702|AI634439|AA152496|AI539557|AI123490|AI613215|
AI318363|AW105672|AA843483|AI366889|AW181938|AI813801|AI433695|AA934772|
N72230|AI760632|BE858965|AW058302|AI760087|AI682077|AA886672|AI350384|AW

10

243848|AW300574|BE466359|AI859529|AI921588|BF062899|BE855597|BE617708 poliA = 2 poliA = 3

TGTTTATATAACTGTGTTTCGTTTTTGTGTGTTCCGTCCCGTCGTCCTTGTAGACTCTCATC
CTCGTGTGTTTTGGACCCCTCCAGGGGTGACATCGGGTCTTGTGTTTCAGCTCTCCTGGACT
GTTATTCCTTGTCCGCGTGTTCGTGTTAGACATTGTCCACGATCTGTATCATGCCTATGT
CTCATTTCGGTCTCTTATTTTCAGCGTGAACACTATAGTTCCAAGTTTGTTCGGATAATTC

TGATTCTTGTCAACAGCGTGAGATTTCAACAGAACTTGTTCGGAACAAATACTCACTTAA
AACTTCAGCAGAAGAAAAATTACTTAGTCCTTAGGCCAACCAATTTAACTGCAGTGTGAT
GTTTCACAGGCCCTTCCTACATTTAGAAATCGTCACACAGCTGTGATAAGAGTAGATTATT
TTACTATGAAATAATTCTGAATAGATGAAAGCATAAAATGTGAGAACTGAATGTATTAT
TCAGGAAGAATACTGAGTGCCTTCATTTAACTAAAGTTGAATGTAAAAGTCAATTTGCAC
TTCTTTATAATCCTCTGGTTTAGAATTATAAATTGTAAAACCTTGATAATTGTCATTTA
ATTATATTTTCAGGTGTCCTGAACAGGTCAGTACTGACTCTACATTGGGCAGCCTTTAAATAT
GATTCTTTGTAATGCTAAATAGCCTTTTTTTCTCTTTTACTGCAACTTAATATTTCTAT
TTAGAACACAGAAAATGAAAATATTTAGAATAAGTTGTACATTTGATGACAAATAAATCA
CTATTAAAAATAAAAAAAAAAAAAAAAAAAAAA

5 >Hs.199354_single1 AI669760 poliA = 1 poliA = 2

AGGAACCCCTGTGGGAAAGGTTTTAAACCTAAACAGTGCCCCCTTTGGCTCCTCCTCCCT
TGGCGGAATGGGTTCTCGGACCATGTGCATTTTCANTGGGCCATGGGATTTACATTTCTCT
GCATCCCCAGGTGGTTTGATCCCTGCCAGGGCCCCCTTCCTTCCTGCTCATGGTTTTTCAGG
GGGCCTGATCATGGAAGTAAGGGGGTTGGGCCTTCCTTTTGGGGGTGAACCCCTGACTC
CATCCCCCTATTGCCCCCTAACCAATCATGCAAACTTTCCCCCCTGGGGTAATTCAC
CAGTTAAAAAAGCTTTTTTTAAATGTTTTGTTTTGGGGGGGGGAGGGCCCCCTTTTT
GTTTTTTTAAGGAGTTGGTTTTGGTTTTGGCTGATGTTTTGTTTTTAACATGCCCCCA
GTTTGTAAGGCCAAAGGTAAA
AAAAAAA

10 >Hs.162020_contig1 AW2911891AA505872 poliA = 2 poliA = 3

TAAGCTTTAAAGGCTCTGTGTTAGGGCATAGTCTAGAAACATGGGGCCCCAAGGGCACCGG
GAAACTTACAAAGGGAAGAGATGGAAGTGGGAGGGTTCAAGCTACCAGTTCCATCTCTC
CATGTTTTAGAGAATTGGGGCACTAAGTCAGCCAGGTAAAGTCAGGTGAGAGAGGGCCC
GGATGAAGCATGAGATGCAGAGGGACAGTGCCTGAATGGAGACCTTGGGTAGCACCACG
TGAGCGGCAGAGGTGGGGTGGATGTGGCTGATGTCAGGGAGAGAATGGGGAGCATGCAC
AGGGCTCAGTCTTATACATACATTGAAAATCCTTTAGCCTTTCAAAGATTATTAACCCAA
ATCACCTTTCTTGCTTACTCCAGATGCCTCAGCCTCTGATATAATTGCTAAGTATCTGCC
GTGTTAAAAATAAACATTTGAGAATCAAAAAAAAAAAAAAAAAAAAAA

>Hs.30743_ARNm_3 gij18201906|ref|NM_006115.2| antígeno en melanoma expresado preferentemente en Homo sapiens (PRAME), ARNm poliA = 3

15

GCTTCAGGGTACAGCTCCCCCGCAGCCAGAAGCCGGGCTGCAGCGCCTCAGCACCGCTC
 CGGGACACCCCCACCCGCTTCCCAGGCGTGACCTGTCAACAGCAACTTCGCGGTGTGGTGA
 ACTCTCTGAGGAAAAACCATTTTGATTATTACTCTCAGACGTGCGTGGCAACAAGTGACT
 GAGACCTAGAAATCCAAGCGTTGGAGGTCCTGAGGCCAGCCTAAGTCGCTTCAAAATGGA
 ACGAAGGCGTTTGTGGGGTTCCATTAGAGCCGATACATCAGCATGAGTGTGTGGACAAG
 CCCACGGAGACTTGTGGAGCTGGCAGGGCAGAGCCTGCTGAAGGATGAGGCCCTGGCCAT
 TGCCGCCCTGGAGTTGCTGCCAGGGAGCTCTTCCCGCCACTCTTCATGGCAGCCTTTGA
 CGGGAGACACAGCCAGACCCTGAAGGCAATGGTGCAGGCCCTGGCCCTTACCTGCCTCCC
 TCTGGGAGTGCTGATGAAGGGACAACATCTTACCTGGAGACCCTTCAAAGCTGTGCTTGA
 TGGACTTGATGTGCTCCTTGGCCAGGAGGTTGCCCCAGGAGGTGGAAACTTCAAGTGCT
 GGATTTACGGAAGAACTCTCATCAGGACTTCTGGACTGTATGGTCTGGAAACAGGGCCAG
 TCTGTACTCATTTCCAGAGCCAGAAGCAGCTCAGCCCATGACAAAGAAGCGAAAAGTAGA
 TGGTTTGAGCACAGAGGCAGAGCAGCCCTTCATTCCAGTAGAGGTGCTCGTAGACCTGTT
 CCTCAAGGAAGGTGCCTGTGATGAATTGTTCTCTACCTCATTGAGAAAGTGAAGCGAAA
 GAAAAATGTACTACGCCTGTGCTGTAAGAAGCTGAAGATTTTGTCAATGCCCATGCAGGA
 TATCAAGATGATCCTGAAAATGGTGCAGCTGGACTCTATTGAAGATTTGGAAGTGACTTG
 TACCTGGAAGCTACCCACCTTGGCGAAATTTTCTCCTTACCTGGGCCAGATGATTAATCT
 GCGTAGACTCCTCCTCTCCACATCCATGCATCTTCTTACATTTCCCCGGAGAAGGAAGA
 GCAGTATATCGCCAGTTTACCTCTCAGTTCCTCAGTCTGCAGTGCCTGCAGGCTCTCTA
 TGTGGACTCTTTATTTTCTTCTAGAGGCCGCTGGATCAGTTGCTCAGGCACGTGATGAA
 CCCCTTGGAAACCCTCTCAATAACTTACTGCCGGCTTTCGGAAGGGGATGTGATGCATCT
 GTCCCAGAGTCCCAGCGTCAGTCAGCTAAGTGTCTGAGTCTAAGTGGGGTCATGCTGAC
 CGATGTAAGTCCCAGCCCTCCAAGCTCTGCTGGAGAGAGCCTCTGCCACCCTCCAGGA
 CCTGGTCTTTGATGAGTGTGGGATCACGGATGATCAGCTCCTTGCCCTCCTGCCTTCCCT
 GAGCCACTGCTCCCAGCTTACAACCTTAAGCTTCTACGGGAATTCATCTCCATATCTGC
 CTTGCAGAGTCTCCTGCAGCACCTCATCGGGCTGAGCAATCTGACCCACGTGCTGTATCC
 TGTCCCCCTGGAGAGTTATGAGGACATCCATGGTACCCTCCACCTGGAGAGGCTTGCTTA
 TCTGCATGCCAGGCTCAGGGAGTTGCTGTGTGAGTTGGGGCGGCCAGCATGGTCTGGCT
 TAGTGCCAAACCCTGTCTCACTGTGGGGACAGAACCTTCTATGACCCGGAGCCCATCCT
 GTGCCCCGTGTTTCATGCCTAACTAGCTGGGTGCACATATCAAATGCTTCATTCTGCATAC

TTGGACACTAAAGCCAGGATGTGCATGCATCTTGAAGCAACAAAGCAGCCACAGTTTCAG
 ACAAATGTTTCAAGTGTGAGTGAGGAAAACATGTTTCAAGTGAGGAAAAACATTAGACAAAT
 GTTCAAGTGAGGAAAAAAGGGGAAGTTGGGGATAGGCAGATGTTGACTTGAGGAGTTAAT
 GTGATCTTTGGGGAGATACATCTTATAGAGTTAGAAATAGAACTCTGAATTTCTAAAGGGA
 GATTCTGGCTTGGGAAGTACATGTAGGAGTTAATCCCTGTGTAGACTGTTGTAAAGAAAC
 TGTGAAAAATAAGAGAAGCAATGTGAAGCAAAAAAAAAAAAAAAAAAAAA

>Hs.271580_contig1

Al632869|AW338875|AW613773|Al982899|AW193151|BE206353|BE208200| Al811548|AW264021
 poliA = 2 poliA = 3

5

AACACAGCCCTACCAANCAATGATGACCAGTGGAAAAACAATGAAGTCACCAAACCCTGGA
 CAGGGCTCATGCTCCAGGACAANTTGTGTGGCGTAAATGGTCCATCAGACTGGCAAAAA
 TACACATCTGCCTTCCGGACTGAGAATAATGATGCTGACTATCCCTGGCCTCGTCAATGC
 TGTGTTATGAACAATCTTAAAGAACCTCTCAACCTGGAGGCTTGTAACCTAGGCGTGCT
 GGTTTTTATCACAATCAGGGCTGCTATGAACTGATCTCTGGTCCAATGAACCGACACGCC
 TGGGGGGTTGCCTGGTTTGGATTGCCATTCTCTGCTGGACTTTTGGGTTCTCCTGGGT
 ACCATGTTTCTACTGGAGCAGAATTGAATATTAAGCATAAAGTGTGACCACCATACCTCCT
 TCCCCGAGTGACTCTGGATTGTTGCTGGAACAGCTCTCTCCTAATATTCACGTTTGT
 GCCCCACACTAACGTGTGTCTTACATTGCCAAGTCAGATGGTACGGACTTCCTTTAGG
 ATCTCAGGCTTCTGCAGTTCTCATGACTCTTACTTTTCTCCTAGTCTAGCATCTTGCAA
 CATTATATAGACTGTTGAAAGGAGAATTTGAAAAATGCATAATAACTACTTCCATCCCT
 GCTTATTTTTTAATTTGGGAAAAATAAATACATTCTGAAGGAAAAAAAAAAAA

>Hs.69360_ARNm_2 gi|14250609|gb|BC008764.1|BC008764 clon de Homo sapiens MGC:1266 IMAGE:3347571
 poliA = 3

10

GGCACGAGGGCGAAATTGAGGTTTCTTGGTATTGCGCGTTTCTCTTCCTTGCTGACTCTC
 CGAATGGCCATGGACTCGTCGCTTCAGGCCCCGCTGTTTCCCGGTCTCGCTATCAAGATC
 CAACGCAGTAATGGTTTAAATTCACAGTGCCAATGTAAGGACTGTGAACCTGGAGAAATCC
 TGTGTTTTCAGTGGAATGGGCAGAGGAGGTGCCACAAAGGGCAAAGAGATTGATTTTGAT
 GATGTGGCTGCAATAAACCCAGAACTCTTACAGCTTCTTCCCTTACATCCGAAGGACAAT
 CTGCCCCTTGCAGGAAAATGTAACAATCCAGAAACAAAAACGGAGATCCGTCAACTCCAAA
 ATTCCTGCTCCAAAAGAAAGTCTTCGAAGCCGCTCCACTCGCATGTCCACTGTCTCAGAG
 CTTGCGATCACGGCTCAGGAGAATGACATGGAGGTGGAGCTGCCTGCAGCTGCAAACTCC
 CGCAAGCAGTTTTTCAGTTCTCTCCTGCCCCACTAGGCCTTCCTGCCCTGCAGTGGCTGAA
 ATACCATTTGAGGATGGTCAGCGAGGAGATGGAAGAGCAAGTCCATTCCATCCGAGGCAGC
 TCTTCTGCAAAACCCTGTGAACCTCAGTTCGGAGGAAATCATGTCTTGTGAAGGAAGTGGA
 AAAATGAAGAACAGCGAGAAGAGAAGAAGGCCCAGAACTCTGAAATGAGAATGAAGAGA
 GCTCAGGAGTATGACAGTAGTTTTCCAACTGGGAATTTGCCCGAATGATTAAAGAATTT
 CGGGCTACTTTGGAATGTATCCACTTACTATGACTGATCCTATCGAAGAGCACAGAATA
 TGTGTCTGTGTTAGGAAACGCCCCACTGAATAAGCAAGAATTGGCCAAGAAAGAAATTTGAT
 GTGATTTCCATTCTTAGCAAGTGTCTCTCTTGGTACATGAACCCAAGTTGAAAGTGGAC
 TTAACAAAGTATCTGGAGAACCAAGCATTCTGCTTTGACTTTGCATTTGATGAAACAGCT
 TCGAATGAAGTTGTCTACAGGTTTCACAGCAAGGCCACTGGTACAGACAATCTTTGAAGGT
 GGAAAAGCAACTTGTTTTGCATATGGCCAGACAGGAAGTGGCAAGACACATACTATGGGC
 GGAGACCTCTCTGGGAAAGCCCAGAATGCATCCAAAGGGATCTATGCCATGGCCTCCCGG
 GACGTCTTCTCTCTGAAGAATCAACCCTGCTACCGGAAGTTGGGCCTGGAAGTCTATGTG
 ACATTTCTTCGAGATCTACAATGGGAAGCTGTTTGACCTGCTCAACAAGAAGGCCAAGCTG
 CGCGTGCTGGAGGACGGCAAGCAACAGGTGCAAGTGGTGGGGCTGCAGGAGCATCTGTTT
 AACTCTGCTGATGATGTATCAAGATGATCGACATGGGCAGCGCCTGCAGAACCTCTGGG
 CAGACATTTGCCAACTCCAATTCTCTCCCGCTCCACGCGTGCTTCCAAATTATTCTTCGA
 GCTAAAGGGAGAATGCATGGCAAGTTCTCTTTGGTAGATCTGGCAGGGAATGAGCGAGGC
 GCGGACACTTCCAGTGCTGACCGGCAGACCCCGCATGGAGGGCGCAGAAATCAACAAGAGT
 CTCTTAGCCCTGAAGGAGTGATCAGGGCCCTGGGACAGAACAAGGCTCACACCCCGTTC
 CGTGAGAGCAAGCTGACACAGGTGCTGAGGGACTCCTTCATTGGGGAGAACTCTAGGACT
 TGCATGATTGCCACGATCTCACCAGGCATAAGCTCCTGTGAATATACTTTAAACACCCTG
 AGATATGCAGACAGGGTCAAGGAGCTGAGCCCCCAAGTGGGCCAGTGGAGAGCAGTTG
 ATTCAAATGGAAACAGAAGAGATGGAAGCCTGCTCTAACGGGGCGCTGATTCCAGGCAAT
 TTATCCAAGGAAGAGGAGGAAGTGTCTTCCAGATGTCCAGCTTTAACGAAGCCATGACT
 CAGATCAGGGAGCTGGAGGAGAAGGCTATGGAAGAGCTCAAGGAGATCATACAGCAAGGA
 CCAGACTGGCTTGAGCTCTCTGAGATGACCGAGCAGCCAGACTATGACCTGGAGACCTTT
 GTGAACAAAGCGGAATCTGCTCTGGCCCAGCAAGCCAAGCATTCTCAGCCCTGCCAGAT
 GTCATCAAGGCCCTTGCGCCTGGCCATGCAGCTGGAAGAGCAGGCTAGCAGACAAATAAGC
 AGCAAGAAACGGCCCCAGTGACGACTGCAAATAAAAAATCTGTTTGGTTTGACACCCAGCC

TCTTCCCTGGCCCTCCCCAGAGAACTTTGGGTACCTGGTGGGTCTAGGCAGGGTCTGAGC
 TGGGACAGGTTCTGGTAAATGCCAAGTATGGGGGCATCTGGGCCCAGGGCAGCTGGGGAG
 GGGGTGAGAGTGACATGGGACACTCCTTTTCTGTTCTCAGTTGTGCGCCCTCACGAGAGG
 AAGGAGCTCTTAGTTACCCCTTTTGTGTTGCCCTTCTTCCATCAAGGGGAATGTTCTCAG
 CATAGAGCTTTCTCCGCAGCATCCTGCCTGCGTGGACTGGCTGCTAATGGAGAGCTCCCT
 GGGGTTGTCTGGCTCTGGGGAGAGAGAGCGAGCCCTTAGTACAGCTATCTGCTGGCTCT
 AAACCTTCTACGCCTTTGGGCCGAGCACTGAATGTCTTGTACTTTAAAAAATGTTTCTG
 AGACCTCTTTCTACTTTACTGTCTCCCTAGAGATCCTAGAGGATCCCTACTGTTTTCTGT
 TTTATGTGTTTATACATTGTATGTAACAATAAAGAGAAAAAATAAAAAAAAAAAAAA
 AAAAAAAAAAAAAA

>Hs.30827_contig1 H07885|N39347|W85913|AA583408|W86449 poliA = 2 poliA = 3

ATCGGACTTCGGTNAACTNTGGCAAGGATTGGACAGNCTAGGTAGGCTAAATGTGTGCTC
TGTCCCTGTTTGCTTCAACAGAGGAGCAAGCCTCAGCTGAGAAGGAGGGCACNTGGAACA
CCTAGCTCCTCCCGTGATTCCCCAAACCCATAACATTCTTCCATAGGGCTGGAACCACTG
CCCCGTCTTGACAGGGATGAAAAGTGAACCCCTCAGGTACAGAGAGGCCAGAGTTGAGGT
TCTGCCACTTCTGTCCCTGGGGAGCCACTCAAGTTACCAGGGCTACCGGCTGAAATAAA
TCTTTTCCGGGTAGGGTCAAGGGCAGTGTGTTCGAAGGCAACTGATGTAGGCCAGTTGCG
TGACTCCAGGTTTGTCTGTACTCAGTGGGTCCAATCACCTGGCATTGATCACCTGGCA
TTGATCAGCACCCACCCACCCCTGAGGCTTGCCAGCCCCCAGGCCCTCAGATCCCTGC
TCTTCTGCCTTTCTGCCCCATGTGTACCCAGCACCCAAGGTTCAGTGACACAGGGTGG
TTTGGAGCTGGTCACTGTCTATAGCAGCTGTGATTTTACAAGGAAGGGTGCTGCAGGGGGA
CCTGGTTGATGGGGAGTGGGAAGGGGAAGGAATAAAGAGATCTTCCTCAGGTAAAAAAA
AAAAAAA

>Hs.211593_contig2

BF592799|AI570478|AA234440|R40214|BE501078|AW593784|AI184050|AI284161|W7

5 2149|AW780437|AI247981|AW241273|H60824 poliA = 2 poliA = 3

ACCTCGTTTGCTCCAGTTACTTCTTATCTGGAGCAGTAATGTAGTCCACTTCACTCATG
CCTACCCCGCGTGTCTCGTCTCCTGACATGTCTCACAGACGCTCCTGAAGTTAGGTCATT
ACCTAACCCATAGTTATTTACCTTGAAAAGATGGGTCTCCGCACTTGGAAGGTTTCAAGA
CTTGATACTGCAATAAATTATGGCTCTTCACTGGGCGCCAAGTGTGATCAACGAAATG
CTTGTGAATCAGGGGCAACCGAGTACAGACGCTCTCAAGACTGAAACGGCCCCATTGCC
TGGTCTAGTAGCGGATCTCACTCAGCCGCAGACAAGTAATCACTAACCCGTTTATTCTA
TTCCTATCTGTGGATGTGTAATGGCTGGGGGGCCAGCCCTGGATAGGTTTATTATGGGAA
TTCTTTACAATAAACATAGCTTGTAAGTGTGAGATCTACAAATCCATTCACTGATTGGG
CATGAAATCCATGGTCAAGAGGACAAGTGGAAAGTGTGAGAGGGAAGGTTTGTAGACACCT
TCGCTTGTATCTTGTCAAGATAGAAAAGATAGTATCATTTTACCCTTGCCAGTAAAAAC
CTTTCATCCACCCATTCTCAGCAGACTCCAGTATTGGCACAGTCACTCACTGCCATTCT
CACACTATAACAAGAAAAGAAATGAAGTGCATAAGTCTCCTGGGAAAAGAACCTTAACCC
CTTCTCGTGCCATGACTGGTGATTTTATGACTCATAAGCCCTCCGTAGGCATCATTTCAA
GATCAATGGCCCATGCATGCTGTTTGAGCAGTCAATTGAGTTGAATTAGAATTTCAACC
ATACATTTTAAAGGTATTTGTGCTGTGTGTATTTTGTGATAAAATGTTGTGACTTCATGG
CAACAGGTTGGATGTGTAAAAATGGAATAAAAAAAGAGTCAAAAAAAGAGTCAAAAAA
AATT

>Hs.155097_ARNm_1 gj|15080385|gb|BC011949.1|BC011949 clon de Homo sapiens MGC:9006
10 IMAGE:3863603 poliA = 3

GGCGCCCAAGCCGCCGCCGCCAGATCGGTGCCGATTCCTGCCCTGCCCGACCGCCAGCG
CGACCATGTCCCATCACTGGGGGTACGGCAAACACAACGGACCTGAGCACTGGCATAAGG
ACTTCCCCATTGCCAAGGGAGAGCGCCAGTCCCCTGTTGACATCGACACTCATAAGCCA
AGTATGACCCCTCCCTGAAGCCCCTGTCTGTTTCTATGATCAAGCAACTTCCCTGAGGA
TCCTCAACAATGGTCACTGCTTTCAACGTGGAGTTTGATGACTCTCAGGACAAAGCAGTGC
TCAAGGGAGGACCCCTGGATGGCACTTACAGATTGATTCACTTTCACTTTCACTGGGGTT
CACTTGATGGACAAGGTTTCAAGCATACTGTGGATAAAAAGAAATATGCTGCAGAACTTC
ACTTGGTTTCACTGGAAACACCAATATGGGGATTTTGGGAAAGCTGTGCAGCAACCTGATG
GACTGGCCGTTCTAGGTATTTTTTTGAAGGTTGGCAGCGCTAAACCGGGCCTTCAGAAAG
TTGTGATGTGCTGGATTCCATTAAAAACAAGGGCAAGAGTGTGACTTCACAACTTTG
CAGCTCGTGGCCTCCTTCTGAATCCCTGGATTACTGGACCTACCCAGGCTCACTGACCA
CCCCCTCTTCTGGAATGTGTGACCTGGATTGTGCTCAAGGAACCCATCAGCGTCAGCA
GCGAGCAGGTGTTGAAATTCGTAAGTAACTTCACTTCAATGGGGAGGGTGAACCCGAAGAAC
TGATGGTGGACAACCTGGCGCCAGCTCAGCCACTGAAGAACAGGCAATCAAAGCTTCTT
TCAAATAAGATGGTCCCATAGTCTGTATCCAAATAATGAATCTTCGGGTGTTTCCCTTTA
GCTAAGCACAGATCTACCTTGGTGTATTGGACCCCTGGTTGCTTTGTGTCTAGTTTCTAG

ACCCTTCATCTCTTACTTGATAGACTTACTAATAAAATGTGAAGACTAGACCAATTGTCA
 TGCTTGACACAACCTGCTGTGGCTGGTTGGTGCTTTGTTTATGGTAGTAGTTTCTGTAA
 CACAGAATATAGGATAAGAAATAAGAATAAAGTACCTTGACTTTGTTTACAGCATGTAGG
 GTGATGAGCACTCACAATTGTTGACTAAAATGCTGCCTTTAAAACATAGGAAAGTAGAAT
 GGTGAGTGCAAATCCATAGCACAAAGATAAAATTGAGCTAGTTAAGGCAAATCAGGTAAAA
 TAGTCATGATTCTATGTAATGTAAACCAGAAAAAATAAATGTTTCATGATTTCAAGATGTT
 ATATTAAAGAAAACTTTAAAAATTATTATATATTTATAGCAAAGTTATCTTAAATATGA
 ATTCTGTTGTAATTTAATGACTTTTGAATTACAGAGATATAAATGAAGTATTATCTGTAA
 AAATTGTTATAATTAGAGTTGTGATACAGAGTATATTTCCATTTCAGACAATATATCATAA
 CTTAATAAATATTGTATTTTAGATATATTTCTCTAATAAAATTTCAGAATTCATAAAAAAA
 AAAAAAA

>Hs.5163_ARNm_1 gi|15990433|gb|BC015582.1|BC015582 clon de Homo sapiens MGC:23280 IMAGE:4637504
 poliA = 3

5

GGCACGAGGCATGGAGGCGCTGCTGCTGGGCGCGGGGTTGCTGCTGGGCGCTTACGTGCT
 TGTCTACTACAACCTGGTGAAGGCCCGCCGTGCGGCGGCATGGGCAACCTGCGGGGCCG
 CACGGCCGTGGTCACGGGTGAGTGCGGAGGCGGGTGAGTGCGAGCTGGCGGGCGCGCGG
 AGAGGAGGCCCGGGCCGGCGGTAGCAGCGGCCCGCGGGCTCAGCTCAGCTCGGCTCCCGC
 CCGCGGTCCGCGAGGCGCCAACAGCGGCATCGGAAAGATGACGGCGCTGGAGCTGGCGCGC
 CGGGGAGCGCGCGTGGTGCTGGCCTGCCGAGCCAGGAGCGCGGGGAGGCGGCTGCCTTC
 GACCTCCGCCAGGAGAGTGGAACAATGAGGTTCATCTTCATGGCCTTGGACTTGGCCAGT
 CTGGCCTCGGTGCGGGCCTTTGCCACTGCCTTTCTGAGCTCTGAGCCACGGTTGGACATC
 CTCATCCACAATGCCGGTATCAGTTCTGTGGCCGACCCGTGAGGCGTTTAACTGCTG
 CTTCCGGGTGAACCATATCGGTCCCTTTCTGCTGACACATCTGCTGCTGCCTTGCTGAAG
 GCATGTGCCCCCTAGCCGCGTGGTGGTGGTAGCCTCAGCTGCCCACTGTGCGGGACGTCTT
 GACTTCAAACGCCTGGACCGCCAGTGGTGGGCTGGCGGCAGGAGCTGCGGGCATATGCT
 GACACTAAGCTGGCTAATGTACTGTTTGCCCGGGAGCTCGCCAACCAGCTTGAGGCCACT
 GGCGTAACTGCTATGCAGCCACCCAGGGCCTGTGAACCTCGGAGCTGTTCTGCGCCAT
 GTTCCTGGATGGCTGCGCCCACTTTTGCGCCCAATTGGCTTGGCTGGTGCTCCGGGCACCA
 AGAGGGGGTGCCAGACACCCCTGTATTGTGCTCTACAAGAGGGCATCGAGCCCTCAGT
 GGGAGATATTTTGCCAAC TGCCATGTGGAAGAGGTGCCTCCAGCTGCCCGAGACGACCGG
 GCAGCCCATCGGCTATGGGAGGCCAGCAAGAGGCTGGCAGGGCTTGGGCCTGGGGAGGAT
 GCTGAACCCGATGAAGACCCCCAGTCTGAGGACTCAGAGGCCCATCTTCTCTAAGCACC
 CCCCACCTGAGGAGCCACAGTTTCTCAACCTTACCCAGCCCTCAGAGCTCACCAGAT
 TTGTCTAAGATGACGCACCGAATTCAGGCTAAAGTTGAGCCTGAGATCCAGCTCTCCTAA
 CCTCAGGCCAGGATGCTTGCCATGGCACTTCATGGTCCTTGAAAACCTCGGATGTGTGC
 GAGGCCATGCCCTGGACACTGACGGGTTTGTGATCTTGACCTCCGTGGTTACTTTCTGGG
 GCCCAAGCTGTGCCCTGGACATCTCTTTTCTGTTGAAGGAATAATGGGTGATTATTT
 CTTCTGAGAGTGACAGTAACCCAGATGGAGAGATAGGGGTATGCTAGACACTGTGCTT
 CTCGGAAATTTGGATGTAGTATTTTCAGGCCCCACCTTATTGATTCTGATCAGCTCTGG
 AGCAGAGGCAGGGAGTTTGCATGTGATGCACTGCCAACATTGAGAATTAGTGAACCTGAT
 CCCTTTGCAACCGTCTAGCTAGGTAGTTAAATTACCCCATGTTAATGAAGCGGAATTAG
 GCTCCCGAGCTAAGGGACTCGCCTAGGGTCTCACAGTGAGTAGGAGGAGGGCCTGGGATC
 TGAACCAAGGGTCTGAGGCCAGGGCCGACTGCCGTAAGATGGGTGCTGAGAAGTGAGTC
 AGGGCAGGGCAGCTGGTATCGAGGTGCCCATGGGAGTAAGGGGACGCCCTTCCGGGCGGA
 TGCAGGGCTGGGGTCATCTGTATCTGAAGCCCTCGGAATAAAGCGCTTGACCGCCAAA
 AA

>Hs.55150_ARNm_1 gi|17068414|gb|BC017586.1|BC017586 clon de Homo sapiens MGC:26610
 IMAGE:4837506 poliA = 3

10

AGCGGTGGAGAAAAGGCAGAACCCAGAGTAGAGATTGACAGTGAGCTGAGCCAATCAGGCT
GTGAATCTGCAGCAGTGATCCCAGGTCCTCCAATTAATACTAAGAGAGTGGACCAGGGCC
CCTGAGGAAGACAGATGGCAGGGACAGCGCGCCATGACCGAGAGATGGCGATCCAGGCCA
AGAAAAAGCTCACCACGGCCACCAACCCATTGAAAGACTCCGACTGCAGTGCCTGGCCA
GGGGCTCTGCTGGGATCAAAGGACTTGGCAGAGTGTTTAGAATTATGGATGACGATAATA
ATCGAACCCTTGATTTTAAAGAATTTATGAAAGGGTTAAATGATTATGCTGTGGTCATGG
AAAAAGAAGAGGTGGAAGAACTTTTCCGGAGGTTTGATAAAGATGGAAATGGAACAATAG
ACTTCAATGAATTTCTTCTCACATTAAGACCTCCAATGTCCAGAGCCAGAAAAGAGGTAA
TCATGCAAGCTTTTAGAAAGTTAGACAAGATCGAGATGGTGTTATAACAATCGAAGACC
TTCGTGAAGTATATAATGCAAAACACCACCCAAAGATACCAATGGGGAATGGAGTGAGG
AACAAGTATTTAGGAAATTTCTGGATAACTTTGATTACCCCTATGACAAAGATGGATTGG
TGACCCCTGAGGAGTTCATGAACTACTATGCAGGTGTGAGCGCATCCATTGACACTGATG
TGTACTTCATCATCATGATGAGAACCGCCTGGAAGCTTTAAGCACATGACCTGGGGACCA

GGCCCTGGGACAGCCATGTGGCTCCAAATGACTAAATGTCAGCTCAAAAACCGAATCGT
ATTTGATTTCACTCATCCTAATGTTTTTCTGTGTCAAAATATTGCATTTTCTGGGG
CCAAAAACAGGCAGAAATAAAGACATTGAGTAGTCAAAAAAAAAAAAAAAAAA

>Hs.170177_contig3

AI620495|AW291989|AA780896|AA976262|AI298326|BF111862|AW591523|AI922518|
AI480280|BF589437|AA600354|AI886238|AA035599|H90049|BF112011|N52601|AI57
0965|AI565367|AW768847|H90073|BE504361|N45292|AI632075|AA679729|AW168052
|AI978827|AI968410|AI669255|N45300|AI651256|AI698970|AI521256|AW078614|A
I802070|AI885947|AI342534|AI653624|AW243936|T16586|R15989|AI289789|AI871_636|AI718785|AW148847
poliA = 2 poliA = 3

[illegible]

>Hs.184601 ARNm 5 g|4426639|gb|AF104032.1|AF104032 Homo sapiens poliA = 2

GCGGCGCGCACACTGCTCGCTGGGCGCGGCTCCCGGGTGTCCAGGCCCGCGCGGTGCG
 CAGAGCATGGCGGGTGC GGCGCCGAAGCGGCGCGCTAGCGGCGCGCGCGCGGAGGAG
 AAGGAAGAGGCGCGGGAGAAGATGCTGGCCGCCAAGAGCGCGGACGGCTCGGCGCGCGCA
 GCGAGGGCGAGGGCGTGACCTGCAGCGGAACATCACGCTGCTCAACGGCGTGCCATC
 ATCGTGGGGACCATTATCGGCTCGGGCATCTTCGTGACGCCCACGGGCGTGCTCAAGGAG
 GCAGGCTCGCCGGGGCTGGCGCTGGTGGTGTGGGCGCGTGCGGCGTCTTCTCCATCGTG
 GCGCGCTCTGCTACGCGGAGCTCGGCACCACCATCTCCAAATCGGGCGGCGACTACGCC
 TACATGCTGGAGGTCTACGGCTCGCTGCCCGCCTTCCTCAAGCTCTGGATCGAGCTGCTC
 ATCATCCGGCCTTCATCGCAGTACATCGTGGCCCTGGTCTTCGCCACCTACCTGCTCAAG
 CCGCTCTTCCCCACCTGCCCGGTGCCCGAGGAGGCAGCCAAGCTCGTGGCCTGCCCTCTGC
 GTGCTGCTGCTCACGGCCGTGAAGTGTACAGCGTGAAGGCCGCCACCCGGGTCCAGGAT
 GCCTTTGCCGCGGCCAAGCTCCTGGCCCTGGCCCTGATCATCCTGCTGGGCTTCGTCCAG
 ATCGGGAAGGGTGATGTGTCCAATCTAGATCCCAACTTCTCATTGGAAGGCACCAACTG
 GATGTGGGGAACATTGTGCTGGCATTATACAGCGGCCTCTTGCCTATGGAGGATGGAAT
 TACTTGAATTTTCGTACAGAGGAAATGATCAACCCCTACAGAAACCTGCCCTGGCCATC
 ATCATCTCCCTGCCCATCGTGACGCTGGTGTACGTGCTGACCAACCTGGCCTACTTCACC
 ACCCTGTCCACCGAGCAGATGCTGTGCTCGAGGCGGTGGCCGTGGACTTCGGGAACAT
 CACCTGGGCGTCATGTCTGGATCATCCCCGTCTTCGTGGGCTGTCTGCTTCGGCTCC
 GTCAATGGGTCCCTGTTCACATCTCCAGGCTCTTCTTCGTGGGGTCCCGGAAGGCCAC
 CTGCCCTCCATCCTCTCCATGATCCACCCACAGCTCCTCACCCCGTGCCGTCCCTCGTG
 TTCACGTGTGTGATGACGCTGCTCTACGCCCTTCTCCAAGGACATCTTCTCCGTATCAAC
 TTCTTCAGCTTCTTCAACTGGCTCTGCGTGGCCCTGGCCATCATCGGCATGATCTGGCTG
 CGCCACAGAAAGCCTGAGCTTGAGCGGCCCATCAAGGTGAACCTGGCCCTGCCTGTGTTC
 TTCATCCTGGCCTGCCTCTTCTGATCGCCGTCTCCTTCTGGAAGACACCCGTGGAGTGT
 GGCATCGGCTTACCATCATCCTCAGCGGGCTGCCCGTCTACTTCTTCGGGGTCTGGTGG
 AAAACAAGCCCAAGTGGCTCCTCCAGGGCATCTTCTCCACGACCGTCTGTGTGTCAGAAG
 CTCATGCAGGTGGTCCCCAGGAGACATAGCCAGGAGGCCGAGTGGCTGCCCGAGGAGCA
 TGCGCAGAGGCCAGTTAAAGTAGATCACCTCCTCGAACCCTCCGGTCCCCGCAACCC
 ACAGCTCAGCTGCCCATCCAGTCCCTCGCCGTCCCTCCAGGTGGGCGAGTGGAGGCTG
 CTGTGAAAACCTCTGGTACGAATCTCATCCCTCAACTGAGGGCCAGGGACCCAGGTGTGCC
 TGTGCTCCTGCCCAGGAGCAGCTTTGGTCTCCTTGGGCCCTTTTTCCTTCCCTCCTTT
 GTTTACTTATATATATATATTTTTTTTAAACTTAAATTTTGGGTCAACTTGACACCCTAAG
 ATGATTTTTTTAAGGAGCTGGGGGAAGGCAGGAGCCTTCTTCTCCTGCCCCAAGGGCCC
 AGACCCTGGGCAAAACAGAGCTACTGAGACTTGAACCTCATTGCTACGACAGACTTGCAC
 TGAAGCCGGACAGCTGCCCAGACACATGGGCTTGTGACATTCTGTGAAAACCAACCCTGTG

GGCTTATGTCTCTGCCTTAGGGTTTGCAGAGTGGAACTCAGCCGTAGGGTGGCACTGGG
 AGGGGGTGGGGGATCTGGGCAAGGTGGGTGATTCTCTCAGGAGGTGCTTGAGGCCCCGA
 TGGACTCCTGACCATAATCCTAGCCCTGAGACACCATCCTGAGCCAGGGAACAGCCCCAG
 GGTGGGGGGTGGCCGCATCTCCCTAGCTCACCAGGCTGGCCTCTGGGCAGTGTGGCC
 TCTTGGCTATTTCTGTGTCCAGTTTGGAGGCTGAGTTCTGGTTCATGCAGACAAAGCCC
 TGTCTTTCAGTCTTCTAGAAACAGAGACAAGAAAGGCAGACACACCGCGGCCAGGCACCC
 ATGTGGGCGCCACCCCTGGGCTCCACACAGCAGTGTCCCTGCCCCAGAGGTCGCAGCTA
 CCTCAGCCTCCAATGCATTGGCCTCTGTACCGCCCGGCAGCCCCCTTCTGGCCGGTGTCTG
 GGTTCCCACTCCCGGCTAGGCACCTCCCGCTCTCCCTGTACGCTCATGTCTCTGTCTCT
 GGTCTGATGCCCCGTGTCTAGGAGACAGAGCCAAGCACTGCTCACGTCTCTGCCGCTG
 CGTTTGGAGGCCCCCTGGGCTCTCACCAGTCCCCACCGCCTGCAGAGAGGGAAGTAGGG
 CACCCCTTGTCTTCTGTGTCCCGTGAATTTTTTTCGCTATGGGAGGCAGCCGAGGCCTG
 GCCAATGCGGCCCCACTTCTGAGCTGTGCTGCTCCATGGCAGCAGCCAAGGACCCCC
 AGAACAAGAAGACCCCCCGCAGGATCCCTCCTGAGCTCGGGGGGCTCTGCCTTCTCAGG
 CCCCAGGCTTCCCTTCTCCCCAGCCAGAGGTGGAGCCAAGTGGTCCAGCGTCACTCCAGT
 GCTCAGCTGTGGCTGGAGGAGCTGGCCTGTGGCACAGCCCTGAGTGTCCCAAGCCGGGAG
 CCAACGAAGCCGGACACGGCTTCACTGACCAGCGGCTGCTCAAGCCGAAGCTCTCAGCA
 AGTGCCAGCGGAGCCTGCCGCCCCCACCTGGGCACCGGGACCCCCCTCACCATCCAGTGG
 GCGCGAGAAACCTGATGAACAGTTTGGGGACTCAGGACCAGATGTCCGTCTCTTTGCT
 TGAGGAATGAAGACCTTTATTCACCCCTGCCCGTTGCTTCCCGCTGCACATGGACAGAC
 TTCACAGCGTCTGCTCATAGGACCTGCATCCTTCTGGGGACGAATTCACCTCGTCCAAG
 GGACAGCCCACGGTCTGGAGGCCGAGGACCACAGCAGGCAGGTGGACTGACTGTGTGG
 GCAAGACCTCTTCCCTCTGGGCTGTCTCTTGGCTGCAAATAAGGACAGCAGCTGGTGC
 CCCACCTGCTGTGTCATTGCTGTGTGAATCCAGGAGGCAGTGGACATCGTAGGCAGCCA
 CGGCCCGGGTCCAGGAGAAAGTGTCTCCCTGGAGGCACGCACCACTGCTTCCCACTGGGGC
 CGGCGGGGCCCCACGCACGACGTGAGCCTTACCTTCCCGCCTCGGCTAGGGGTCTCTCGG
 GATGCCGTTCTGTTCCAACCTCCTGCTCTGGGACGTGGACATGCCTCAAGGATACAGGGA
 GCGCGCGGCTCTCGACGGCACGCACCTTGCCTGTTGGCTGCTGCGGCTGTGGGCGAGCAT
 GGGGGCTGCCAGCGTCTGTTGTGGAAAGTAGCTGCTAGTGAAATGGCTGGGGCCGCTGGG
 GTCCGTCTTCACTGCGCAGGTCTCTTCTGGGCGTCTGAGCTGGGGTGGGAGCTCTTCC
 GCAGAAGGTTGGTGGGGGTCCAGTCTGTGATCCTTGGTGTGTGTGCCCCACTCCAGCC
 TGGGGACCCCACTTCAAGAGGTAGGGGCCGTGTCCCGCGGTGCTGACTGAGGCTGCTTC
 CCCCTCCCCCTCCTGCTGTGCTGGAATCCACAGGGACCAGGGCCACCGCAGGGGACTGT
 CTCAGAAGACTTGATTTTTCCGTCCCTTTTTCTCCACACTCCACTGACAAACGTCCCCAG
 CGGTTTCCACTTGTGGGCTTCAAGTGTTTTCAAGCACAACCCACCACAACAAGCAAGTGC
 ATTTTCAGTCGTTGTGCTTTTTTGTGTTTGTGCTAACGTCTTACTAATTTAAAGATGCTGT
 CGGCACCATGTTTATTTATTTCCAGTGGTTCATGCTCAGCCTTGCTGCTCTGCGTGGCGCA
 GGTGCCATGCTGCTCCCTGTCTGTGTCCCAGCCACGCAGGGCCATCCACTGTGACGTGCG
 GCCGACCAGGCTGGACACCCCTCTGCCGAGTAATGACGTGTGTGGCTGGGACCTTCTTTAT
 TCTGTGTTAATGGCTAACCTGTTACACTGGGCTGGGTTGGGTAGGGTGTCTGGCTTTTT
 TGTGGGGTTTTTTATTTTAAAGAAACACTCAATCATCCTAAAAAAAAAAAAAAAAAAAAAA
 AA
 AA

>Hs.351972_singletel AA865917 poliA = 2 poliA = 3

GGGACTTGGAAAGGGGAACTGGGATTTGGGGAGGGGCTGGAGGACTTCCGCACGCTTCCA
 CCTCTTCGACCTCCACTGCGCCCCACCTCCCTGCCTGTGTGTGTTATTTCAAAGGAAAA
 GAACAAAAGGAATAAATTTTCTAAGCTCTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

5

>Hs.5366_ARNm_2 gi|15277845|gb|BC012926.1|BC012926 clon de Homo sapiens MGC:16817 IMAGE:3853503
 poliA = 3

GCAGGCTCTGCCTGTGGCCACTAGCAGAGAAGCTGCTGTCTTCCACCACCAGCACC
 CCACCTGCTCCAAGACCAGCCTCCTGGGGGGACCAGGCACCCGGCCTTCACTGGCACC
 GGGAGCCGTCCTCAGCAGCGTCAACATGTCAAGGCCAGCAGCAGAGCCATTTACTTGCA
 CCGGAAGGAGTACTCCAGAACCTCACCTCAGAGCCCACCTCCTGCAGCACAGGGTGGA
 GCACTTGATGACATGCAAGCAGGGGAGTCAGAGAGTCCAGGGGCCGAGGATGCCTTGCA
 GAAGCTGTTGAGATGGATGCACAGGGCCGGGTGTGGAGCCAAGACTTGATCCTGCAGGT
 CAGGGACGGCTGGCTGCAGCTGCTGGACATTGAGACCAAGGAGGAGCTGGACTCTTACCG
 CCTAGACAGCATCCAGGCCATGAATGTGGCGCTCAACACATGTTCTTACAATCCATCCT
 GTCCATCACCGTGACAGGAGCCGGGCCTGCCAGGCACTAGCACTCTGCTCTTCCAGTGCCA
 GGAAGTSGGGGCAGAGCGACTGAAGACCAGCCTGCAGAAGGCTCTGGAGGAAGAGCTGGA
 GCAAAGCAGACCTCGACTTGGAGGCCTTCAGCCAGGCCAGGACAGATGGAGGGGGCCTGC
 TATGGAAAGGCCGCTCCCTATGGAGCAGGCACGCTATCTGGAGCCGGGGATCCCTCCAGA

ACAGCCCCACCAGAGGACCCTAGAGCACAGCCTCCACCATCCCCAAGGCCCTGCCACG
 CCACACCAAGTGGCCGAGAACCAAGTGCCCTTACTCTGCCTCCTCCAAGGCGGTCTCTTC
 CCCCAGGACCCAGAGAGGGACGAGGAAGTGCTGAACCATGTCTAAGGGACATTGAGCT
 GTTCATGGGAAAGCTGGAGAAGGCCCAGGCAAGACCAGCAGGAAGAAGAAATTTGGGAA
 AAAAAACAAGGACCAGGGAGGTCTCACCCAGGCACAGTACATTGACTGCTTCCAGAAGAT
 CAAGTACAGCTTCAACCTCCTGGGAAGGCTGGCCACCTGGCTGAAGGAGACAAGTGCCCC
 TGAGCTCGTACACATCCTCTTCAAGTCCCTGAACTTCATCCTGGCCAGGTGCCCTGAGGC
 TGGCCTAGCAGCCCAAGTGATCTCACCCCTCCTCACCCCTAAAGCTATCAACCTGCTACA
 GTCTGTCTAAGCCACCTGAGAGTAACCTTTGGATGGGGTTGGGGCCAGCCTGGACCAC
 TAGCCGGGCGGACTGGACAGGCGATGAGCCCTGCCCTACCAACCCACATTCTCGGATGA
 CTGGCAACTTCCAGAGCCCTCCAGCCAAGCACCCCTTAGGATACCAGGACCCTGTTTCCCT
 TCGGCGGGGAAGTCATAGGTTAGGGAGCACCTCACACTTTCCTCAGGAGAAGACACACAA
 CCATGACCCTCAGCCTGGGGACCCCACTCCAGGCCCTCCAGCCCCAAACCTGCCAGCC
 AGCCCTGAAAATGCAAGTCTTGACGAGTTTGAAGCTAGGAACCCACGGGAACCTGACTGT
 GGTCCAGGGAGAGAAGCTGGAGGTTCTGGACCACAGCAAGCGGTGGTGGCTGGTGAAGAA
 TGAGGCGGGACGGAGCGGCTACATTCCAAGCAACATCCTGGAGCCCCTACAGCCGGGGAC
 CCCTGGGACCCAGGGCCAGTCACCCCTCTCGGGTTCCAATGCTTCGACTTAGCTCGAGGCC
 TGAAGAGGTACAGACTGGCTGCAGGCAGAGAACTTCTCCACTGCCACGGTGAGGACACT
 TGGGTCCCTGACGGGGAGCCAGCTACTTCGCATAAGACCTGGGGAGCTACAGATGCTATG
 TCCACAGGAGGCCCCACGAATCCTGTCCCGGCTGGAGGCTGTGAGAAGGATGCTGGGGAT
 AAGCCCTTAGGCACCAGCTTAGACACCTCCAAGAACCAGGCCCCGCTGATGCAAGATGGC
 AGATCTGATACCCATTAGAGCCCCGAGAATTCCTCTTCTGGATCCCAGTTTGACGAAAC
 CCCACACCCAGCTCACACAGCAAAAACAATGGACAGGCCCCAGAGGGTGAAGCAACAGT
 GTCCCTTCTGGCTGTGTTGGAGCCTCCCCAGTAACCACCTATTTATTTTACCTCTTTCCC
 AAACCTGGAGCATTTATGCCTAGGCTGTCAAGAATCTGTTTCACTCCCTCTCCTTCTCAA
 TAAAGCATCTTCAAGCTTGAAAAAAAAAAAAAAAAA

>Hs.18140_contigl

AI685931|AA410954|T97707|AA706873|AI911572|AW614616|AA548520|AW027764|BF
 511251|AI914294|AW151688 poliA = 1 poliA = 1

5

CCTTCCATTGAATTCCACCAGACACATTTCAGGTTANCTTCGTAATGTCTTCATATGAGTA
TCAATCAACACCTTCCCCAACTCAATTGTACTAGGTTGTAGAGCACAAGGATGGTCTCGT
GCTGCTCTGTGGCACCTGTGCCTACACTGCTCTGAGCTTTGAGGAGGCTGCTCTCTTTGC
TGACCCCATGATCTTTTCTGCCCTTCTGTTAAGGGCATTTGGCCACAGCAACGGGGCAAAT
GCCCCAAGCTGGCTGTAAGTGACCCATCCCTTTGGCTCCCATGATTAGACCAAGGAGAGG
CATGGGGTCCAGCTGAGCCATTGAGAACCATTCTTTAGCATTTTTCCACTCAAAGGTTAGA
GATGAGATTTTCTCTTCCCAAGGCTACCTCTGGCCATGGTTCCAGCTTCATGGGGGCAAT
GGGATTAGGAAAATGAGGTCAACCTGCAAAAGGAAAGCAGATGCAAGAGATGGAGACAGAA
TGGGGGTGTCCTGGGGATCTTGAGCCTGAATTCATTGGCACAAAAGGCAGCAGCATCCT
CACTGTATCTGCAGTCCATTTGGACTCAATAAAAACTTTGAAAGTCACATGTGTTATGGA
ATTCTTCTCAGTGACACATTTCATCTGTGCTCAGTTGTCCCAGCAAGGGTCAGCCCCCTCA
TACCCCTGCAGCATCCGCTGCTATGAAGCAGAGCTGTAAACGCCCTCCCTGTGTATAGGA
AAAGCTACATGGAGCAAATCCTCCTGCCTGAAGAAAGTGCATCTCAGCATCACTTCAGCTG
TCGGGGCATTTGTGGGGAGAACCAGACCACCTCTGCGGAAGGCAGCAGACCCTCTTCCAG
CCATGGATGGAGTTGAATTCTCTATAAACGGTTCCACCAGCAAACCACCAATACATCCAT
TGTTTGCCTAGAGAGAAATTTAAAAATAAATAAATGTTCACTTAT

>Hs.133196_contig2
BF224381|BE467992|AW137689|AI695045|AW207361|BF445141|AA405473 poliA = 2 WARN poliA = 3

5

TGCGGCCGCGGCATGAAAGGCGGCGAGGAGAGGCAGCACTGCTGCTCTTGACTTCTGAGC
AGGGCTTAGAGAGCCTGCCCCGGCTTAAGCCGAGCTGCTGGTGCTGACCTGAGCGCCGA
GTCCGCGAGCTCTGAGTCCGGAGCCTCCAGCCGTGGAGCCGTGGGATGAGGGGGGCGTT
GGGGGACAGGGCAAAGTCGATCTTGTTGTACAGCCGCCGATCCTAGCGCGGAGCTGCG
AGCCTGACCGGCCGCGTCTGGCATGGTCAGAGAAAGAATTTCTTTTCCCAACTCCGGCT
TTTGGTTTTGTGTGCCACCTTGCGCAACTCCGGAGCCAGCCGACCCACATGGATTCTC
AACAGGTGGCCGCGCACATCTTCTGAGCCTCGCTCTCTCATCTGAAAGTGGAGTGTAAGTC
CAAGAAGATTCAATTTAGACAAAGAAGGTGGAAGGAAAGGACTTTCTGGGCCAGCAAGTCG
GATGACCACCTCCAAAGGGGCAGAGGAGGGCCCATTTTGTGAAGAAGAAATCAACTACCC
GGAAAACGCCACAGGAGGACATGTTTTCTGCAGATGTAGTTGCCCTAGAAACAGAAGAGTA
TGGGGGTGTGAATGTCTTCTTTTTGGGGGCAACACTATGTCCTTTTCTTTTCTAGAT
ACAGTTAATTCCTGGAAATTTTAGCGAGTTTGTCTTGTGGATATTTTGAACAATAAAGA
GTGAAAATCAAAAAA

>Hs.63325_ARNm_5_gij15451939|ref|NM_019894.1| proteasa transmembrana de Homo sapiens, serina 4 (TMPRSS4), ARNm poliA = 3

10

CCCAATCACTCCTGGAATACACAGAGAGAGGCAGCAGCTTGCTCAGCGGACAAGGATGCT
 GGGCGTGAGGGACCAAGGCCTGCCCTGCACTCGGGCCTCCTCCAGCCAGTGCTGACCAGG
 GACTTCTGACCTGCTGGCCAGCCAGGACCTGTGTGGGGAGGCCCTCCTGCTGCCTTGGGG
 TGACAATCTCAGCTCCAGGCTACAGGGAGACCGGGAGGATCACAGAGCCAGCATGTTACA
 GGATCCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAACCCCTGCGCAAACCCCG
 TATCCCCATGGAGACCTTCAGAAAGGTGGGGATCCCCATCATCATAGCACTACTGAGCCT
 GGGAGTATCATCATTGTGGTTGTCTCATCAAGGTGATTCTGGATAAATACTACTTCTT
 CTGCGGGCAGCCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGACTG
 TCCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGGCCTGCAGTGGC
 AGTCCGCCTCTCCAAGGACCGATCCACACTGCAGGTGCTGGACTCGGCCACAGGGAAGT
 GTTCTCTGCCTGTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGGCAGAT
 GGGCTACAGCAGCAAACCCACTTTTCAGAGCTGTGGAGATTGGCCCAGACCAGGATCTGGA
 TGTGTGTGAAATCACAGAAAACAGCCAGGAGCTTCGCATGCGGAACCTCAAGTGGGGCCCTG
 TCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTGCTGTGGGAAGAGCCTGAAGACCCC
 CCGTGTGGTGGGTGGGGAGGAGGCCTCTGTGGATTCTTGCCCTTGGCAGGTGAGCATCCA
 GTACGACAAACAGCACGTCTGTGGAGGGAGCATCCTGGACCCCCACTGGGTCTCACGGC
 AGCCCACTGCTTCAGGAAACATACCGATGTGTCAACTGGAAGGTGCGGGCAGGCTCAGA
 CAAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAGATCATCATCATTGAATTCAACCC
 CATGTACCCCAAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCCCACTCACTTTCTC
 AGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCC
 ACTCTGGATCATTGGATGGGGCTTTACGAAGCAGAATGGAGGGAAGATGTCTGACATACT
 GCTGCAGGCCTCAGTCCAGGTCAATTGACAGCACACGGTGCAATGCAGACGATGCGTACCA
 GGGGGAAGTCACCGAGAAGATGATGTGTGCAGGCATCCCGGAAGGGGGTGTGGACACCTG
 CCAGGGTGACAGTGGTGGGCCCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGGCAT
 CGTTAGCTGGGGCTATGGCTGCGGGGGCCCCGAGCACCCAGGAGTATACACCAAGGTCTC
 AGCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTGTAATGCTGCTGCCCTT
 TGCAGTGCTGGGAGCCGCTTCTTCTGCCCCGCCCCACCTGGGGATCCCCCAAAGTCAGA
 CACAGAGCAAGAGTCCCCTTGGGTACACCCCTCTGCCACAGCCTCAGCATTTCTTGGAG
 CAGCAAAGGGCCTCAATTCTGTAAAGAGACCCCTCGCAGCCCAGAGGCGCCAGAGGAAGT
 CAGCAGCCCTAGCTCGGCCACACTTGGTGCTCCAGCATCCCAGGGAGAGACACAGCCCA
 CTGAACAAGGTCTCAGGGGTATTGCTAAGCCAAGAAGGAACCTTCCCACTACTGAATG
 GAAGCAGGCTGTCTTGTAAAAGCCAGATCACTGTGGGCTGGAGAGGAGAAGGAAAGGT
 CTGCGCCAGCCCTGTCCGTCTTACCCATCCCCAAGCCTACTAGAGCAAGAAACAGTTG
 TAATATAAAATGCACTGCCCTACTGTTGGTATGACTACCGTTACCTACTGTTGTATTGT
 TATTACAGCTATGGCCACTATTATTAAAGAGCTGTGTAACATCAAAAAAAAAAAAAAAAAA
 AAAA

>Hs.250692_ARNm_2 gi|184223|gb|M95585.1|HUMHLF ARNm de factor de leucemia hepática (HLF) humana, cds completa poliA = 3

TTTTCAATTTTGAACATTTTGCAAAACGAGGGGTTCGAGGCAGGTGAGAGCATCCTGCA
 CGTCGCCGGGGAGCCCGCGGGCACTTGGCGCGCTCTCCTGGGACCGTCTGCACTGGAAAC
 CCGAAAGTTTTTTTTTAATATATATTTTTATGCAGATGTATTTATAAAGATATAAGTAAT
 TTTTTCTTCCCTTTTCTCCACCGCCTTGAGAGCGAGTACTTTTGGCAAAGGACGGAGGA
 AAAGCTCAGCAACATTTTAGGGGGCGGTTGTTTCTTCTTTCTTATTCTTTTTTAAGGG
 GAAAAAATTTGAGTGCATCGCGATGGAGAAAAATGTCCCGACCGCTCCCCCTGAATCCAC
 CTTTATCCCGCCTCCCTACGGCGTGCTCAGGTCCCTGCTGGAGAACCCGCTGAAGCTCCC
 CCTTCACCACGAAGACGCATTTAGTAAAGATAAAGACAAAGAAAAGAAGCTGGATGATGA
 GAGTAACAGCCCGACGGTCCCCAGTCGGCATTCTGGGGCCTACCTTATGGGACAAAAC
 CCTTCCCTATGACGGAGATACTTCCAGTTGGAATACATGGACCTGGAGGAGTTTTTGTG
 AGAAAATGGCATTTCCCCCAGCCCATCTCAGCATGACCACAGCCCTCACCTCCTGGGCT
 GCAGCCAGCTTCCTCGGCTGCCCCCTCGGTCTGACCTCAGCAGCCGGGCTCTGCACC
 CCTTCACCCTGGCATCCCATCTCCGAAGTGTATGCAGAGCCCCATCAGACCAGGTGAGT
 GTTGCCAGCAAACCGCAATACACCAAGTCCCATTGATCCTGACACCATCCAGGTCCCAGT
 GGGTTATGAGCCAGACCCAGCAGATCTTGCCCTTTCCAGCATCCCTGGCCAGGAAATGTT
 TGACCCTCGCAAACGCAAGTTCTCTGAGGAAGAACTGAAGCCACAGCCCATGATCAAGAA
 AGCTCGCAAAGTCTTCATCCCTGATGACCTGAAGGATGACAAGTACTGGGCAAGGCGCAG
 AAAGAACAACATGGCAGCCAAGCGCTCCCGCGACGCCCCGGAGGCTGAAAGAGAACCAGAT
 CGCCATCCGGGCTCGTTCTTGGAGAAGGAGAACTCGGCCCCCTCCGCCAGGAGGTGGCTGA
 CTTGAGGAAGGAGCTGGGCAAATGCAAGAACATACTTGCCAAGTATGAGGCCAGGCACGG
 GCCCCGTAGGATGGCATTTTTTGACGGCTGGCTTTGGAATAGATGGACAGTTTGTTTCCT

GTCTGATAGCACCACACGCAAACCAACCTTTCTGACATCAGCACTTTACCAGAGGCATAA
 ACACCAACTGACTCCCATTGTTGGTGTGCATCTGTGTGTGTGTGCGTGTATATGTGCTTGTG
 CTCATGTGTGTGGTCAGCGGTATGTGCGTGTGCGTGTTCCTTTGCTCTTGCCATTTTAAAG
 GTAGCCCTCTCATCGTCTTTTAGTTCCAAAGAAAGGTGCCATGTCTTTACTAGACTG
 AGGAGCCCTCTCGCGGTCTCCCATCCCTCCCTCCTTCACTCCTGCGCTCCTCAGCTTTG
 CTTTCATGTTTCGAGCTTACCTACTCTTCCAGGACTCTCTGCTTGGATTCACTAAAAAGGGC
 CCTGGTAAAATAGTGGATCTCAGTTTTTAAGAGTACAAGCTCTTGTCTGTCTTAGTCCG
 TAAGTTACCATGCTAATGAGGTGCACACAATAACTTAGCACTACTCCGCAGCTCTAGTCC
 TTTATAAGTTGCTTTTCTCTTACTTTTTCAGTTTTGGTGATAATCGTCTTCAAATTAAGTG
 CTGTTTAGATTTATTAGATCCCATATTTACTTACTGCTATCTACTAAGTTTCTTTTAAAT
 TCTACCAACCCAGATAAGTAAGAGTACTATTAAATAGAACACAGAGTGTGTTTTTGCAC
 GTCTGTACCTAAAGCAATAATCCTATTGTACGCTAGAGCATGCTGCCTGAGTATTACTAG
 TGGACGTAGGATATTTTCCCTACCTAAGAATTTCACTGTCTTTTAAAAAACAAAAAGTAA
 AGTAATGCATTTGAGCATGGCCAGACTATTCCCTAGGACAAGGAAGCAGAGGGAAATGGG
 AGGTCTAAGGATGAGGGGTTAATTTATCAGTACATGAGCCAAAACTGCGTCTTGGATTA
 GCCTTTGACATTGATGTGTTGCGTTTTGTTGTTCCCTTCCCTCACACCCTGCCTCGCCC
 CCACTTTTCTAGTTAACTTTTTCCATATCCCTCTTGACATTCAAACAGTTACTTAAGAT
 TCAGTTTTCCCACTTTTTGGTAATATATATATTTTTGTGAATTATACTTTGTTGTTTTTA
 AAAAGAAAATCAGTTGATTAAGTTAATAAGTTGATGTTTTCTAAGGCCCTTTTTCTAGT
 GGTGTCATTTTTGAATGCCTCATAAATTAATGATCTGAAGCTTATGTTTCTTATCTCT
 GTTTGCTTTTGAACGTATGTGCTCTTATAAAGTGGACTTCTGAAAAATGAATGTAAAGA
 CACTGGTGTATCTCAGAAGGGGATGGTGTGTGTCACAACTGTGGTTAATCCAATCAATTT
 AAATGTTTACTATAGACCAAAAGGAGAGATTATTAATCGTTTAAATGTTTATACAGAGTA
 ATTATAGGAAGTTCTTTTTTGTACAGTATTTTTTCAGATATAAATACTGACAATGTATTT
 GGAAGACATATATTATATATAGAAAAGAGGAGAGGAAACTATTCCATGTTTTAAATTA
 TATAGCAAAGATATATATTACCAATGTTGTACAGAGAAGAAGTGCTTGGGGGTTTTTGA
 AGTCTTTAATATTTTAAGCCCTATCACTGACACATCAGCATGTTTTCTGCTTTAAATTA
 AATTTTATGACAGTATCGAGGCTTGTGATGACGAATCCTGCTCTAAAATACACAAGGAGC
 TTTCTTGTCTTCTTATTAGGCCCTCAGAAAGAAGTCAGTTAACGTACCCCAAAGCACAAAA
 TGGATTTTGTAGTCAAATATTTATTGGATGATACAGTGTTTTTTAGGAAAAGCATCTGCCAC
 AAAAAATGTTCACTTCGAAATTTCTGAGTTCTTGGAAATGGCACGTTGCTGCCAGTGCCCCAG
 ACAGTTCTTTTCTACCTGCGGGCCCGCACGTTTTATGAGGTTGATATCGGTGCTATGTG
 TTTGGTTTTATAATTTGATAGATGTTTACTTTAAAGATGATTGTTCTTTTGTTCATTAA
 GTTGTAATAATGTCAAGAAATTTCTGCTGTACGACAAAGAAACATTTTACGCTAGATTAA
 ATATCCTTTTCATCAATGGGATTTTCTAGTTTCTGCTTTCAGAGTATCTAATCCTTTAAT
 GATCTGGTGGTCTCCTCGTCAATCCATCAGCAATGCTTCTCTCATAGTGTATAGACTTG
 GGAAACCCCAACAGTAGGATATTTCTACAAGGTGTTCAATTTGTGTCACAAGCTGTAGATA
 CAGCAAGAGATGGGGGTGATTGGAATTGCAATACATTGTTTCAGGTGAATAATAAATCA
 AAACTTTTGAATCTTAAGCAGAGATAAATAAAGATAGCAATATGAGACACAGGTGGA
 CGTAGAGTTGGCCTTTTTACAGGCAAGAGGCGAATTGTAGAATTGTTAGATGGCAATAG
 TCATTAAAAACATAGAAAAATGATGTCTTTAAGTGGAGAATTGTGGAAGGATTGTAACAT
 GGACCATCCAAATTTATGGCCGTATCAAATGGTAGCTGAAAAAATATATTTGAGCACTG
 GTCTCTCTTGAATTAGATGTTTATATCAAATGAGCATCTCAAATGTTTTCTGCAGAAAA
 AAATAAAAAGATTCTAATAAAAAAA

>Hs.250726 single4 AW298545 poliA = 2 poliA = 3

TTCCTTCCCTCCCTCCNTTCTCAGGAGCCGCCAGTCCCCAAGTTGGCTGTGGTTGGGCA
 CCTGGTTTGGGTCTGTCAGAGCTGGGCTCAGGCCCTGGGCTCTGAACCTGTGAACCTTG
 CTGTGTTACGAACTTTCTTCTCTGAGGGCCCTGAACCTCTCCTTTTCTTCTTTTGG
 GGGTGGGGTTAACTTTATTTTCTTCTTCCCTGTATCTGCCTCTCCCTTCCCTCAATTTCC
 TGTTTTAAACTGAATGGCACGAAATGTTTTCTTCAACTCGGAGATTCTGTATGGAGA
 GAATCAATTTCTATATTTGCAATAAATTTCTTATTTAAAGCTAAAAAAAAAAAAAAAAA

5

>Hs.79217_ARNm_2 gi|16306657|gb|BC001504.1|BC001504 clon de Homo sapiens MGC:2273 IMAGE:3505512 poliA = 3

GGCACGAGGGCCATCTGTGGGGGCTTTGGGCCAGGGGTCTCCGGACAGCATGAGCGTGGG
 CTTTCATCGGCGCTGGCCAGCTGGCTTTTGCCCTGGCCAAGGGCTTCACAGCAGCAGGCGT
 CTTGGCTGCCCACAAGATAATGGCTAGCTCCCCAGACATGGACCTGGCCACAGTTTCTGCG
 TCTCAGGAAGATGGGGGTGAAGTTGACACCCACACAAGGAGACGGTGCAGCACAGTGA
 TGTGCTCTTCTGGCTGTGAAGCCACACATCATCCCCCTTCATCCTGGATGAAATAGGCGC
 CGACATTGAGGACAGACACATTGTGGTGTCTGCGCGGCGGGCGTCACCATCAGCTCCAT
 TGAGAAGAAGCTGTGAGCGTTTCGGCCAGCCCCAGGGTCATCCGCTGCATGACCAACAC
 TCCAGTCGTGGTGCGGGAGGGGGCCACCGTGTATGCCACAGGCACGCACGCCCAGGTGGA

GGACGGGAGGCTCATGGAGCAGCTGCTGAGCAGCGTGGGCTTCTGCACGGAGGTGGAAGA
 GGACCTGATTGATGCCGTACGGGGCTCAGTGGCAGCGGCCCCGCTACGCATTACAGC
 CCTGGATGCCCCTGGCTGATGGGGGCGTGAAGATGGGACTTCCAAGGCGCCTGGCAGTCCG
 CCTCGGGGCCAGGCCCTCTGGGGGCTGCCAAGATGCTGCTGCACTCAGAACAGCACCC
 AGGCCAGCTCAAGGACAACGTGAGCTCTCCTGGTGGGGCCACCATCCATGCCTTGCATGT
 GCTGGAGAGTGGGGGCTTCCGCTCCCTGCTCATCAACGCTGTGGAGGCCTCCTGCATCCG
 CACACGGGAGCTGCAGTCCATGGCTGACAGGAGCAGGTGTACCAGCCGCCATCAAGAA
 GACCATCCTGGACAAGGTGAAGCTGGACTCCCCCTGCAGGGACCGCTCTGTGCGCTTCTGG
 CCACACCAAGCTGTCCCCCGCAGCCTGGCCCCAGCGGGCAAGGATTGACACGTCCTGCC
 TGACCACCATCCTGCCACCACCTTCTCTTCTTGTCACTAGGGGGACTAGGGGGTCCCC
 AAAGTGGCCCACTTTCTGTGGCTCTGATCAGCGCAGGGGGCCAGCCAGGGACATAGCCAGG
 GAGGGGCCACATCACTTCCCACTGGAAATCTCTGTGGTCTGCAAGTGCTTCCAGCCCAG
 AACAGGGGTGGATTCCCCAACCTCAACCTCCTTTCTTCTCTGCTCCCAAACCATGTCAGG
 ACCACCTTCTCTAGAGCTCGGGAGCCCGGAGGGTCTTACCCACTCCTACTCCAGTATC
 AGCTGGCACGGGCTCCTTCTGAGAGCAAAGGTCAAGGACCCCTCTGTGAAGGCTCAGC
 AGAGGTGGGATCCCACGCCCCCTCCCGGGCCCCCTCCCTGCCCTCCATTGAGGGAGAAACCT
 CTCCTTCCCGTGTGAGAAGGGCCAGAGGGTCCAGGCATCCCAAGTCCAGCGTGAAGGGCC
 ACAGCCCCCTCTTGGCTGCCAAGCACGCAGATCCCATGGACATTTGGGGAAAGGGCTCCTT
 GGGCTGCTGGTGAATCTCTGTGGCCACCACCTCCTGCTCCTGACCTCCCTGGGAGGGTGC
 TATCAGTTCTGTCTGGCCCTTTTCAGTTTATAAGTTGGTTTCCAGCCCCCAGTGTCTTG
 ACTTCTGTCTGCCACATGAGGAGGGAGGCCCTGCCTGTGTGGGAGGGTGGTTACTGTGGG
 TGGAATAGTGGAGGCCTTCAACTGATTAGACAAGGCCCGCCACATCTTGGAGGGCATCT
 GCCTTACTGATTAAATGTCAATGTAATCTAAAAAAAAAAAAAAAAAAAAA

>Hs.47986_ARNm_1 gi|13279253|gb|BC004331.1|BC004331 Homo sapiens MGC:10940 IMAGE:3630835 poliA
 = 3

GATAAATGCGGAGGGACGGTCCAGCTTFTAGCTCTCTGCTCGCCGCCGCCGCTGTGCGCCG
 CACCTCCTCTGATCTACGAAAGTCATGTTACCCAACACCGGGAGGCTGGCAGGATGTACA
 GTTTTTATCACAGGTGCAAGCCGTGGCATTGGCAAAGCTATTGCATTGAAAGCAGCAAAG
 GATGGAGCAAATATTGTTATTGCTGCAAGACCGCCAGCCACATCCAAAACCTTCTAGGC
 ACAATCTATACTGCTGCTGAAGAAATTGAAGCAGTTGGAGGAAAGCCCTTGCCATGTATT
 GTTGATGTGAGAGATGAACAGCAGATCAGTGTGCTGAGTGGAGAAAGCCATCAAGAAATTT
 GGAGCTTATAACCATTGCTAAGTATGGTATGTCTATGTATGTGCTTGGAATGGCAGAAGAA
 TTTAAAGGTGAAATTGCAGTCAATGCATTATGGCCTAAAACAGCCATACACACTGCTGCT
 ATGGATATGCTGGGAGGACCTGGTATCGAAAGCCAGTGTAGAAAAGTTGATATCATTGCA
 GATGCAGCATATTCCATTTTCCAAAAGCCAAAAGTTTTACTGGCAACTTTGTCTATTGAT
 GAAAATATCTTAAAGAAGAAGGAATAGAAAATTTTGACGTTTATGCAATTAAACCAGGT
 CATCCTTTGCAACCAGATTCTTCTTAGATGAATACCCAGAAGCAGTTAGCAAGAAAGTG
 GAATCAACTGGTGCTGTTCCAGAATTCAAAGAAGAGAAAAGTGCAGCTGCAACCAAAACCA
 CGTTCTGGAGCTGTGGAAGAAACATTTAGAATTGTTAAGGACTCTCTCAGTGATGATGTT
 GTTAAAGCCACTCAAGCAATCTATCTGTTTGAAGTCTCCGGTGAAGATGGTGGCAGTGG
 TTTCTTGATCTGAAAAGCAAGGGTGGGAATGTCCGATATGGAGAGCCTTCTGATCAGGCA
 GATGTGGTGATGAGTATGACTACTGATGACTTTGTAAAAATGTTTTAGGGAAACTAAAA
 CCAACAATGGCATTTCATGTCAGGGAATTTGAAGATTAAAGGTAACATGGCCCTAGCAATC
 AAATTGGAGAAGCTAATGAATCAGATGAATGCCAGACTGTGAAGGAAAATATAAAAAAA
 AGTCGACTGCTATGCTCAAAAAGTAAAAAAGCTCAACAGTTAAATCTAATGTTTGT
 TCTTTCTGTTATATTATAAGGATATGCACGTTTGTCTGGAAAAGATAGAATTTGTCTC
 TAAAAGACTTGAAATTGTAATTAAATGGCAAGCTAATCAAACATAAGCTTCATTAAGTG
 GGATTCTAAGACAGTCTGTGTTTTATATTCAAGGGTTTAAACCTTTGAGCCTTACATC
 TCATTCACTGTCTTTCTCCAAGAAAAGTATTTTGGGCGGACAGTCAGATCAAGCAGTAA
 ATTAGCTCTTTCAAATCTTCTTGTCTATGTAATGAAGCTAGTCTGTTTTAAATTTTTTA
 GTTTTGGATTGTATACTAATGAAAATCTTAATGATGTTTTTGTATTTTTATATACTTATTT
 TAAAGAAAATCTTATATAGTACATTTTACAAAATTATAAAAAATGAATTAGTACTGGCG
 AGGACTAAATGAAACATAATTTTTTCATTTTGATACTAGCTTTCCAGGTGGACTTAGCC
 ATAGGAAAATATTACTAATGTAATTTAACAATTTGCTGCATGTATTCCATTTAAAAATAT
 GTTTAAATTTGTCTTAAACAAAATAATTTTTCTCCCTAGGAGTATGCATTTGGCTACAGTG
 TTTTGAACAGAAACCTTAGAATAGGTGATTGGTATGGGCTGAACTGTGTATCCCCCAAT
 TCATTTGTTGAGGTCTTAACCTCCATTTCTTTTGAATGTGACTGTTCCGAGATGAGGCCT
 TTAAGAGGTGACTTAAGTTCAAAGGAGGCTGTTAGTCTAATCCAACATGGTGTCTTTG
 GACATAAGAGATACCAGCAATGTGTGCACAGAACAAAGACCAGGAGAGGACACAGTGAGA
 AGGCAGTTATCTGCAAGCAAAGAGAGAGGCTTCAGAAGAAACAAAATCACCAGCACCTTG
 ATCTTTGACTTCTAATCTCCAGAATAGTGAGAAATAAATTTCTGTTGTTAAGCCGTCCAC

TGTGGGAGGCCGACGCAGGAGGATTGCTTGAGGCCAGGAGTTCAAGGCCAGCCTGGACAA
 CATAGTAAGACCCTATCTCTACCCCCCTAATAAATTAATTTAAAAGCCCCCAATCTGT
 GGTATTTTATTATGGCAGCCCTAGCAAGCTAATACAGTGGTTTGAGAGGCTGGGAGGGTT
 GAGGGGAAGATAAACTTTTAAAAAGCTCTTATCTTTCATTTCAATCAGTTAAAAATACTT
 GCTCAGTGTAACAATTTTGCTTCTCAGCTTCCACTCTAATATTGTTGTGCCATTAGCAA
 TTTAGCTAATCCTGACATTTCTTAGATTTCATAATGTTAGGAGCATTAAATCTGTATTTTA
 CAAGTTAGGAAGCAGAGGATCAGAGATGGGAAAGGACTAGCCCAAGGCCAACATTAACAA
 GCCCTCTAACAAAACTTTACAATACATTTATGTTGAATGGAACTCCAAGATCTCACCTC
 TCCATCCAGGAATGGAGTCCATGTAATCAAAGTGAACCTAAAAATAGGACAGTTTCAACA
 AGTCAGGAGATTACAGCAACTGATCAAAGGGAGTCCAGTCAACGTGAGCAAGCGTGATT
 ATGATGAGGAAGCCCCCTCTGCTTTAATCCACACAAGGAACGTAACCTGAAGTAACCTGA
 TGTTAACCATCTGCTGTGTCTACTATGCTGTTTCCCTTGTTTCTGCTAGTGCTGCTTTAC
 AAATGCAGACCATTCTATCATACCTGGCAGGGCTTCTGTTTTATTTTGTAGGCTGGATGC
 TACCCAGTTTCATGAATCGCTAATAAAGCCAAATTAGATCTTTAAAAAATAAAAAA
 AAA

>Hs.94367_ARNm_1 gij10440200|dbj|AK027147.1|AK027147 ADNc de Homo sapiens: FLJ23494 fis, clon
LNG01885 poliA = 3

TATTTAAAAGTACCCCATGGATGGACCTCCAAATGAGTTTAGGGTAATTGCGCTTAAAATA
 TTAGGACCAAAGTACATTTATTTTATAGATGGAGGAGGGGAGGAGACGAGTGGGGACCAG
 CTTGACATCCAGTCTTCACCTGGACATATGGAAAGAACAAATGTGCGATCTGCTCGTTCC
 CTCTGAAGGTCTCTGTACGTATTTCTCTCTCTCTCCAGAGCATAATAACCAATGACTG
 CTCTCAGAAAAGGTACTGTGACCACCACCTTGCTTGGCTCTCCAACTTCCTCCCCATTTC
 CTCTTGACTCCTGTTTGCCATAACACCTTCTGTCCCCCTAGCCTTGCCCTCAGGTCCCCGAC
 GAATCCTGCCCCTTAATCTGTGGGGGTGGTAGGTGGCACTGGTTTGAAGAGCTTACTGGAT
 CTCCCTCAGTGAGTCAGCCTGGAGTTGTGTTTGA AAAACACAGGCCCTGACTGTGGCTGT
 AAGACCTCCAGACACCACCTGCTGCTGCCTATCATCATCTTCAGGTGCTGGGCTCCCCCT
 GTGGGCCCTCGTCTGCCCCGCCCTCTGCTGCAGCTGTCCCATGGGCGCCCCGCCCTCTCTGAC
 ACCACAAGAGAGCCCATCTAGATTCCAGGAAAAAACTCATCTTTATTTGCCTTCTTCCCA
 CTGAAGGTAAAAGCAACATTAATAACCACAACAAATACTTAGTGAGTGCTTACTATTATT
 CATTTAATTGTAGGCCCTTCCATCCCTGGCCATGATGAGAGACATGCCATAGCTTACTCC
 TAAAGAGACCTGAGGACACACGTGCACAAACATATTGGGCATATCATCAATGGCATCAAA
 ACTGATTTTCCCTGTCTACCCAGAACAGGCCTGAGGGAGAGGGAAAAGCGGATACCCACC
 TGTGTCTGCTGTTTGCCTGCCAAGTCCAGGAACAGTCCATACAGCCCCTGCTGCATCCCACG
 ACGCTGTCACAAAGCAGGAGTTTCATCCGAGGCCAAGGTATGGAGAACTGAGGCCAGAA
 ATTGATGTCCAGAATGCTTTGCTCTTAGCCACTGTACTATTATGGCATATTTTATCTTTA
 TGTATTGCATCATTTTCATGGATTCAAGTTTATCAATGTCTTTGACAAGTTTAAAAATCT
 GTCTGCTAAAAATCTATCAAATACATTAAGGAAAAGTCCCACTGGCACATCTCCCACACC
 AGATGTTAATTATTCATACTGCATGACTGAGGATTTTGGAGGCAGAGAGAGATTTCATCTG
 CAATATTGGAAACACCAATGGAGGTCTATGTCAACACAGAATTTATACAGCAGCTGGTGC
 TAGTCAGAGCTAATGACAGAATTTTCAGTTTAAATAAAAAGACCCCCCACTGAGCACACCAT
 CTTGAAAAAAGTATACTTATCAAACAGCTTTCAATCAGTTCAAGAGAGACACCTTAATTG
 GGGAGAGGAAGAATTGCAGAGTAGTTTGTAAATCATGCCAATTCAGATCAATAACTGCAT
 GTCTGTTCTTTGGTAGAAATAGCTTTTGTCTTTATATTAAAGTAATCACATATATATTCTCT
 CTATTTGGATAAGGAAACCTTCGCTTTATTTGACAATGTATAATGATATACTCTTCTAAT
 TCACCTCTGTGTCTTCACAATAAACATGAGTAAAATTTAGACAAGTGATGGTAAAGGTCA
 ATATAATTATTTATTTTAAAAATAAATTTTGTATCTAACAGGAAAGCAGTTCTTATGAAA
 TTTTATATTTTCAAAAATTGTTTTGTTCAATAAAAATTTTATGAGTAAAGTTAAAAAAA
 AAAAAAAAAAAAAAAAAAAAAA

>Hs.49215_contig1

BI493248|N66529|AA452255|BI492877|AW196683|AI963900|BF478125|AI421654|BE 466675 poliA = 1 poliA = 1

GGGTACCTGGTGGGGCCAATCACCGAGCCATGAACATCAGTAACGTACTCTAAAGACCAA
 GGCTACGATGGCTATGATGGTCAGAATTACTACCACCACCAGTGAAGCTCCAGCCTGGGA
 TGAATTCATCCATTCCTGGCTTTGCATCCGGCTACCATTTTCGAAGTTCAACTCAGGAAGG
 TGCAATATAACAAATGTGCATATTATAATGAGGAATGGTACTACCGTTCCAGATTTTCTG
 TAATTGCTTCTGCAAAGTAATAGGCTTCTTGTCCCTTTTTTTCTGGCATGTTATGGAAT
 GATCATTTGTAAATCAGGACCATTATCAAGCAGTACACCAACTCATAAGATCAAATTTCA
 TTGAATGGTTTGGAGTTGTAGCTCTATAAATAGTAGTTTTTAACATGCCTGTAGTATTGC
 TAACTGCAAAAACATACTCTTTGTACAAGAAGTGCTTCTAAGAATTTTCATTGACATTAAT
 GACACTGTATACAATAAATGTGTAGTTTCTTAATCGCACTACCTATGCAACACTGTGTAT
 TAGGTTTATCATCCTCATGTATTTTTATGTGACCTGTATGTATATTCTAATCTACGAGTT

TTATCACAATAAAAAATGCAATCCTTCAA

5

>Hs.281587_contig2

R61469|R15891|AA007214|R61471|AI014624|N69765|AW592075|H09780|AA709038|A
 I335898|AI559229|F09750|R49594|H11055|T72573|AA935558|AA988654|AA826438| AI002431|AI299721
 poliA = 1 poliA = 2

10

AAGGTGGGCTTTTCATTGTGATTTTGTCTGTTCAGTAATATAGGAGCACATTTTGGCC
 ATTGTAATTACAGGGAACAAAGGGATTGCGGACACATATCTGGACTTCTTTTCTCCCTT
 ATTGTTGTGGAAGAGACACTAGAAATGCTCAAACACCTGCAATATACAGAATATACACAA
 TTTTATTCCAGTATTTCCCTAACATATGGTTTAAAAATTATTCAGGTATACAGTGTATGC
 AATTCCTGCATTATCACAGAGGAACAACCTCTTTTAAAAAATAAATAGGTCAGCCATTT
 TTATTAACGTGCAAAAACCTTTATCACTCTAACATGCTCTAGGTAGTTGAGGAAAAGAGGT
 CTGATCACTGTTTGTATTTTATTTTCTTTGTGGGAACATTTACCTGCTGAGTGTACATG
 AATTTGCTTTCTATAAAAGGCTTTTATGAGTTTACAGTAGAATCAGTGGAAGGAAGAGTT
 AATAAGGGCTGTTTTTAAAAAAACAAACAAACAAACAAATAATTAAAAAAAATTT
 TTACATTCCTTCTTATTCTCTAACTACACTTGGGAAGTGCACCTCAGATAAGTTTGCAGT
 GTGACTGAGAGATGAAGGAAATCCATAGAAAAGGTCCTCTTAGTGAACAAAATTTAGTTA
 TTAACCTTTATAGCTATGAAATTTCCCGGGCATTGTTTTTGTTCAAACAGACTTTAACC
 TCTGCATCATACTTAACCTGCGACATGCGTACAGTATGCATATTTTGTTTGAAAAAA
 ATGTTTCGTTCAGTCTGTGAAGAATATCAAAAATAATAAGGTATTGCTTAATAAAAT
 TGCTAGAATTGTTTAGCAGTACATGCACAATATTTACTAGATTCTTTGTTTTAATAGTG
 TTTTGTGAGACTGAAAATCTTAAAATGGTCTGCGCAAATACAAAAAAAAGAAAACACC
 AAAAAAAA

>Hs.79378_ARNm_1 gi|16306528|ref|NM_003914.2| ciclina A1 de Homo sapiens (CCNA1), ARNm poliA = 3

GGTGTGTGTTCCGGACACATAGAAAGATAACGACGGGAAGAGCGGGGCCCGCTTTGGGGTC
 CAGGCAGGTTTTGGGGCCTCCTGTCTGGTGGGAGGAGCGCAGCGCAGCACCTGCTCG
 TCACTTGGGATGGAGACCGGCTTTCCCGCAATCATGTACCCTGGATCTTTTATTTGGGGGC
 TGGGGAGAAGAGTATCTCAGCTGGGAAGGACCGGGGCTCCAGATTTCGTCTTCCAGCAG
 CAGCCCGTGGAGTCTGAAGCAATGCACTGCAGCAACCCCAAGAGTGGAGTTGTGCTGGCT
 ACAGTGGCCCGAGGTCCCGATGCTTGTCTAGATACTCACCAGAGCCCCGCTGGGCCAGGAT
 CCCCCGAGAGGACAGTGCTAGGGCTGCTAACTGCAAATGGGCAGTACAGGAGGACCTGT
 GGCCAGGGGATCACAAGAATCAGGTGTTATTCTGGATCAGAAAATGCCTTCCCTCCAGCT
 GGAAAGAAAGCACTCCCTGACTGTGGGGTCCAAGAGCCCCCAAGCAAGGGTTTGACATC
 TACATGGATGAAGTAGAGCAGGGGACAGAGACAGCTGCTCGGTGAGAGAGGGGATGGCA
 TTTGAGGATGTGTATGAAGTAGACACCGGCACACTCAAGTCAGACCTGCACTTCTGCTG
 GATTTCAACACAGTTTCCCTATGCTGGTAGATTCTCTCTCTCCAGTCTGAAGAT
 ATATCCAGTCTTGGCACAGATGTGATAAATGTGACTGAATATGCTGAAGAAATTTATCAG
 TACCTTAGGGAAGCTGAAATAAGGCACAGACCCAAAGCACACTACATGAAGAAGCAGCCA
 GACATCACGGAAGGCATGCGCACGATTTCTGGTGGACTGGCTGGTGGAGGTTGGGGAAGAA
 TATAAACTTCGAGCAGAGACCCTGTATCTGGCTGTCAACTTCCTGGACAGGTTCCCTTTCA
 TGTATGTCTGTTCTGAGAGGGAACTGCAGCTCGTAGGAACAGCAGCTATGCTTTTGGCT
 TCGAAATATGAAGAGATATATCCTCCTGAAGTAGACGAGTTTGTCTATATCACCGATGAT
 ACATACACAAAACGACAACCTGTTAAAAATGGAACACTTGCTTCTGAAAGTTCTAGCTTTT
 GATCTGACAGTACCAACCAACCAACAGTTTCTCCTTCAGTACTTGAGGCGACAAGGAGTG
 TGCGTCAAGGACTGAGAACCTGGCTAAGTACGTAGCAGAGCTGAGTCTACTTGAAGCAGAT
 CCATTCCTGAAATATCTTCTTCACTGATAGCTGCAGCAGCTTTTTGCTGGCAAACTAT
 ACTGTGAACAAGCACTTTTGGCCAGAAACCTTGTGCTGCATTTACAGGGTATTCAATTAAGT
 GAAATGTGCTTGCCTGAGTGAGCTTCATAAAGCGTACCTTGATATACCCCATCGACCT
 CAGCAAGCAATTAGGGAGAAGTACAAGGCTTCAAAGTACCTGTGTGTGTCCTCATGGAG
 CCACCTGCAGTTCTTCTTCTACAATAAGTTTCTGAATGGAAGCACTTCCAGAACTTCACC
 TCCATATCAGAAGTGCCAATAATCGTCATAGGCTTCTGCACGTTGGATCAACTAATGTTG
 TTTACAATATAGATGACATTTTAAAAATGTAAATGAATTTAGTTTCCCTTAGACTTTAGT
 AGTTTGTAAATATAGTCCAACATTTTAAACAATAAACTGCTTGTCTTATGACAAAAAA
 AAA

5

>Hs.156469_contig2 AI341378|AI670817|AI701687|AI335022|AW235883|AI948598|AA446356 poliA = 2 poliA = 3

TCCAAGCCATTAAGGACTGTGGAACCTTGCTATGATCATGGACGTGCTGTATGGTGGCGTT
 TGTATGCAGGAATTGATACAGATCCTGAGCTAAAATACCCAAAAGGTGCTGGGCGAGTT
 GCTTTCTCCAATCAGCAGAGCTATATTGCTGCCATTAGTGTCTGGTTTGTTCAGCTTCAG
 CATGGTGTATATTGATAAACGTGTGGAGGTAAAGCCATATGTGCTAGATGACCAGATGTGT

10

GATGAATGCCAGGGCGCACGCTGTGGTGGAAAATTTGCTCCCTTTTTTTGTGCCAATGTC
 ACTTGCCCTGCAGTATTACTGTGAGTTTTGTGGGGCAAATATCCACTCTCGTGTGGACGT
 GAGTTCCATAAGCCATTGGTAAAGGAAGGTGCTGATCGCCACGTCAGATCCACTTCCGC
 TGGAACTAAGAATAGCAAACCTGGCCTCTGTTTAAACAAGGAAAGAAAGGGTGCATGTGGCT
 TACTGTGTCTGAAGATACTGACATGCAGAAGAAATAAGTGCATTCTTCTGCTTTTCACCC
 CAGCTATCAATACATGCATCTTTATCAGCAGCCAAAACACTACAAGCCTCTTGTTTTTCA
 CCAAAACCCTACATCTCAGGCTTACTAATTTTTGTGATATTTTCATGTTCAAATAAAATG
 TTTTTTTGTATTTTCAAAAAAAAAAAAAAAAAAAAAA

>Hs.6631_ARNm_1 gi|7020430|dbj|AK000380.1|AK000380 ADNc de Homo sapiens FLJ20373 fis, clon
 HEP19740 poliA = 3

5

CTCGATGTAGAGGGGTTGGTAGCAGACAGGTGGTTACATTAGAATAGTCACACAACTGT
 TCAGTGTTCAGGAACCTTTTCTTGGGGGTGGGGGAGTTTCCCTTTCTAAAAATGCAAT
 GCACTAAAACTATTTAAGAATGTAGTTAATCTGCTTATTCATAAAGTGGGCATCTTCT
 GTGTTTTAGGTGTAATATCGAAGTCCCTGGCTTTTCTCGTTTTCTCACTTGCTCTCTTGT
 CTCTGTTTTTTTAAACCAATTTTACTTTATGAATATATTCATGACATTTGTAATAAATGT
 CTTGAGAAAGAATTTGTTTCATGGCTTCATGGTCATCACTCAAGCTCCCGTAAGGATATT
 ACCGTCTCAGGAAAGGATCAGGACTCCATGTACAGTCCTGCCATCTTACTTTCCTCTTG
 TCGAGTTCTGAGTGGAAATAACTGCATTATGGCTGCTTTAACCTCAGTCATCAAAAGAAA
 CTTGCTGTTTTTTAGGCTTGATCTTTTCTTTGTGGTTAATTTTCTGTATATTTGTGAA
 AATGGGGGATTTTCCCTCTGCTCCACCCACCTAAACACAGCAGCCATTTGTACCTGTTT
 GCTTTCCCATCCCACTTGGCACCCACTCTGACCTCTTGTGAGTTTCTGTTCCTGTTCTGTTCA
 TCTTTTGAAGGCGCTTCTTTGAGCTACAAACATCTGGTAAGACAAGTACATCCACT
 CATGAATGCAGACACAGCAGCTGGTGGTTTTGTGTATACCTGTAAAGACAAGCTGAGAGG
 CTTACTTTTTTGGGGAAGTAAAGAAGATGGAAATGGATGTTTCATTTGTATGAGTTTGGGA
 GCAGTGTGAAGGCCAAAGCCGCTACTGGTTTGTAGTTAACCTAGAGAAGGTTGAAAAA
 TTAATCCTACCTTTAAAGGGATTGAGGTAGGCTGGATTCCATCGCCACAGGACTTTAGT
 TAGAATTAAATTCCTGCTTGTAATTTATATCCATGTTTAGGCTTTTCATAAGATGAAACA
 TGCCACAGTGAACACACTCGTGTACATATCAAGAGAAGAAGGAAAGGCACAGGTGGAGAA
 CAGTAAAAGGTGGGCAGATGTCTTTGAAGAAATGCTCAATGTCTGATGCTAAGTGGGAGA
 AGGCAGAGAACAAGGATGTGGCATAATGGTCTTAACATTATCCAAAGACTTGAAGCTCC
 ATGTCTGTAAAGTCAAATGTTACACAAAAAAATGCAAATGGTGTTCATTGGAATTACC
 AAGTGCTTAGAACTTGCTGGCTTTCCCATAGGTGGTAAAGGGGTCTGAGCTCACACCGAG
 TTGTGCTTGGCTTGCTTGTGTCAGCTCCAGGCACCCGGTGGGCACTCTGGTGGTGTGTTGTG
 GTGAATGAATTGAATCCATTGTTGGGCTTAAGTTACTGAAATTGGAACACCCCTTTGTCC
 TTCTCGGCGGGGCTTCCCTGGTCTGTGCTTTACTTGGCTTTTTTCTTCCCGTCTTAGCC
 TCACCCCTTGTCAACCAGATTGAGTTGCTATAGCTTGATGCAGGGACCCAGTGAAGTTT
 CTCCGTAAAGATTGGGAGTCGTCGAAATGTTTAGATTCTTTTAGGAAAGGAATTATTTT
 CCCCCCTTTACAGGGTAGTAACCTCTCCACAGAAGTCCAATATGGCAAAATTACACAA
 GAAACAGTATTGCAATGACACCATTACATAAGGAACATTGAACTGTTAGAGGAGTGCTC
 TTCCAAACAAAACAAAAATGTCTCTAGGTTTAGTCAAGAGCTTTCACAAGTAATAACCTTT
 CTGTATTAAATCAGAGTAACCCCTTCTGTATTGAGTGCAGTGTTTTTACTCTTTTCTC
 ATGCACATGTTACGTTGGAGAAAATGTTTACAAAAATGGTTTTGTACACTAATGCGCAC
 CACATATTTATGGTATATTTTAAAGTGACTTTTTATGGGTTATTTAGGTTTTCTGCTTAGT
 TGTAGCACACTTACCCTAATTTTGCCAATTATTAATTTGCTAAATAGTAATACAAATGAC
 AACTGCATTAAATTTACTAATTATAAAGCTGCAAGCAGACTGGTGGCAAGTACACAGCC
 CTTTTTTTTGTCAGTGCTAATTTGTCTACTGTGATTATGAAAATTACTGTTGTCCCCCA
 CCCTTTTTTCTTAAATAAAGTAAAAATGACACCCTAAAAA
 AAAAAA

>Hs.155977_contig1 AI309080|AI313045 poliA = 1 WARN poliA = 1

TATACGGCTGCTAGAAGACGACAGAAGGTGGCTTGGGGGTGGATATCTTTGGGTTGCTGG
 AAAAGGTGTGGGAAGGTTTCAGGATGGTGGGAGGGACTGAGGTCCCTGAGGTGAAGAGGCC
 CTTGGTCTGACGGGTTTGACCCGTGCTTGGACCCTTGGAGCAGTGTTGTGTGAACCTGCT
 CTAGAATCTGCCTTCTCCGTTGTCAATAAAGCCTCCCCCTCATGACCTAAAAA
 AAAAAA

10

ES 2 550 652 T3

>Hs.95197_ARNm_4 gi|5817138|emb|AL110274.1|HSM800829 ARNm de Homo sapiens; ADNc DKFZp564I0272 (del clon DKFZp564I0272) poliA = 3

GAGCAGGAAAATATATACCCTAAACAGAACTCTTACTTGTTTTATGAGCAAGTCTGAGT
GAGTCCTAAAATGGCTGGCGAAGAGCTACCAATACTGACTGACAGGTCACCTTAAAGCCT
CTAGGTGTGCCAAGTTTGATTTATCTTAGGGACTAGAACCTAGTCTTCTAAATGTGATT
TGCCTTGCTGTTTCGTCCTGATGTGAAGCTAACCACACAGAGAGATTGGGCTGCATCAGT
AATGATATGCATACCTTTCGTGCATCAGTGAGCTTCTTCCCTGTTAACTGTATGACCACA

AAATTTAGCTGGAGTAAATAAATATGCGACAGAAATCCTGGAACAAGATGGTGAAATTGC
 TTAAGAATCGAGACTTCAGGGCTCAATGACCTCTGAGCATGTTTCCCAAAGTGTGACCCA
 CATGACCATCTGTCTCTCAGTCTCCTGGTCCCTCCGTAGAGCTTCTGAAACTGAATCTTT
 GTGGGGTGGGGTAGCGTTCAAGAATCAAAAGTTGAACCAAGCTCTTTGGGTGATACTTA
 TGTATACTGAGGTTTCAGGAAGTCTGGAGAGATGACTGGGCACCAAGAGGATGACAGTGGA
 CTCAGCTGGCATCCCTTAGCTGTTTATGTCAGAGCTGAGTGGGCACCTCTGTCTCTGAC
 CCCAGCTTCAGTGCTCTTTATCTCCTCCATGCCTCCTCAGTCGTGCTGCTCTAAGACTGC
 TTAAGTGGCTTTCTTTCATGTCCTGGGCACAGAGCAGTTCTTTTGGTAGCAGATTGAGTC
 CACTTCCCCCGTGACAGATCACTGCTCAGGACCCAGAGAGGAGCAGCTCTGCTCCAGCA
 GGGTTTTCCATTGCATCACACACCCAAACGGTAGGATCCAACAGTCACACTTGAAAGCAA
 CCATAATTGTGAGGTTTCTGATGCTGTAGACTTCCTTACATTTCACAACTAGTTAGA
 GAGTCACATGGGGGTGAAGTGTGGCTCGCGACCTGCCCAACAAGTGCCTGCAGAAAGCCA
 GGAAACAAAGGAGTAAATTCACTTCAAATGGGATGCACATGGTGTCCGTGATGAAGAGAC
 ACATTGAGAAATTGCCCAAGGACAGGAAATGACCAGAGAGAGCCAGAGCTGAGCTGGTAA
 TAAAGAGACTCCGAGACTGAGTGGAGTTAATGAGGGAAGCATGCAACGAGTGGGGCAATT
 TCAGTTGGTTTTCTCTCATTGCTTTAAGCGAAATGAACATACCGACAGGAGAACAGCCTG
 CTTGCCCCAGTCTCTCCTTGGCCGCCCTCTGTTGTCCCTGTCAACTCAGGTGCCACGGT
 GCTCAGAGGAGGTGCTGGCAAAGCCCCCTGGAGCCTTATGTAGGCCATGGGGCTCTTAA
 AGGAACCTGAATGAATCATTACAGCAGGTCTCTCTTGTAAAGCCAGCCACAGTAACCTC
 GTACACTGACTGTTTCAAAGACAGCCTTTCTTAATCATTAAATTGTTTCATATTCAAAT
 ATATCTCCTAATTGTTTTTATTTTTTCTGATCTAGAAGATATGACAACAGGGTAGAACT
 TGGGAAGAGGGAATAGGAAGCTCGCCCTTCTCTCCTTCCCTCCTCCTCCTCTCTACTTTCCT
 TCCTTCTCTGGTCATCAGGTACCTTCTTTGTGCTGCTGTTGTAGGCTACACCCATGTT
 TGGTGGAAGGCAAAAAGAAAAATCAGTAGGATACAACCTCAGTAGGGAAGACAGAGATATT
 CAAGCCCCCTTGTCTCCAGTGTGATAAGTGTGGTGGTTGAGGTGTGAACAAGGGGCTCT
 GTGAACAGAGAGGACGAAAGAGGAGCTCCTCCTGAGGCTGTTGGGAAAAGCATCACTGAA
 GAGTGACTTTCAGAAGAAGAGAAGAAAAAGAGGAGAACATGCGTGATTTTATAATGAAT
 AGATTAGATAAGGGGAAAAAGGCATTTAAACAAGGCAAAAAGAACAGGAGAATAGAGAA
 GAGATGTGGAGGAGAAGGAGCACTGTAGTAAACACGCAGAAGGACAGGAACACTTAGACA
 TGCAACCCACTCCACCCCTCCGTCTTGGGGGAGGAAAGCACACTACTGTCCCAAAGAACT
 AATACTGAACCAAGTGTGCTGCCTTGTGGAGAGAGGCATGGCCAAGGCGTTTCAGAGACCTGGG
 CCTGGTCCCACCGCTGCCACAGCACTCAGCCTCTGAGCACAGCCTGGGGTCATCTGTGT
 GCCCTCTGGCAAGGCTGATGGTAGTTCTCTGAGTAATTGAGAGTCATTGCTGTCTGTG
 CAGTATTGGTAAACAAGTCACCTTTTAACTTTAACTACTTTTAAAAAATCTTTAAAGTT
 TTAAGAAACTTCTTTAAAAAATACTCATGAGATGACAGTTTCTCTGACCCTCAGAGGAA
 GGCTGGGCTGCGCATACGTGAGGAATTTTACATGAACATCCAGGACTTGCTGTTTCGCA
 GGTGATAAAGTGCACCTCCCAGGACTCCCGCTGCACTCACATGCAGCTCCCTGGACTTC
 TGGTATCTGACCCGGCCCATTTCTGTGTTTTCAGGGGAGAATTTGGCTTGGGGAGTACTC
 AGAAGTTAAGACGGTGACAGTAAAGATCCCCAGAAGAACTCCTAAGAAGGCCAAGAAGG
 AGGATGAAGCCAGCCTGCACGTCTGTCCCTCTCTGCTTTCTCTGTAGGGCCCAGCTCTC
 AGGAATACAAAGTTGAGCCACGGTCTTACTTAAAGATTGAAAAGATAACATGTAGGCCA
 GGCAGGTCACTGCACAACTAAAGCAAAACAGCTGGGTACAGTTTCTTGGCACTCTGTAA
 GGGCCACCTTAATCATACCAAATATTGGGGAAAGTGGGATAAAGGGAGGAGGAGCTA
 GCAGACACATCCAGTATCTCCTTCTGGAGCACAGGATGAAATAAGGGAGCTGTATTATT
 CATGTCTTTGTACAAAGAACTTCTCTCAAGGAAAGGTGACCTTTCTCTCTCTCTCAT
 TTTCTCTCTTCCAGGCCCTCCTCGCTCACCCACCCCTCCCTCTCTCTCAAGGAGATGTCA
 GCTGAGCTCATTCTGGGGCAGATGTTTGGGGCCGGGAACAATTTTCAAGGTGTAAAGCC
 AAATATCATTTTATGTTATCCATTTCTTCAAAGCAAAACATGAAATGGTTTTAGCTAGA
 GTCAGACCAGAATGAAAATGCCAGGAGCTGGTACACTACAGATGTAGTAAGAACCCTGGGA
 TATTCTGACCCAATCTGGTTTTCTTTTACCCATAAATAACATGAATGAAAAAGATTGG
 GACAATAGAGACTGGAAGTCATCATGTGCAGTTTACCGCTTCTGAGCTTGCTGCGAGTTTT
 GGGGTGTGTGTATTAGATTCTTCTCAGTTATTCTGGAATAAGGCAAGGAGTGGGTG
 TTTTTCATAGCTAGATAAGATCTTTTCAAAGTTTTTCTTAGAACCACCAAAAAACAAT
 CCGAGTAGGCCCCGAGAATTGATAATGCTGGATGCCTTGCAGACATCATTGATTCTTAA
 TATTGGGCAACAATTATTATTAAATGAATTATTCTGTAGTTGGAATCTGTACTCTCTGA
 ACCTCTACACCAATAACTGCTGCAGGTGTGATTTTGGTCTGTGCACACTGTACATCTATCA
 TAATGTGCCCTGTATCTATTGGCAGTGACCTTGGAAAATCTGGCCAAGCCTAGGGGTTTC
 CTTTTCCATTTGCCAAGTTCATTGTGCCAGGACTGCCGTGCTCCACTGAGCTCCTCTGT
 CACACCCCATTTCTTGGCCCTCACTGGGCAGGCCATGGCCTACAGCTTGCAGGGAGTAAAG
 CAGGCCCGCCTCCCTTTCTTCCCATCCACATACTCCTCTCTGCTTTCCAGTGACTCCAC
 CAGTTTGATGTGGGAAGTGTAGCTTCTTTCTTCTTCCATCCCTTCTTCCATCTTTCC

AGCTGTCAAATCCAATCCAGTCTCTAACCTAAATGCAGATCATTTATTTAAAAGTACCAA
ACATAACCCAGAGTATGTGGAATATGGGCAACATATATATAGCCTTCTGTATTTAACGAT
CTTCTGCTTCTTAACCGTACCAGTTTTCTATTTATAACTCTTATCTATCCATGATGTTTT
AAAGTCTCCACTTGCTGTTATTTACAAACGACAGTGCATTTCAGCAGCCCAGTGCCGTGAG
CCCTGACAGATGCCGTATTTCTGAGTGTCTCCATGTGAATGCTGCCCTCCTGTAGCATGT
GTCCAAGTGGACATAGCCACTAACCAACTAGTTACCTTTGGACTGCAACAAAAAATGTGA
AAATGAAGATTTATTTCTTTTAATTTACTTAAAAAGAAACCTCTGTGCTAGCAATAAAGC
ATTTATATTGTGCAAAAAAAAAAAAAAAAAAAAAAAC

>Hs.48956_contig1 N64339|AI569513|AI694073 poli = 1 poliA = 1

TGAAAATTTATATAACTGTTGTTGATAAGGAACATTATCCAGGAATTGATACGTTTATTA
GGAAAAGATATTTTTATAGGCTTGGATGTTTCTAGTTCTGACTTTGAATTTATATAAAGT
ATTTTATAATGACTGGTCTTCTTACCTGGAAAAACATGCGATGTTAGTTTITAGAAATTA
CACCACAAGTATCTAAATTTGGAACCTTACAAAGGGTCTATCTTGTAATATTTGTTTTGCA
TTGTCTGTTGGCAAATTTGTGAACGTGTCATGATACGCTTAAGGTGGAAGTGTTCATTGC
ACAATATATTTTTACTGCTTTCTGAATGTAGACGGAACAGTGTGGAAGCAGAAGGCTTTT
TTAACTCATCCGTTTGCCAATCATTGCAACAACCTGAAATGTGGATGTGATTGCCCTCAAT
AAAGCTCGTCCCCATTGCTTAAGCCTTCAAAAA

5

>Hs.118825_ARNm_10 gi|1495484|emb|X9675.1|HSSAPKK3 ARNm de H. sapiens para quinasa de quinasa
MAP poliA = 3

CTTTTAGCTGCCAGCCCTGGCCCATCATGTAGCTGCAGCACAGCCTTCCCTAACGTTGCA
ACTGGGGGAAAAATCACTTTCAGTCTGTTTTGCAAGGTGTGCATTTCCATCTTGATTCC
CTGAAAGTCCATCTGCTGCATCGGTCAAGAGAACTCCACTTGCATGAAGATTGCACGCC
TGCAGCTTGCATCTTTGTTGCAAACTAGCTACAGAAGAGAAGCAAGGCAAGTCTTTTG
TGCTCCCCCTCCCCCATCAAAGGAAAGGGGAAAATGTCTCAGTCGAAAGGCAAGAAGCGAA
ACCCTGGCCTTAAATTTCCAAAAGAAGCATTGTAACAACCTCAGACCAGTTCCACACCAC
CTAGAGATTTAGACTCCAAGGCTTGCAATTTCTATTGGAAATCAGAACTTTGAGGTGAAGG
CAGATGACCTGGAGCCTATAATGGAACGGGACGAGGTGCGTACGGGGTGGTGGAGAAGA
TGCGGCACGTGCCAGCGGGCAGATCATGGCAGTGAAGCGGATCCGAGCCACAGTAAATA
GCCAGGAACAGAAACGGCTACTGATGGATTTGGATATTTCCATGAGGACGGTGGACTGTC
CATTCACTGTACCTTTTATGGCGCACTGTTTCGGGAGGGTGTATGTGTGGATCTGCATGG
AGCTCATGGATACATCACTAGATAAATTTCTACAAACAAGTTATTGATAAAGGCCAGACAA
TTCCAGAGGACATCTTAGGGAAAAATAGCAGTTTCTATTGTAAAGCATTAGAACATTTAC
ATAGTAAGCTGTCTGTCATTCACAGAGACGTCAAGCCTTCTAATGTACTCATCAATGCTC
TCGGTCAAGTGAAGATGTGCGATTTTGGAAATCAGTGGCTACTTGGTGGACTCTGTTGCTA
AAACAATTGATGCAGGTTGCAAACCATACATGGCCCTGAAAGAATAAACCCAGAGCTCA
ACCAGAAGGGATACAGTGTGAAGTCTGACATTTGGAGTCTGGGCATCACGATGATTGAGT
TGGCCATCCTTCGATTTCCCTATGATTTCATGGGGAACCTCATTTCAGCAGCTCAAACAGG
TGGTAGAGGAGCCATCGCCACAACCTCCAGCAGACAAGTTCTCTGCAGAGTTTGTGACT
TTACCTCACAGTGTCTTAAAGAAGAATTCAAAGAACGGCCTACATACCCAGAGCTAATGC
AACATCCATTTTTTACCCTACATGAATCCAAAGGAACAGATGTGGCATCTTTGTAAAAC
TGATTCTTGGAGACTAAAAAGCAGTGGACTTAATCGGTTGACCCTACTGTGGATTGGTGG
GTTTCGGGGTGAAGCAAGTTCACTACAGCATCAATAGAAAGTCATCTTTGAGATAATTTA
ACCCTGCCTCTCAGAGGGTTTTCTCTCCCAATTTTCTTTTTACTCCCCCTCTTAAGGGGG
CCTTGGAAATCTATAGTATAGAATGAAGTGTCTAGATGGATGAATTATGATAAAGGCTTAG
GACTTCAAAGGTGATTAATATTTAATGATGTGTTCATATGAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAA

10

>Hs.135118_contig3
AI683181|AI082848|AW770198|AI333188|AI873435|AW169942|AI806302|AW340718|BF196955|AA909720
poliA = 1 poliA = 2

15

CAGTCCCACCATGTATTTTGCTTTGTTTCTAAAAAGCTTTTAAAACTGTTATTTAATA
CCAAAGGGAGGAATCGTATGGGTCTTCTGCCACCGTTGTGACTAAGAATGCACAGGGA
CTTGGTTCTCGTTGCACCTTTTTTTAGTAACATGTTTCATGGGGACCCACTGTACAGCCC
TTCATTCGTCTGTGTGTCAGTTTGGCCTGGCCTGACACTGGCTGCCCCAGCGGGGACCACGG
AAGCAGAGTGAGAGCCTTCGCTGAGTCAATGCTACCTTCAGCCCCAGACGCATCCCATTT
CCATGCTCTTCCATGCTCACTGCTCATGCACTTTTTACACGGTTTCTTCCAAACAGCCCGG
TCTTGATGCAGGAGAGTCTGGAAGGAAGAAAAATGTTTTAGTTTCAAAATTCAAAGGA
AAAAGTTGAGGACTTATTTTGTCTGTCAAGATTGCAAGAACATGTAAAATGTACGGAGC
TTCATAATACGTTATATTGTTCCGAAGCAGCTCGTTGAGAAACATTTGTTTTCAATAACA
TTTTAGCTTAAAAAAA

>Hs.171857_ARNm_1 gi|13161080|gb|AF332224.1|AF332224 ARNm de proteína de testículo de Homo sapiens,
cds parcial poliA = 3

5

TCACCTCGTGGCGTAGGGGAGAGGTAACACCCGAGAAGAGGCAGCGGCGGTGGCNCAGAGA
CGATTGGTGCCAAACAGGGCAGAACGCAACTCAGCTCTGGGTTTGTGAATAGCACAAATGG
AAGAAGCTGGACTTTGTGGGTTAAGAGAGAAAGCAGATATGTTGTGTAACCTCTGAATCAC
ATGATATTCTTCAACATCAAGACTCAAATTGCAGTGCCACAAGTAATAAACATTTATTGG
AAGATGAAGAAGGCCGTGACTTTATAACAAAGAACAGGAGTTGGGTGAGCCCAGTGCAC
GCACACAAGAGTCAAGAAGGGAGCTTCCTGAGCAAGAAGTAGCCCCCTCCGTCTGGTCAGC
AAGCTTTACAATTGCAACAGGAACAAAGAAAAAGTCTTAGGAAAAGAAGTTTTATTATTG
ATGCAAGCCCTAAACACTCTTCCGACTCCAGAGGAGAAGCTGGCAGCTCTCTGTAAGAA
ATATGCTGATCTTGGAAATTACCTCTTCTATAGAAGAGTTTGTGTTTGAAGTATACGATT
TGAAACAAAATTCTTTTTTTGGAGACTATGGAACATTCTCAACAGGGAAACCTACTAG
ACTTTGTAAAGCAAATAATGGAAGAGATACAGAACTTTTGAAGAATCATGGGAAATTTT
TATAATTAAATAAATGCTAAAATTCTGTTTGTGAAACATTTATGGGAATTATCACTGAC
AGTTTTTGTACACTTTCAAATAGTGTTAAAGCAGCAACTCCATGTTGTAAATGCACAAAA
CAAAATATTAGTTAATAATCAACTCCAAGAATAAAGCTGTAACAATAATAGTTAAAAAAA
A

>Hs.18910_ARNm_3 gi|12804464|gb|BC001639.1|BC001639.1| clon de Homo sapiens MGC:1944
IMAGE:2959372 poliA = 3

10

GGACACGAGGGGTACAGCAGCCGCCAGACTTCTCTGCCGAAGTCCGAGCCCCCTCCCGGGGCTG
GAGGGGGGCAAGCGGGTTCCGAGGTGCAAAGCCTGGTGCCCCGAGCCTGCGGAGCTCGG
GGCCAGCATGGCCCCCAGCGTGC AACAGGCGTACCGGAGGCGCTGGTGGATGGCCTGCAC
GGCTGTGCTGGAGAACCTCTTCTTCTCTGCTGTACTCCTGGGCTGGGGCTCCCTGTTGAT
CATTCTGAAGAACGAGGGCTTCTATTCCAGCACGTGCCCAGCTGAGAGCAGCACCAACAC
CACCCAGGATGAGCAGCGCAGGTGGCCAGGCTGTGACCAGCAGGACGAGATGCTCAACCT
GGGCTTACCATTGGTTCCCTTCGTGCTCAGCGCCACCACCCTGCCACTGGGGATCCTCAT
GGACCGCTTTGGCCCCCGACCCGTGCGGCTGGTTGGCAGTGCCCTGCTTCACTGCGTCTTG
CACCTCATGGCCCTGGCCTCCCGGGACGTGGAAGCTCTGTCTCCGTTGATATTCTTSGC
GCTGTCCCTGAATGGCTTTGGTGGCATCTGCCTAACGTTCACTTCACTCACGCTGCCCAA
CATGTTTGGGAACCTGCGCTCCACGTTAATGGCCCTCATGATTGGCTCTTACGCCTCTTC
TGCCATTACGTTCCAGGAATCAAGCTGATCTACGATGCCGGTGTGGCCTTCGTGGTCAT
CATGTTACCTGGTCTGGCCTGGCCTGCCTTATCTTTCTGAAGTGCACCCTCAACTGGCC
CATCGAAGCCTTTCTGCCCTTGAGGAAGTCAATTACACGAAGAAGATCAAGCTGAGTGG
CTGGCCCTGGACCAAGAAGGTGACAGGTGACCTTCTTACACCATGTGACCACCATGGG
CAGAGGCTCAGCCAGAAGGGCCCCAGCCTGGAGGACGGTTCGGATGCTTTCATGTCAAC
CCAGGATGTTTCGGGGCACCTCAGAAAACCTTCTGAGAGGTCTGTCCCCTTACGCAAGAG
CCTCTGCTCCCCCACTTTCTGTGGAGCCTCCTCACCATGGGCATGACCCAGCTGCGGAT
CATCTTCTACATGGCTGCTGTGAACAAGATGCTGGAGTACCTTGTGACTGGTGGCCAGGA
GCATGAGACAAATGAACAGCAACAAAAGGTGGCAGAGACAGTTGGGTTCTACTCTCCGT
CTTCGGGGCCATGCAGCTGTTGTGCCTTCTCACTGCCCCCTCATTGGCTACATCATGGA
CTGGCGGATCAAGGACTGCGTGGACGCCCCAACTCAGGGCACGTGCTCTCGGAGATGCCAG
GGACGGGGTTGCTACCAAATCCATCAGACCACGCTACTGCAAGATCCAAAAGCTCACCAA
TGCCATCAGTGCCCTTACCCTGACCAACCTGCTGCTTGTGGGTTTTTGGCATCACCTGTCT
CATCAACAACCTTACACCTCCAGTTTGTGACCTTTGTCTGCACACCATTGTTTCGAGGTTT
CTTCCACTCAGCCTGTGGGAGTCTCTATGCTGCAGTGTTCCCATCCAACCCTTTGGGAC
GCTGACAGGCCTGCAGTCCCTCATCAGTGCTGTGTTTCGCCCTTGCTTCAGCAGCCACTTTT
CATGGCGATGGTGGGACCCCTGAAAGGAGAGCCCTTCTGGGTGAATCTGGGCCTCCTGCT
ATTCTCACTCCTGGGATTCTTGTGCTTCTTACCTCTTCTATTACCGTGCCCGGCTCCA
GCAGGAGTACGCCGCCAATGGGATGGGCCCACTGAAAGGTGCTTAGCGGCTCTGAGGTGAC
CGCATAGACTTCTCAGACCAAGGGACCTGGATGACAGGCAATCAAGGCCTGAGCAACCAA
AAGGAGTGCCCCATATGGCTTTTCTACCTGTAAACATGCACATAGAGCCATGGCCGTAGAT
TTATAAATACCAAGAGAAGTTCTATTTTTGTAAAGACTGCAAAAAGGAGGAAAAAAAAACC
TTCAAAAACGCCCCCTAAGTCAACGCTCCATTGACTGAAGACAGTCCCTATCCTAGAGGG
GTTGAGCTTTCTTCTCCTTGGGTTGGAGGAGACCAGGGTGCCCTTATCTCCTTCTAGC
GGTCTGCCTCCTGGTACCTCTTGGGGGGATCGGCAAAACAGGCTACCCCTGAGGTCCCATG
TGCCATGAGTGAGCACAACATGCTGTCTGTGTATGTGTGAATGTGAGAGAGACACAGC
CCTCCCTTTTCAGAAAGGAAAGGGGCGCTGAGGTGCGAGCTGTGCTCCTGGGTAGGGGTTGGG
GTCGGCCCTTCCAGGGCCAGGAGGGCAGGTTCCCTCTCTGTGTGCTGCTGCTTGCAGTCT
TTAGAGGAAATAAAAAAGGGAAGTGAGAAAAA

>Hs.194774_ARNm_1 gj|16306633|gb|BC001492.1|BC001492 clon de Homo sapiens MGC:1774
IMAGE:3510004 poliA = 3

5

GGCACGAGGGAGGCGGCGGCTCCAGCCGGCGCGGCGCGAGGCTCGGCGGTGGGATCCGGC
GGGCGGTGCTAGCTCCGCGCTCCCTGCCTCGCTCGCTGCCGGGGGCGGTCCGAAGGCGCG

GCGCGAAGCCCGGGTGGCCCGAGGGCGCGATGGCTGCTCCTGTCCCGTGGGCCTGCTGTG
 CTGTGCTTGCCGCCGCCGCCGAGTTGTCTACGCCCAGAGACACAGTCCACAGGAGGCAC
 CCCATGTGCAGTACGAGCGCCTGGGCTCTGACGTGACACTGCCATGTGGGACAGCAAAC
 GGGATGCTGCGGTGACGTGGCGGGTAAATGGGACAGACCTGGCCCCCTGACCTGCTCAACG
 GCTCTCAGCTGGTGTCTCCATGGCCTGGAACCTGGGCCACAGTGGCCTCTACGCCTGCTTCC
 ACCGTGACTCCTGGCACCTGCGCCACCAAGTCCTGCTGCATGTGGGCTTGCCGCCGCGGG
 AGCCTGTGCTCAGCTGCCGCTCCAACACTTACCCCAAGGGCTTCTACTGCAGCTGGCATC
 TGCCCAACCCACCTACATTCCCAACACCTTCAATGTGACTGTGCTGCATGGCTCCAAAA
 TTATGGTCTGTGAGAAGGACCCAGCCCTCAAGAACCGCTGCCACATTGCTACATGCACC
 TGTCTCCACCATCAAGTACAAGGTCTCCATAAGTGTGAGCAATGCCCTGGGCCACAATG
 CCACAGCTATCACCTTTGACGAGTTACCATTGTGAAGCCTGATCCTCCAGAAAATGTGG
 TAGCCCGGCCAGTGCCCAAGCAACCTCGCCGGCTGGAGGTGACGTGGCAGACCCCTCGA
 CCTGGCCTGACCTGAGTCTTTTCTCTCAAGTTCTTTCTGCGCTACCGACCCCTCATCC
 TGGACCAGTGGCAGCATGTGGAGCTGTCCGACGGCACAGCACACCATCACAGATGCCT
 ACGCCGGGAAGGAGTACATTATCCAGGTGGCAGCCAAGGACAATGAGATTGGGACATGGA
 GTGACTGGAGCGTAGCCGCCACGCTACGCCCTGGACTGAGGAACCGCGACACCTCACCA
 CGGAGGCCAGGCTGCGGAGACCACGACCAGCACACCAGCTCCCTGGCACCCCACTA
 CCACGAAGATCTGTGACCCTGGGGAGCTGGGCAGCGGCGGGGGACCCCTCGGCACCCCTTCT
 TGGTCAGCGTCCCATCACTCTGGCCCTGGCTGCCGCTGCCGCCACTGCCAGCAGTCTCT
 TGATCTGAGCCCGGCACCCCATGAGGACATGCAGAGCACCTGCAGAGGAGCAGGAGGCCG
 GAGCTGAGCCTGCAGACCCCGTTTCTATTTTGCACACGGGCAGGAGGACCTTTTGCATT
 CTCTTCAGACACAATTTGTGGAGACCCCGCGGGCCCGGGCCTGCCGCCCCAGCCCTG
 CCGCACCAAGCTGGCCCTCCTTCTCCTCAGGGGAGGTGGGCCATGCAGCTAACCCACC
 CACCAAAGACCCCTCACCCCTGGCCCTTGGGCTGGACCTCCAATGCCAGCGACTCCCA
 GGAGCCCTTGGGGGACGTGAGGGGAGCCTCTCACATCCGATTTCTCCTCCTGCCCCAGCC
 TCCTGTCTATCCCAGGCTCTCTGTTGCCACCATCAGATTATAAGCTCCTGATGCTGGGGG
 GGCCAGCCATCCCCCTCCCCCAGCACCCACAATTTTCAGTCCCTCCTCTGCCCCTG
 TTTTGTATAACCCCTCCCCCTGACCCTGCTCCTATCCACAGTATTTAATGCCCTGTCAGTC
 CCTTCTAGTCTGACTCAATGGTAACTTGCTGTATTTGAATTTTTTATAGATGTATATACA
 GGGTGGGGGAGTGGGCGGTTCTCATTAACGTACCATTTTCATGAAAAAAAAAAAAAAAA
 AAA

>Hs.127428_ARNm_2 gj|16306818|gb|BC006537.1|BC006537 clon de Homo sapiens MGC:1934
 IMAGE:2987903 poliA = 3

GGCACGAGGAGTTTCATAATTTCCGTGGGTGCGGCCGGGGCGGGCCAGGCGCTGGGCACGG
 TGATGGCCACCACTGGGGCCCTGGGCAACTACTACGTGGACTCGTTCTGCTGGGCGCCG
 ACGCCGCGGATGAGCTGAGCGTTGGCCGCTATGCGCCGGGGACCCCTGGGCCAGCCTCCCC
 GGCAGGCGGCGACGCTGCCGAGCACCCGACTTCAGCCCGTGCAGCTTCCAGTCCAAGG
 CGACGGTGTTTTGGCGCCTCGTGGAACCCAGTGACGCGGCGGGCGCCAACGCTGTACCCG
 CTGCGGTGTACCAACCACCATCACCAACCACCCCTACGTGCACCCCCAGGCGCCCGTGGCGG
 CGGCGGCGCCGGACGGCAGGTACATGCGCTCTGGCTGGAGCCCACGCCCGGTGCGCTCT
 CTTTCGCGGGCTTGCCCTCCAGCCGGCCTTATGGCATTAAACCTGAACCGCTGTGCGCCA
 GAAGGGGTGACTGTCCCACGCTTGACACTCACACTTTGTCCCTGACTGACTATGCTTGTG
 GTTCTCTCCAGTTGATAGAGAAAAACAACCCAGCGAAGGCGCCTTCTCTGAAAACAATG
 CTGAGAATGAGAGCGGCGGAGACAAGCCCCCATCGATCCCAATAACCCAGCAGCCAACT
 GGCTTCATGCGCGCTCCACTCGGAAAAAGCGGTGCCCTATACAAAACACCAGACCCTGG
 AACTGGAGAAAGAGTTTCTGTTCACATGTACCTCACCAGGGACCGCAGGTACGAGGTGG
 CTCGACTGCTCAACCTCACCGAGAGGCAGGTCAAGATCTGGTTCCAGAACCGCAGGATGA
 AAATGAAGAAAATCAACAAAGACCGAGCAAAAGACGAGTGATGCCATTTGGGCTTATTTA
 GAAAAAGGGTAAGCTAGAGAGAAAAAGAAAGAACTGTCCGTCCCCCTTCCGCCTTCTCC
 CTTTTCTACCCCCACCCTAGCCTCCACCATCCCCGCACAAAGCGGCTCTAAACCTCAGG
 CCACATCTTTTCCAAGGCAAAACCTGTTTCAGGCTGGCTCGTAGGCCTGCCGCTTTGATGG
 AGGAGGTATTGTAAGCTTTCATTTTCTATAAGAAAAAGGAAAAGTTGAGGGGGGGGCAT
 TAGTGCTGATAGCTGTGTGTGTAGCTTGTATATATATTTTTTAAAAATCTACCTGTTCCT
 GACTTAAACAAAAGGAAAGAACTACCTTTTTATAATGCACAACTGTTGATGGTAGGCT
 GTATAGTTTTTAGTCTGTGTAGTTAATTTAATTTGCAGTTTGTGCGGCAGATTGCTCTGC
 CAAGATACTTGAACACTGTGTTTTATTGTGGTAATTATGTTTTGTGATTCAAACCTTCTGT
 GTACTGGGTGATGCACCCATTTGTGATTGTGGAAGATAGAATTCAATTTGAACCTCAGGTTG
 TTTATGAGGGGAAAAAACAGTTGCATAGAGTATAGCTCTGTAGTGGAATATGTCTTCTG
 TATAACTAGGCTGTAAACCTATGATTGTAAAGTAGCTGTAAGAATTTCCCAGTGAAATAA
 AAAAAAATTTTAAGTGTTCTCGGGGATGCATAGATTTCATCATTTTCTCCACCTTAAAAAT
 GCGGGCATTTAAGTCTGTCCATTATCTATATAGTCCTGTCTTGTCTATTGTATATATAAT

CTATATGATTAAAGAAAAATATGCATAATCAGACAAGCTTGAATATTGTTTTTGCACCAGA
 CGAACAGTGAGGAAATTCGGAGCTATACATATGTGCAGAAGGTTACTACCTAGGGTTTAT
 GCTTAATTTTAAATCGGAGGAAATGAATGCTGATTGTAACGGAGTTAATTTTATTGATAAT
 AAATTATACACTATGAAACCGCCATTGGGCTACTGTAGATTTGTATCCTTGATGAATCTG
 GGGTTTCCATCAGACTGAACTTACACTGTATATTTTGCAATAGTTACCTCAAGGCCTACT
 GACCAAATTGTTGTGTTGAGATGATATTTAACTTTTTGCCAAATAAAATATATTGATTCT
 TTTCTAAAAA

>Hs.126852_contig1

AI802118|BF197404|BF224434|AA931964|AW236083|AI253119|AW614335|AI671372|

5 AI793240|AW006851|AI953604|AI640505|AI633982|AW195809|AI493069|AW058576| AW293622 poliA = 2
poliA = 3

AAACCAGTGTATCCAGTCATGGAAAAGAAGGAGGAAGATGGCACCCCTGGAGCGGGGGCAC
 TGGAAACAACAAGATGGAGTTTGTGCTGTCTAGTGGCTGGGGAGATCATTGGCTTAGGCAAC
 GTCTGGAGGTTTCCCTATCTCTGCTACAAAAATGGGGGAGGTGAGATGAGAGCCCTTGTG
 CCACCCACCCACTCCTGGAAGGAGGATACTTCCATCTCCTGCACTTACGGCCCCCTCTGG
 GGAGTCCCATAGATGTATAGAATTCTGGAGGTAGGAGGACGCTTGGAGGTCATTAAGGAC
 ACTCTGTAAGAGACTAAGACCTAGAAAGGTTACGTGACTATCCCAGGGCTCTTCTATTA
 TAACGTGGCATCGTAGAAATATGAGCACAAAGCTGGAACCAGGTGGATGAGAGTTTGGATT
 CTGGCTCTGCTACTTAACACTCTGTGTGATCTTGGACAAGTTACTTAAGCTCTCAGAGCA
 TCAATTGCCGCTCCTGCAAATTGAGATAATAATGCCCTGCCCTTCAAGGTCATTGTAAGGA
 TTAGAGACAATGTGTGTAAAGCACTTAATAAATAGTAGCTCTGCTGATGATGACGTTGAT
 AACCAAACCTGTTCTGTGGTCTTAAGTAATAAATAGTAGCTCTGCTGATGATGACGTTGAT
 AACCAAACCTGTTCTGTGGTCTTAAGTAATAAGTAGTAGCTCTGTTGATGATGACGTTGAT
 AACCAAACCTGTTCTGTGGTCTTAAGTAATAAGTAGTAGCTCTGCTGATGATGACGTTGAT
 AACCAAACCTGTTCTGTGGTCTTAAGTAATAAATAGTAGCTCTGCTGATGATGACGTTGAT
 AACCAAACCTGTTCTGTGGTCTTAAGTAATAAATAGTAGCTCTGCTGATGATGACGTTGAT
 AACCAAACCTGTTCTGTGGTCTTAAGTAATAAATAGTAGCTCTGCTGATGATGACGTTGAT
 AACCAAACCTGTTCTGTGGTCTTAAGTAATAAATAGTAGCTCTGCTGATGATGACGTTGAT
 AACCAAACCTGTTCTGTGGTCTTAAGTAATAAATAGTAGCTCTGCTGATGATGACGTTGAT
 AAAAAAAAAAAAAAAAAAAAAA

>Hs.28149_ARNm_1 gi|14714936|gb|BC010626.1|BC010626 clon de Homo sapiens MGC:17687
 IMAGE:3865868 polIA = 3

GGAAGACATCAGGATGTACCATCTGCCCTTCTGTCTGGACCCAGGGTACGTCCCATGAGC
 GCGGCCGAGCTGCGTTCGAGGGCAGCAGAGCGTGTCTGCACTGCTCAGGGACCCGGAATCTG
 CAGTTTCTCCTGCACTGTTTTACCTTTGGCCAGACGGGCTCTGGGAAGACCTACACCCCT
 GACTGGACCCCTCCCCAGGGGGAGGGGGTGCCTGTACCCCCAGCCTGGCTGGCATCAT
 GCAGAGGACCTTCGCTGGCTGTTGGACCGCGTGCAGCACCTGGGTGCCCCCTGTACCCCT
 TCGCGCCTCTTATCTGGAGATCTACAATGAGCAGGTTTCGGGACTTGCTGAGCCTGGGGTC
 TCCCCGGCCCCCTCCCTGTTCTGCTGGAACAAGACTCGGGGCTTCTATGTGGAGCAGCTGCG
 GGTGGTGGAAATTTGGGAGTCTGGAGGCCCTGATGGAACCTTTGCAAACGGGTCTCAGCCG
 TCGAAGGAACCTCAGCCACACCCCTGAACCAGGCCTCCAGCCGAAGCCATGCCCTGCTCAC
 CCTTTACATCAGCCGTCAAACCTGCCAGCAGATGCCTTCTGTGGACCCCTGGGGAGCCCCC
 TGTGGTGGGAAGCTGTGCTTTGTGGACCTGGCAGGCAGTGAGAAGGTAGCAGCCACGGG
 ATCCCGTGGGGAGCTGATGCTTGAGGCTAACAGCATCAACCGAAGCCTGCTGGCCCTGGG
 TCACTGCATCTCCCTGCTGCTGGACCCACAGCGGAAGCAGAGCCACATCCCTTTCCGGGA
 CAGCAAGCTCACCAAGTTGCTGGCAGACTCACTGGGAGGGCGCGGGGTACCCCTCATGGT
 GGCCTGCGTGTCCCCCTCAGCCAGTGCCTTCTGAGACTCTCAGCACCCCTGCGATATGC
 AAGCCGAGCTCAGCGGGTCAACACCCGACCACAGGCCCCCAAGTCTCCTGTGGCAAAGCA
 GCCCCAGCGTTTGGAGACAGAGATGCTGCAGCTCCAGGAGGAGAACCGTCTCGCTGCAGTT
 CCAGCTGGACCAAATGGACTGCAAGGCCTCAGGGCTCAGTGGAGCCCGGGTGGCCTGGGC
 CCAGCGGAACCTGTACGGGATGCTACAGGAGTTTCATGCTAGAGAATGAGAGGCTCAGGAA
 AGAAAAGAGCCAGCTGCAGAATAGCCGAGACCTGGCCCAAGATGAGCAGCGCATCCTGGC
 CCAGCAGGTCCATGCACTAGAGAGGCGTCTCCTCTCTGCTGCTACCATCACCAGCAGGG
 TCCTGGCCTGACCCACCGTGTCCCTGCTTGATGGCCCCAGCTCCCCCTTGCCATGCACT
 GCCACCCCTCTACTCCTGCCCTGCTGCCACATCTGCCCACTGTGTGAGTGCCCTGGC
 CCACTGGGCCTGCCTGCCAGGGGAGCACCACTGCCCAAGGTGTGGACCCCTGAGGCCTC
 AGGTGGCAGGCCCCATCTGCCCGGCCCCACCCCTGGGCACCCCATGCAGCCCTGGCTC
 TGCCAAGTGCCCAAGAGAGAGGAGTCAAGTGACTGGACTCAGACCCGAGTCTCTGGCAGA
 GATGTTGACGGAGGAGGAGGTGGTACCTTCTGCACCTCCCTGCTGTGAGGCCCCCGAA
 GACATCACAGGGCTCAGAGGTGGGGCCGGGGTTCCAAACCTGGCCAGAGACTGGAGGC
 CCTCAGAGACCAGATTGSCAGCTCCCTGCGACGTGGCCGAGCCAGCCACCCCTGCAGTGA
 GGGCGCACGGAGCCAGGCCAAGTCTCCTCCCTCCCATTTGAAGGCCAAGTGGGAACCCAGG

AGACTGCTGTGTGACCTCAGACTGGGCTCCACACTCTTGGGCTTCAGTCTGCCCATCTGC
 TGAATGGAGACAGCAGCTGCTACTCCACCTGCAGCTGGGCTAGGGCGGGGACTGGGGGT
 GCTATTTAGGGGAACAAGGGGATTACAGGAGAAACCAGGCAGCAGGGGATGAAATACATGA
 ATAAAGAGAGGCATCAGCTCCAAAAAAAAAAAAAAAAAAAAA

>Hs.35453_ARNm_3 gi|7018494|emb|AL157475.1|HSM802461 ARNm de Homo sapiens; ADNc DKFZp761G151 (del clon DKFZp761G151); cds parcial poliA = 3

CTCCCCCTGAGAGAGGCTGGGCAGCACCCCCCTTCTGCCAGGAGTGCCAGCCAAGGTGCC
AGACCCCTGTCCAGTGGCAAGCTGGAAGGCTTTTCAGAGCATCGATGAAGCTATAGCCTGG
CTCAGGAAGGAACTGACGGAGATGCGGCTGCAGGACCAGCAACTGGCCAGACAGCTCATG
CGCCTGCGTGGCGACATCAACAAGCTGAAAATCGAACACACCTGCCGCCTCCACAGGAGG
ATGCTCAACGATGCCACCTACGAGCTGGAGGAGCGGGATGAGCTGGCCGACCTCTTCTGT
GACTCCCCCTCTGCCTCCTCCTTCAGCCTCTCCACACCACTCAAGCTTATTGGCGTGACC
AAGATGAACATCAACTCTCGGAGGTTCTCTCTCTGCTGAGGAGCCCTCAGACTGGGCGGA
GGGGCTGGAGCGGAGGGCTTGGGCTGGAGGGGTGTGAGAGGAAGCTGAGGCCAAGTTACT
CCAGTGGGTCTCCCGGAGGCAGGGGTCCCTGGGACTGGGACTCAAGGGCCCCAGGACCT
ATTCAGTGGTGCTCTCCACCCAGGGGCCCTGGGTGTGGATGCCAGTGTCTCTGTGACTG
GCTCTTGCTTACTACCCAAAGAGCTCTGCAGAAGGGCCGCTCCAACCAAGATGTTAAAGG
AGACCTGGGTTCACCATTAATCCATCCCTCCACGGTCACGTTCTGTTCCTGGAATCA
CTGGTGCTATGAACTGGGATTCCCAAAGGGAGGCCCCCAACAAAGCTGTCAATTTTTGCA
GAAGGCTGTCCCGCAAGGGCCTTGGGGGAAATTAGGCATGTCAGATGTGCCTGTCTCAG
TGCTGTTGCTGTCTCTAAGTATTGTCTCAAATTCACCTAAGTACATGACTCAGCAACA
TTGACAGGGAGCTACTAGGAAGGGAAAATCGAAAGGCATGACAAATGGGCACCTGGGGAC
GCAGCCCCAGTGGCTGGCAGCCAGTGTCTCTGGTGAGCCTGACACTACAAGGCTGTGTAA
ATTGTAAATTCTGGCGTGTGCTGGGACATGTGATGGGGGCACTAGCGTAGCTTGGGTGCA
ACAAGCACAGATGTCCCCATTGTCTCCCTGGCCACATGCATCTCCAAAGAGCCTCTTCA
CTGCCACCCACACCCAGGGTGACAGCCTGGGAGACCCTGGTGACTGAACCAGGCAGGT
CCTGAAAGCATTTCCTATACTGAATTCTCTGCAGGGGCGTGACCGGGGCCTCTGGTG
GATTCTGGTGGTGTACCTTACTGCCCTCTCTGGAAGACAATCTAGGGAGCCCAGAGGC
CCATCCTGAGCCTCCTCTGAGATTTTGTGCTGACCTAAACAACCTAGTTTTTAAGACT
GTTACTGATGTGTGTTCACTTGTGTAGTAAGTATGATTTTGTCCAAATGCGGAAGCCACTT
GTGTAGGTCAACTACAGTGCCTAGGATTTTGAATTTTAAGAGTTTCTCCCTCCCAACAGGCT
TGAGGATCAGCAAGTTAAGACCCAGCAGGTTAGGGAGGTCAGTCTGGGGTCATACGGCA
TGGCAGGGGTCCCTCGGCCAGACCCGTAGAATCCTGAGATAAGGAGTGTTCCTGACCTTT
GGTGTCTATCTAGTCGAGTCTCTCATTAGTAAAGGAGCAAAGTGAAACCTGGGGGAGGAG
AAGGACTTCCCTCAGGTTGCACAGCTGTTTAGGCTATAGAATATTGATGTGTGAAACCAT
TATTGATAATGCCTAGTAGATCACATGTCAATGAACTTGAACCCCAAAGATGGTCTGTGAT
GCTTTTGCCAAACCCGCACACTGCCAACCCCTCTACTCTCCACCTCAGCCCCACCCACAT
CTCCCAGAGTATTGCAATTCAGAACATTTGGGTCAAGGTGGAGCAAGGCACTGACAGTGG
CCCCACAGGGCATGTGTCACTAATCACTGTCCCATGGTCTACGCACGGCATCTGGCTGCT
CTGTCTACTGTGACTTCTTCTGTGTAATCTCAGTGGGGGCCGCTGTCCACCCACACATCG
TGACCCACATAGGGGAGAGGTTGCTTTTCTTTTGTGGGCTGAGAGTAGGACAATGCAAT
GAATGATCTCTAGTAGACAGAAAAGAACTTGGTCTCTTTTTTAAATTTCAAAGAGCCAG
AAGTTCTATGCCTCCTTCAAAGTAGGCAGAACACGCAGCCAAGATCTACTGTCTGCCAT
GCTCTGTGCAATGAAGTCTGCAGGCCTGAGGACCATGTACTGCTGTCTCTCCTCAGAGCT
CTGCACAAACACTGCCAAGTCTGAAGACGCATTCTTTTCTGCCAACCTCTTTCCAGAT
AAGCCCTTGAGGTCTCGGGCTGACCTACACACACACACACACACACACACACACACAC
ACACCCACACACACACACACACAGCAGAGAGAACATGCCATAAACATCCTTGAACCCATG
CAGGAAAGCCCATCCCATATTCTGAAAAAATGCCAAATTAGGTTTTTCTTTTCTTTTGGGA
AATCAGTCATTACAGTAACCGAAACCATTTGGGTTTCAGCGAAAATGGAAAGATTTAGCTGA
ATGTAGTCAGTCCAATTAAGTTGGATGCAACTGAGTGATTTAGTTGCTTGGGTAACCCAG
TGCTTGCTTGCTTTCTTCACTCTCTGGGTGGAACTAAGATCAAGACACATGTTTGGGGA
TAAGTTAAATGTCTGAGCTATTTTGTCTCGGTTTATCCTAAGAGAACTTTATTATGGGATG
AGGAGGTGACCCAAGATGAGAAGTGGAGGGGGACAGCGATGTTTTCTAAACATCGTCCAG
TGTTGACTGGCTTCTTACTTTGCACAGTGAACACAACCTAACCACATTAATTCAGCTTTG
TGAAGTCCCTGCTCTCTGTGGGTTCTATGAGTCAGCAGCAACATTGGCCTAACCTCCGTC
CCAGCCTCCTGGCTCACCACATGTGTACAGTGTCTGTTTGCAGTTGTACTCATTATCCATC
CATCTCTCTGCCATCCCCAAGCATCGCTGGGTGTAAACGCAAACTCTCCACCGACACTG
CCATGCGTGGTTCATGTCTTGATGCCTTCAGGGGCTCAGTAGCTATCAAAGAGGCCTGGAG
GGCCTGGGCAGGCTTGACGATGCCTGACCGAGTTCAAGACCCACACCTGTAGCAATACC
AAGTGCTATTACATAATCAATGGACGATTTATACTTTTATTTTATGATTATTTGTTTC
TATATTGCTGTTAGAAAAAGTGAAATAAAAATACTTCAAAGAAAAA

AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAA

>Hs.180570_contig1 R08175|AA707224|AA699986|R11209|W89099|T98002|AA494546 poliA = 2 poliA = 3

TGAAGGACCGCGATCCTAAAGAGATTGAATGGGACGACCTGGCCCAGCTGCCCTTCCTGA
CCATGTGCGTGAAGGAGAGCCTGAGGTTACATCCCCAGCTCCCTTCATCTCCCGATGCT
GCACCCAGGACATTGTTCTCCAGATGGCCGAGTCATCCCCAAGGGCATTACCTGCCTCA
TCGATATTATAGGGGTCCATCACAACCCAACTGTGTGGCCGGATCCTGAGTCTACGACCC
CTTCCGCTTTGACCCAGAGAACAGCAAGGGGAGGTCACCTCTGGCTTTTAATTCCCTTCT
CCGCAGGGCCCAGGAACTGCATCGGGCCAGCGTTTCCCATGGCGGAGATGAAAGTGGTTC
CTGGCGTTGATGCTGCTGCACTTCCGGTTCCTGCCAGACCACACTGAGCCCCGCAGGAAG
CTGGAAGTGAATGATGCGGCCGAGGGCGGGCTTTGGCTGCGGGTGGAGCCCCTGAATGTA
GGCTTGCAGTGACTTTCTGACCCATCCACCTGTTTTTTTGCAGATTGTCATGAATAAAAC
GGTGCTGTCACCTCAAAAAAAAAAAAAANNNAAAA

5

>Hs.196270_ARNm_1 gi|11545416|gb|AF283645.1|AF283645 cromosoma 8 map 8q21 de Homo sapiens poliA = 3

GAGTCCTCTCGTTGGTCCCAGGAGGTGGGGTTGCGCTCACAAGGGGCGACCGTCGCCACGG
 TGGCGGCCACTGCATCGCGTCCCACCTCCGCGGCCCTGGGCGCCGTGGTGTGACGGGGCC
 CCGAGCCTATGACGGGCCAGGGCCAGTCGGCGTCCGGGTGCTCGGCGTGGAGCACGGTAT
 TCCGCCACGTCCGGTATGAGAACCCTGATAGCGGGCGTGAGCGGCGGCGTCTTATCCAACC
 TTGCGCTGCATCCGCTCGACCTCGTGAAGATCCGCTTCGCCGTGAGTGATGGATTGGAAC
 TGAGACCGAAATATAATGGAATTTTACATTGCTTGACTACCATTTGGAAACTTGATGGAC
 TACGGGGACTTTATCAAGGAGTAACCCCAAATATATGGGGTGACAGGTTTATCCTGGGGAC
 TCTACTTTTCTTTTACAATGCCATCAAGTCATATAAAACAGAAGGAAGAGCTGAACATT
 TAGAGGCAACAGAATACCTTGTCTCAGCTGCTGAAGCTGGAGCCATGACCCTCTGCATTA
 CAAACCCATTATGGGTAACAAAACCTCGCCTTATGTTACAGTATGATGCTGTTGTTAACT
 CCCCACACCGACAATATAAAGGAATGTTTGATACACTGTGAAAATATATAAGTATGAAG
 GTGTGCGTGGATTATATAAGGGATTTGTTCTGGGCTGTTTGGAAACATCGCATGGTGCCC
 TTCAGTTTATGGCATATGAATTGCTGAAGTTGAAGTACAACCAGCATATCAATAGATTAC
 CAGAAGCCAGTTGAGCACAGTAGAATATATATCTGTTGACGCACTATCCAAAATATTTG
 CTGTCGACGCAACATACCCATATCAAGTCGTAAGAGCTCGTCTTCAGGATCAACACATGT
 TTTACAGTGGTGTAATAGATGTAATCACAAGACATGGAGGAAAGAAGGCGTCGGTGGAT
 TTTACAAGGGAATTGCTCCTAATTTGATTAGAGTGACTCCAGCCTGCTGTATTACCTTTG
 TGGTATATGAAAACGCTCTCACATTTTTTACTTGACCTTAGAGAAAAGAGAAAGTAAGCTC
 AAAGAGGACAATTCCAGTATATCTGCCCAAGGCAGCAACAAGCTCTTTTGTGTTTAAAGC
 ATAAAAGAAGAAATCTGCATAGAAACATGGCTCATATTCGAAATTGCTCTATAGTCATTA
 GAAGCCAGAGAACTGCTAAGTCTCCTGCAATGTTTTTCTTGCTTTTTTGCCCTCCCCATAT
 ATATGGAACCTTGGCTACCTCTGCCTGAAATGGCTGCCATCAACACAATGTTAAAACCTGAC
 ACGAAGGATAGAGTTTACAGATTTCTACGTTTTATTGGTGGAAGCTGATTTGCAACATT
 TGCTAAATGGATTAGATGAATGTAATCTTTTTTGTGAGCTTACTTGCCCTGGATTGCTTTA
 AAATTAACCTTTGTGCAATACCAAGAAAATAGCTCTTTAAAAGAATGTCTTTGTATGTCT
 CAAGGTAAATTAAGGATTTACTGAATAAGGTGTTGACCAATCCAGACCATTTTATTTTA
 TTTTTTTATTTATTTATTTTTTGAGATGGAGTCTTGCTTTGTGCGCCAGGCTGGAGTGCA
 GTGGCGTGATCTCAGCTCACTGCAACCTCCACCTCCCGGTTTACGCCATTCTCCTGCCT
 CAGCCTCCTGAGTAGCTGGGACTACAGGCACCTGCCACCACGCTGGCTAACTTTTTTTTT
 ATATTTTGAGTAGAAATGGGGTTTACCATGTTAGCCAGGATGGTCTCAATCTCCTGACC
 TTGTGATCCGCTGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACTGCGC
 CTGGCCAGACCATTTTAGAATTGGGAAATTTTAGTGAGAAAAATGCACGTGTAATATGC
 TTTAGTTTTAATTCAGTTGGGATGCACTACCTAGCGAAAAATTGAGAACTATATACTTCT
 CAGAGAAATATCTGACATCTATTGTCATTCCATTGCTATTTTTTTTCCCCAGAGACTTCC
 ATAATTTAAAATAAAATCCTAGATCCAGTTCTTGTTTTTTGGCATAAATACTTAATCTAT
 TTTAAATTTATAAAATCTGAGCTTCTAGGATCCAGCTGTGTCAACCTTTATTTAGCATAT
 ATAATATAAAATCACTTATTACAGATGCTAAATAGATCACCTTTTACAGATGCTGAAATG
 TTTGGGATATGTTTGTGACAAGGTAAATGGAAATGAGAACTTTATACTTCAGTTTTC
 GATATATGGATCTAGATCCCAAATAAATGATTAATCTTCATGTTTCTCAAATTCAGGT
 TGAAATACAAATTAATAGCCTTTATTGATTTTACTTTTATGAGTCATTGTAGACATCTAT
 AAATATAAAAAGGGCTGTACCCAAAGGATGCCAGAATACTAGTATTTTTTATTTATCGTAA
 ACATCCACGAGTGCTGTTGCACTACCATCTATTTGTTGTAAATAAAAGTGTTGTTTTCAA
 AAAAAAAAAAAAAA

>Hs.9030_ARNm_3 gj|12652600|gb|BC00045.1|BC00045 clon de Homo sapiens MGC:2032 IMAGE:3504527
 poliA = 3

CTAGAGGGGCGGAAAGTAACAAGGAGGTGGGGGTACAAATCCTCAGCTCCTGCTCCGCA

AGCACTAACCTGCTCTGAAGTGAGCCAGGCAGCTCTGGCCATCTTTTCCCAGCCACAGAA
 TCAGGTGATGGTCCAGAATTAAGAGCTGTCACCTGTGTCAATCACTCACAATGGAAGAAA
 TGAAGAAGACTGCCATCCGGCTGCCCCAAGGCAACAGAAAGCCTATAAAGACGGAATGGA
 ATTCCCGGTGTGTCTTTTTCACCTACTTCCAAGGGGACATCAGCAGCGTAGTGGATGAAC
 ACTTCTCCAGAGCTCTGAGCAATATCAAGAGCCCCCAGGAATTGACCCCCCTCGAGTCAGA
 GTGAAGGTGTGATGCTGAAAAACGATGATAGCATGTCTCCAAATCAGTGGCGTTACTCGT
 CTCCATGGACAAAGCCACAACCAGAAGTACCTGTACAAACCGTGCCGCCAACTGCAACT
 TGCATGTGCCTGGTCCCATGGCTGTGAATCAGTTCTCACCGTCCCTGGCTAGGAGGGCCT
 CTGTTCCGGCCTGGGGAGCTGTGGCATTCTCCTCCCTGGCGGGCACCAGCTCCTTAGAGC
 CTGGCTACTCTCATCCCTTCCCGCTCGGCACCTGGTTCAGAGCCCCAGCCTGATGGGA
 AACGTGAGCCTCTCCTAAGTCTCCTCCAGCAAGACAGATGCCTAGCCCGTCCCTCAGGAAT
 CTGCCGCCAGGGAGAATGGCAACCCTGGCCAGATAGCTGGAAGCACAGGGTTGCTCTTCA
 ACCTGCCCTCCCGGCTCAGTTCACTATAAGAACTATATGTATCTCGTGGATCTGCCAGTA
 CCAGCCTTCCAAATGAAACTCTTTCAGAGTTAGAGACACCTGGGAAATACTCACTTACAC
 CACCAAACCACTGGGGCCACCCACATCGATACCTGCAGCATCTTTAGTCAAGTTGGAGGA
 GAAAGACAACACTTGGTCTAAGACACGGCAGCAAGACATCCCTGCATATTGTTCCAGATA
 AAAATGAAAGCTGCTCACACCCACTTGCCTCCCCAATCTGTTAAACAGCTTCGTGTCTAG
 TATGAGCTCAGTACTTGCCCTGTGAAAATCCCAGAAGCCCCCGCTGTCAATGTTCCCAT
 CCACACCCTGCTTGCTCCTGTGTAAACAGCTCAGATGATGAATAATAATAAACTGTACTT
 TTTTGGATGGTGAAAAAAAAAAAAAAAAAAAAA

>Hs.1282_ARNm_3 gj|4559405|ref|NM_000065.1| componente 6 de complemento de Homo sapiens (C6),
 ARNm poliA = 1

TTGCCTTGTGTTAGCTAGCAATAAGAAAAAGAGCTTTGTTTGGATTAAACATATATACCCT
 CTTCAATTCTGCATACCTATTTTTTCCCAATAATTGTCAGCTTAGGTCCGAGGACACCAC
 AAACCTCTGCTTAAAGGGCCTGGAGGCTCTCAAGGCATGGCCAGACGCTCTGTCTTGTACT
 TCATCCTGCTGAATGCTCTGATCAACAAGGGCCAAGCCTGCTTCTGTGATCACTATGTCAT
 GGACTCAGTGGACCAGCTGCTCAAAAACTTGCAATTCTGGAACCCAGAGCAGACACAGAC
 AAATAGTAGTAGATAAGTACTACCAGGAAAACTTTTGTGAACAGATTTGCAGCAAGCAGG
 AGACTAGAGAATGTAACCTGGCAAAGATGCCCCATCAACTGCCCTCCTGGGAGATTTTGGAC
 CATGGTCAGACTGTGACCCTTGTATTGAAAAACAGTCTAAAGTTAGATCTGTCTTGCCTC
 CCAGTCAGTTTGGGGGACAGCCATGCACTGAGCCTCTGGTAGCCTTTCAACCATGCATTTC
 CATCTAAGCTCTGCAAAATTGAAGAGGCTGACTGCAAGAATAAATTTGCTGTGACAGTG
 GCCGCTGCATTGCCAGAAAGTTAGAATGCAATGGAGAAAATGACTGTGGAGACAATTCAG
 ATGAAAGGGACTGTGGGAGGACAAAGGCAGTATGCACACGGAAGTATAATCCCATCCCTA
 GTGTACAGTTGATGGGCAATGGGTTTCATTTTCTGGCAGGAGAGCCCAGAGGAGAAGTCC
 TTGATAACTCTTTCACTGGAGGAATATGTAAACTGTCAAAGCAGTAGGACAAGTAATC
 CATAACCGTGTTCGGCCAATCTGGAATAATGTCGGCTTTGAGGTACAACTGCAGAAAGATG
 ACTTGAAAACAGATTTCTACAAGGATTTAACTTCTCTTGGACACAATGAAAATCAACAAG
 GCTCATTCTCAAGTCAGGGGGGGAGCTCTTTCAGTGTACCAATTTTTTATTCTCCTCAAAGA
 GAAGTGAAAATATCAACCATAATTTCTGCCTTCAAACAAGCCATTCAAGCCTCTCACAAAA
 AGGATTTCTAGTTTTATTAGGATCCATAAAGTGATGAAAGTCTTAAACTTCACAACGAAAG
 CTAAGATCTGCACCTTTCTGATGTCTTTTGAAGCACTTAACCATCTGCCTCTAGAAT
 ACAACTCTGCTTTGTACAGCCGAATATTCGATGACTTTGGGACTCATTACTTCACCTCTG
 GCTCCCTGGGAGGCGTGTATGACCTTCTCTATCAGTTTAGCAGTGAGGAAGTAAAGAACT
 CAGGTTTAAACGAGGAAGAAGCCAAACACTGTGTGAGGATTGAAACAAAGAAACGCGTTT
 TATTTGCTAAGAAAACAAAAGTGGAAACATAGGTGCACCACCAACAAGCTGTCAGAGAAAC
 ATGAAGGTTTCATTTATACAGGGAGCAGAGAAATCCATATCCCTGATTGAGGTGGAAGGA
 GTGAATATGGAGCAGCTTTGGCATGGGAGAAAGGGAGCTCTGGTCTGGAGGAGAAGACAT
 TTTCTGAGTGGTTAGAATCAGTGAAAGGAAAAATCCTGCTGTGATTGACTTTGAGCTTGCCC
 CCATCGTGGACTTGGTAAGAAACATCCCCTGTGTCAGTGACAAAACGGAACAACCTCAGGA
 AAGCTTTGCAAGAGTATGCAGCCAAGTTCGATCCTTGCCAGTGTGCTCCATGCCCTAATA
 ATGGCCGACCCACCCTCTCAGGGACTGAATGTCTGTGTGTGTGTCAGAGTGGCACCTATG
 GTGAGAACTGTGAGAAACAGTCTCCAGATTATAAATCCAATGCAGTAGACGGACAGTGGG
 GTTGTGGTCTTCTGAGTACCTGTGATGCTACTTATAAGAGATCGAGAACCCGAGAAT
 GCAATAATCCTGCCCCCAACGAGGAGGGAAACGCTGTGAGGGGGAGAAGCCGACAAGAGG
 AAGACTGCACATTTTCAATCATGGAACAATGGACAACCATGTATCAATGATGATGAAG
 AAATGAAAGAGGTGATCTTCTGAGATAGAAGCAGATTCGGGGTGTCTCAGCCAGTTC
 CTCCAGAAAATGGATTTATCCGGAATGAAAAGCAACTATACTTGGTTGGAGAAGATGTTG
 AAATTTTCATGCCTTACTGGCTTTGAACTGTTGGATACCAGTACTTCAGATGCTTACCAG
 ACGGGACCTGGAGACAAGGGGATGTGGAATGCCAACGGACGGAGTGCATCAAGCCAGTTG
 TGCAGGAAGTCTGACAATTACACCATTTTCAGAGATTGTATAGAATTGGTGAATCCATTG

AGCTAACTTGCCCCAAAGGCTTTGTTGTTGCTGGGCCATCAAGGTACACATGCCAGGGGA
 ATTCTTGGACACCACCCATTTCAAACCTCTCTCACCTGTGAAAAAGATACTCTAACAAAT
 TAAAAGGCCATTGTCTAGCTGGGACAGAAACAATCAGGATCTGAATGCATTTGTATGTCTC
 CAGAAGAAGACTGTAGCCATCATTCAGAAGATCTCTGTGTGTTTGACACAGACTCCAACG
 ATTACTTTACTTCACCCGCTTGTAAGTTTTTGGCTGAGAAATGTTTAAATAATCAGCAAC
 TCCATTTTCTACATATTGGTTCCTGCCAAGACGGCCGCCAGTTAGAATGGGGTCTTGAAA
 GGACAAGACTTTTCATCCAACAGCACAAAGAAAGAATCCTGTGGCTATGACACCTGCTATG
 ACTGGGAAAAATGTTTCAGCCTCCACTTCCAAATGTGTCTGCCTATTGCCCCACAGTGCT
 TCAAGGGTGGAAACCAACTCTACTGTGTCAAAATGGGATCATCAACAAGTGAGAAAACAT
 TGAACATCTGTGAAGTGGGAACATAAGATGTGCAAACAGGAAGATGGAAATACTGCATC
 CTGGAAAGTGTGTTGGCCTAGCACAAATTACTGTCTAGGCCCAGCACAAATGAACAGATTTACC
 ATCCCGAAGAACCAACTCTACAAATGAGAATCTTGCACAAACAGCAGACTGGCATGCT
 CAAAGTTACTGACAAAAATTATTTCTGTGTAGTTTGAGATCATTATTCTCCCTGACTCT
 CCTGTTTGGGCATGTCTTATTCAGTTCAGCTCATGACGCCCTGTAGCATACCCCTAGGT
 ACCAACTTCCACAGCAGTCTCGTAAATCTCCTGTTTACATTGTACAAAAATAATGTGAC
 TTCTGAGGCCCTTATGTAGCCTGTGACATTAAGCATTCTCACAATTAGAAATAAGATAA
 AACCCATAATTTTCTTCAATGAGTTAATAAACAGAAATCTCCAGAACCTCTGAAACACAT
 TCTTGAAGCCAGCTTTCATATCTTCATTCACAAATAATTTCTGAGTGTGTATACAGGA
 TGTCAAGTACTGACCAAAGTCTTGAGAACTCGGCAGATAATAAACAGACAAAAGCCTTT
 GCCTTCATGAAGCATACATTCATTCAGGGGTAGACACACAAAAAATGAAATAAACAGGTA
 AAATATGTAGC

>Hs.268562_ARNm_2 gij|15341874|gb|BC013117.1|BC013117 clon de Homo sapiens MGC:8711
 IMAGE:3882749 poliA = 3

5

CTCTCCTCGCCCGCTGGGTGCTGAAGTTGGGCGGATGGCAGCAAACCGGCTCCGCTAGAG
 GACCGAGCCGCCCAGCCCCGCTCCCCGGACCCATCGGCGCGCTGCCACACCTCCAGGC
 GACCGGCCAACTGGGTCTTGAAGTAGCTGAAATGCGAAAAAGGCAGCAGTCCCAAATGA
 AGGAACACCTGCCGTGTCTCAAGCTCCTGGAAACCAGAGGCCCAACAACACCTGTTGCTT
 TTGTGTGTTGCTGTTGTTGTCAGCTGCTCCTGCCTCACTGTGAGGAATGAAGAAAGAGGGGA
 AAATGCGGGAAGACCCACACACACTACAAAAATGGAGAGTATCCAGGTCTTAGAGGAATG
 CCAAAACCCCACTGCAGAGGAAGTCTTGTCTGGTCTCAAAATTTTGACAAGATGATGAA
 GGCCCCAGCAGGAAGAAACCTTTTCAGAGAGTTCTCCGAACAGAATACAGTGAAGAGAA
 CCTACTTTTCTGGCTTGCTTGTGAAGACTTAAAGAAGGAGCAGAACAAAAAAGTAATTGA
 AGAAAAGGCTAGGATGATATATGAAGATTACATTTCTATACTATCACCAAAAGAGTTCAG
 TCTTGATCTCGAGTTAGAGAGGTGATCAATAGAAATCTGTTGGATCCCAATCCTCACAT
 GTATGAAGATGCCCAACTTCAGATATATACTTTAATGCACAGAGATTCTTTTCCAAGGTT
 TTTGAACTCTCAAATTTATAAGTCATTTGTTGAAAGTACTGCTGGCTCTTCTTCTGAATC
 TTAATGTTTCATTTAAAAACAATCATTTTGGAGGGCTGAGATGGGAAATAAAAGTAGTTAA
 ATAACATCAGAACTGAGTTCTGGAGAACTACAGTTTAGCATTCCTCAGGCTACTGTGA
 AAACACAACCGTTATGGTCTTTGTCTCCATTTTATCAAGGTTTCCATGGTTAAGTTTG
 GAGAAAATACCACACAAAACAATGAATTGCCAAATGTTTGTGTTTATTCAAGACTCATTCT
 TACTTGCAAGCAAAGTGTATTTGTAGTCTTATGAACAGTCTCCTCGTGTATCTCCAGAGA
 CTGCATGTGCAAAGTAAAATGCTTCATTTGCCACATAGTTGTTGTAATATTTAATCCAGT
 AGCATAACTTATATCTGTATTTAAGGACTTTTGTGCAATATGGTCTTAAGAAATAATTGC
 CAAAAAATCGGCCATGGTCTGCAATTTTAAACATAATCTAAGACAGAAAAAAGCAATT
 TTTACTATGTAACAAATGGTATTCAACATTCTATATACTGTGTTTAGTACACTAATTTTGA
 AGCCAATATTTCTGTACATGAAAAAGAGCTATTTATCTCTGTTTGTGGAATCCTAAT
 GGGGATTCCTCTGGTTGTTCACTGCCAAACTGTGGCATTTCATTACAGGAGAGTTTAC
 TATGCTAAAAGCAAAAAACAAAAAAGGGGAAGAGGAAAAAAGCAAAAAACA
 ATTTGAAGATATCCTATCTCAATGACAAATCAAAAGAGTGATATTGCTTTTAACTGTAAT
 AGAAGAAAATGAATTTATGTATATATCAGATGTCCAATACTGTAATTAATTTATTAAAGA
 CTGGCTCTCCAGTTTAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG

>Hs.151301_ARNm_3 gij|16041747|gb|BC015754.1|BC015754 clon de Homo sapiens MGC:23085
 IMAGE:4862492 poliA = 3

10

AAAAGAACCAGGATTGCATTTGAAGTTAAGCTGCAAAAAACCAGTCGATCAACAGATTTT
CGAGTCCCACAGTCAATATGCACCATGTTTAATGTTATGGTTGATGCCAAAGCTCAATCA
ACAAAACTTTGCAGCATGGAAATGGGCCAAGAGTTTGCTAAAATGTGGCATCAATACCAT
TCAAAAATAGACGAACTAATTGAAGAACTGTTAAAGAAATGATAACACTCTTGGTTGCA
AAGTTCGTTACTATCTTGGAAGGAGTGCTGGCAAAATTATCCAGATATGACGAAGGGACT
TTGTTTTCTCTTTTCTGTCATTTACCGTGAAGGCAGCTTCCAAATATGTGGATGTACCT
AAACCCGGGATGGACGTGGCCGACGCCTACGTGACTTTCGTCCGCCATTCTCAGGATGTC
CTGCGTGATAAGGTCAATGAGGAGATGTACATAGAAAGGTTATTTGATCAATGGTACAAC

AGCTCCATGAACGTGATCTGCACCTGGTTGACGGACCGGATGGACTTACAGCTTCATATT
TATCAGTTGAAAACTAATTAGGATGGTAAAGAAAACTACAGAGATTTCCGATTGCAA
GGGGTCTCTGGACTCCACCTTAAACAGCAAGACCTATGAAACGATCCGGAACCGTCTCACT
GTGGAGGAAGCCACAGCATCAGTGAGTGAAGGTGGGGGACTGCAGGGCATCAGCATGAAG
GACAGCGATGAGGAAGACGAAGAAGACGATTAGACCATTGTTGCTCTAGAGTCTGCTGGGA
CAGAGTCTGTAAATCAGTGCATGTCCTTAGTCTGTTAGTTAAACCCATTAGGAATTTTCT
GTCAACTACCATGCCCATGAGATGTTTATCAATACAAC TGCCATTTTAGCTATGTGGTAC
CAAGATTAGCAAATGACCTTCATATCCACTGATTTCTTGATGTCCATGTCTATATGTTTA
CAAGCAATATGGAGCACCATTCTTTAAATACTGTTTCATGGAGAATACATAGTCTAACCAC
TAGGCGTGTCCCTGTTATCAGCAAAGATCAATGATGCTTCATTCATGTACTATGTATGCA
TTGGTGGTAAATGGATGTGAGGGCAAGTACATCAAGTACATTCACTCTGTTTCACGTATG
TGGATGCCAGTTAATTAAATGAGTACGTAAATAAATTAATTAACACATAGATCTGCTT
TGTGTTTTTATTTTTATTTTTTGAAAAACAAAAGGCCAAGTCTCCAACAATTAACTTTTGA
TGCTTTCTGTTCCCTAAAACCAAAAAATGAACCCCTTGTTGTCGTTGTTAACCCATCCTT
TCATTTACTCATATAATTAGCCAAAAAAAAGGATGGCTACATACCAATGGATTGATTC
TCTTAATTGCCACGGCAAGGGGGCGATCCTATCATGACTTAACATCAAGCGCGCAGTTCA
AACTACTGTCTTCTGTCAAAGTTTTCTCCTCTTAAATGTTATTTTGCTTTTACGTCTCA
ACTGTGTATGTAAAAAAACGAATATTTAAATTACAACCTAGACTAAAAATGTGTTTAT
AATAAGATGTGGATATTTCTCTTCAGTAGATTGTAACCAATAATTTAAATTAATTTGTTC
CACTGTTTTTTTATATCTGTCTGTCATGTCATTGTCATTGTCATCTGTAAGTGCACAACCCCTGGG
GTTTGCTGCAGAGCTATTTCTTTCCATGTAAAGTAGTGGATCCATCTTGCTTTTGCCTTA
TATAAAGCCTACAGTTATGGAAGTGTGGAAAACTGTGGCTTCTCAATAAATATTTCAGATG
TCCTAAGAAAAA
AAAAAA

>Hs.111_contig1 AA946776|AW24238|H24274|AI078616 poliA = 1 poliA = 2

ACCTGAACTGTCTAAGATATTCTAAGCAAAGTTGACAAAGACAATTCTCCACTTGAGCCC
TAAAAAATGTAACTACTATAAAGGTTTCACGCGGTGGTTCTTATTGATTGCTGTGTCAT
CACATCAGCTCCACTGTTGCCAAACTTTGTGCGATGCATAATGTATGATGGAGGCTTGGA
TGGAATATGCTGATTTTGTCTGCACTTAAAGGCTTCTCCTCCTGGAGGGCTGCCTAGG
GCCACTTGCTTGATTTATCATGAGAGAAGAGGAGAGAGAGAGAGACTGAGCGCTAGGAGT
GTGTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATGTGTGTAGCGGGAGATGTGG
GCGGAGCGAGAGCAAAAGGACTGCGGCCTGATGCATGCTGGAAAAAGACACGCTTTTCAT
TTCTGATCAGTTGTACTTCATCCTATATCAGCACAGCTGCCATACTTCGACTTATCAGGA
TTCTGGCTGGTGGCCTGCGGAGGGTGCAGTCTTACTTAAAGACTTTTCAGTTAATTCTC
ACTGGTATCATCGCAGTGAACCTTAAAGCAAAGACCTCTTAGTAAAAAATAAAAAAATAA
A

5

>Hs.150753 contig1 AI123582|AI288234 poliA = 0 poliA = 0

GCTTCTCTTTAAAATTGACCCAAGGCATGAGCCACTGCGCCTGGCCAGCAAATGCTTTTT
GTGCAGAATACACTTCTTTCAGGCATTGTCAGGTGCTGTTTTGTTTAAGCTCTAACTCAC
CCCTGGAATACAGGGGAATGATGACAACCAGCCCAGCCAGGCCTGACTCATCATGGTCAC
ATCCAGCCCCCACCCTGGCCAACTAACCCTGCAGGCTCCTCTTCCAGACTCACCAGGG
GGCCTCGAGGCCCCGGCATCTCCCTTGGCCCTGGGTGTGGGTTTTACAAGACTGTGTCTT
TCATGACATCATAGCCAACCATGTGAGAAGAAGGAGAAGGCCCCCTTCTTCATTAAT
CTGAAAA

>Hs.82109_ARNm_1 gi|14250611|gb|BC008765.1|BC008765 clon de Homo sapiens MGC:1622 IMAGE:3347793
poliA = 3

GGCAGAGGAAGGGCCTGTGGGTTTATTATAAGGCGGAGCTCGGCGGGAGAGGTGCGGGC
CGAATCCGAGCCGAGCGGAGAGGAATCCGGCAGTAGAGAGCGGACTCCAGCCGGCGGACC
CTGCAGCCCTCGCTGGGACAGCGGCGCGCTGGGCAGGCGCCCAAGAGAGCATCGAGCAG
CGGAACCCGCGAAGCCGGCCCGCAGCCGCGACCCGCGCAGCCTGCCGCTCTCCCGCCGCC
GGTCCGGGCGAGCATGAGGCGCGCGCGCTCTGGCTCTGGCTGTGCGCGCTGGCGCTGAGC
CTGCAGCCGGCCCTGCCGCAAATTGTGGCTACTAATTTGCCCCCTGAAGATCAAGATGGC
TCTGGGGATGACTCTGACAACTTCTCCGGCTCAGGTGCAGGTGCTTTGCAAGATATCACC
TTGTACAGCAGACCCCTCCACTTGGAAAGGACACGCAGCTCCTGACGGCTATTTCCACG
TCTCCAGAACCCACCGGCTTGGAGGCTACAGCTGCCTCCACCTCCACCCTGCCGGCTGGA
GAGGGGCCCCAAGGAGGGAGAGGCTGTAGTCTGCCAGAAGTGGAGCCTGGCCTCACC GCC
CGGGAGCAGGAGGCCACCCCCGACCCAGGGAGACCACACAGCTCCCGACCACTCATCAG
GCCTCAACGACCACAGCCACCACGGCCCGAGGAGCCCGCCACCTCCCAACCCACAGGGAC
ATGCAGCCTGGCCACCATGAGACCTCAACCCCTGCAGGACCCAGCCAAGCTGACCTTCAC
ACTCCCCACACAGAGGATGGAGGTCCTTCTGCCACCGAGAGGGCTGCTGAGGATGGAGCC
TCCAGTCAGCTCCAGCAGCAGAGGGCTCTGGGGAGCAGGACTTCACCTTTGAAACCTCG

GGGGAGAATACGGCTGTAGTGGCCGTGGAGCCTGACCGCCGGAACCACTCCCCAGTGGAT
CAGGGGGCCACGGGGCCCTCACAGGGCCTCCTGGACAGGAAAGAGGTGCTGGGAGGGGTC
ATTGCCGTAGGCCTCGTGGGGCTCATCTTTGCTGTGTGCCTGGTGGGTTTCATGCTGTAC
CGCATGAAGAAGAAGGACGAAGGCAGCTACTCCTTGGAGGAGCCGAAACAAGCCAACGGC
GGGGCCTACCAGAAGCCCAACAGGAGGAATTCTATGCCTGACGCGGGAGCCATGCG
CCCCCTCCGCCCTGCCACTCACTAGGCCCCCACTTGCCCTCTTCCTTGAAGAAGTGCAGGC
CCTGGCCTCCCTGCCACCAGGCCACCTCCCCAGCAATCCAGCCCCCTCTGGTCTGCTCCTG
CCCACGGAGTCGTGGGGTGTGCTGGGAGCTCCACTCTGCTTCTCTGACTTCTGCCTGGAG
ACTTAGGGCACCAGGGTTTCTCGCATAGGACCTTTCCACCACAGCCAGCACCTGGCATC
GCACCAATCTGACTCGGTTTCTCCAACTGAAGCAGCCTCTCCCCAGGTCCAGCTCTGGA
GGGGAGGGGGATCCGACTGCTTTGGACCTAAATGGCCTCATGTGGCTGGAAGATCCTGCG
GGTGGGGCTTGGGGCTCACACACCTGTAGCACTTACTGGTAGGACCAAGCATCTTGGGGG
GGTGGCCGCTGAGTGGCAGGGGACAGGAGTCCACTTTGTTTCGTGGGGAGGTCTAATCTA
GATATCGACTTGTTTTTGACATGTTTTCTCTAGTTCTTTGTTTCATAGCCCAGTAGACCT
TGTTACTTCTGAGGTAAGTTAAGTAAGTTGATTCCGGTATCCCCCATCTTGCTTCCCTAA
TCTATGGTTCGGGAGACAGCATCAGGGTTAAGAAGACTTTTTTTTTTTTTTTTTTAACT
AGGAGAACCAATCTGGAAGCCAAAATGTAGGCTTAGTTTGTGTGTTGTCTCTTGAGTTT
GTCGCTCATGTGTGCAACAGGGTATGGACTATCTGTCTGGTGGCCCCGTTTCTGGTGGTC
TGTTGGCAGGCTGGCCAGTCCAGGCTGCCGTGGGGCCGCGCCTCTTCAAGCAGTCGTG
CCTGTGTCCATGCGCTCAGGGCCATGCTGAGGCCTGGGCCGCTGCCACGTTGGAGAAGCC
CGTGTGAGAAGTGAATGCTGGGACTCAGCCTTCAGACAGAGAGGACTGTAGGGAGGGCGG
CAGGGGCTGGAGATCTCTGACAGACCACGCCCCGTCTGCTGTGGCGCCGTCTCCAGG
GGCTGCTTCTCTCGGAAATTGACGAGGGGTGTCTTGGGCAGAGCTGGCTCTGAGCGCCT
CCATCCAAGGCCAGGTTCTCCGTTAGCTCCTGTGGCCCCACCTGGGGCCCTGGGCTGGAA
TCAGGAATATTTTCCAAAGAGTGATAGTCTTTTGCTTTTGGCAAACTCTACTTAATCCA
ATGGGTTTTTCCCTGTACAGTAGATTTTCCAAATGTAATAAACTTTAATATAAAGTAAAA
AAAAAAAAAAAAAAAAAAAAAAAAA

5

>Hs.44276_ARNm_2 gi|12654896|gb|BC001293.1|BC001293 clon de Homo sapiens MGC:5259 IMAGE:3458115
poliA = 3

CGGATGGGGAAAAAAGATGTCAGCTCCTCCGCTGTAGTATTGCTCCTTAAAAACCCC
 TCTCTCTGAAAATGACATGCCCTCGCAATGTAACCTCCGAACCTCGTACGCGGAGCCCTTGG
 CTGCGCCCGGCGGAGGAGAGCGCTATAGCCGGAGCGCAGGCATGTATATGCAGTCTGGGA
 GTGACTTCAATTGCGGGGTGATGAGGGGCTGCGGGCTCGCGCCCTCGCTCTCCAAGAGGG
 ACGAGGGCAGCAGCCCCAGCCTCGCCCTCAACACCTATCCGTCTACCTCTCGCAGCTGG
 ACTCCTGGGGCGACCCCAAAGCCGCCTATCGCCTGGAACAACCTGTTGGCAGGCCGCTGT
 CCTCCTGCTCCTACCCACCTAGTGTCAAGGAGGAGAATGTCTGCTGCATGTACAGCGCAG
 AGAAGCGGGCGAAAAAGTGGCCCCGAGGCAGCTCTCTACTCCCAACCCCTTGCCGGAGTCCT
 GCCTTGGGGAGCACGAGGTACCCGTGCCCGAGCTACTACCGCGCCAGCCCCGAGCTACTCCG
 CGCTGGACAAGACGCCCCACTGTTCTGGGGCCAACGACTTCGAAGCCCCCTTTCGAGCAGC
 GGGCCAGTCTCAACCCGCGCGCCGAACATCTGGAATCGCCTCAGCTGGGGGGCAAAGTGA
 GTTTCCTTGAGACCCCCAAGTCCGACAGCCAGACCCCCAGCCCCAATGAAATCAAGACGG
 AGCAGAGCCTGGCGGGCCCTAAAGGGAGCCCCCTCGGAGAGCGAAAAGGAGAGGGCCAAAG
 CTGCCGACTCCAGCCAGACACCTCGGATAACGAAGCGAAAGAGGAGATAAAGGCAGAAA
 ACACCACAGGAAATTGGCTGACAGCAAAGAGCGGAAGGAAGAAGAGGTGCCCCCTATACTA
 AACACCAGACGCTGGAATTGGAGAAAGAATTTCTGTTCAATATGTATTTGACCGGAGAGC
 GCCGCCTGGAGATTAGCAAGACCATTAAACCTTACAGACAGACAAGTCAAAATCTGGTTTC
 AAAATCGCAGAATGAAACTCAAGAAAAATGAACCGAGAGAATCGGATCCGGGAACTGACCT
 CCAATTTTAATTTACCTGAGAGCGCGGCCTCTCCTCCTCCCTTCCCGCTCCTTCCTCTC
 CCGCCCCCTCCTCCCTTTGTGCCTGGTGATATATTTTTTTTTCTCCTGAGTATAAATG
 CAATGCGACTGCAAAAAAGGCAAAGACCTCAGACTCTCCTTCCAAGGGACCTGTGGTTCCG
 TGCTGCGAAGATGCTTCCACTTAAAGCATGAGAAATGGGGTGCCGGGATGTGGGGTGTGG
 TGTGTGCCCTCATAGATGGGGGTGGGAGTGTGGCTGGTGTGTGTGTCAAACCCCTCACTCA
 CCCACGCACTCACACACAGCATTCTGTTCTCCATGCAAAGTTAAGATCGAATCCATCCGC
 TTGTAGGGGAAAAAAGGAAAAAATTAACCAGAGAGGGTCTGTAATCTCGCAGAGCACA
 GGCAGAATCGTTCCCTTCTGCTGCATTTCTCCTTAGACTAATAGACGTTTGGAAAGT
 TCGGCTAGTGTTCGTGTGTTTGTGCTAGCACCCAGAGCCTCCACCAAACCCCTCTCCATGT
 CTTTACCTCCCAGTCGCTCTAAGAAATCTGCTTGAAGTCTCGTATTTGTACTGCTTTCTGC
 TTTTC'TCCCACCCCTCCTAGCACCCCCACATCCCCCATCTAGTAACATCTCAGAAATTC
 ATCCAGAGGAACAAAAAATTAAAAATAGAACATAGCAAAGCAAAGACAGAATGCCCCC
 CCCAAATATTGTCTGTCCCTGTCTGGGAGTTGTGTTATTTAAAGATATTCTGTATGTTG
 TATCTTTTGCATGTAGCTTCCTTAATGGAGAAAAAATCCTAATAAATTTCCAGAATC

ATAATCCTCAA
 AAAAAAAAAA

>Hs.2142_ARNm_4 gi|13325274|gb|BC004453.1|BC004453 clon de Homo sapiens MGC:4303 IMAGE:2819400
 poliA = 3

GCAGTGGCCACGAGAGGCAGGCTGGCTGGGACATGAGGTTGGCAGAGGGCAGGCAAGCTG
 GCCCTTGGTGGGCTCGTCTGAGCACTCGGAGGCACTCCTATGCTTGGAAAGCTCGCTA
 TGCTGCTGTGGGTCCAGCAGGCGCTGCTCGCTTGCTCCTCCCCACACTCCTGGCACAGG
 GAGAAGCCAGGAGGAGCCGAAACACCACCAGGCCCGCTCTGCTGAGGCTGTGGATTACC
 TTTTGACCAACTACAGGAAGGGTGTGCGCCCCGTGAGGGACTGGAGGAAGCCAACCACCG
 TATCCATFGACGTCAATTGTCTATGCCATCCTCAACGTGGATGAGAAGAATCAGGTGCTGA
 CCACCTACATCTGGTACCGGCAGTACTGGACTGATGAGTTTCTCCAGTGGAAACCTGAGG
 ACTTTGACAACATCACCAAGTTGTCCATCCCCACGGACAGCATCTGGGTCCCGGACATTC
 TCATCAATGAGTTTGGTGGATGTGGGGAAGTCTCCAAATATCCCGTACGTGTATATTGGC
 ATCAAGGCGAAGTTTCAAGACTACAAGCCCCCTTCAGGTGGTGGTACTGCCTGTAGCCTCGACA
 TCTACAACCTCCCCCTTCGATGTCCAGAACTGCTCGCTGACCTTCACCAGTTGGCTGCACA
 CCATCCAGGACATCAACATCTCTTTGTGGCGCTTGCCAGAAAAGGTGAAATCCGACAGGA
 GTGCTTTCATGAACAGGAGAGTGGGAGTTGCTGGGGTGTGCTGCCCTACTTTCCGGAGT
 TCAGCATGGAAAGCAGTAATACTATGCAGAAATGAAGTTCTATGTGGTTCATCCGCCGGC
 GGCCCCCTCTTCTATGTGGTTCAGCCTGCTACTGCCAGCATCTTCCTCATGGTTCATGGACA
 TCGTGGGCTTCTACCTGCCCCCCAACAGTGGCGAGAGGGTCTCTTTCAAGATTACACTCC
 TCCTGGGCTACTCGGTCTTCCTGATCATCGTTTCTGACACGCTGCCGGCCACTGCCATCG
 GCACTCCTCTCATTTGGTGTCTACTTTGTGGTGTGCATGGCTCTGCTGGTGATAAGTTTGG
 CCGAGACCATCTTCATTGTGCGGCTGGTGCACAAGCAAGACCTGCAGCAGCCCGTGCTG
 CTTGGCTGCGTCACCTGGTTCTGGAGAGAATCGCCTGGCTACTTTGCCTGAGGGAGCAGT
 CAACTTCCCAGAGGCCCCCAGCCACCTCCCAAGCCACCAAGACTGATGACTGCTCAGCCA
 TGGGAAACCACTGCAGCCACATGGGAGGACCCAGGACTTCGAGAAGAGCCCGAGGGACA
 GATGTAGCCCTCCCCCACCACCTCGGGAGGCGCTCGCTGGCGGTGTGTGGGCTGCTGCAGG
 AGCTGTCTTCCATCCGGCAATTCTTGGAAAAGCGGGATGAGATCCGAGAGGTGGCCCCGAG
 ACTGGCTGCGCGTGGGCTCCGTGCTGGACAAGCTGCTATTCCACATTTACCTGCTGGCGG
 TGCTGGCCTACAGCATCACCTGGTTATGCTCTGGTCCATCTGGCAGTACGCTTGAGTGG
 GTACAGCCCAGTGGAGGAGGGGGTACAGTCTGGTTAGGTGGGGACAGAGGATTTCTGCT
 TAGGCCCCCTCAGGACCCAGGGAATGCCAGGGACATTTTCAAGACACAGACAAAGTCCCGT
 GCCCTGTTTCCAATGCCAATTCATCTCAGCAATCACAAGCCAAGGTCTGAACCCCTCCAC
 CAAAACTGGGTGTTCAAGGCCCTTACACCCCTGTCCCACCCCCAGCAGCTCACCATGGC
 TTTAAACATGCTCTCTTAGATCAGGAGAACTCGGGCACTCCCTAAGTCCACTCTAGTT
 GTGGACTTTTCCCATTTGACCCTCACCTGAATAAGGGACTTTGGAATTCTGCTTCTCTT
 CACAACCTTTGCTTTTAGGTTGAAGGCAAAACCACTCTCTACTACACAGGCTGTGATACT
 CTGTACGAGGCTTCTCTAACCCTAGTGTCTTTTTTTTCTTTCACCTCACTTGTGGCAGCT
 TCCCTGAACACTCATCCCCATCAGATGATGGGAGTGGGAAGAATAAAATGCAGTGAAAC
 CCTAAAAA

>Hs.180908_contig1 AA846824|AW611680|AA846182|AA846342|AA846360 poliA = 2 poliA = 3

TCCTCGCTCCTCTACCCCATAAAATTCCTTACAAATGCAAAAATTCGAGATAGAAGAAGC
 CGTCCCTGAAATTGCTGTCTAATTCACCGGAAACCTCTCCATAAACAAGGAGAAACGA
 ATGCACACGCATTTTGTCTAAGAAGCCCCGGATTAAAGATTTAAGGATACAAGCTGAAAGA
 AAAATGAAAAATGCTTCTCCGCGCGTCAATCGAGGGGTGGATGCGCCACGCAGCTGAGC
 CCAGCTCACAGCCACGCGTAAGACCAAAAGCTGCCATGGGTTCTGCGCGCGGAGACCTCA
 GAGCCGAAGAGAGAAGTCCCCGCGTCAGAAACGCTGCGGATGCCAGGTCTTGAAAATGCT
 GACTTCTGAGGCTAAGAATTATTTCAAAGACAAAAAGAAAGACTGGTGAGGAGGCCCTTC
 CGGTGCAAGGGCGCCTATCCGCTAATTTGGATGGGGAAGTAGGGATTATTCGTTAAAT
 TCAATCGCGAGCACCAAGTCGGAAGTGGCCGGGATGGAGAAGGGCAACCCCCACCTTTAG
 AAAATAAAAGATCTCGAAGGCCAAAAA

5

>Hs.89436_ARNm_1_gj|16507959|ref|NM_004063.2| cadherina 17 de Homo sapiens, cadherina LI (hígado-intestino) (CDH17), ARNm poliA = 1

AGGGAGTGTTCCCGGGGGAGATACTCCAGTCGTAGCAAGAGTCTCGACCACTGAATGGAA
GAAAAGGACTTTTAACCACCATTTTGTGACTTACAGAAAGGAATTTGAATAAAGAAAAC
ATGATACTTCAGGCCCATCTTCACTCCCTGTGTCTTCTTATGCTTTATTTGGCAACTGGA
TATGGCCAAGAGGGGAAGTTTAGTGGACCCCTGAAACCCATGACATTTTCTATTTATGAA
GGCCAAGAACCGAGTCAAATTATATTCAGTTTAAGGCCAATCCTCCTGCTGTGACTTTT
GAACTAACTGGGGAGACAGACAACATATTTGTGATAGAACGGGAGGGACTTCTGTATTAC
AACAGAGCCTTGGACAGGGAAACAAGATCTACTACAATCTCCAGGTTGCAGCCCTGGAC

GCTAATGGAATTATAGTGGAGGGTCCAGTCCCTATCACCATAGAAGTGAAGGACATCAAC
 GACAATCGACCCACGTTTCTCCAGTCAAAGTACGAAGGCTCAGTAAGGCAGAACTCTCGC
 CCAGGAAAGCCCTTCTTGTATGTCAATGCCACAGACCTGGATGATCCGGCCACTCCCAAT
 GGCCAGCTTTATTACCAGATTGTCAATCCAGCTTCCCATGATCAACAATGTCACTGTACTTT
 CAGATCAACAACAAAACGGGAGCCATCTCTCTTACCCGAGAGGGATCTCAGGAATTGAAT
 CCTGCTAAGAATCCTTCTTATAATCTGGTGATCTCAGTGAAGGACATGGGAGGCCAGAGT
 GAGAATTCCTTTCAGTGATACCACATCTGTGGATATCATAGTGACAGAGAATATTTTGAAA
 GCACCAAACCTGTGGAGATGGTGGAAAACCTCAACTGATCCTCACCCTCATCAAAATCACT
 CAGGTGCGGTGGAATGATCCCGGTGCACAATATTCCTTAGTTGACAAAGAGAAGCTGCCA
 AGATTCCCATTTTCAATTGACCAGGAAGGAGATATTTACGTGACTCAGCCCTTGGACCGA
 GAAGAAAAGGATGCATATGTTTTTATGCAAGTGGCAAAGGATGAGTACGGAAAACCACTT
 TCATATCCCGTGGAAATTCATGTAAAAGTTAAAGATATTAATGATAATCCACCTACATGT
 CCGTCACCAGTAACCGTATTTGAGGTCCAGGAGAATGAACGACTGGGTAACAGTATCGGG
 ACCCTTACTGCACATGACAGGGATGAAGAAAATACTGCCAACAGTTTTCTAAACTACAGG
 ATTTGTGGAGCAAACCTCCCAAACCTTCCCATGGATGGACTCTTCTAATCCAAACCTATGCT
 GGAATGTTACAGTTAGCTAAACAGTCTTTGAAGAAGCAAGATACTCCTCAGTACAACCTTA
 ACGATAGAGGTGTCTGACAAAGATTTCAAGACCTTTGTTTTGTGCAAATCAACGTTATT
 GATATCAATGATCAGATCCCATCTTTGAAAAATCAGATTATGGAAACCTGACTCTTGCT
 GAAGACACAAACATTGGGTCCACCATCTTAACCATCCAGGCCACTGATGCTGATGAGCCA
 TTTACTGGGAGTTCTAAAATTCTGTATCATATCATAAAGGGAGACAGTGAGGGACGCTG
 GGGGTTGACACAGATCCCATACCAACACCGGATATGTCATAATTAAAAAGCCTCTTGAT
 TTTGAAACAGCAGCTGTTTCCAACATTGTGTTCAAAGCAGAAAATCCTGAGCCTCTAGTG
 TTTGGTGTGAAGTACAATGCAAGTTCTTTTGCCAAGTTCACGCTTATTGTGACAGATGTG
 AATGAAGCACCTCAATTTTCCCAACACGTATTCGAAGCGAAAGTCAGTGAGGATGTAGCT
 ATAGGCACTAAAGTGGGCAATGTGACTGCCAAGGATCCAGAAGGTCTGGACATAAGCTAT
 TCACTGAGGGGAGACACAAGAGGTGGCTTAAATTTGACCACGTGACTGGTGAGATCTTT
 AGTGTGGCTCCATTGGACAGAGAAGCCGGAAGTCCATATCGGGTACAAGTGGTGGCCACA
 GAAGTAGGGGGGTCTTCTTGGAGCTCTGTGTGAGAGTTCCACCTGATCCTTATGGATGTG
 AATGACAACCTTCCAGGCTAGCCAAGGACTACACGGGCTTGTTCTTCTGCCATCCCTC
 AGTGACCTGGAAGTCTCATTTCGAGGCTACTGATGATGATCAGCACTTATTTGGGGT
 CCCCATTTTACATTTTCCCTCGGCAGTGGAAGCTTACAAAACGACTGGGAAGTTTCCAAA
 ATCAATGGTACTCATGCCGACTGTCTACAGGCACACAGAGTTTGAGGAGAGGGAGTAT
 GTCGTCTTGATCCGCATCAATGATGGGGGTCCGCCACCCTTGGGAAGGCATTGTTTCTTTA
 CCAGTTACATTCTGCAGTTGTGTGGAAGGAAGTTGTTTCCGGCCAGCAGGTCAACAGACT
 GGGATACCCACTGTGGGCATGGCAGTTGGTATAGTGTGACCACCCTTCTGGTGATTGGT
 ATAATTTTAGCAGTTGTGTTTATCCGCATAAAGAAGGATAAAGGCAAAGATAATGTTGAA
 AGTGCTCAAGCATCTGAAGTCAAACCTCTGAGAAGCTGAATTTGAAAAGGAATGTTTGAA
 TTTATATAGCAAGTGCTATTTTCAGCAACAACCATCTCATCCTATTACTTTTCATCTAACG
 TGCATTATAATTTTTTAAACAGATATTCCTCTTGCTCTTAAATATTTGCTAAATATTTT
 TTTTTGAGGTGGAGTCTTGCTCTGTGCGCCAGGCTGGAGTACAGTGGTGTGATCCAGC
 TCACTGCAACCTCCGCCTCCTGGGTTACATGATTCTCCTGCCTCAGCTTCTTAAGTAGC
 TGGGTTTACAGGCACCCACCACCATGCCCAGCTAATTTTTGTATTTTTAATAGAGACGGG
 GTTTCGCCATTGCGCCAGGCTGGTCTTGAACTCCTGACGTCAAGTGATCTGCCTGCCTTG
 GTCTCCCAATACAGGCATGAACCACTGCACCCACCCTACTTAGATATTTTCATGTGCTATAG
 ACATTAGAGAGATTTTTTCATTTTTCCATGACATTTTTCTCTCTGCAAATGGCTTAGCTA
 CTTGTGTTTTTCCCTTTTGGGGCAAGACAGACTCATTAATATTCTGTACATTTTTTCTT
 TATCAAGGAGATATATCAGTGTGTCTCATAGAACTGCCTGGATTCCATTATGTTTTTT
 CTGATCCATCCTGTGTCTCCCTTCATCCTTGACTCCTTTGGTATTTTCACTGAATTTCAA
 CATTTGTGAGAGAAGAAAACGTGAGGACTCAGGAATAAATAAATAAATAAAGAACGCC
 TTTTCCCTTAGTATTAACAGAAATGTTTTCTGTGTCTTAACCATCTTTAATCAATGTGAC
 ATGTTGCTCTTTGGCTGAAATCTTCAACTTGGAATGACACAGACCCACAGAAGGTGTT
 CAAACACAACCTACTCTGCAAACCTTGGTAAAGGAACCAAGTCAGCTGGCCAGATTTCTCT
 ACTACCTGCCATGCATACATGCTGCGCATGTTTTCTTCATTCTGATGTTAGTAAAGTTTT
 GGTATTATATATTTAACATGTGGAAGAAAACAAGACATGAAAAGAGTGGTGACAAATCA
 AGAATAAACACTGGTTGTAGTCAGTTTTGTTGTGTTAA

>Hs.151544_ARNm_8_gi|3153107|emb|AL023657.1|HSDSHP ADNc de SH2D1A de Homo sapiens, conocido anteriormente como DSHP polIA = 3

AAATCCTTCTTCCAATGTTCTCTCCCTCTCTGTATGAACCCCTGTGTTGGGGGGCAGAAGA
 TGGAAAGCCCTTGGCAAGCTCGATCGAACCAAGCTACTAAATTGCTGAGCTCGTTTTAACT
 GAAGTGTGAGAAGGAGGTTTAAGGCAAGTAGACAAACATCCTGTTGTTGGGGTGCTTCTCT
 CTTTTTGCACATCTGGCTGAACTGGGAGTCAGGTGGTTGACTTGTGCCTGGCTGCAGTA

GCAGCGGCATCTCCCTTGACAGTTCTCCTCCTCGGCCTGCCAAGAGTCCACCAGGCCA
 TGGACGCAGTGGCTGTGTATCATGGCAAAATCAGCAGGGAAACCGGCGAGAAGCTCCTGC
 TTGCCACTGGGCTGGATGGCAGCTATTTGCTGAGGGACAGCGAGAGCGTGCCAGGCGTGT
 ACTGCCATATGTGTGCTGTATCACGGTTACATTTATACATACCGAGTGTCCAGACAGAAA
 CAGGTTCTTGGAGTGTCTGAGACAGCACCTGGGGTACATAAAAGATATTTCCGGAAAATAA
 AAAATGCTCATTTTCAGCATTTTCAGAAGCCAGATCAAGGCATTGTAATACCTCTGCAGTATC
 CAGTTGAGAAGAAGTCCTCAGCTAGAAGTACACAAGGTACTACAGGGGATAAGAGAAGATC
 CTGATGTCTGCCTGAAAGCCCCATGAAGAAAAATAAAACACCTTGTAATTTATTTCTAT
 AATTTAAATATATGCTAAGTCTTATATATTTGTAGATAATACAGTTTCGGTGAGCTACAAAT
 GCATTTCTAAAGCCATTGTAGTCTGTAAATGGAAGCATCTAGCATGTGCTCAAAGCTGAA
 ATGGACTTTTGTACATAGTGAGGAGCTTTGAAACGAGGATTGGGAAAAAGTAATTCGGTA
 GGTTATTTTTCAGTTATTATATTTACAAATGGGAAACAAAGGATAATGAATACTTTATAA
 AGGATTAATGTCAATTCTTGCCAAATATAAATAAAATAATCCTCAGTTTTTGTGAAAAG
 CTCCATTTTTAGTGAAATATTATTTTATAGCTACTAATTTTAAATGTCTTGCTTGATTG
 TATGGTGGGAAGTTGGCTGGTGTCCCTTGTCTTTGCCAAGTTCTCCACTAGCTATGGTGT
 CATAGGCTCTTTTGGGATTTTTGAAGCTGTATACTGTGTGCTAAAACAAGCACTAAACAA
 AGAGTGAAGGATTTATGTTTAATTCTGAAAGCAACCTTCTTGCCCTAGTGTCTGATATTG
 GACAGTAAAATCCACAGACCAACCTGGAGTTGAAAATCTTATAATTTAAAATATGCTCTA
 AACATGTTTATCGTATTTGATGCTACAGGATTTGAAATTGTATTACAAATCCAATGAAAT
 GAGTTTTTCTTTTCATTTACCTCTGCCCCAGTTGTTTCTACTACATGGAAGACCTCATTT
 TGAAGGGAATTTTCAGCAGCTGCAGCTCATGAGTAACGTGATTTGTAACAAGCCTCCTTTT
 AAAGTAACCCTACAAAACCACTGGAAAGTTTATGGTTGTATTATTTTTTAAAAAAATTC
 AAGTGATTGAAACCTACACGAGATACAGAATTTTATGCGGCATTTTCTTCTCACATTTAT
 ATTTTTGTGATTTTGTGATTGATTATATGTCATTTGCTACAGGGCTCACAGAATTCATT
 CACTCAACAAACATAATAGGGCGCTGAGGGCATAGAAGTAAAAACACCTGGTCCCTGCTC
 TCAGTTCACTGTCTTGTGGACGAGAAAAAGAAACAATAACGATAAAAGACAGTGAAAGAA
 AATAACGATAAAAGACAGTGAAAGAAAAATAACAAATAAAAGACAAGGAAAAAATAACAAATG
 AAAGTTGATAAGTACATGATAAGCGAGGTTCCCGTGTGTAGGTAGATCTGGTCTTTAGA
 GGCAGATAGATAGGTGAGTGCAAATACTCTGGTCCATGGGCCATATGAAAAGGCTAAGCT
 TCACGTGTAATAATAACTGGGAATTCTGGATTGTGTATGGGTGTTGGTGAACCTTGGTTT
 TAATTAGTGAACCTGCTGAGAGACAGAGCTATTCTCCATGTACTGGCAAGACCTGATTTCT
 GAGCATTTAATATGGATGCCGTGGGAGTACAAAAGTGGAGTGTGGCCTGAGTAATGCATT
 ATGGGTGGTTTACCATTTCTTGAGGTAAAAGCATCACATGAACCTGTAAAGGAATTTAAA
 AATCCTACTTTTATAATAAGTTGCATAGGTTTAATAATTTTAAATTATATGGCTTGAGTT
 TAAATTGTAATAGGCGTAACATAATTTAACTCTATAATGTGTTTATTCTGGAATAATCCT
 AAACATATGAATTATGTTTGCATGTTCACTTCCAAGAGCCTTTTTTTGAAAAAAGCTTT
 TTTTGAATCATCAAGTCTTTCACATTTAAATAAAGTGTGTTGAAAGCTTTATTTAAAAAAA
 AA
 AAGAAAAAA

>Hs.1657_contig4

5 AW473119|AA164586|AI540656|AI758480|AI810941|AI978964|AI675862|AI784397|
 AW591562|AW514102|AI888116|AI983175|AI634735|AI669577|AI202659|AI910598|
 AI961352|AI565481|AI886254|AI538838|AA291749|AW571455|AI370308|AI274727|
 AW473925|AW514787|AI273871|AW470552|AI524356|AI888281|AW089672|AI952766|
 AW440601|AI654044|AW438839|AI972926 poliA = 2 poliA = 3

AATTGTTTTCTAAGTAATTGCTGCCTCTATTATGGCACTTCATTTTTGCACTGTCTTTTG
AGATTCAAGAAAAATTTCTATTCTTTTTTTTTGTCATCCAATTGTGCCTGAACCTTTTAAAT
ATGTAAATGCTGCCATGTTCCAAACCCATCGTCAGTGTGTGTGTTTAGAGCTGTGCACCC
TAGAAACAACATATTGTCCCATGAGCAGGTGCCTGAGACACAGACCCCTTTGCATTACACA
GAGAGGTCAATTGGTTATAGAGACTTGAATTAATAAGTGACATTATGCCAGTTTCTGTTCT
CTCACAGGTGATAAAACAATGCTTTTTTGTGCACTACATACTCTTCAGTGTAGAGCTCTTGT
TTTATGGGAAAAGGCTCAAATGCCAAATTGTGTTTGTATGGATTAATATGCCCTTTTGCCG
ATGCATACTATTACTGATGTGACTCGGTTTTGTGCGAGCTTTGCTTTGTTAATGAAACA
CACTTGTAACCTCTTTTGCACTTTGAAAAAGAATCCAGCGGGATGCTCGAGCACCTGTA
AACAATTTTCTCAACCTATTGATGTTCAAATAAAGAATTAAACTAAAAAAAAAAAAAAAAA
A

>Hs.35984_ARNm_1 gi|6049161|gb|AF133587.1|AF133587 cromosoma 22 map 22q11.2 de Homo sapiens poliA
= 3

5

GGCGCCGCGGACGCTGCTGGAGTCGCCTGGCAACGATGTGCGCTGGCAACTGAATAGGTT
GGCCAGTGGCGCGGGCTACTGGAAGCAGAAAGGGCTGCGGAGGCAGTGAGTGGTTTCTGC
AGAGCTTCATTTGGAAGGCCTCTGTAGTTGGGGAAAGATGGCCCATTCCCAGAACTCCT

TGGAGCTTCCCATTAAACATCAATGCCACCCAGATTACCACTGCCTATGGCCATCGGGCCC
TGCCCAAGCTGAAGGAGGAGCTGCAGTCAGAGGACCTCCAGACGAGGCAGAAAGCCCTCA
TGGCCCTGTGTGACCTCATGCATGACCCCGAGTGTATCTACAAGGCCATGAACATAGGCT
GTATGGAGAACCTGAAAGCTTTGCTGAAGGATAGCAACAGTATGGTGCGCATAAAGACCA
CCGAGGTGCTCCACATCACGGCAAGCCATAGCGTGGGCAGATACGCCCTTTCTAGAGCACG
ACATCGTCCTTGCCCTGTCTTCCCTGCTGAATGACCCAGCCAGTCTGCCGGGGGAACC
TGTACAAGGCATACATGCAGCTGGTCCAGGTGCCTAGAGGGGCCCCAAGAGATCATCAGCA
AAGGTCTGATTTCTCACTGGTATGGAAGCTGCAGGTGGAGGTGGAGGAGGAGGAGTTCC
AGGAGTTTCATCTGGACACACTGGTCTCTGCTGCAGGAGGATGCCACCGAGGCCCTGG
GCAGCAATGTGGTGCTTGTCTGAAGCAGAAGCTCCTCAGCGCCAACCAGAACATCCGCA
GCAAGGCCGCCCCGTGCGCTCCTTAATGTTCAGCATATCTCGAGAGGGCAAGAAACAGGTGT
GTCATTTTGACGTCACTCCCATCCTGGTCCATCTGCTGAAAGACCCAGTGGAGCATGTGA
AGTCTAACGCTGCCGCTGCCCTGATGTTGCCACAGTGATCACTGAAGGGAAGTATGCGG
CCCTGGAGGCACAAGCCATCGGCCCTGCTCCTGGAGCTGCTGCACTCCCCCATGACCATAG
CGCGCCTGAATGCCACCAAGGCCCTTACCATGCTGGCAGAGGCCCCCGAGGGCCGCAAGG
CCCTGCAGACGCACGTGCCCACTTTCCGTGCCATGGAGGTGGAGACTTACGAAAAGCCTC
AAGTGGCCGAAGCCTTACAGCGGCAGCCCGGATCGCCATCAGTGTCTCGAGTTCAAAC
CCTGAGCCCTTCATTCACCTCTGTGAGTGAATAAATGTGCTAAGTCTCTTTAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAA

>Hs.334534_ARNm_2 gi|17389403|gb|C017742.1|BC017742 Homo sapiens, clon IMAGE:4391536, ARNm poliA
= 3

10

AGAGCAGTAAGCTTGTGATAAAGGCCAATTCAGGTAGCTCTTGAAGGTGATAGCCATCT
 ACTTTCCAGTGGCTGCCAACACAGGGAGTGCCAGTTAACACTGGAAGGATTAAGGCAAG
 GTCCCTTCTCTTGAGACTCCCTCTGAGATCTGAAAAATGAAGTGGCTTAGGAACATCAG
 CAGTGAAGAACTGCCAAGAGTTGGTGAAGGTTGTCTCTTCCGAGGGCCTTCTGAAGACAG
 GGCTCTTGAACAGACAAGTGGAAGGGCTGTACCAGGGATAAAGGAAAGAAGTGCCTGTCC
 AGCAGGGAGCTTGAATTTAAGTTCATGTATGAAGTCATTGGCTCTATCTGCATTTTCT
 GTCATTCTCTTCATTTGTTTTAAGGTGGAATAATTTCTTACAGTTGATGCAAAGTATCAA
 CTACTTTACCCTACCTTCTCCCTTTTAGATGGGTTCTTCTGAGTTTGGAGTCTTGTA
 TGATTATCAGTATTCCCCTGTCAAAATCAAATCTATTACAGGTTTCTTCACTGTTGAGAAC
 ACCTAAATGTTTTATTTTTGAGAAAGTGGGGACAGAGTCTCACTATGTCACCCAGGCTGG
 AGTGCAATGGCATGATCTCAGCTCACTGCAACCTTCGCCTCCTGGGTTCAAGCGATTCTC
 CTGCCCTCCGCCTCCTGAGTAGCTGGGATTATAGGCACGCACCACCACGCCAGCTAATTT
 TTTGTATTTTTAGTAGAGACAGAGTTTCACCATGTTGGCCAGGCTGGTCTTGAACCTCTG
 ACCTTGTGATCCACCCACCTCGGCCTCCAGAGTGTCTGGGATTACAGGCATGAGCCACCA
 CGCTTGGCTAAGAACACCTAAATTTTATGTTTCTTGGCTCAAAAACAGTTCATTTCT
 AATGTTGTCCTCACAAGAAGGCTAATGGTGGTGAGACAGCAGGGGAGGAGGAAGAGCTG
 TGGTTTGTAACTTGTTCAACTCAGGCAATAAGCGATTTTAGCTTTATTTAAAGTCTTCTG
 TCCAGCTTTAAGCACTTTGTAAAGACATGGCTGAAAGTAGCTTTTCTATCAGAATTGCAGA
 TAGTCATGTTGGGCTAACAGTCAATTGGATATATTCCTTTACCTCACATGACCCCAGCAA
 CTGTGGTGGTATCTAGAGGTGAAACAGGCAAGTGAAATGGACACCTCTGCTGTGAATGTT
 TTAGAGAAGGAAATTCAAAAATGTTGTAACCTGAAAGCACTGTTGAATATGGGTATCGGC
 TTTCTTTTTCTACTTTGACTCTTAACATTATCAGTCAACTTCCACATTAATGAAAGTTGAC
 CATAGTTATTTCCAAATAAAAAGAAACCACTCTTACCAGGTCTTGGACTGTGATGTCAT
 ATTATTCAGTTTTATGCTTGTTCCTGAGCAGAACTCATAAGAGTGACATAGTCAGCTGCT
 GACGGCACCTCAGCCACGCCACTCTTACTCAGTTCAGTGGGTGTGCTTGCGTGGTAGGAT
 GTGGTGCAGCCCTCTCTACGCTCTTCTATTTTTGGTATATTTCTATCTAACCTTCAAT
 AGCTTCCAATTCTTTTTTCTTGGACTGGCTTCATTCTGAATTTGTGCTAAAAATAATCTT
 TCATAAAGAGACCTCAGTTTATAGCGTAACAGACTACACAATGCACTGATGTTTTCATAA
 TGTTTAAGGGACCCACTGCAAGAAGCTTGCTGCCCTCTTTAATTGTATTCATTTAGATT
 TTGATTTTCCATGTTAAGAAGGTGAGGTCCATGTTGGTGCCCTTCAGAGTAGAGAACCAT
 GTAAACATTAGGAATGAACAGAGGCCTTAGGAATGAATAGAGAGTTTGCTTATACAATT
 TCCTGTTACAAAGCTCTCCCTCTCATGCAAAGTAGGGAACACCTTTTGAGCATCTTTGAA
 TTTGACAAATGGTGTCTGTTGCAACACTTTTTTTTTGAGATGAAGTCTCGCGTGTGTAC
 CCGGGCTGGAGTGAGTGGCGTGATCTCGGCTCACTGCAACTTCCACCTCCTGGGTCCCA
 GCAGTTCTCCTGCCTCAGCCTCCCAAGTAGCTGAGATTACAGGCGCCTGCCACCCACCT
 GGCTGATTTTTGTAATTTTAGTAGAGACGGGGTTTACCATGTTGGCCAGGCTGATTAAC
 TCCTGACCTCAGGTGATCCACCTTTCTCGGCCTCCCAAAGTGCTGGGATTACGGGTGTGA
 GCCACCGTGCCCGGCTGCAACACATTTAATTGACAACACTAGGGCTGTTGTACAAAA
 TAGTAATGATAGCCATGGAAGTTTACCTTATTCTGTGAGAAGTGTCTTAACTTATTA
 AGTGTCTAAACTAAGGTTTAGTGCTTTTTTAAAGGAAAGTTGTCCCAGGATTCATCCTAA

AGAAAGCAAAAGTTAATTCAACTGATCCACCAATGGAATTAGATGGGTAGAGTTGGGTTT
 TTGAGTTTACCACCACTTAGTTCCTCACTGAATTTTGTAATTCCTGTGTTTGCATCCTC
 TGTTCTATTCTGCCCTTGCTCTGTGTCTCAGTCATTTGACTTAGAAAGTGCCTTC
 AAAAGGACCCTGTTCACTGCTGCACTTTTCAATGAATTTAAATTTATTTCTGTTCTAAAA
 AAAAAAAAAA

>Hs.60162_ARNm_1 gj|10437644|dbj|AK025181.1|AK025181 ADNc de Homo sapiens: FLJ21528 fis, clon COL05977 poliA = 3

TGATCAACAACGTGTCTAGCTCCCAGTCAGAGAGAAAGGGCCTCTTCAGTCTGTCTCAGGAG
 ACTGGGAGAAACAGCATAAAGGACCCCAAGGAAGGGAGAGGTACCCTGGGTGAGGCGC
 TTGTGGAGAGAGGGCTTCGCATGTAAAGTGACGTGAGGAAAAATAGAACAGAAAAAAGC
 CAGGGCCAGCCAGAGGCACCTGAGAAGAATCAGACCCACAGCTCAGCCAGCCCTGGCA
 CAGAGAAGAGACAGGCCTGGCAGCACCCAGGGACCCCTTTCTCAGCCTCCACCTGCAG
 GACAGCAGGAGCACTGATGCGCTGAAGGTACGTTCTGGAGTCTGGAAGCAGCAGAACTGA
 AGGAACTAAACACGGGTGTCTGGGAAGACCCCTCAAGCTGCAGTAAAGCCCAGGACTGAA
 TTGGCCACCTGAGGCCAAGGGTGGCACTCCAACCTCCTCCTAAAGGCTGGCTAGAGCCAC
 AGGAAAAGGGCCAGAAGCCAGAGAAAGGGCAAAGGTGGACCCCTGCCTCCAAACCTCCTCT
 GGAGACTGACCTCCTCTTTCTGTGCCTTATTGTTTCTCCCTCTTCTCTTTGTTGCGCCAC
 TGGGCGGTGACCTCAGGGATCCTGGCCTAACCTGGTGATTGTGCAGGCAACTGTGTCCGA
 GAAGACCCTTCTCTGGAAGATTGAACCCCAATTCAGCCATGGTGACTCCTTTGATGTCAA
 ACTGGTAAGGGCTGAGCCGTGGGCACAGGATACCACTCCTTCCAGCTCTTCTGCTGTGAC
 CTGCCCATGGAAGTCCCTGTGGACACGAAATCCTGTTTGGATCATCTAACTGGAGGCTCT
 CTGTCTTTCACCTCCACGCGCCCTCTTGACCCAGGAGGTTTCAGGGGAGGAAGTACGCCA
 CTCTCCACTGGCACCCCTCCTTGGCCTACACAGAGTCACCCCTGAGCCCTCAATGTGTGC
 TGAGGTGGGCCCTGCTCTCTGCAGGGGTATGGAGAGAAATAGCTTGGGGTGTGTGAGGC
 CCCGAAGAAGCTGGGCCTGTCTTCTCCATCGAGGCGATCCTAAAGAGGCCTGCCAGGAG
 GAGTGATATGGACAGACCAGAAGGGCCAGGTGAAGAGGGCCCCGGAGAAGCTGCGGCCTC
 AGGCTCTGGGCTAGAAAAGCCTCCAAAGGACCAGCCCAGGAAGGAAGGAAGAGCAAGCG
 GAGGGTTCGTACCACCTTCACCACTGAGCAGCTGCATGAGCTGGAGAAGATCTTCCACTT
 TACCCACTACCCAGACGTTTACATCCGCAGCCAGCTGGCAGCCAGGATCAACCTCCCAGA
 AGCTCGGGTGCAGATCTGGTTCCAGAATCAGCGAGCCAAGTGGCGGAAGCAGGAGAAGAT
 TGGCAACCTGGGGGCTCCACAGCAGCTGAGTGAAGCCAGTGTGGTCTCTGCCACAAATCT
 GGATGTGGCTGGGCCACGTTGACATCCACTGCTCTGCGCAGGCTGGCTCCTCCACGAG
 CTGTTGTCCATCGGCTCAAGATCAGCTGGCCTCTGCCTGGTTCCCTGCCTGGATCACCTT
 CCTCCCAGCGCACCCATGGGAAACACAGCCTGTCCCAGGTCTTCCCATCCATCAAACTTG
 CATCCCTGTGCTATGCATCCTTCCACCTCCACA'CCCCAAATGGGGCAGCATCTGTGCTAC
 TTCAACATAGAGATTGGACATGCTCTCCCCAAATGAGCCACTTTCTCTCCAGGTGAAGG
 CAGGTAGCAGATGTGCCCTGGGCCTCTGGGGAAATCGATCTCACAATCCAAAATGGCCC
 ACAGCCCAGGAAGCTACCCTGAACATGCCAGTTGGAAGGCTGCACCAGACTCAAAAGCAA
 ACTAAACAATAAAGGACAGCTCTCTTCTCTCTCTGGCTAAAGCTGCTCTCTCTGGTTGAGAA
 GACAGGCTGGATGAGATCTCAGGCCGAGCTCTGAAATAGGGAGGTAATCCTCCAGCACCT
 GTGTTTCTCTAACTTGCTGTGTGACCTCCAGCCGGTCACTCACCTCTCTGGACCTCAT
 CTGTAAGAGGAGCCAGCTGGATAAGATGATTTCTGAAGACGCTTCCATGGTGGGCACTGA
 GGCACAGAGGAGGCCAAGGAGAGGTTGTTTGTTCATGCATGCATTATCCGTGACACATG
 AGTACCTACTGAGGACTCCATAAACAGAACGGGATACAGAGATAAACAAATTTGGGTTCTG
 TCCACGTTTGTCAAAGGTGGTGCTGGCCCCACCTCTGAAAGCAGAACACTTGCTCAACAA
 CCTTGCTGTTGGCCCAAGTCTAACACATTCTTTATGACTGTGAGCATCTCAGAGTGAGAG
 AAAATGTAGAAAGTTTTTTAAATTCTAAACAGGATTTAGTGTCTTTAGTTATCTTGCTG
 GATGGGAAAGGGATGTTGTCTATTTCTGGCACAATGAAAAGTAGGACGGAAGCTCCTTTT
 CATTCAGTTTATCTTTCCAGGATATATGAAAAGGGACCAGCTGGAAGACAGCCTCACTC
 TGTCTCGAAAAGCCTGAGCTTTCAATTCAACTCCCTATTTCCATGCAAAGACGCTGGGCAA
 ACCACATGTTCTGTCTGAGCCTCAGTTTTCTATCCATAAAATGAAGGTAGCCAGGCCTG
 CCTCAAAGAGCATTCAGGAGGCTCTGAGAGGACATGAGAGTATTTTGCAAAGTGAGGGCA
 AGGCCAGTGTGGAGTGATATTGTTATTCCAAGATTCCACTGCAAAGTGGCTGCTTTGG
 ATGCCAGCCCAGGATGAGTAGTTCTGTTCTCAGGGAGGTCATCCGCTGAGCATCCCTTC
 TGCACAGATGTCTCTGATTCTTGTCCTTGACAGGTGGAGGACAGGGCCTGCTCCCTAAGC
 TGGGAAGCCTGGAATGACCTCTTGACAAGCCTAAATTCAGGAATCTTCCCCAAATCCC
 AGATCCTCTGCAATCTACCTGCACCCCTGACCCACCCAGGAGTTGGACCGGGAGTTGGGA
 AGCCTAGGTCTTAGTCTTACTCCTTCTAATTTGCTGTGTAACCTTACCATTAACTCTCT
 CTGGGTCTCAGTTTTCTCATCTGTATTGGAGGTAGCAGTGCTAGCTCTGCCTTCAGGCAT
 GCAATATGCCAGAACTACAGACAACAGCCCACAGGATGCAAAAGTGCTTTGCCATCTTAA
 AAATGCCAGATCACTCAGAGCCTATGAATGTGGATATCAACACCAGGTCTCTAGCACCGC

TGGATGAAAGGAGAAGGCTAGAGGCTGAGGGAGGAAAGAGCAGTTAAACAAACAAAGGCAG
 TAGCTCATCACTTGGGTAGCAGGTACCCATTTTAGGACCC'TCACTCAAATGTGCAAAAT
 AAAATTTCTATCATTTTGTCTATAAAAAAAAAAAAAAAAAAAAAA

APPENDIX 2

>NM_004967

GAGTGAGTGAGAGGGCAGAGGAAATACTCAATCTGTGCCACTCACTGCCTTGAGCCTGCT
 TCCTCACTCCAGGACTGCCAGAGGCTCACTCCCTTGAGCCTGCTTCCTCACTCCAGGACT
 GCCAGAGGAAGCAATCACCAAAATGAAGACTGCTTTAATTTTGCTCAGCATTTTGGGAAT
 GGCCGTGTGCTTTCTCAATGAAAAATTTGCATCGAAGAGTCAAAATAGAGGATTCTGAAGA
 AAATGGGGTCTTTAAGTACAGGCCACGATATTATCTTTACAAGCATGCCTACTTTTATCC
 TCATTTAAAACGATTTCCAGTTCAGGGCAGTAGTGAATCATCCGAAGAAAATGGAGATGA
 CAGTTCAGAAGAGGAGGAGGAAGAAGAGGAGACTTCAAATGAAGGAGAAAACAATGAAGA
 ATCGAATGAAGATGAAGACTCTGAGGCTGAGAATACCACACTTTCTGCTACAACACTGGG
 CTATGGAGAGGACGCCACGCCTGGCACAGGGTATACAGGGTTAGCTGCAATCCAGCTTCC
 CAAGAAGGCTGGGGATATAACAAACAAAGCTACAAAAGAGAAAGGAAAGTGATGAAGAAGA
 AGAGGAGGAAGAGGAAGGAAATGAAAACGAAGAAAGCGAAGCAGAAGTGGATGAAAACGA
 ACAAGGCATAAACGGCACCAGTACCAACAGCACAGAGGCAGAAAACGGCAACGGCAGCAG
 CGGAGGAGACAATGGAGAAGAAGGGGAAGAAGAAAGTGTCACTGGAGCCAATGCAGAAGG
 CACCACAGAGACCGGAGGGCAGGGCAAGGGCACCTCGAAGACAACAACCTCTCCAAATGG
 TGGGTTTGAACCTACAACCCACCACAAGTCTATAGAACCCTTCCCACCTTTTGGGAA
 AACCACCACCGTTGAATACGAGGGGGAGTACGAATACACGGGCGTCAATGAATACGACAA
 TGGATATGAAATCTATGAAAGTGAGAACGGGGAACCTCGTGGGGACAATACCGAGCCTA
 TGAAGATGAGTACAGCTACTTTAAAGGACAAGGCTACGATGGCTATGATGGTCAGAATTA
 CTACCACCACAGTGAAGCTCCAGCCTG

5

>NM_002847

CTGATGGCGGCCCTCAGTGCCTATGCTGCCAGAGGCCCCACAGCTCCCCCGGGGAGGGC
AGCCTGGAGCCACAGTACCTTCTGCGTGACCCCTCAAGAATGCCCAGGCCTTTGCTGGCA
CCAGCCGCCCCCAGAAGTGGCCTTCACTCTGGGAGATTCCGAAGACCCCTCCAGCACA
GGCGATGGAGCAGGATTACATACCCTCCTGAAGGACCTGCAGAGGCAGCCGGCTGAGGTG
AGGGGCCTGAGTGGCCTGGAGCTGGACGGCATGGCTGAGCTGATGGCTGGCCTGATGCAA
GGCGTGGACCATGGAGTAGCTCGAGGCAGCCCTGGGAGAGCGGCCCTGGGAGAGTCTGGA
GAACAGGCGGATGGCCCCAAGGCCACCCTCCGTGGAGACAGCTTTCAGATGACGGAGTG
CAGGACGACGATGATAGACTTTACCAAGAGGTCCATCGTCTGAGTGCCACACTCGGGGGC
CTCCTGACAGGACCAGGGTCTCGACTCTTACCTGGAGCCCTCCCCCTTGCAAGGCCCTC
GACATGGAGAGGAAGAAGTCCGAGCACCCTGAGTCTTCCCTGTCTTCAGAAGAGGAGACT
GCCGGAGTGGAGAACGTCAAGAGCCAGACGTATTCAAAGATCTGCTGGGGCAGCAGCCG
CATTCGGAGCCCGGGGCCGTGCGTTTGGGGAGCTCCAAAACAGATGCCTGGGCCCTCG
AAGGAGGAGCAGAGCCTTCAGCGGGTGCTCAGGAGGCCCTCAGCGACGGCCTGCAATTG
GAGGTCCAGCCTTCCGAGGAAGAGGCGCGGGCTACATCGTGACAGACAGAGACCCCTG
CGCCCCGAGGAAGGAAGGCGGCTGGTGGAGGACGTGCGCCGCTCCTGCAGGTGCCCAGC
AGTGCGTTCGCTGACGTGGAGGTTCTCGGACCAGCAGTGACCTTCAAAGTGAGCGCCAAT
GTCCAAAACCTTGACCACTGAGGATGTGGAGAAGGCCACAGTTGACAACAAAGACAAATG
GAGGAAACCTCTGGACTGAAAATTCTTCAAACCGGAGTCGGGTGAAAAGCAAACCTAAG
TTCTTGCCCTCCTCAGGCGGAGCAAGAAGACTCCACCAAGTTTCATCGCGCTCACCTTGGTC
TCCCTCGCCTGCATCCTGGGCGTCTCCTGGCCTCTGGCCTCATCTACTGCCTCCGCCAT
AGCTCTCAGCACAGGCTGAAGGAGAAGCTCTCGGGACTAGGGGGCGACCCAGGTGCAGAT

GCCACTGCCGCTACCAGGAGCTGTGCCGCCAGCGTATGGCCACGCGGCCACCAGACCGA
 CCTGAGGGCCCCGCACACGTCACGCATCAGCAGCGTCTCATCCCAGTTTCAGCGACGGGCCG
 ATCCCCAGCCCCCTCCGCACGCAGCAGCGCCTCATCCTGGTCCGAGGAGCCTGTGCAGTCC
 AACATGGACATCTCCACCGGCCACATGATCCTGTCTACATGGAGGACCACCTGAAGAAC
 AAGAACC GGCTGGAGAAGGAGTGGGAAGCGCTGTGCGCCTACCAGGCGGAGCCCCAACAGC
 TCGTTTCGTGGCCCCAGAGGGAGGAGAACGTGCCCAAGAACCGCTCCCTGGCTGTGCTGACC
 TATGACCACTCCCGGGTCTGTGCTGAAGGCGGAGAACAGCCACAGCCACTCAGACTACATC
 AACGCTAGCCCCATCATGGATCACGACCCGAGGAACCCCGGTACATCGCCACCCAGGGA
 CCGCTGCCCGCCACCGTGGCTGACTTTTGGCAGATGGTGTGGGAGAGCGGCTGCGTGGTG
 ATCGTCATGCTGACACCCCTCGCGGAGAACGGCGTCCGGCAGTGCTACCACTACTGGCCG
 GATGAAGGCTCCAATCTCTACCACATCTATGAGGTGAACCTGGTCTCCGAGCACATCTGG
 TGTGAGGACTTCCTGGTGAGGAGCTTCTATCTGAAGAACCTGCAGACCAACGAGACGCGC
 ACCGTGACGCAGTTCCACTTCCTGAGTTGGTATGACCGAGGAGTCCCTTCCTCCTCAAGG
 TCCCTCCTGGACTTCCGCAGAAAAAGTAAACAAGTGCTACAGGGGCCGTTCTTGCTCCAATA
 ATTGTTTCATTGCAGTGACGGTGCAGGCCGAGCGGCACCTACGTCTTGATCGACATGGTT
 CTCAACAAGATGGCCAAAGGTGCTAAAGAGATTGATATCGCAGCGACCCCTGGAGCACTTG
 AGGGACCAGAGACCCGGCATGGTCCAGACGAAGGAGCAGTTTGAGTTTCGCGCTGACAGCC
 GTGGCTGAGGAGGTGAACGCCATCCTCAAGGCCCTTCCCCAGTGAGCGGCAGCCTCAGGG
 GCCTCAGGGGAGCCCCACCCACGGATGTTGTGAGGAATCATGATCTGACTTTAATTGT
 GTGTCTTCTATTATAACTGCATAGTAATAGGGCCCTTAGCTCTCCGCTAGTCAGCGCAGT
 TTAGCAGTTAAAGTGATTTTGTGTTAATCAAAACAATAAAGAGAGATTGTGAGAAA
 AATCCAGTTACGGGTGGAGGGGAATCGGTTTCATCAATTTTCACTTGCTTAAAAAAAATAC
 TTTTCTTAAAGCACCCGTTACCTTCTTGGTTGAAGTTGTGTTAACAATGCAGTAGCCA
 GCACGTTTCGAGGCGGTTTCCAGGAAGAGTGTTGCTTGTCTGCTGCTGCTGCTGCTGCTG
 GATCCACTGTGCAGGAGTGGCCGGGGAAAGCTGGCAGCACTCAGTGAGGCCGCCCGGCACA
 CAAGGCACGTTTGGCATTCTCTTTGAGAGAGTTTATCATTGGGAGAAGCCGCGGGGACA
 GAACTGAACGTCCTGCAGCTTCGGGGCAAGTGAGACAATCACAGCTCCTCGCTGCGTCTC
 CATCAACACTGCGCCGGGTACCATGGACGGCCCCGTGAGCCACACCTGTGAGCCCAAGCA
 GAGTGATTGAGGGGCTCCCCGGGGGAGACACCTGTGACCCCATGAGTAGTGCCCACTT
 GAGGCTGGCACTCCCCTGACCTCACCTTTGCAAGTTACAGATGCACCCCAACATTGAGA
 TGTGTTTTTAATGTTAAATATATGATTCTACGTTATGAAAACAGATGCCCCCGTGAATG
 CTTACCTGTGAGATAACCACAACCAGGAAGAACAATCTGGGCATTGAGCAAGCTATGAG
 GGTCCCCGGGAGCACACGAACCTTGCCAGGCCCCCGCTGGCTCCTCCAGGCACGTCCCGG
 ACCTGTGGGGCCCCAGAGAGGGGACATTTCCCTCCTGGGAGAGAAGGAGATCAGGGCAAC
 TCGGAGAGGGCTGCGAGCATTTCCCTCCCGGAGAGGAGATCAGGGCGACCTGCACGCAC
 TCGGTAGAGCCTGGAAGGGAAGTGAGAAACCAGCCGACCGGCCCTGCCCCCTTCCCGGG
 ATCACTTAATGAACCACGTGTTTGGACATCATGTAAACCTAAGCACGTAGAGATGATTTCG
 GATTTGACAAAATAACATTTGAGTATCCGATTCGCCATCACCCCTACCCAGAAATAGG
 ACAATTCACTTCATTGACCAGGATGATCACATGGAAGGCGGCGCAGAGGCAGCTGTGTGG
 GCTGCAGATTTCTGTGTGGGGTTCAGCGTAGAAAACGCACCTCCATCCCGCCCTTCCCA
 CAGCATTCCTCCATCTTAGATAGATGGTACTCTCCAAAGGCCCTACCAGAGGGGAACACGG
 CCTACTGAGCGGACAGAATGATGCCAAAATATTGCTTATGTCTCTACATGGTATTGTAAT
 GAATATCTGCTTTAATATAGCTATCATTTCTTTTCCAAAATTACTTCTCTCTATCTGGAA
 TTTAATTAATCGAAATGAATTTATCTGAATATAGGAAGCATATGCCTACTTGTAATTTCT
 AACTCCTTATGTTTGAAGAGAAACCTCCGGTGTGAGATATACAAATATATTTAATTGTGT
 CATATTAACTTCTGATTCAAAAAAAA

>BC002551

GGCACGAGGCCACGAGCTGTTGTGCATCCAGAGGTGGAATTGGGGCCCGGCATTCCCTCC
 TCGTCCCGGGCTGGCCCTTGCCCCACCCTGCAACTCCTGGTTGAGATGGGCTCAGCCAA
 GAGCGTCCCACTCACACCAGCGCGGCCTCCGCCGACACAAGCATCTGGCTCGAGTGGC
 GGACCCCCGTTACCTAGTGCTGGCATCCTGCGCACTCCCATCCAGGTGGAGAGCTCTCC
 ACAGCCAGGCCTACCAGCAGGGGAGCAACTGGAGGGTCTTAAACATGCCAGGACTCAGA
 TCCCCGCTCTCCTACTCTTGGTATTGCACGGACACCTATGAAGACCAGCAGTGGAGACCC
 CCCAAGCCCACTGGTGAAACAGCTGAGTGAAGTATTTGAAACTGAAGACTCTAAATCAAA
 TCTTCCCCCAGAGCCTGTTCTGCCCCCAGAGGCACCTTTATCTTCTGAATTGGACTTGCC
 TCTGGGTACCCAGTTATCTGTTGAGGAACAGATGCCACCTTGAACCAGACTGAGTTCCC
 CTCCAAACAGGTGTTTTCCAAGGAGGAAGCAAGACAGCCACAGAAACCCCTGTGGCCAG
 CCAGAGCTCCGACAAGCCCTCAAGGGACCCTGAGACTCCAGATCTTCAGGTTCTATGCG
 CAATAGATGGAAACCAACAGCAGCAAGGTACTAGGGAGATCCCCCCTCACCATCCTGCA
 GGATGACAACCTCCCCTGGCACCCCTGACACTACGACAGGGTAAGCGGCCTTCACCCCTAAG

TGAAAATGTTAGTGAACTAAAGGAAGGAGCCATTCTTGGAAGTGGACGACTTCTGAAAAC
 TGGAGGACGAGCATGGGAGCAAGGCCAGGACCATGACAAGGAAAATCAGCACTTTCCTT
 GGTGGAGAGCTAGGCCCTGCATGGCCCCAGCAATGCAGTCACCCAGGGCCTGGTGATATC
 TGTGTCTCTCACCCCTTCTTCCCAGGGATACTGAGGAATGGCTTGTTTCTTAGACTC
 CTCCTCAGCTACCAAACCTGGGACTCACAGCTTTATTGGGCTTCTTTGTGTCTTGTGTGT
 TTCTTTTATATTAAAGGAAGTAATTTTAAATGTTACTTTAAAAAGGTAAAAA
 AAAAAA

>AL039118

GCATTGCTAGTAAAGGTGCCCAAGAAATTATTTTGGCCATTTATTGTTTTGTCTTTTCT
 TTAAAGAACTGTTTTTTTTTCTTTTGTCTTACTTTTAGACCAAAGATTGGGTTCTAGAAAA
 TGCACCTTGGTATACCTAAGTATTAAACAAACAAAAAGGAAAGTTGTTTCAGTTGGCAACA
 CTGCCCATTCAATTGAATCAGAAGGGGACAAAAATTAACGATTGCCTTCAGTTTGTGTTGT
 GTATATTTTGATGTATGTGGTCACTAACAGGTCACTTTTATTTTTCTAAATGTAGTGAA
 ATGTTAATACCTATTGTACTTATAGGTAAACCTTGCAAATATGTAACCTGTGTTGCGCAA
 ATGCCGCATAAATTTGAGTGATTGTTAATGTTGTCTTAAATTTCTTGATTGTGATACTG
 TGGTCATATGCCCGTGTTTGTCACTTACAAAAATGTTTACTATGAACACACAGAAATAAA
 AAATAGGCTAAATTCATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

5

>NM_000198

GAGGCAGTAAGGACTTGGACTCCTCTGTCCAGCTTTTAACAATCTAAGTTACGGTTACCC
 TCTTCTGGGTACGCTAGAATCAGATCTGCTCTCCAGCATCTTCTGTTTCCTGGCAAGTG
 TTTCTGTACTTTGGATTGGCCACGATGGGCTGGAGCTGCCTTGTGACAGGAGCAGGAG
 GGCTTCTGGGTACAGAGGATCGTCCGCCTGTTGGTGGGAAGAGAAGGAACTGAAGGAGATCA
 GGGCCTTGGACAAGGCCCTTCAGACCAGAATTGAGAGAGGAATTTTCTAAGCTCCAGAACA
 GGACCAAGCTGACTGTACTTGAAGGAGACATTCTGGATGAGCCATTCTGAAAAGAGCCT
 GCCAGGACGTCTCGGTCTCATCCACACCGCCTGTATCATTGATGTCTTTGGTGTCACTC
 ACAGAGAGTCCATCATGAATGTCAATGTGAAAGGTACCCAGCTACTGTTGGAGGCCTGTG
 TCCAAGCCAGTGTGCCAGTCTTCATCTACACCAGTAGCATAGAGGTAGCCGGGCCCCAACT
 CCTACAAGGAAATCATCCAGAACGGCCACGAAGAAGAGCCTCTGGAAAACACATGGCCCA
 CTCCATACCCGTACAGCAAAAAGCTTGCTGAGAAGGCTGTGCTGGCGGCTAATGGGTGGA
 ATCTAAAAATGGTGTATACCTTGTACACTTGTGCGTTAAGACCCACATATATCTATGGGG
 AAGGAGGCCCATTCCTTTCTGCCAGTATAAATGAGGCCCTGAACAACAATGGGATCCTGT
 CAAGTGTGGAAAGTTCTCTACAGTCAACCCAGTCTATGTTGGCAACGTGGCCTGGGCCC
 ACATTCTGGCCTTGGGGCTCTGCGGGACCCCAAGAAGGCCCCCAAGTGTCCGAGGTCAAT
 TCTATTACATCTCAGATGACACGCCTCACCAAAGCTATGATAACCTTAATTACATCCTGA
 GCAAAGAGTTGGCCTCCGCCTTGATTCCAGATGGAGCCTTCCTTTAACCTGATGTACT
 GGATTGGCTTCTTGCTGGAAGTAGTGAGCTTCTACTCAGCCCAATTTACTCCTATCAAC
 CCCCCTTCAACCGCCACACAGTCACATTATCAAATAGTGTGTTACCTTCTCTTACAAGA
 AGGCTCAGCGAGATCTGGCGTATAAGCCACTCTACAGCTGGGAGGAAGCCAAGCAGAAAA
 CCGTGGAGTGGGTTGGTTCCCTTGTGGACCGGCACAAGGAGACCCTGAAGTCCAAGACTC
 AGTGATTTAAGGATGACAGAGATGTGCATGTGGGTATTGTTAGGAAATGTCATCAAACCTC
 CACCCACCTGGCTTCATACAGAAGGCAACAGGGGCACAAGCCCAGGTCCTGCTGCCTCTC
 TTTCACACAATGCCCCAATTACTGTCTTCTTCATGTGCATCAAATCTGCACAGTCACTGG
 CCCAACCAGAACTTTCTGTCTAATCATACACCAGAAGACAAACAATATGATTTGCTGTT
 ACCAAATCTCAGTGGCTGATTCTGAACAATTGTGGTCTCTCTTAACCTTGAGGTTCTCTTT
 TGACTAATAGAGCTCCATTTCCCCTCTTAAATGAGAAAGCATTTCCTTTCTCTTAATCT
 CCTATTCCTTCACACAGTTCAACATAAAGAGCAATAAATGTTTTAATGCTTAA

>H05388

AAATTTTGACCCCATATAAAGAAATGTGTTATGTATGTTGTGCCTCCTTAGAGACATAAA
 TTTAGTGTCAAAACATGGGAGATGGCTTACTCAGAAGCATACTCCACTTAACATACCATG
 GCCTGAGCTAAGTACCATGTCTGTTTGTGTCTTATTTTTAAATATTTTCTTTGTCCACA
 TGGGCGGTTGACCTTAGAGTTAAGGCGGTTGCTTTTTTGAAGAAATCACCAAAGTTTCTG
 GGAAACTATGTTCAAGGTTGAAATGGAGAGTAGATTTAATTTTATTTGTCTTGTAGGGAA
 GAAATCTTCCTTTGAACCGCTTTTCTTGCTTTTTCCCTTTTTCCCAAAGTACAGG
 TTCTTATCTGCAAGGTTCAAGTTGCTTAGACATTGTTTTCCAGTATCTGCAGGGCCAGT
 CAGTTGTACAGAAGTTGGAATATTCTGTTCCAGAATTAAAGAAGTTTTTAGATTATGAAA
 TATTATGATAATAAAGCTATATTTCTGAAAAAAAAAAAA

5

>NM_004062

GAAGGAGCTCTCTTCTTGCTTGGCAGCTGGACCAAGGGAGCCAGTCTTGGGCGCTGGAGG
 GCCTGTCTTGACCATGGTCCCTGCCTGGCTGTGGCTGCTTTGTGTCTCCGTCCCCCAGGC
 TCTCCCCAAGGCCAGCCTGCAGAGCTGTCTGTGGAAAGTTCCAGAAAACATGGGTGGAAA
 TTTCCCTTTATACCTGACCAAGTTGCCGCTGCCCCGTGAGGGGGCTGAAGGCCAGATCGT
 GCTGTGAGGGGACTCAGGCAAGGCAACTGAGGGGCCATTTGCTATGGATCCAGATTCTGG
 CTTCTGTGCTGGTGACCAGGGCCCTGGACCGAGAGGAGCAGGCAGAGTACCAGCTACAGGT
 CACCCTGGAGATGCAGGATGGACATGTCTTGTGGGGTCCACAGCCTGTGCTTGTGCACGT
 GAAGGATGAGAATGACCAGGTGCCCCATTTCTCTCAAGCCATCTACAGAGCTCGGCTGAG
 CCGGGGTACCAGGCCCTGGCATCCCCCTTCTCTTCTTCTGAGGCTTCAGACCGGGATGAGCC
 AGGCACAGCCAACCTCGGATCTTCGATTCCACATCCTGAGCCAGGCTCCAGCCCAGCCTTC
 CCCAGACATGTTCCAGCTGGAGCCTCGGCTGGGGGCTCTGGCCCTCAGCCCCAAGGGGAG
 CACCAGCCTTGACCACGCCCTGGAGAGGACCTACCAGCTGTTGGTACAGGTCAAGGACAT
 GGGTGACCAGGCCCTCAGGCCACCAGGCCACTGCCACCGTGGAAGTCTCCATCATAGAGAG
 CACCTGGGTGTCCCTAGAGCCTATCCACCTGGCAGAGAATCTCAAAGTCTTATACCCGCA
 CCACATGGCCCAGGTACACTGGAGTGGGGGTGATGTGCACTATCACCTGGAGAGCCATCC
 CCGGGGACCTTTGAAGTGAATGCAGAGGGAAACCTCTACGTGACCAGAGAGCTGGACAG
 AGAAGCCCAGGCTGAGTACCTGCTCCAGGTGCGGGCTCAGAATTCCTATGGCGAGGACTA
 TGCGGGCCCCCTCTGGAGCTGCACGTGCTGGTGTGATGGATGAGAATGACAACGTGCCTATCTG
 CCCTCCCCGTGACCCACAGTCAGCATCCCTGAGCTCAGTCCACCAGGTACTGAAGTGAC
 TAGACTGTGACAGAGGATGCAGATGCCCCCGGCTCCCCCAATTCCACGTTGTGTATCA
 GCTCCTGAGCCCTGAGCCTGAGGATGGGGTAGAGGGGAGAGCCTTCCAGGTGGACCCAC
 TTCAGGCAGTGTGACGCTGGGGGTGCTCCCACTCCGAGCAGGCCAGAACATCCTGCTTCT
 GGTGCTGGCCATGGACCTGGCAGGCGCAGAGGGTGGCTTCAGCAGCACGTGTGAAGTCGA
 AGTCGCAGTCACAGATATCAATGATCACGCCCTGAGTTTCATCACTTCCCAGATTGGGCC
 TATAAGCCTCCCTGAGGATGTGGAGCCCGGACTCTGGTGGCCATGCTAACAGCCATTGA
 TGCTGACCTCGAGCCCGCCTTCCGCCTCATGGATTTTGCCATTGAGAGGGGAGACACAGA
 AGGGACTTTTGGCCTGGATTGGGAGCCAGACTCTGGGCATGTTAGACTCAGACTCTGCAA
 GAACCTCAGTTATGAGGCAGCTCCAAGTCAATGAGGTGGTGGTGGTGGTGCAGAGTGTGGC
 GAAGCTGGTGGGGCCAGGCCAGGCCCTGGAGCCACCGCCACGGTGACTGTGCTAGTGGGA
 GAGAGTGATGCCACCCCCCAAGTTGGACCAGGAGAGCTACGAGGCCAGTGTCCCATCAG
 TGCCCCAGCCGGCTCTTTCCTGCTGACCATCCAGCCCTCCGACCCCATCAGCCGAACCTT
 CAGGTTCTCCCTAGTCAATGACTCAGAGGGCTGGCTCTGCATTGAGAAATTCTCCGGGGA
 GGTGCACACCGCCAGTCCCTGCAGGGCGCCAGCCTGGGGACACCTACACGGTGTCTGT
 GGAGGCCCAGGATACAGATGAGCCGAGACTGAGCGCTTCTGCACCCCTGGTGTATCCACTT
 CCTAAAGGCCCTCCTGCCCCAGCCCTGACTCTTGCCCCCTGTGCCCTCCCAATACCTCTG
 CACACCCCGCCAAGACCATGGCTTGATCGTGAGTGGACCCAGCAAGGACCCGATCTGGC
 CAGTGGGCACGGTCCCTACAGCTTACCCCTTGGTCCCAACCCACGGTGCAACGGGATTG
 GCGCCTCCAGACTCTCAATGGTTCCCATGCCTACCTCACCTTGGCCCTGCATTGGGTGGA
 GCCACGTGAACACATAATCCCCGTGGTGGTGCAGCCACAATGCCCAGATGTGGCAGCTCCT
 GGTTCCAGTGATCGTGTGCTGCAACGTGGAGGGGAGTGCATGCGCAAGGTGGGCCG
 CATGAAGGGCATGCCCACGAAGCTGTCGGCAGTGGGCATCCTTGTAGGCACCTTGGTAGC
 AATAGGAATCTTCTCATCTCATTTCACCCACTGGACCATGTCAAGGAAGAAGGACCC
 GGATCAACCAGCAGACAGCGTGGCCCTGAAGGCGACTGTCTGAATGGCCCAGGCAGCTCT
 AGCTGGGAGCTTGGCCTCTGGCTCCATCTGAGTCCCCTGGGAGAGAGCCCAGCACCCAAG
 ATCCAGCAGGGGACAGGACAGAGTAGAAGCCCTCCATCTGCCCTGGGGTGGAGGCACCA
 TCACCATCACAGGCATGTCTGCAGAGCCTGGACACCAACTTTATGGACTGCCCATGGGA
 GTGCTCCAAATGTCAGGGTGTTCGCCCAATAATAAGCCCCAGAGAACTGGGCTGGGGCC
 TATGGGATTGGTA

>AA782845

TCTTTACCTATGTGAAGCGAGGTGACGTGATACGTCACTGGCGCCGTCTTATAATTTAGA
 TGTA AAAATCTTTAGAAACAAATAAACTCTCTATATATGTGTATGTCTGTGTACAAAA
 AATGACAGAGCTGATGGCCAGTGTATACAGAGCGTGGCCCGCGGTGTACAATACCCATAT
 AAGGTACATTGTGCAGGAGGGGAATTGCTGGCTGCTTTTACTTCCCTGACCAAGACTGAAA
 AATTATTTACTGAAATCTGTAAACCTTTTTATGAACTTTTAAGCACCAGGCTGTTTACT
 TACACAATTTAGGTCTGCCAGAAAATTTCTATCTGTGATAGATCTGTAAAGAGGGTCAGGG
 GTTAGAGTTTACTATTTTGAAGTTTACATTGTTACATATGAAATGGAAACATTATTTTG
 AAACGTTGTCATAACCCAATGGTGCATTCTGTAACCATGGAGTCTTCTGTTTCCCTGGGGG
 AAAGGGGCATTTCATGACCTGAACCTTTTGTAGCAAATTATTATTCTCAGTTTCCATTACCTG
 TTTGGCCAAACAGATTAATAAAATATTTGAAAAAGAAGCAATAAAAAAAAAA

>AI457360

CTGAGAAAGTCCGGTCCCTATAAGGGGACATCAGTGGAGACCTGCTCCGTGCTGTGAGN
 ACAAGAGGCACCATAACAAGNAAGCTCCAGTTGAGGTGCGACAGGCACTCGCCNAAGTCC
 NTGATGGCTTCGTCCAGTACTCACAAAACGGCTCCCCCGGCTGGTCTTTCACACGCACC
 GAGCCATGAGGAGCTGGCGCCTCTGAGAGCCTCTTCTGCCCTACTACCCGCCAGACTCA
 GAGGCCAGGAGGCATGCCCTGGGGCCACAGGGAGGTGAGGTGGGCTGGATGCCACACAG
 ATGGTCTCCGTGCTGGCTCACTGAAGAGCTGAGCCTGTGGCTGGCCTCAGAATCAGGCTG
 GGTGCAGTGGCTCACACCTGTAATCCCAGCATTTTGGGAGGCTGAGTGAGAGGATCACTT
 GAGCTCAGGAGTTTCGAGACCAGCCTGGCCAAACATGGCAACACCCCATTTCTACAAAAAAT
 TTGTAAATTTAGCCAGGCATGGTGGCGCACGCTGTAGTCCAGCTGCTTGGGAGGCTGA
 GGTGGGAGAATCACTTGAGCCCAGGAGTTTCGAGGCTGCAGTGAGCCAGGATCATGCCACT
 GCACTCCAGCCTGGTCCACAGAGAGACACTGTCAACCCCTTTCCCCCACAAGACTGGCAG
 AGGCTGGGCAGCCTGGGGCTGATGAAGCAGAGATGTTGCTGGATCCCAGGCCCTGGCAC
 CCCTCAGGAAATACAAGAAAAAGAATATTCACATCTGTTAATGTGCATAAAGCCAAGGA
 AAGGACAGTTCCGAATTCAAAAA AAAAAAAAAAAAAAAAAA

5

>BF446419

TTTTTTTTTTTTTTTTTAAATATTTAACTTATTTATTTAACAAGTAGAAGGGAATCCAT
 TGCTAGCTTTTCTGTGTTGGTGTCTAATATTTGGGTAGGGTGGGGGATCCCCAACATCA
 GGTCCCCTGAGATAGCTGGTCATTGGGCTGATCATTTGCCAGAATCTTCTTCTCCTGGGGT
 CTGGCCCCCAAAATGCCTAACCCAGGACCTTGGGAATTCTACTCATCCAAATGATAAT
 TCCAAATGCTGTTACCCAAGGTTAGGGTGTGTAAGGAAGGTAGAGGGTGGGGCTTCAGGT
 CTCACGGCTTCCCTAACCAACCCCTCTTCTCTTGGCCCAGCCTGGTTCCCCCACTTCCA
 CTCCCTCTACTCTCTCTAGGACTGGGCTGATGAAGGCACTGCCAAAATTTCCCTTACC
 CCCAACTTTCCCTTACCCCAACTTTCCCAACAGCTCCACAACCCCTGTTTGGAGCTACT
 GCAGGACCAGAAGCACAAAGTGCGGTTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGT
 ATATCTGTGCTTGGGGAATCTCACACAGAACTCAGGAGCACCCCTGCCTGAGCTAAGG
 GAGGTCTTATCTCTCAGGGGGGGTTTAAGTGCCGTTTGCAATAATGTCGTCTTATTTATT
 TAGCGGGGTGAATATTTTATACTGTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACA
 AATTAAAGGCTTCTTATATGTTTAAAAA

10

>BC006819

GCCTTATAAAGCACCAAGAGGCTGCCAGTGGGACATTTCTCGGCCCTGCCAGCCCCCAG
 GAGGAAGGTGGGTCTGAATCTAGCACCATGACGGAAGTAGAGACAGCCATGGGCATGATC
 ATAGACGTCCTTTCCCGATATTGCGGCAGCGAGGGCAGCACGACGACCCCTGACCAAGGGG
 GAGCTCAAGGTGCTGATGGAGAAGGAGCTACCAGGCTTCCCTGCAGAGTGGAAGACAAAG
 GATGCCGTGGATAAATTGCTCAAGGACCTGGACGCCAATGGAGATGCCAGGTGGACTTC
 AGTGAGTTCATCGTGTTCGTGGCTGCAATCACGTCTGCCTGTCACAAGTACTTTGAGAAG
 GCAGGACTCAAATGATGCCCTGGAGATGTCACAGATTCCTGGCAGAGCCATGGTCCCAGG
 CTTCCCAAAGTGTTTGTGGCAATTATTCCTAGGCTGAGCCTGCTCATGTACCTCTGA
 TTAATAAATGCTTATGAAATGAAAAA AAAAAAAAAA

>AA765597

CCAGCAAAGTCTCTTTTGACCACACGCTTTATCCGAGATGCTTAGAAGTATATTTGGCTG
TTTTATTTGCATCTTTGATTAAGATGTCTATCATTGTAAAAAGGTATTCAAAACAAAAGT
GTACTCTTTTATTATTATGAATCACATTGTACTGAGCTGTGAAGTCAGTGTTTTAAAAAT
GTAGAGTTTATTCATGGAGCATGCCATTGAGGTTTGGATGGTGGCAGGTAAAAACAGAAAG
GCAAGATGTCATCTGACATTAGGCTACTTATAAATAAATGTTTATCTAGCTTTTATTTCA
TGCCCTAATGAATAAAACATGCTTCGAAAAAGAAAGTAAAAAAAAAAAAACAAA

>X78202

GGCGAGAGAGACGCTCCCGCTCGCCGCCAGCTCTGATTGGCCCAGCGGTAGGAAAGGTTA
AACCAAAAATTTTTTTACAGCCCTAGTGTGCGCCTGTAGCTCGGAAAATTAATTGTGGCT
ATAGCCGCCTCGATCGCTGTCTCCCCAGCCTCGCCGCGGACGCTCCGGGACGCGCCCGCC
CGCCGCCCGGTTCTCCCCCCTTTGGGCTGGTGTCTGCTGCTGTGACTGCTGCTGCCA
AAGGAGGAGGAGGAGGAGGAAGCAGCGGGGGGGGGAGCGGTGGGTGTGGGGGAAACCAAG
AGTACAGTGGACGAGGACTCACCCCGCGTGGTGTCTTTTTTCTTCTTTCTTTCTTTTC
CTTTTTTTTTTTTTTTCTAATTCCTGAGGGGTGGTTGCTGCTTTTGCTACATGACTTGC
CAGCGCCCGAGCCTGCGGTCCAACCTGCGCTGCTGCCGAGCGCTCAGTGCCCGCCGCTGCC
GCCCCTGCCCCCGCGCCCCGTTCCGGCACCCACCGTCCCGCCCCCGCCCGCGCCGCT
GTCCCGCTCCCGCGCCCGCCGCCGTTTCCCCCGACGACTGGGTGATGCTGGACATGG

5

GAGATAGGAAAGAGGTGAAAATGATCCCCAAGTCCTCGTTTCAGCATCAACAGCCTGGTGC
 CCGAGGGCCTCCAGAACGACAACCACCACGCGAGCCACGGCCACCACAACAGCCACCACC
 CCCAGCACCACCACCACCACCACCACCATCACCACCACCCGCGCCGCCCCGCCCCGCAAC
 CGCCGCCGCCGCCGAGCAGCAGCAGCCGCCGCCGCCGAGACGCGGGGCCCCGGCGCC
 GACGACGACGAGGCCCCAGCAGTTGTTGTTCCGCCGCGCACGCACACGGCGGCTGAGG
 GCCAACGGCAGCTGGCGCAAGGCGACCGCGCGCGCGGGGATCTGCCCCGTGCGGCCGG
 ACGAGAAGGAGAAGGCCCGCGCCGGGGGGGAGGAGAAGAAGGGGGCGGGCGAGGGCGGCA
 AGGACGGGGAGGGGGGCAAGGAGGGCGAGAAGAAGAACGGCAAGTACGAGAAGCCGCCGT
 TCAGCTACAACGCGCTCATCATGATGGCCATGCGGCAGAGCCCCGAGAAGCGGCTCACGC
 TCAACGGCATCTACGAGTTCATCATGAAGAACTTCCCTTACTACCGCGAGAACAAGCAGG
 GCTGGCAGAACTCCATCCGCCACAATCTGTCCCTCAACAAGTGCTTCGTGAAGGTGCCGC
 GCCACTACGACGACCCGGGCAAGGGCAACTACTGGATGCTGGACCCGTGAGCGACGACG
 TGTTCAATCGGCGGCACCACGGGCAAGCTGCGGCGCTCCACCACCTCGCCGGCCAAGCCGG
 CCTTCAAGCGGTGCCGCGCTCACCTCCACCGGCTCACCTTCATGGACGCGCCGGCTC
 CCTTACTGCGCCCTTCTGTCCCTGCACCAACCCCGCCAGCAGCACTTTGA
 GTTACAACGGGACCACGTGCGCCTACCCAGCCACCCCATGCCCTACAGCTCCGTGTTGA
 CTCAAAACCTCGCTGGGCAACAACCACTCCCTCCCTCCACCGCCAACGGGCTGAGCGTGGACC
 GGCTGGTCAACGGGGGAATCCCGTACGCCACGCACCACCTCACGGCCGCCGCGCTAACCG
 CCTCGGTGCCCTGCGGCCTGCTGGTGCCCTGCTCTGGGACCTACTCCCTCAACCCCTGCT
 CCGTCAACCTGCTCGCGGGCCAGACCAGTTACTTTTTCCCCACGTCCCGCACCCGTCAA
 TGACTTCGAGAGCAGCAGTCCATGAGCGCCAGGGCCGCGTCTCTCTCCACGTGCGCCGG
 CAGGCCCCCTCGACCCCTGCCCTGTGAGTCTTTAAGACCCTCTTTGCCAAGTTTACGA
 CGGGACTGTCTGGGGGACTGTCTGATTATTTACACATCAAAATCAGGGGTCTTCTTCCA
 ACCCTTTAATACATTAAACATCCCTGGGACCAGACTGTAAGTGAACGTTTTACACACATTT
 GCATTGTAAATGATAATTAAAAAATAAGTCCAGGTATTTTTTTATTAAGCCCCCCCCCTCC
 CATTTCTGTACGTTTGTTCAGTCTCTAGGGTTGTTTATTATTCTAACAAGGTGTGGAGTG
 TCAGCGAGGTGCAATGTGGGGAGAATACATTGTAGAAATATAAGGTTTGGAAAGTCAAAATTA
 TAGTAGAATGTGTATCTAAATAGTGACTGCTTTGCCATTTTATTCAAACCTGACAAAGTCT
 ATCTCTAAGAGCCGCCAGATTTCCATGTGTGCAGTATTATAAGTTATCATGGAACATAT
 GGTGGACGCAGACCTTGAGAAACAACCTAAATTATGGGGAGAATTTTAAATGTTAAACTG
 TAATTTGTATTTAAAAAGCATTTCGTAGTAAAGGTGCCCAAGAAATTATTTTGGCCATTTA
 TTGTTTTCTCCTTTTCTTTAAAGAACTGTTTTTTTTCTTTTGTCTTACTTTTAGACCAA
 GATTGGGCGGTTCTAGAAAATGCGCCTTGGTATACTAAGTATTAAAAACAAACAAAAAGGA
 AAGTTGTTTCAGTTAACGCTGCCCATTCATTTGAATCAGAAGGGGACAAAATTAACGATT
 GCCTTCAGTTTGTGTTGTGTATATTTTGATGTATGTGGTCACTAACAGGTCACTTTTATT
 TTTTCTAAATGTAGTGAAATGTTAATACCTATTGTACTTATAGGTAAACCTTGCAAATAT
 GTAACCTGTGTTGCGCAAATGCCGCATAAAATTTGAGTGATTGTTAATGTTGTCTTAAAT
 TTCTTGATTGTGACTATGTGGTCATATGCCCGTGTTTGTCACTTACAAAAATGTTTACTA
 TGAACACACATAAATAAAAAATAG

>AK026790

GGCACGAGGCTGCCTGCCCCCGGGTGGGGCTGCGGCTCTGGCCTCCAGGCCCATCCTC
AACAGCTACCCAGCCAAACACCAAGGCCACAAGGGGACCCGGCCTAGGAGGCAGGAAGC
CAAGGTACAGAGAGCAGCCTGGCCCTCACCAGTGCAGCAAGCTGGGGCAGCAAGGCTGACA
GTTGCTGCATGCCCAGGGCAGGGTGTGGTACTGGCACCACAAGTTCAGCATGGCAGAGCTG
GCCAACAGCTTGTCCCCGATCTGCCTCCAGCCCCAAGATGCCTACAGCCCCCAGGCCCTT
TCGGCAGCACTGCCTCTGCCCACCTGCCCTTTAAGAGACTCCAGGGCTGCTCCTGTCTATGC
AGCGAAGGTTTTGTCTGTTTCAAAGTTCGAGACTCAACTTGAGGGACTGTTTTTGACAAT
CCCCGCTGACCTCCGCTCCTCGTGGCGCCCTGGCCCTACACCCAGCCTGGCCCAGGGCCG
GCTTTGCCTGGTGAGGCTGGAGGGAGCACCAGGACCTGCTGTCTGCTGTCTCAGCCCCCTCCT
GGTGTCTGGTGCCCTGATGCTGTGCCCTTGTACCCATTGAGCTGCAAGAGGGACCAAGAGG
GGGCCACGCAGCCAGCCAGATGCCCTGGCCCTGTGCTGGGGCAGACAACGCTGCAGAGCCC
AGGGAGCCTGGCGCTAGGACGTGCGTCCTTGTGACACTGGCCTGTCTGAACTCACCTGGC
CTGGGAAGCACCGTCTGCCCGGGCCCAAGCCCTGCCCTCCAGAGTCCAGAGCCAGGAAG
GGGCTGCTGAGGGCGAGCATCCTGCTGGGCTCTCTGCCCGGGCCACCCCTCCAAGGGGCT
GGCCTGTGAGCCTTGACTGGGATTCATGATGTGGAGGCCCCCAACTTCCAGAAGCAGCTG
GTACTCTGCTCACACAAGCGACTGGGCCGGCCGGCCCTGGACCCCTAGACCCCGAGCCGC
CTGCCGACTGCCTGCACAGGGAGAGCAGTTGAGGCCCGGGCAGGGCCCCCACACCAGACC
CCAACATAGCTTCCCCACCCAGGCACCCCTCCCGGGGCAGCAGGCGTGGGAGTCAGGGC
TGCATGCTCCTCCCTCCCACCTCACAGCGGCCCTTAGGCAAGTCATTTTCTGTCTATCAC
AAGGTCGCCTCTGCCTAGTCAGGTCTGGCGTCCAGAGTAAGGATGTGCGGCCCCAGGC
CCCCGCACACCTCCCTCAGCACCAAGACCGGGACCCCCCACCACGTGTCTCATTGTGG
CTGCCATATGGAATCCCCGGCCCTTGTGTGCAGGCCAGGCCCTTCCACTGATTTTTTAAAGT
GAACCATTTGCTGGATCTCAGATCTGTGGCATCTAAGGCCTAGCAGGGGTGGGCACACGG
GTCAACCCGAGGCCCATACCAAGACTCTGTTCCTGCCCTAGGCCAGTCTCAAAGGAAGCC
ACAAGGCGCGGGGGCCACTGAGGAAGGAAATGTTCATTTTCAATTTGTCCAAAACACCTT
AAGTTTTAAGTATATTAATCTTGATGCTTTTTTAATATTGCTTTTTAACTTGCTGAGATT
TAGAAATACTGTTATAAAAACTTTTTTAATTTCTGTATTTTTTTTCTGTATTGTATCTTC
ATGGGACATTAGGGGTTTTCTATGGTAAGCACACCTATGGTTTTGGTAAAAACATTATCA
AATATATATCCAGACGGTTCTTCCCTAGAAGAAAAACAAGTCTTTACACCTGATAAAATA
TTTTGCGAAGAGAGGTGTTCTTTTTCTTACTGGTGCTGAAAGGAAGGATGGATAACGAG
GAGAAAATAAACTGTGAGGCTCAAAAAAAAAAAAAAAAAA

>R45389

CCTGCCCTTCTCTATATGTACCATCTCCAAAAACCATGTACATCTCCAAAACTGGAGTA
GAAAGTTAGATTGCTCAACTACAACCTCTTAGAACTCTATAGCTCTGACATACAGATTC
ACACTCTCCTCTATTTGCTAAGTATGTAAAGAATGTTTCTTTTAAATGTTCTCTTTTG
AGAACAACCTGCTTATTTGTTATAAAAGCATTTGGTTAAATGATGTCATCATAAAGAACA
GTGGCTTTGTTTCAATACATATTTTTGAGATGATTATCTAGAAGCCAGATTAATAAAATC
AGCTTGTGACCTTGCTAAGCATATAAACTGGAAATTCAGATACATTCAAATTTATGGGTT
CATTTAAAAGTGTTCTACCTTTTGGGTATGAGACTAATATCACTAATTCCTCAATAGTTA

TCATGGCTCTATCTTAATTAATTAGAAAAATATGTGTGTTTAATTCTTTGAGAATTAAAAAT
AGAGAATATTAACAGAGGGTTAAAACTGCTTCAACTCCAATAAGATAAAGGAAGCTCAA
AATCTATGAGCTGAGTGTTCAATTAGCTTTGCCCTACTGAGTTCAATTTTATGTCAATACA
ACAGTGGATCAGACAGTACGACTTTGAACTGGTGAATGTAAACAATGTTTTTACCTAA
GCTGCTTTTGAAGAAGTATGCTTGCTGCTAACTAAAGTTTTGGATGTATCGATTTAGAG
AACCAATTAATACCTGCAAAATAAAGCATACTGTGGTACTTCTGTTTGATCTAGTATGTG
TGATTTTAGATTGATGGATTAAAAATTAATAAAGATCATACATTCCATACCAAAAAAAAAA
AAAAAAA

5

>BC006811

CCAGAAGCCTGCATTTCTGCATTCTGCTTAATTCCTTTCTTAGATTTGAAAGAAGCCA
 AACTAAACCACAAATATACAACAAGGCCATTTTCTCAAACGAGAGTCAGCCTTTAACGA
 AATGACCATGGTTGACACAGAGATGCCATTCTGGCCCACCAACTTTGGGATCAGCTCCGT
 GGATCTCTCCGTAATGGAAGACCACTCCCACCTCCTTTGATATCAAGCCCTTCACTACTGT
 TGA CTTCCTCCAGCATTTCTACTCCACATTACGAAGACATTCCATTACAGAAGACAGATCC
 AGTGGTTGCAGATTACAAGTATGACCTGAAACTTCAAGAGTACCAAGTGCAATCAAAGT
 GGAGCCTGCATCTCCACCTTATTATTCTGAGAAGACTCAGCTCTACAATAAGCCTCATGA
 AGAGCCTTCCAACCTCCCTCATGGCAATTGAATGTGCTGTCTGTGGAGATAAAGCTTCTGG
 ATTTCACTATGGAGTTCATGCTTGTGAAGGATGCAAGGGTTTCTTCCGGAGAACAATCAG
 ATTGAAGCTTATCTATGACAGATGTGATCTTAACTGTGCGATCCACAAAAAAGTAGAAA
 TAAATGTCAGTACTGTGCTTTCAGAAATGCCTTGAGTGGGGATGTCTCATAATGCCAT
 CAGGTTTGGGCGGATGCCACAGGCCGAGAAGGAGAAGCTGTTGGCGGAGATCTCCAGTGA
 TATCGACCAGCTGAATCCAGAGTCCGCTGACCTCCGGGCCCTGGCAAAACATTTGTATGA
 CTCATACATAAAGTCTTCCCGCTGACCAAAGCAAAGGCGAGGGCGATCTTGACAGGAAA
 GACAACAGACAAATCACCATTCTGTTATCTATGACATGAATTCTTAAATGATGGGAGAAGA
 TAAAATCAAGTTCAAACACATCACCCCCCTGCAGGAGCAGAGCAAAGAGGTGGCCATCCG
 CATCTTTTCAAGGCTGCCAGTTTCTGCTCCGTGGAGGCTGTGCAGGAGATCACAGAGTATGC
 CAAAAGCATTCTCGTGTGTAATCTTGACTTGAACGACCAAGTAACTCTCCTCAAATA
 TGGAGTCCACGAGATCATTTACACAATGCTGGCCTCCTTGATGAATAAAGATGGGGTTCT
 CATATCCGAGGGCCAAGGCTTCATGACAAGGGAGTTTCTAAAGAGCCTGCGAAAGCCTTT
 TGGTGACTTTATGGAGCCCAAGTTTGAGTTTGTCTGTGAAGTTCAATGCACTGGAATTAGA
 TGACAGCGACTTGGCAATATTTATTGCTGTCTATTATTCTCAGTGGAGACCGCCAGGTTT
 GCTGAAGTGTGAAGCCCATTTGAAGACATTCAAGACAACCTGCTACAAGCCCTGGAGCTCCA
 GCTGAAGCTGAACCAACCTGAGTCCCTCAGAGCTGTTTGCCAAGCTGCTCCAGAAAATGAC
 AGACCTCAGACAGATTGTACGGAACACGCTGCAGCTACTGCAGGTGATCAAGAAGACGGA
 GACAGACATGAGTCTTCAACCGCTCCTGCAGGAGATCTACAAGGACTTGTACTAGCAGAG
 AGTCTTGAGCCACTGCCAACATTTCCCTTCTTCCAGTTGCACTATTCTGAGGGAAAATCT
 GACACCTAAGAAATTTACTGTGAAAAAGCATTTTAAAAAGAAAAGGTTTTAGAATATGAT
 CTATTTTATGCATATTGTTTATAAAGACACATTTACAATTTACTTTTAATATTTAAAAATT
 ACCATATTATGAAAAAAAAAAAAAAAAA

>X05615

GCAGTGGTTTTCTCCTCCTTCCCTCCCAGGAAGGGCCAGGAAAATGGCCCTGGTCTCGGAGA
 TCTTCAACCTGTGCTGGCCTCCATCTGCTGGGTGTGCGCCAATATCTTCGAGTACCAGGTTG
 ATGCCAGCCCCCTTCGTCCCTGTGAGCTGCAGAGGGAAACGGCCTTTCTGAAGCAAGCAG
 ACTACGTGCCCCAGTGTGCAGAGGATGGCAGCTTCCAGACTGTCCAGTGCCAGAACGACG
 GCCGCTCCTGCTGGTGTGTGGGTGCCAACGGCAGTGAAGTGCTGGGCAGCAGGCAGCCAG
 GACGGCCTGTGGCTGTCTGTCTATTTTGTGAGCTACAGAAACAGCAGATCTTACTGAGTG
 GCTACATTAACAGCACAGACACCTCCTACCTCCCTCAGTGTGAGGATTCAGGGGACTACG
 CGCCTGTTTCAAGTGTGATGTGCAGCATGTCCAGTGTGCTGGTGTGTGGACGCAGAGGGGATGG
 AGGTGTATGGGACCCGCCAGCTGGGGAGGCCAAAGCGATGTCCAAGGAGCTGTGAAATAA
 GAAATCGTCTGCTTCTCCACGGGGTGGGAGATAAGTCACCACCCAGTGTCTGCGGAGG
 GAGAGTTTATGCCTGTCCAGTGCAAATTTGTCAACACCACAGACATGATGATTTTGTATC
 TGGTCCACAGCTACAACAGGTTTCCAGATGCATTTGTGACCTTCAGTTCTTCCAGAGGA
 GGTTCCTGAGGTATCTGGGTATTTGCCACTGTGCTGACAGCCAAGGGCGGGAACCTGGCTG
 AGACAGGTTTGGAGTTGTTACTGGATGAAATTTATGACACCATTTTGTGCTGGCCTGGACC
 TTCCTTCCACCTTCACTGAAACCACCTGTACCGGATACTGCAGAGACGGTTCTCTCGCAG
 TTCAATCAGTCATCTCTGGCAGATTCCGATGCCCCACAAATGTGAAGTGGAGCGGTTTA
 CAGCAACCAGCTTTGGTCACCCCTATGTTCCAAGCTGCCGCCGAAATGGCGACTATCAGG
 CGGTGCAGTGCCAGACGGAAGGGCCCTGCTGGTGTGTGGACGCCAGGGGAAGGAAATGC
 ATGGAACCCGGCAGCAAGGGGAGCCGCATCTTGTGCTGAAGGCCAATCTTGTGCTCCTCG

AAAGGCAGCAGGCCTTGTCCAGACTCTACTTTGGGACCTCAGGCTACTTCAGCCAGCACG
 ACCTGTTCTCTTCCCCAGAGAAAAGATGGGCCTCTCCAAGAGTAGCCAGATTGCCCACAT
 CCTGCCACCCACGATCAAGGAGCTCTTTGTGGACTCTGGGCTTCTCCGCCCAATGGTGG
 AGGGACAGAGCCAAACAGTTTTCTGTCTCAGAAAATCTTCTCAAAGAAGCCATCCGAGCAA
 TTTTTCCTTCCCGAGGGCTGGCTCGTCTTGCCCTCAGTTTACCACCAACCCAAAGAGAC
 TCCAGCAAAACCTTTTGGAGGGAAATTTTGGTGAATGTTGGCCAGTTTAACTTGTCTG
 GAGCCCTTGGCACAAGAGGCACATTTAAGTTTCAAGTCAATTTTCCAGCAACTTGGTCTTG
 CAAGCTTCTTGAATGGAGGGAGACAAGAAGATTTGGCCAAAGCCACTCTCTGTGGGATTAG
 ATTCAAATTCTTCCACAGGAACCCCTGAAGCTGCTAAGAAGGATGGTACTATGAATAAGC
 CAACTGTGGGCAGCTTTGGCTTTGAAATTAACCTACAAGAGAACCAAAATGCCCTCAAAT
 TCCTTGCTTCTCTCTGGAGCTTCCAGAATTCCTTCTCTTCTTGCAACATGCTATCTCTG
 TGCCAGAAGATGTGGCAAGAGATTTAGGTGATGTGATGGAAACGGTACTCGACTCCCAGA
 CCTGTGAGCAGACACCTGAAAGGCTATTTGTCCCATCATGCACGACAGAAGGAAGCTATG
 AGGATGTCCAATGCTTTTCCGAGAGTGTGGTGTGTGAATTCTTGGGGCAAAGAGCTTC
 CAGGCTCAAGAGTCAGAGATGGACAGCCAGGTGCCCCACAGACTGTGAAAAGCAAAGGG
 CTGCGTCAGCAAGCCTCATGGGCAGCCAGCCTGCTGGCTCCACCTTGTTTGTCCCTGCTT
 GTACTAGTGAGGGACATTTCTGCCTGTCCAGTGTCTCAACTCAGAGTGTACTGTGTG
 ATGCTGAGGGTCAGGCCATTCTGGAAGTCAAGTGAATAGGGAAAGCCCAAGAAATGCC
 CCACGCCCTGTCAATTACAGTCTGAGCAAGCTTTCTCAGGACGGTGCAGGCCCTGCTCT
 CTAAGTCCAGCATGCTACCCACCTTTCCGACACCTACATCCACAGTGCAGCACCGATG
 GGCAGTGGAGACAAGTGCAATGCAATGGGCCTCCTGAGCAGGTCTTCGAGTTGTACCAAC
 GATGGGAGGCTCAGAACAGGGCCAGGATCTGACGCTTCCAAAGCTGCTAGTGAAGATCA
 TGAGCTACAGAGAAGCAGCTTCCGGAAGCTTCAGTCTCTTTATTCAAAGTCTGTATGAGG
 CTGGCCAGCAAGATGTCTTCCCGTGTGTGTGCAATACCTTCTCTGCAAGATGTCCAC
 TAGCAGCACTGGAAGGGAAACGGCCCCAGCCAGGGAGAATATCCTCCTGGAGCCCTACC
 TCTTCTGGCAGATCTTAAATGGCCAAGTCAAGCAATACCCGGGGTCTACTCAGACTTCA
 GCACTCCTTTGGCACATTTTGATCTTCGGAAGTGTGTGTGTGATGAGGCTGGCCAAAG
 AACTGGAAGGAATGCGGTCTGAGCCAAGCAAGCTCCCAACGTGTCTTGGCTCCTGTGAGG
 AAGCAAAGCTCCGTGTACTGCAGTTCATTAGGGAAACGGAAGAGATTGTTTCAGCTTCCA
 ACAGTTCTCGGTTCCTCTGGGGGAGAGTTTCTGTTGGCCAAGGGAAATCCGGCTGAGGA
 ATGAGGACCTCGGCCTTCTCCTCGCTCTTCCCGCCCCGGGAGGCTTTCGCGGAGTTTCTGC
 GTGGGAGTGATTACGCCATTTCGCTTGGCGGCTCAGTCTACCTTAAGCTTCTATCAGAGAC
 GCGCTTTTCCCGGACGACTCGGCTGGAGCATCCGCCCTTCTGCGGTTCGGGCCCCCTACA
 TGCCACAGTGTGATGCGTTTGGAAAGTGGGAGCCTGTGCAAGTGCACGCTGGGACTGGGC
 ACTGCTGGTGTGTAGATGAGAAAGGAGGTTTCATCCCTGGCTCACTGACTGCCCGCTCTC
 TGCAGATTCCACAGTGGCCGACAACCTGCGAGAAATCTCGAACCAGTGGGCTGCTTTCCA
 GTTGGAAACAGGCTAGATCCCAAGAAAACCCATCTCCAAAAGACCTGTTTCTGCCAGCCT
 GCCTAGAAACAGGAGAATATGCCAGGCTGCAGGCATCGGGGGCTGGCACCTGGTGTGTGG
 ACCCTGCATCAGGAGAAGAGTTGCGGCCTGGCTCGAGCAGCAGTGCAGTGCCTCAAGCC
 TCTGCAATGTGCTCAAGAGTGGAGTCTCTCTAGGAGAGTCAAGCCAGGCTATGTCCCAG
 CCTGCAGGGCAGAGGATGGGGGCTTTTCCCAAGTGCATGTGACCAGGCCCCAGGGCAGCT
 GCTGGTGTGTCATGGACAGCGGAGAAGAGGTGCTTGGGACCGCGGTGACCGGGGGCCAGC
 CCGCTGTGAGAGCCCGCGGTGTCCGCTGCCATTCAACGCGTTCGGAGGTGGTTGGTGGAA
 CAATCCTGTGTGAGACAATCTCGGGCCCCACAGGCTCTGCCATGCAGCAGTGCCAATTGC
 GTGCGCCGCAAGGCTCTGGAGCGTGTTCACACAGGGCCATTGATATGAGCCTGGAGA
 TCGGAGCTTGGGAGTCAAGCTGCTCAGCCCCGGGCTGCCAACGGCCCCAGCTGTGGC
 AGACCATCCAGACCCAAGGGCACTTTTCAAGTCTCCAGCTCCCCGCGGGCAAGATGTGCAGTG
 CTGACTACGCGGGTTTGTGTCAGACTTTCCAGGTTTTCATATTGGATGAGCTGACAGCCC
 GCGGCTTCTGCCAGATCCAGGTGAAGACTTTTGGCACCCCTGGTTTCCATTCTGTCTGCA
 ACAACTCCTCTGTGCAAGTGGGTGTCTGACCAGGGAGCGTTTAGGAGTGAATGTTACAT
 GGAAATCACGGCTTGAGGACATCCAGTGGCTTCTCTTCTGACTTACATGACATTGAGA
 GAGCCTTGGTGGGCAAGGATCTCCTTGGGCGCTTCAAGATCTGATCCAGAGTGGCTCAT
 FCCAGCTTCTGACTCCAAGAGCTTCCAGCGGAAACCATCCGCTTCTCCAAGGGG
 ACCACTTTGGCACCTCTCTTAGGACACGGTTTGGGTGCTCGGAAGGATTCTACCAAGTCT
 TGACAAGTGAAGCCAGTCAAGACGAGTGGGATGCGTTAAGTGCCATGAAGGAAGCTATT
 CCCAAGATGAGGAATGCATTCTTGTCTGTGGATTCTACCAAGAACAGGCAAGGAGCT
 TGGCCTGTGTCCCATGTCTGTGGGCAGAACGACCATTTCTGCGGAGCTTTTCAAGCAGA
 CTCACTGTGTCACTGACTGTGAGGGAACGAAGCAGGCTGCAATGTGACCAGAATGGCC
 AGTATCGAGCCAGCCAGAAGGACAGGGGAGTGGGAAGGCTTCTGTGTGACGCGGAGG
 GCGGAGGCTGCCATGGTGGGAACAGAGGCCCTCTTGAAGTCAAGTGTGTTGATGA
 TGCAGAAGTTTGAAGGTTCCAGAATCAAAGGTGATCTTCGACGCCAATGCTCCTGTGG

CTGTCAGATCCAAAGTTCTTGATTCTGAGTTCCCCGTGATGCAGTGCTTGACAGATTGCA
 CAGAGGACGAGGCCTGCAGCTTCTTCACCGTGTCCACGACGGAGCCAGAGATTCTCTGTG
 ATTTCTATGCTTGGACAAGTGACAATGTTGCCTGCATGACTTCTGACCAGAAACGAGATG
 CACTGGGGAACCTCAAAGGCCACCAGCTTTGGAAAGTCTTCGCTGCCAGGTGAAAGTGAGGA
 GCCATGGTCAAGATTCTCCAGCTGTGTATTTGAAAAAGGGCCAAGGATCCACCACAACAC
 TTCAGAAACGCTTTGAACCCACTGGTTTCCAAAACATGCTTTCTGGATTGTACAACCCCA
 TTGTGTTTCTCAGCCTCAGGAGCCAATCTAACCGATGCTCACCTCTTCTGTCTTCTTGCAT
 GCGACCGTGATCTGTGTTGCGATGGCTTCGTCTCACACAGGTTCAAGGAGGTGCCATCA
 TCTGTGGGTTGCTGAGCTCACCCAGTGTCTGTCTTTGTAATGTCAAAGACTGGATGGATC
 CCTCTGAAGCCTGGGCTAATGCTACATGTCTCTGGTGTGACATATGACCAGGAGAGCCACC
 AGGTGATATTGCGTCTTGGAGACCAGGAGTTTCATCAAGAGTCTGACACCCTTAGAAGGAA
 CTCAGACACCTTTACCAATTTTCAGCAGGTTTATCTCTGGAAGATTCTGACATGGGGT
 CTCGGCTGAGTCTATGGGATGTAGAAAAAACACAGTGCCAGGCCAGCATCTCCAACAG
 AAGCAGGTTTGAACAACAGAACTTTTCTCCCTGTGGACCTCAACCAGGTCAATTGTCAATG
 GAAATCAATCACTATCCAGCCAGAAGCACTGGCTTTTCAAGCACCTGTTTTCAGCCAGC
 AGGCAAACTATGGTGCCTTTCTCGTTGTGTGCAGGAGCACTCTTCTGTGAGCTCGCAG
 AGATAACAGAGAGTGCATCCTTGTACTTCACCTGCACCCTCTACCCAGAGGCACAGGTGT
 GTGATGACATCATGGAGTCCAATACCCAGGGCTGCAGACTGATCCTGCCTCAGATGCCAA
 AGGCCCTGTTCGGAAGAAAGTTATACTGGAAGATAAAGTGAAGAACTTTTACACTCGCC
 TGCCGTTCCAAAAACTGATGGGGATATCCATTAGAAATAAAGTGCCCATGTCTGAAAAAT
 CTATTTCTAATGGGTTCTTTGAATGTGAACGACGGTGCATGCGGACCCATGCTGCACTG
 GCTTTGGATTCTAAATGTTTCCAGTTAAAAGGAGGAGAGGTGACATGTCTCACTCTGA
 ACAGCTTGGGAATTAGATGTGCAGTGAGGAGAATGGAGGAGCCTGGCGCATTTTGGACT
 GTGGCTCTCCTGACATTGAAGTCCACACCTATCCCTTCGGATGGTACCAGAAGCCATTG
 CTCAAAAATAATGCTCCCAGTTTTTGGCCCTTTGGTTGTTCTGCCTTCCCTCACAGAGAAAG
 TGTCTCTGGAATCGTGGCAGTCCCTGGCCCTCTCTTCAGTGGTTGTTGATCCATCCATTA
 GGCACTTTGATGTTGCCCATGTGAGCACTGCTGCCACCAGCAATTTCTCTGTGTCCGAG
 ACCTCTGTTTGTGGAATGTTCCCAACATGAGGCCTGTCTCATCACCCTCTGCAAACCC
 AACTCGGGCTGTGAGATGTATGTTCTATGTGATACTCAAAGCTGCACACATAGTCTGC
 AGGGTCGGAACCTGCCGACTTCTGCTTCGTGAAGAGGCCACCCACATCTACCGGAAGCCAG
 GAATCTCTCTGCTCAGCTATGAGCATCTGTACCTTCTGTGCCCATTTCCACCCATGGCC
 GGCTGCTGGGCAGGTCCCAGGCCATCCAGGTGGGTACCTCATGGAAGCAAGTGGACCAGT
 TCCTTGGAGTTCCATATGCTGCCCCCGCCCTGGCAGAGAGGCACTTCCAGGCACAGAGC
 CCTTGAACCTGGACAGGCTCCTGGGATGCCAGCAAGCCAAGGGCCAGCTGCTGGCAGCCAG
 GCACCAGAACATCCACGTCTCCTGGAGTCAGTGAAGATTGTTTGTATCTCAATGTGTTCA
 TCCCTCAGAATGTGGCCCTAACGCGTCTGTGCTGGTGTCTTCCACAACACCATGGACA
 GGGAGGAGAGTGAAGGATGGCCGGCTATCGACGGCTCCTTCTTGGCTGCTGTTGGCAACC
 TCATCGTGTCACTGCCAGCTACCGAGTGGGTGTCTTCGGCTTCTGTAGTTCTGGATCCG
 GAGAGGTGAGTGGCAACTGGGGGCTGCTGGACCAGGTGGCGGCTCTGACCTGGGTGCAGA
 CCCACATCCGAGGATTTGGCGGGACCTCGGCGCGTGTCCCTGGCAGCAGACCGTGGCG
 GGGCTGATGTGGCCAGCATCCACCTTCTCACGGCCAGGGCCACCACTCCCACTTTTCC
 GGAGAGCTGTGCTGATGGGAGGCTCCGCACTCTCCCGGCCCGCTCATCAGCCATGAGA
 GGGCTCAGCAGCAGCAATTGCTTTGGCAAGGAGGTGAGTGGCCCATGTCTATCCAGCC
 AAGAAGTGGTGTCTGCTCCGCCAGAAGCCTGCCAATGTCTCAATGATGCCAGACCA
 AGCTCCTGGCCGTGAGTGGCCCTTTCCACTACTGGGGTCTGTGATCGATGGCCACTTCC
 TCCGTGAGCCTCCAGCCAGAGCACTGAAGAGGTCTTTATGGGTAGAGGTGATCTGCTCA
 TTGGGAGTTCTCAGGACGACGGGCTCATCAACAGAGCAAGGCTGTGAAGCAATTTGAGG
 AAAGTCCAGGCCGGACAGTAGCAAAACAGCCTTTTACCAGGCACTGCAGAAATCTCTGG
 GTGGCGAGGACTCAGATGCCCGCGTTCGAGGCTGCTGCTACATGGTATTACTCTCTGGAGC
 ACTCCACGGATGACTATGCCTCCTTCTCCCGGGCTCTGGAGAATGCCACCCGGGACTACT
 TTATCATCTGCCCTATAATCGACATGGCCAGTGCCCTGGGCAAAGAGGGCCCGAGGAAACG
 TCTTCATGTACCATGCTCCTGAAAACCTACGGCCATGGCAGCCTGGAGCTGTGGCGGATG
 TTCAGTTTGGCTTGGGGCTTCCCTTCTACCCAGCCTACGAGGGGCAAGTTTCTCTGGAGG
 AGAAGAGCCTGTGCTGAAAATCATGCAGTACTTTTCCCACTTCATCAGATCAGGAAATC
 CCAACTACCCTTATGAGTTCTCACGGAAAGTACCCACATTTGCAACCCCTGGCCTGACT
 TTGTACCCCGTGTGCTGGTGGAGAGAACTACAAGGAGTTCAAGTGTGCTCCCCAATCGAC
 AGGGCCTGAAGAAAGCCGACTGCTCCTTCTGGTCCAAGTACATCTCGTCTCTGAAGACAT
 CTGCAGATGGAGCCAAGGGCGGGCAGTCAGCAGAGAGTGAAGAGGAGGAGTTGACGGCTG
 GATCTGGGCTAAGAGAAGATCTCCTAAGCCTCCAGGAACCAGGCTCTAAGACCTACAGCA
 AGTGACCAGCCCTTGAGCTCCCCAAAACCTCACCCGAGGCTGCCCACTATGGTCACTCT
 TTTCTCTAAATAGTTACTTACCTTCAATAAAGTATCTACATGCGGTG

ES 2 550 652 T3

>X79676

AGATCTCTCCAGATCACACTGTACGTGTACCTAGCACATCTCGAGAACTCCTTTGGGGCC
 GTCTGGGGCCCGGGAAGGAAGCCTGAGTTCTCAAGATTCCAGGACTGAGAGTGCCAGCTT
 GTCTCAAAGCCAGGTCAATGGTTTCTTTGCCAGCCATTAGGTGACCAAACCTGGCAGGA
 ATCACAGCATGGCAGCCCTTCCCCATCTGTAATATCCAAAGCCACCGAGAAAGAGACTTT
 CACTGATAGTAACCAAGCAAAACTAAAAAGCCAGGCATTCTGATGTAACTGATTACTC
 AGACCGTGGAGATTTCAGACATGGATGAAGCCACTTACTCCAGCAGTCAGGATCATCAAAC
 ACCAAAACAGGAATCTTCTCTTCAGTGAATACATCCAACAAGATGAATTTTAAACCTTT
 TCCTTCATCACCTCCTAGGTCTGGAGATATCTTTGAGGTTGAACTGGCTAAAAATGATAA
 CAGCTTGGGGATAAGTGTACGGGAGGTGTGAATACGAGTGTACAGACATGGTGGCATTTA
 TGTGAAAGCTGTTATTCCCCAGGGAGCAGCAGAGTCTGATGGTAGAATTCACAAAGGTGA
 TCGCGTCTAGCTGTCAATGGAGTTAGTCTAGAAGGAGCCACCCATAAGCAAGCTGTGGA
 AACACTGAGAAATACAGGACAGGTGGTTTCATCTGTTATTAGAAAAGGGACAATCTCCAAC
 ATCTAAAGAACATGTCCCGTAACCCACAGTGTACCTTTTCAGATCAGAATGCCCAAGG
 TCAAGGCCCAGAAAAAGTGAAGAAAACAACCTCAGGTCAAAGACTACAGCTTTTGTCACTGA
 AGAAAATACATTTGAGGTAAAATTATTTAAAAATAGCTCAGGTCTAGGATTAGTTTTTTC
 TCGAGAAGATAATCTTATACCGGAGCAAATTAATGCCAGCATAGTAAGGGTTAAAAAGCT
 CTTTCTGGACAGCCAGCAGCAGAAAGTGGAAAAATTGATGTAGGAGATGTTATCTTGAA
 AGTGAATGGAGCCTCTTTGAAAGGACTATCTCAGCAGGAAGTCATATCTGCTCTCAGGGG
 AACTGCTCCAGAAGTATTCTTGCTTCTCTGCAGACCTCCACCTGGTGTGCTACCGGAAAT
 TGATACTGCGCTTTTGACCCCACTTCAGTCTCCAGCACAAGTACTTCCAAACAGCAGTAA
 AGACTCTTCTCAGCCATCATGTGTGGAGCAAAGCACCAGCTCAGATGAAATGAAATGTC
 AGACAAAAGCAAAAAACAGTGCAAGTCCCCATCCAGAAAAAGACAGTTACAGTGACAGCAG
 TGGGAGTGGAGAAGATGACTTAGTGACAGCTCCAGCAAACATATCAAATTCGACCTGGAG
 TTCAGCTTTGCATCAGACTCTAAGCAACATGGTATCACAGGCACAGAGTCATCATGAAGC
 ACCAAGAGTCAAGAAGATACCATTGTACCATGTTTTACTATCCTCAGGAAAAGGCCCAA
 TAAACCAGAGTTTGAGGACAGTAATCCTTCCCTCTACCACCGGATATGGCTCCTGGGCA
 GAGTTATCAACCCCAATCAGAATCTGCTTCTCTAGTTCGATGGATAAGTATCATATACA
 TCACATTTCTGAACCAACTAGACAAGAAAAGTGGACACCTTTGAAAAATGACTTGGAATA
 TCACCTTGAAGACTTTGAAGTGAAGTAGAAGTCCCTCATTACCCTAATTAAATCAGAAAA
 AGGAAGCCTGGGTTTTACAGTAACCAAAGGCAATCAGAGAATTGGTTGTTATGTTTCATGA
 TGTCATACAGGATCCAGCCAAAAGTGATGGAAGGCTAAAACCTGGGGACCGGCTCATAAA
 GGTTAATGATACAGATGTTACTAATATGACTCATACAGATGCAGTTAATCTGCTCCGGGG
 ATCCAAAACAGTCAGATTAGTTATTGGACGAGTTCTAGAATTACCCAGAATACCAATGTT
 GCCTCATTTGCTACCGGACATAACACTAACGTGCAACAAAGAGGAGTTGGGTTTTTCTCTT
 ATGTGGAGGTCTGACAGCCTTTATCAAGTGGTATATATATTAGTGATATTAAATCCAAGGTC
 CGTCGCAGCCATTGAGGGTAATCTCCAGCTATTAGATGTCATCCATTATGTGAACGGAGT
 CAGCACACAAGGAATGACCTTGGAGGAAGTTAACAGAGCATTAGACATGTCACCTTCCTTC
 ATTGGTATTGAAAGCAACAAGAAATGATCTTCCAGTGGTCCCAGCTCAAAGAGGTCTGC
 TGTTTTAGCTCCAAAGTCAACCAAAGGCAATGGTTCTACAGTGTGGGGTCTTGCAGCCA
 GCCTGCCCTCACTCCTAATGATTATTCTCCACGGTTGCTGGGGAAGAAATAAATGAAAT
 ATCGTACCCCAAAGGAAATGTTCTACTTATCAGATAAAGGGATCACCAAACCTTGACTCT
 GCCCAAAGAATCTTATATACAAGAAGATGACATTTATGATGATTTCCCAAGAAGCTGAAGT
 TATCCAGTCTCTGCTGGATGTTGTGGATGAGGAGTCCCAGAATCTTTTAAACGAAAATAA
 TGCAGCAGGATACTCCTGTGGTCCAGGTACATTAAAGATGAATGGGAAGTTATCAGAAGA
 GAGAACAGAAGATACAGACTGCGATGGTTACCTTTACCTGAGTATTTTACTGAGGCCAC
 CAAAATGAATGGCTGTGAAGAATATTGTGAAGAAAAAGTAAAAAGTGAAAGCTTAATTCA
 GAAGCCACAAGAAAAGAAGACTGATGATGATGAATAACATGGGGAAATGATGAGTTGCC
 AATAGAGAGAACAACCATGAAGATTCTGATAAAGATCATTCCTTTCTGACAAACGATGA
 GCTCGCTGTACTCCCTGTCTGTCAAAGTGCTTCCCTCTGGTAAATACACGGGCGCCAACCTT
 AAAATCAGTCATTTCGAGTCTGCGGGTTGCTAGATCAGGAATTCCTTCTAAGGAGCTGGA
 GAATCTTCAAGAATTAAACCTTTGGATCAGTGTCTAATTGGGCAAACTAAGGAAAACAG
 AAGGAAGAACAGATATAAAAAATATACTTCCCTATGATGCTACAAGAGTGCCCTCTTGAGA
 TGAAGGTGGCTATATCAATGCCAGCTTCATTAAGATAACAGTTGGGAAAGAAGAGTTTCGT
 TTACATTGCCCTGCCAAGGACCACTGCCTACAACCTGTTGGAGACTTCTGGCAGATGATTTG
 GGAGCAAAAATCCACAGTGATAGCCATGATGACTCAAGAAGTAGAAGGAGAAAAAATCAA
 ATGCCAGCGCTATTGGCCCAACATCCTAGGCAAAACAACATGGTCAGCAACAGACTTCG
 ACTGGCTCTTGTGAGAATGCAGCAGCTGAAGGGCTTGTGGTGAGGGCAATGACCCTTGA
 AGATATTAGACCAGAGAGGTGCGCCATATTTCTCATCTGAATTTCACTGCCTGGCCAGA
 CCATGATACACCTTCTCAACCAGATGATCTGCTTACTTTTATCTCCTACATGAGACACAT

CCACAGATCAGGCCCAATCATTACGCACTGCAGTGCTGGCATTGGACGTTTCAGGGACCCCT
GATTTGCATAGATGTGTTCTGGGATTAATCAGTCAGGATCTTGATTTTGACATCTCTGA
TTTGGTGCGCTGCATGAGACTACAAAGACACGGAATGGTTTCAGACAGAGGATCAATATAT
TTTCTGCTATCAAGTCATCCTTTATGTCTGACACGTCTTCAAGCAGAAGAAGAGCAAAA
ACAGCAGCCTCAGCTTCTGAAGTGACATGAAAAGAGCCTCTGGATGCATTTCCATTTCTC
TCCTTAACCTCCAGCAGACTCCTGCTCTCTATCCAAAATAAAGATCACAGAGCAGCAAGT
TCATACAACATGCATGTTCTCCTCTATCTTAGAGGGGTATTCTTCTTGAAAAATAAAAAAT
ATTGAAATGCTGTATTTTACAGCTACTTTAACCTATGATAATTATTTACAAAATTTTAA
CACTAACCAACAATGCAGATCTTAGGGATGATTAAAGGCAGCATTGATGATAGCAGAC
ATTGTTACAAGGACATGGTGAGTCTATTTTTAATGCACCAATCTTGTTTATAGCAAAAAT
GTTTTCCAATATTTTAAATAAGTAGTTATTTATAGGCATACTTGAAACCAGTATTTAAGC
TTTAAATGACAGTAATATTGGCATAGAAAAAGTAGCAAATGTTTACTGTATCAATTTCT
AATGTTTACTATATAGAATTTCCCTGTAATATTTATATACTTTTTCATGAAAAATGGAGT
TATCAGTTATCTGTTTGTACTGCATCATCTGTTTGTAAATCATTATCTCACTTTGTAAAT
AAAAACACACCTTAAACATGAACAAGCCAAAAA

>NM_006142

CCAGGCAGCAGTTAGCCCGCCGCGCCTGTGTGTCCCCAGAGCCATGGAGAGAGCCAGT
CTGATCCAGAAGGCCAAGCTGGCAGAGCAGGCCGAACGCTATGAGGACATGGCAGCCTTC
CCAGGCAGCAGTTAGCCCGCCGCGCCTGTGTGTCCCCAGAGCCATGGAGAGAGCCAGT
CTGATCCAGAAGGCCAAGCTGGCAGAGCAGGCCGAACGCTATGAGGACATGGCAGCCTTC
ATGAAAGGCGCCGTGGAGAAGGGCGAGGAGCTCTCCTGCGAAGAGCGAAACCTGCTCTCA
GTAGCCTATAAGAACGTGGTGGGCGGCCAGAGGGCTGCCTGGAGGGTGCTGTCCAGTATT
GAGCAGAAAAGCAACGAGGAGGGCTCGGAGGAGAAGGGGCCGAGGTGCGTGAGTACCGG
GAGAAGGTGGAGACTGAGCTCCAGGGCGTGTCGACACCGTGCTGGGCCTGCTGGACAGC
CACCTCATCAAGGAGGCCGGGGACGCCGAGAGCCGGGTCTTCTACCTGAAGATGAAGGGT
GACTACTACCGCTACCTGGCCGAGGTGGCCACCGGTGACGACAAGAAGCGCATCATTGAC
TCAGCCCGGTGAGCCTACCAGGAGGCCATGGACATCAGCAAGAAGGAGATGCCGCCACC
AACCCCATCCGCTGGGCCTGGCCCTGAACCTTTCCGTCTTCCACTACGAGATCGCCAAC
AGCCCCGAGGAGGCCATCTCTCTGGCCAAGACCCTTTTCGACGAGGCCATGGCTGATCTG
CACACCTCAGCGAGGACTCCTACAAAGACAGCACCTCATCATGCAGCTGCTGCGAGAC
AACCTGACACTGTGGACGGCCGACAACGCCGGGGAAGAGGGGGGCGAGGCTCCCCAGGAG
CCCCAGAGCTGAGTGTGCCCCGCCACCGCCCCCGCTGCCCCCTCCAGTCCCCGCCCTGC
CGAGAGGACTAGTATGGGGTGGGAGGCCCCACCTTCTCCCTAGGCGCTGTTCTTGCTC
CAAAGGGCTCCGTGGAGAGGGACTGGCAGAGCTGAGGCCACCTGGGGCTGGGGATCCAC
TCTTCTTGACGCTGTTGAGCGCACCTAACCACTGGTCATGCCCCACCCCTGCTCTCCGC
ACCCGCTTCTCCGACCCAGGACCAGGCTACTTCTCCCTCCTCTTGCTCCCTCCTG
CCCCTGCTGCTCTTGATTGCTAGGAATTGAGGAGTGTCTCCGCCTTGTTGGCTGAGAACT
GGACAGTGGCAGGGGCTGGAGATGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGCGCG
CGCGCCAGTGCAAGACCGAGACTGAGGGAAAGCATGTCTGCTGGGTGTGACCATGTTTCC
TCTCAATAAAGTTCCCTGTGACACTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAA

5

>AW445220

CGGCCGCGAGGCCCTGAGATGAGGCTCCAAAGACCCCGACAGGCCCGGCGGGTGGGAGG
CGCGCGCCCCGGGGCGGGCGGGGCTCCCCCTACCGGCCAGACCCGGGGAGAGGCGCGCGG
AGGCTGCGAAGGTTCAGAAAGGGCGGGGAGGGGGCGCCGCGCGCTGACCCTCCCTGGGCA
CCGCTGGGGACGATGGCGCTGCTCGCCTTGCTGCTGGTGGTGGCCCTACCGCGGGTGTGG
ACAGACGCCAACCTGACTGCGAGACAACGAGATCCAGAGGACTCCCAGCGAACGGACGAG
GGTGACAATAGAGTGTGGTGTTCATGTTTGTGAGAGAGAAAACTTTTCGAGTGCCAGAAC
CCAAGGAGGTGCAAATGGACAGAGCCATACTGCGTTATAGCGGCCGTGAAAATATTTCCA
CGTTTTTTCATGGTTGCGAAGCAGTGCTCCGCTGGTTGTGCAGCGATGGAGAGACCCAAG
CCAGAGGAGAAGCGGTTTCTCCTGGAAGAGCCCATGCCCTTCTTTTACCTCAAGTGTGT
AAAATTCGCTACTGCAATTTAGAGGGGCCACCTATCAACTCATCAGTGTTCAAAGAATAT
GCTGGGAGCATGGGTGAGAGCTGTGGTGGGCTGTGGCTGGCCATCCTCCTGCTGCTGGCC
TCCATTGCAGCCGGCCTCAGCCTGTCTTGAGCCACGGGACTGCCACAGACTGAGCCTTCC
GGAGCATGGACTCGCTCCAGACCGTTGTACCTGTTGCATTAAACTTGTCTTCTGTTGAT
TAAAAAAAAAAAAAAAAA

>AK025701

5

TTCAGCCGGAACGTTACTCCGTGTCCACCCGGATCGTGTGTGTGATCGAGGCTGCGGAGA

CGCCTTTACGGGGGGTGTGAGGTGGACGTCTTCGGGAAACTGGGCCGTTCGCCTCCCA
 ATGTCCAGTTTACCTTCCAACAGCCCAAGCCTCTCAGTGTGGAGCCGAGCAGGGACCGC
 AGGCGGGCGGCACCACACTGACCATCCACGGCACCCACCTGGACACGGGCTCCCAGGAGG
 ACGTGCGGGTGACCTCAACGGCGTCCCGTGTAAAGTGACGAAGTTTGGGGCGCAGCTCC
 AGTGTGTCACTGGCCCCCAGGCGACACGGGGCCAGATGCTTCTGGAGGTCTCTACGGGG
 GGTCCCCCGTGGCCAAACCCGGCATCTTCTTCACTACCGCGAAAACCCGTACTGCGAG
 CCTTCGAGCCGCTACGAAGCTTTGCCAGTGGTGGCCGAGCATCAACGTACGGGTGAGG
 GCTTCAGCCTGATCCAGAGGTTTGCCATGGTGGTTCATCGCGGAGCCCTGCAGTCTGGC
 AGCCGCGCGGGAGGCTGAATCCCTGCAGCCCATGACGGTGGTGGGTACAGACTACGTGT
 TCCACAATGACACCAAGGTCGTCTTCTGTTCCTGGCTGTGCCTGAGGAGCCAGAGGTCT
 ACAACCTCACGGTGTGATCGAGATGGACGGGCACCGTGCCTGCTCAGAACAGAGGCCG
 GGGCCTTCGAGTACGTGCCTGACCCACCTTGAGAACTTCACAGGTGGCGTCAAGAAGC
 AGGTCAACAAGCTCATCCACGCCCGGGGACCAATCTGAACAAGGCGATGACGCTGCAGG
 AGGCCGAGGCCCTTCGTGGGTGCCGAGCGCTGCACCATGAAGACGCTGACGGAGACCGACC
 TGTACTGTGAGCCCCGGAGGTGCAGCCCCCGCCCAAGCGGCGGCAGAAACGAGACACCA
 CACACAACCTGCCCGAGTTCAATTGTGAAGTTTCGGCTCTCGCGAGTGGGTGCTGGGCGCG
 TGGAGTACGACACACGGGTGAGCGACGTGCCGCTCAGCCTCATCTTGCCGTGGTTCATCG
 TGCCCATGGTGGTCTCATCGCGGTGTCTGTCTACTGCTACTGGAGGAAGAGCCAGCAGG
 CCGAACGAGAGTATGAGAAGATCAAGTCCCAGCTGGAGGGCCTGGAGGAGAGCGTGCGGG
 ACCGCTGCAAGAAGGAATTCACAGACCTGATGATCGAGATGGAGGACCAGACCAACGACG
 TGCACGAGGCCCGCATCCCCGTGCTGGACTACAAGACCTACACCGACCGCGTCTTCTTCC
 TGCCCTCCAAGGACGGCGACAAGGACGTGATGATCACCGGCAAGCTGGACATCCCCGAGC
 CGCGCGCGCCGGTGGTGGAGCAGGCCCTCTACCAGTTCTCCAACCTGCTGAACAGCAAGT
 CTTTCTCATCAATTCATCCACACCTTGAGAAACAGCGGGAGTTCTCGGCCCGCGCCA
 AGGTCTACTTCGCGTCCCTGCTGACGGTGGCGCTGCACGGGAACTGGAGTACTACACGG
 ACATCATGCACACGCTCTTCTTGAGCTCTTGAGCAGTACGTGGTGGCCAAGAACCCCA
 AGCTGATGCTGCGCAGGTCTGAGACTGTGGTGGAGAGGATGCTGTCCAAGTGGATGTCCA
 TCTGCCTGTACCAGTACCTCAAGGACAGTGCCTGGGGAGCCCCCTGTACAAGCTCTTCAAGG
 CCATCAACATCAGGTGGAAAGGGCCCCGGTGGATGCGGTACAGAAGAAGGCCAAGTACA
 CTCTCAACGACACGGGGCTGCTGGGGATGATGTGGAGTACGACCCCTGACGGTGAAGCG
 TGATCGTGCAGGACGAGGGAGTGGACGCCATCCCGTGAAGGTCTCAACTGTGACACCA
 TCTCCAGGTCAAGGAGAAGATCATGACCAGGTGTACCGTGGGCAGCCCTGCTCCTGCT
 GGCCAGGCCAGACAGCGTGGTCTTGGAGTGGCGTCCGGGCTCCACAGCGCAGATCCTGT
 CGGACCTGGACCTGACGTACAGCGGGAGGGCCGGTGGAAAGCGCGTCAACACCCCTATGC
 ACTACAATGTCCGGATGGAGCCACCTCATCTGTCCAAGGTGGGGTCTCCAGCAGC
 CGGAGGACAGCCAGCAGGACCTGCTGGGGAGCGCCATGCCCTCTGGAGGAGGAGAACC
 GGGTGTGGCACCTGGTGCGGCCGACCGACGAGGTGGACGAGGGCAAGTCCAAGAGAGGCA
 GCGTGAAAGAGAAGGAGCGGACGAAGGCCATCACCGAGATCTACCTGACGCGGCTGCTCT
 CAGTCAAGGGCACACTGCAGCAGTTTGTGGACAACTTCTTCCAGAGCGTGTGGCGCCTG
 GGCACGCGGTGCCACCTGCAGTCAAGTACTTCTTCCACTTCTTGGACGAGCAGGCAGAGA
 AGCACAACATCCAGGATGAAGACACCATCCACATCTGGAAGACGAACAGTTTACCGCTCC
 GGTCTGGGTGAACATCTCAAGAACCCCCACTTCATCTTTGACGTGCATGTCCACGAGG
 TGGTGGACGCCCTCGCTGTCACTCATCGCGCAGACCTTCATGGATGCTGACGCGCACGG
 AGCATAAGCTGAGCCGCGATTCTCCAGCAACAAGCTGCTGTACGCCAAGGAGATCTCCA
 CCTACAAGAAGATGGTGGAGGATTACTACAAGGGGATCCGGCAGATGGTGCAGGTGAGCG
 ACCAGGACATGAACACACACCTGGCAGAGATTTCGGGGCGCACACGGACTCTTGAACA
 CCTCGTGGCACTCCACCAGCTTACCAATACACGCAGAAGTACTATGACGAGATCATCA
 ATGCCTTGGAGGAGGATCTGCCGCCAGAAAGACGACGCTGGCCTTCGCTGCAGCAGA
 TTGCCGCTGCACTGGAGAACAAAGGTCACTGACCTCTGACCTACAATCTCCAGTGTGCTCT
 TGGGACATAGGTACCTGAGGTACCTGAGAGCCCCCTCAGGGGAGGAGGCCGAGTGGCTGTG
 GCTGAGGCCCCACCTCCCTGGAAACGCGCCCCAAGCCGGAGTGGGTGCAGCCGGAACC
 CGCCAGCGTCTAGACTGTAGCATCTTCTCTGAGCAATACCGCCGGGACCGCACACAGC
 ACCAGCCCCAGCCCCAGCTCCCTCCGGCCGAGAACAGCATCGGGTGTTCAGTGTGAG
 TCTCGAGTGATTTGAAATGTGCCTTACGCTGCCACGCTGGGGGACGTGGCCTCCGCT
 CGCCCCACGACCAGGCAAGTTCCAGATCCTATGTCTGTCTGTCTCCACCAGAGATGGGA
 GGAGGAGAAAAAGCGGTACGATGCCCTTCTGACCTCACCGGCCCTCCCAAGGTGCCGGC
 ACTCTGGGTGGACTCACGGCTGTGGGCCCCACGTCAAAGGTCAAGTGAGACGTAGGTCA
 AGTCTTACGTGGGGGCCAGACATCCTGGGGTCTGGTCTGTGACAGAGGCTGCCCTAGA
 GCCCCACCCAGTCCGGGGGACTGGGAGCAGTTCCAAGACCACCCACCCCTTTTTGTAA
 ATCTTGTTCATTGTAAATCAATACAGCGTCTTTTCTACTCCGAAAAAAAAAAAAAAAAA

AAAAAA

>NM_033229

GATGTGGGCACGCCTCAGAGCCAGAAGTTTATGGCTCCCACCTGCTCAATCTGACAGGAA
 GCTTCTGCTCCCCAGTTCTCCCCAGCCACTGTGGTCTACAGATTCCAGGAAACCCATCCC
 CCTGTGACCTCAGGGTGTGCTCTGTTCTCCACCTTAGGGACCAGAAGGAGCCAGGAGTAA
 AGAACTGGCTTACTTGGCCGCCACTGGGAAATTCTGGGTAATTGAGACGCCCTGGAATT
 TGGACCCACTCCGCTGATAGGTGGTGGGCAGGGTTCTAGGGAACACAAGAGGCGGAGCCA
 GGTGGCTTCCCTGTGCTGGCATTCTTGGCTCTCTCTCTCTCTCTTTCTCTCTCTGTCT
 CTCTCTCTCTCTCTGTCTCTCAGCCTTGAAAGCCGTTTCCCTCTGCGATTTCATGTAAGTGT
 GACTCGATTTTCAGGGAAAGGGAACTCGCGTGGGCTGAGGAGACCGGAGTGGACGGGCTGG
 GGAAGGCACCGTGATGCCCCGCAACCCCGTCCCTGAAGGTGGTCCATGAGCTGCCTGCCTG
 TACCTCTGTGCGGGGCCGCTGGAGGATGCGGTGACCGTTCCCTGTGGACACACCTTCTG
 CCGGCTCTGCCTCCCCGCGCTCTCCAGATGGGGGCCCAATCCTCGGGCAAGATCCTGCT
 CTGCCCCGCTCTGCCAAGAGGAGGAGCAGGCAGAGACTCCCATGGCCCCCTGTGCCCTGGG
 CCCGCTCGGAGAACTTACTGCGAGGAGCACGGCGAGAAGATCTACTTCTTCTGCGAGAA
 CGATGCCGAGTTCTCTGTGTGTTCTGCAGGGAGGGTCCCACGCACCAGGCGCACACCGT
 GGGGTTCCCTGGACGAGGCCATTTCAGCCCTACCGGGATCGTCTCAGGAGTCGACTGGAAGC
 TCTGAGCACGGAGAGAGATGAGATTGAGGATGTAAAGTGTCAAGAAGACCAGAAGCTTCA
 AGTGCTGCTGACTCAGATCGAAAGCAAGAAGCATCAGGTGGAAACAGCTTTTGAGAGGCT
 GCAGCAGGAGCTGGAGCAGCAGCGATGTCTCTGCTGGCCAGGCTGAGGGAGCTGGAGCA
 GCAGATTTGGAAGGAGAGGGATGAATATATCACAAAGGTCTCTGAGGAAGTCAACCGGCT
 TGGAGCCCAGGTCAAGGAGCTGGAGGAGAAAGTGTGAGCAGCCAGCAAGTGAGCTTCTACA
 AGATGTGAGAGTCAACCAGAGCAGGTGTGAGATGAAGACTTTTGTGAGTCTTGAGGCCAT
 TTCTCTGACCTTGTCAAGAAGATCCGTGATTTCCACAGGAAAATACTCACCTCCCAGA
 GATGATGAGGATGTTCTCAGAAAACCTGGCGCATCATCTGGAATAGATTTCAGGGGTCAT
 CACTCTGGACCCTCAGACCGCCAGCCGGAGCCTGGTTCTCTCGGAAGACAGGAAGTCAGT
 GAGGTACACCCGGCAGAAGAAGAGCCTGCCAGACAGCCCCCTGCGCTTCGACGGCCTCCC
 GCGGTTCTGGGCTTCCCGGGCTTCTCTCCGGGCGCCACCGCTGGCAGGTTGACCTGCA
 GCTGGGCGACGGCGGGCGGCTGCACGGTGGGGGTGGCCGGGGAGGGGGTGAGGAGGAAGGG
 AGAGATGGGACTCAGCGCCGAGGACGGCGTCTGGGCCGTGATCATCTCGCACCAGCAGTG
 CTGGGCCAGCACCTCCCCGGGCACCGACCTGCCGCTGAGCGAGATCCCGCGCGGCGTGAG
 AGTCGCCCCTGGACTACGAGGCGGGGCAGGTGACCCTCCACAACGCCCAGACCCAGGAGCC
 CATCTTCACCTTCACTGCCTCTTTCTCCGGCAAAGTCTTCCCTTTCTTTGCCGTCTGGAA
 AAAAGGTTCTGCCTTACGCTGAAAGGCTGAAGTGGGGCGCGCGAAGGGCGGCGAAGCGG
 AGACGGCGGCTCTCCGGGATCCAGCTCCGCCCCCTGGCCAGTGTGCGGCCGGGGGCTCCC
 TGTGCCCGCGTGAGGCGAGAGAACAGGGGACTTGAGTCTCGAACAGCGGTTGTTTTTACT
 TTATTTATCTTAGGCCCTCAGCTCCCTGACGTCCTGAGCCTCCCTGTGACGCTCTGGCCT
 TCTCTGCACCTCAGAGTGCAGAACCACAGACGGCTTCGGCTGTGCCTAGGGCAACAGCCA
 ACCTAGGAGCCAGCGGGCTTTCGGGGAAAAAAGAAAAAGACATCTAAATATAAATGTT
 TAAACTGTTTTCAAAATAAAAAAAAAAAAAAAAAAAAA

5 >AV656862

TTTATACATTCTAAATCTCCCCAGTTTCTTTGGGGCTGGAAGATGCAACTTCCATTTAAT
 AGAACTTTGAAATCTTGGGGTAAGGGAGCAGTGGGGGGACTAGGGAGAAGGATAAGAAA
 TAGAATTATTGAAAAGCCCCACCAGGGACCTTCCTGGCCAGAATATGCAGAGTAATTCC
 TGCTGGCTTCACCTTTGAAAGTCCCTCGAAACTATGCAGATGAACTGAGTCTGTTTTTG
 ATATTGTCAGATGTATTCTACCTTGGAGTCCCAACACCTAAACTGGAATTCTTGTATTT
 ACATCTCCTCCACTGTCCCCACACCACCCCTCAATTCTGCTGCCCCCTGCTAATGTTAA
 GCATTTTTCTCTTGTATCATCAGGTTACATTAAAAACAGATACTTACAACTGACTTG
 AAGCACAGATACTTTTACGAATGTGATAAAATATTTTCTTAAGAAAAGGAAAGAGGATGT
 GGGTCAAATAAAACACCGCATGGATGTTGATTGGTGAATACTGGTGTAAAGAAAAGGGAGC
 TCAGGAATTTTTATTACTGTATTTGTAAATGAGTTTGAAGGAATTTGTAAATGCCACTGG
 TACATTTTTAAGGTGACACATTTGCTCCTTATAAAGTTATTAAAAATTACAGGGTAAGCT
 TAAATGACGTTTGCCAGTAGTTTTACTTTATATAATCAATATTGATATTGTTGCTGAACT
 ATGTAACTTTATGATGCATTTTTCAGTCCCTTTTCAGAGCAAATGCTTTTGCAATGGTAG
 TAATGTTTAGTTTAAATTGACTTAATAAATTATTACCTGAGCAAAAAAAAAAAAAAAAAA
 AAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAA
 AACAAATCAATAAACTTAAACAAAAAAAAAATAAAAAAAAAA

>AI499593

GCAGAGATCGCCACATCGTCGGACAAGGTCAAGGACGGGGGCGGCGGGAACGAGGGCTCT
 CCATGCCACCCTGTCCCGGGCCCATAGCCGGGCAAGCCCTAGGAGGCAGCCGGGCGTCCG
 CCGGCCCCGGCGCCGTACGCTCGCCCTCGGCGCAGTGTCTTTTCCAGGCGGGACGGTG
 CTGTCCCGGCTCTCTACTACACCGCGCCCTTCTATCCCGGCTACACGAACTATGGCTCC
 TTCGGACACCTTCATGGCCACCCGGGGCCGGGGCCGGGCCCCACACCCGGTCCGGGGTCT
 CATTTCAATGGATTAAACCAGACCGTGTGAACCGAGCGGACGCTTTGGCTAAAGACCCG
 AAAATGTTGCGGAGCCAGTCTCAGCTAGACCTGTGCAAAGACTCTCCCTATGAATTGAAG
 AAAGGTATGTCCGACATTTAACCGGGCTGCGTCCGCTCCCGACTTTTCTAATTTATTAA
 AAACATGGCCTTGGCAGTTATTTTTCCATCACCGAGAGAGAGAGACAGAGAGAGAAATA
 AACTACCCCTCCTATTTCAGAAAGTTTATAGTTTATGGAGATGGATGACATAAAAAATGTAAA
 CATCTCCACACACACAAAAAATGTCTTAACCAACCGAAAAGAAAAATTAAGAAAGGATT
 TGTATTAAATCTTATTCTGTATATTTAATGTAGCATTTTTGTATTTAAATTGATAATTCA
 ATATCTTTGAAGTAAATTATGAAATCAAGACACCTGTACAGGCATTTAATGTTTTTTTGT
 AATATAAATATATACATTTGTGTTTCCCCCAAACCTGTTTCATAGTTAAAAAATACAAGT
 TTAATTTAATTTTTTACACCTATTGATTCTGCTGGGTATGAGCTAAAGTATTACGGAAAG
 GAAACAGGTTATACTCTTAGATTTAAAAAGTGAAAGAACTGCAGGCGCCTTTGTAAAT
 GCAAAATATTTAATTAAAAGAGATTTTAACATAATGAGAGCCACTCATTACTTTTAGAA
 GCCTCAATAAACTGTCCATTGCCTTGGTCAAAAAAAAAAAAAAAAAAAAAA

5

>AI952953

ATATCCAAGAAATTTGGACACCTATACCTACAGAATAATGAAATAGAAAAGATGAATCTN
 ACAGTGATGTCTCTTCTATTGACCCACTACATTACCACCATTTAACATACATTCTGTGTG
 GACCAAAATAAACTAAAAGAACCAATAAGCTCATACATCTTCTTCTGCTTCCCTCATATA
 CACACTATTTATTATGGTGAACAACGAAGCACTAATGGTCAAACAATACAATAAGACC
 CAAGTTTTTCAGGAGATTTCCAGATGATGATGATGAAAGTGAAGATCACGATGATCCTGAC
 AATGCTCATGAGAGCCCAGAACAAGAAGGAGCAGAAGGGCACCTTTGACCTTCATTATTAT
 GAAATCAAGAATAGCAAGAACTATATAGGTATACACTTACGACTTCACAAAACCTATA
 CTTAATATAGTAAATCTAAGTAAACATGTATTACTCAAAGTAATATATTTAGAATTATGT
 ATTAGTATAAGATCAGAATTGAATTTAAGTTGTTGGTGACATCTGCATCATTTTCATAGGA
 TTGAACCTTACTCAAAATAATGTAAATCTTTAAAAATATAAATTAGAATGACAAGTGGGA
 ATCATAAATTAACGTTAATGGTTTCTTATGCTCTTTTAAATATAGAAATATCATGTTA
 AAAAAAA

10

>AK025470

ATGATTGCAACAGTGGATTTTAAAAGTCAATGAATATGAGAAAAACCAAAATGGCTTGAG
 ATCCTAAATAAGATTGAAAACAAAACATACACGAAGCTCAAAAATGGACATGTGTTTAGG
 AAGCAGGCACTGATGAGTGAAGAAAGGACTCTGTTATATGATGGCCTTGTTTACTGGAAA
 ACTGCTACAGGTCGTTTCAAAGATATCCTAGCTCTACTTCTAACTGATGTGCTGCTCTTT
 TTACAAGAAAAAGACCAGAAATACATCTTTGCAGCCGTTGATCAGAAGCCATCAGTTATT
 TCCCTTCAAAGCTTATTGCTAGAGAAGTTGCTAATGAGGAGAGAGGAATGTTTCTGATC
 AGTGCTTCATCTGCTGGTCTGAGATGTATGAAATTCACACCAATTCGAAGGAGGAACGC
 AATAACTGGATGAGACGGATCCAGCAGGCTGTAGAAAGTTGTCCTGAAGAAAAGGGGGA
 AGGACAAGTGAATCTGATGAAGACAAGAGGAAAGCTGAAGCCAGAGTGGCCAAAATTTCAG
 CAATGTCAAGAAATACTCACTAACCAGACCAACAAATTTGTGCGTATTTGGAGGAGAAG
 CTGCATATCTATGCTGAACCTGGGAGAACTGAGCGGATTTGAGGACGTCCATCTAGAGCCC
 CACCTCCTTATTAACCTGACCCAGGCGAGCCTCCCCAGGCAGCCTCATTACTGGCAGCA
 GCACTGAAAGAAGCATTAGTCAAGGAGGGAGAGAAGGAAGAGGCTGTTCCGATGTGGAT
 CCCGGGATCCAGGCTGTGTAACCGACTTGGCCGCTCTCTGATGCAGGGGAGAAGGTGGAA
 TGTAAGAAATTTCCAGGTTCTTCACAATCAGAGATTATACAAGCCATACAGAATTTAACC
 CGTCTCTTATACAGCCTTCAGGCCGCTTGACCATTTCAGGACAGCCACATTGAGATCCAC
 AGGCTGGTTCTCCAGCAGCAGGAGGGCCTGTCTCTCGGCCACTCTATCCTCCGAGGCGGC
 CCCTTGCAAGACCAAGTCTCGCGACGCGGACAGGCAGCATGAGGAGCTGGCCAATGTG
 CACCAGCTTCAGCACCAGCTCCAGCAGGGGACGCGGCGCTGGCTGCGCAGGTGTGAGCAG
 CAGCAGCGGGCGCAGGCGACCAGGGAGAGCTGGCTGCAGGAGCGGGAGCGGGAGTGCCAG
 TCGCAGGAGGAGCTGCTGCTGCGGAGCCGGGCGAGCTGGACCTCCAGCTCCAGGAGTAC
 CAGCACAGCCTGGAGCGGCTGAGGGAGGGCCAGCGCCTGGTGGAGAGGGAGCAGGCGAGG
 ATGCGGGCCAGCAGAGCCTGCTGGGCCACTGGAAGCACGCGCGGAGAGGAGCCTGTCC
 GCGGTGCTCCTTCCGGGTGGCCCCGAGGTAATGGAACCTTAATCGATCTGAGAGTTTATGT
 CATGAAAACCTCATTCTTCATCAATGAAGCTTTAGTACAAATGTCATTTAACACTTTCAAC
 AAAGTGAATCCATCAGTTATCCATCAGGATGCCACTTACCCTACAACCTCAATCTCATTCT
 GACTTGGTGAGGACTAGTGAACATCAAGTAGACCTCAAGGTGGACCCTTCTCAGCCTTCG

AATGTCAGTCACAACTGTGGACAGCCGCTGGTTCCGGCCATCAGATACTTCCTTTCCAT
 GAAAGCAGCAAGGATTCTTGTAATAATGGCTCCAGTATGACAAAGTGCAGTTGTACGTTG
 ACATCTCCCCCGGACTGTGGACTGGAACCACATCTACTTTGAAGGATTTGGACACCTCC
 CACTGAGTCCCCAACCCCCATGACTCAAATTCACACCGCCCTCACTGCAGGCGTTT
 ATAACAGAAGCAAAGCTAAATCTACCGACAAGGACAATGACCAGACAAGATGGGGAACT
 GGAGATGGAGCCAAAGAAAATATTGTTTACCTCTAATTGTGTTGTCAATTTTTTCAAACAA
 AACAAAACACTGGCACTTTTGGGAGAACTTTTGTCTCCATTCTTTATGTATGTGTGAT
 TGTCTGTGTCCAAATTGCTTTAAGAATAATATTTAATATTTCTGGAAGCTCATTTTTTT
 GGCATGAGTCTAATTAAATTATTGAAAGCCACCCTGTTTGTATAATCTTTAACTTATCAA
 ATCTAATTTTCAGATTTCTGGAGGAGAACTAACTTGAATAAGCAGGACTATTTTAAAGT
 TGTTTTGACGCTAGAGTAAATTCATGTACATTTTCTACCCAATCATCTGGATTTCAA
 GATTCCTTTTAAGATCTCAATGAAGCAATTTGGATTTAAAGAGTGGTATTCAAGGGGT
 GAACTTTCACAGTCAGGGCAGTTGCCTCAGTGCCACATAGGCAGAGGAGGATGTGGGAA
 AGGGCTTTCTCAGCTAGTTTTTGTGTGCTCATTCTCTGGGAGCATTTAAAGTGGTGA
 TCTGTTACAGTCACTATTCAACTGGGCAGTGTTGTGATTGGTCAGTCACTGAGCCAGGG
 ATACAGTCCGGACTTGCTTAGTACCTAAGCCTAATGCTGGTGGGGTTTCAAGACATGGTT
 CAGCATCATCTTTTAACAAGGCCAGAGGCCAGAGCCCGCATCAAGTCATTTTGTATGTA
 AATAGTGAACCTTTGTTAGAGCCCTCACTTCTATCAATCAGCTGTCTGTCCCTGCCAGCA
 CCTGGAGCACCACTACCACTCCCTGGAAAGAACCCTTCCCTGCAGTTTTTTAAGGACAA
 AACTGCCCCACTCCTCATTAAGTTTGTGCTGGATACACTTTTCCACAAAGGAAAAGTGG
 CATATCCTGCCTTCCGAGTAGTATGGGTCTCTGTGTGAGAAACCAGGAGATATTTTCATC
 TTGTTCCGAAATACTTGTATGTATTTTGGTGTCAATAAATATCTTGTACCTCATTAAAAA
 AAAAAAAAAAAAA

>NM_006378

CTGAGCCGCATCTGCAATAGCACACTTGCCCGGCCACCTGCTGCCGTGAGCCTTTGCTGC
 TGAAGCCCCCTGGGGTCGCCTCTACCTGATGAGGATGTGCACCCCCATTAGGGGGCTGCTC
 ATGGCCCTTGACAGTGATGTTTGGGACAGCGATGGCATTTCACCCATACCCCGGATCACC
 TGGGAGCACAGAGAGGTGCACCTGGTGCAGTTTCATGAGCCAGACATCTACAACTACTCA
 GCCTTGCTGCTGAGCGAGGACAAGGACACCTTGATACATAGGTGCCCGGGAGGCGGTCTTC
 GCTGTGAACGCACTCAACATCTCCGAGAAGCAGCATGAGGTGTATTGGAAGGTCTCAGAA
 GACAAAAAGCAAAATGTGCAGAAAAGGGGAAATCAAAACAGACAGAGTGCCTCAACTAC
 ATCCGGGTGCTGCAGCCACTCAGCGCCACTTCCCTTTACGTGTGTGGGACCAACGCATTTC
 CAGCCGGCCTGTGACCACCTGAACCTTAACATCCTTTAAGTTTCTGGGGAAAAATGAAGAT
 GGCAAAGGAAGATGTCCCTTTGACCCAGCACACAGCTACACATCCGTGCATGGTTGATGGA
 GAACTTTATTTCGGGGACGTCGTATAATTTTTTGGGAAGTGAACCCATCATCTCCCGAAAT
 TCTTCCACAGTCCTCTGAGGACAGAATATGCAATCCCTTGGCTGAACGAGCCTAGTTTC
 GTGTTTGCTGACGTGATCCGAAAAAGCCAGACAGCCCCGACGGCGAGGATGACAGGGTC
 TACTTCTTCTTACGGAGGTGTCTGTGGAGTATGAGTTTGTGTTTACGGGTGCTGATCCCA
 CGGATAGCAAGAGTGTGCAAGGGGGACAGGGCGGCCTGAGGACCTTGCAGAAAGAAATGG
 ACCTCCTTCTTGAAGCCCGACTCATCTGCTCCCGGCCAGACAGCGGCTTGGTCTTCAAT
 GTGCTGCGGGATGTCTTCGTGCTCAGGTCCCGGGCCTGAAGGTGCCTGTGTTCTATGCA
 CTCTTACCCACAGCTGAACAACGTGGGGCTGTGCGCAGTGTGCGCCTACAACCTGTCC
 ACAGCCGAGGAGGTCTTCTCCACGGGAAGTACATGCAGAGCACCACAGTGGAGCAGTCC
 CACACCAAGTGGGTGCGCTATAATGGCCCGGTACCCAAGCCGCGGCCTGGAGCGTGCATC
 GACAGCGAGGCACGGGCCGCCAACTACACCAGCTCCTTGAATTTGCCAGACAAGACGCTG
 CAGTTTCGTTAAAGACCACCTTTGATGGATGACTCGGTAACCCCAATAGACAACAGGCCC
 AGGTTAATCAAGAAAGATGTGAACACACCCAGATCGTGGTGGACCGGACCCAGGCCCTG
 GATGGGACTGTCTATGATGTCATGTTTGTGACACAGACCGGGGAGCTCTGCACAAAGCC
 ATCAGCCTCGAGCACGCTGTTACATCATCGAGGAGACCCAGCTCTTCCAGGACTTTGAG
 CCAGTCCAGACCCTGCTGCTGTCTTCAAAGAAGGGCAACAGGTTTGTCTATGCTGGCTCT
 AACTCGGGCGTGGTCCAGGCCCGCTGGCCTTCTGTGGGAAGCACGGCACCTGCGAGGAC
 TGTGTGCTGGCGCGGGACCCCTACTGCGCCTGGAGCCCGCCACAGCGACCTGCGTGGCT
 CTGCACCAGACCGAGAGCCCCAGCAGGGGTTTGATTGAGGAGATGAGCGCGGATGCTTCT
 GTGTGCCCCGATAAAAGTAAAGGAAGTTACCGGCAGCATTTTTTCAAGCACGGTGGCACA
 GCGGAAGTGAATGCTCCCAAAATCCAACCTGGCCCGGGTCTTTTGGAAAGTTCCAGAAT
 GCGGTGTTGAAGGCCGAGAGCCCCAAGTACGGTCTTATGGGCAGAAAAAACTTGCTCATC
 TTCAACTTGTCAGAAGGAGACAGTGGGGTGTACAGTGCCTGTCAGAGGAGAGGGTTAAG
 AACAAAACGGTCTTCCAAGTGGTCGCCAAGCACGTCCTGGAAGTGAAGGTGGTTCCAAAG
 CCCGTAGTGGCCCCCACCCTTGTCAGTTGTTAGACAGAAGGTAGTAGGATTGCCACCAA
 GTGTTGGTGGCATCCACCCAAGGTCTTCTCCCCCAACCCAGCCGTGCAGGCCACCTCC

TCCGGGGCCATCACCCCTTCTCCCAAGCCTGCGCCACCGGCACATCCTGCGAACCAAAG
 ATCGTCATCAACACGGTCCCCCAGCTCCACTCGGAGAAAACCATGTATCTTAAGTCCAGC
 GACAACCGCCTCCTCATGTCCCTCTTCCCTCTTCTTTGTTCTCTCCTCTGCCTCTTT
 TTCTACAACCTGCTATAAGGGATACCTGCCAGACAGTGCTTGAAATCCGCTCGGCCCTA
 CTAATTGGGAAGAAGAAGCCCAAGTCAGATTTCTGTGACCGTGAGCAGAGCCTGAAGGAG
 ACGTTAGTAGAGCCAGGGAGCTTCTCCAGCAGAATGGGGAGCACCCCAAGCCAGCCCTG
 GACACCGGCTATGAGACCGAGCAAGACACCATCACCAGCAAAGTCCCCACGGATAGGGAG
 GACTCACAGAGGATCGACGACCTTTCTGCCAGGGACAAGCCCTTTGACGTCAGTGTGAG
 CTGAAGTTCGCTGACTCAGACGCAGATGGAGACTGAGGCCGGCTGTGCATCCCCGCTGGT
 GCCTCGGCTGCGACGTGTCCAGGCGTGGAGAGTTTTGTGTTTCTCCTGTTCAGTATCCGA
 GTCTCGTGCACTGTGCGTAGGTTAGCCCGCATCGTGACACAACCTCAGTCCTCTTGTCT
 TATTTTCTCTTGGGTTGAGCCTGTGACTTGGTTTCTCTTTGTCTTTTGAAAAATGACA
 AGCATTGCATCCAGTCTTGTGTTCCGAAGTCAGTCGGAGTACTTGAAGAAGGCCACGG
 GCGGCACGGAGTTCTTGAGCCCTTTCTGTAGTGGGGGAAAGGTGGCTGGACCTCTGTTGG
 CTGAGAAGAGCATCCCTTCAGCTTCCCTTCCCGTAGCAGCCACTAAAAGATTATTTAAT
 TCCAGATTGGAAATGACATTTTAGTTTATCAGATTGGTAACTTATCGCCTGTTGTCCAGA
 TTGGCACGAACCTTTTCTTCCACTTAATTATTTTTTTAGGATTTTGCTTTGATTGTGTTT
 ATGTTCATGGGTCATTTTTTTTTAGTTACAGAAGCAGTTGTGTTAATATTTAGAAGAAGAT
 GTATATCTTCCAGATTTTGTATATATTTGGCATAAAATACGGCTTACGTTGCTTAAGAT
 TCTCAGGGATAAACTTCTTTTGTCTAAATGCATTCTTTCTGCTTTTAGAAATGTAGACAT
 AAACACTCCCCGGAGCCCACTCACCTTTTTCTTTTTCTTTTTTTTTTTAACTTTATTT
 CCTTGAGGGAAGCATTGTTTTTTGAGAGATTTTCTTTCTGTACTTCGTTTACTTTTCTT
 TTTTTTTAACTTTTACTCTCTCGAAGAAGAGGACCTTCCACATCCACGAGGTGGGTTTT
 GAGCAAGGGAAGGTAGCCTGGATGAGCTGAGTGGAGGCCAGGCTGGCCACAGAGCTGAGATG
 GGAGTGCGGTACAATCTGGAGCCACAGCTGTGCGTCAAGACCTCCTGTGAGACAGATGG
 AACCTTCACAAGGGCGCCTTTGTTCTCTGAACATCTCCTTTCTCTTCTGCTTCAATTG
 CTTACCCACTGCCTGCCAGACTTTCTATCCAGCCTCACTGAGCTGCCCACTACTGGAAG
 GGAAGTGGGCTCGGTGGCCGGGGCGCGAGCTGTGACCACAGCACCTCAAGCATAACGG
 CGCTGTTCTGCACTGTCTTGAAGATGTGAATGGGTGGTACGATTTCAACACTGGTTAA
 TTTCACTCCATCTCCCCGCTTTGTAAATACCCATCGGGAAGAGACTTTTTTTCCATGG
 TGAAGAGCAATAAACTCTGGATGTTTGTGCGGTGTGTGGACAGTCTTATCTTCCAGCAT
 GATAGGATTTGACCATTTTGGTGTAAACATTTGTGTTTTATAAGATTTACCTTGTTTTTA
 TTTTCTACTTTGAATTGTATACATTTGGAAAGTACCCAAATAAATGAGAAGCTTCTATC
 CTTAAAAA

>AA993639

CCCNTCCCCAGAGGCAGGAAAANCAGTNTGCCGAAAGGATAGACTGNGGTGCNGTCTTTC
 CCAAAGTTNTGAACTAGTTTAAAGGTAGCTTAGGATGAAAAATGGAGAATGATTGGGGGT
 TCCAAACCACTTTCTTCTCCCTTGGCTTATATCTCTTCAACATTTGGTGGTCACTGTGG
 GCCTACCCTGGACCTCATCTACTCAGCGAGAATTGGACATGAAGCTAGAGGCAGCTGCCT
 TGGAAAGGAAGTCAGGCTCACTTGGACAGCCAGGCCATGGCAGGAAGAATCCCTTCTCTC
 TTGGGGTCTTGATGGGCATGTGTGATGGGGAAGGAGCAGTCTCCAGCCCTGGGTCTGC
 TCCCCACATCTCTCTAATCCACTTCACCTTTTGCCACCCCTCCCCACCAGAGGCCCTA
 GCCCTTTTGTACCGAAGGCCCCAGAGTGTCTGTGTGAAACCTCTCATTTTCACTG
 TGGCATCAAAATCCACAAAGATGGATTAATTGCACTCTGGTTAATAGCAGCAGCACAAT
 GATTAAATCTATATTCCTATCTTCTCTAGCACCTGGTGTGGGGATGGGGCGGAAGGGT
 GTCTTGAGGGGCAGGGAGGACCCATAAAACAATCCCTCCTGCATTCTCAGGCTAAATAG
 GGCCCCAGTGACTACCTGTTCTTGGCTGTCCCTCTGAAGAGCTCTGCCTTCTCACAGC
 CACCACCAGTTGCCCACTCCCAGGAAAACAGCACATGTTCTTCTCTCTCTGCTTGGAGA
 CTGCGTGTAGTCTTCCATTTCATAACTCATCAGCAGCTCAGTCCTTCTTATGTCTAGTCT
 CAGTTCAATCAGCCAAAGCTCATTTTTTGTCTATCCAAAGTAGAAAGGGTTCTTTTAGAA
 AACTTGAAGAATGTGCTCCTCTTAGCATCTGTTTCTGACTCCAGTTATTTTTAAATA
 AATGATGAATAAAATGCCTGCCCTGAAGGGTCTGGAGGAGTCAGGTATCAAAAAAAA
 AAA

5

>BE552004

TTTTTAAGATGATCTTGCTCCGTCACCCAGGCTGGAGTGCAGTGGCGTAAATCATGGCTTC
CTGCAGCCTCAAACCTCTGGGCTCAATGAGTTCCTTGAGATCTTCCATCCTCAGCTTCCC
AAGTAGCTAGTAGTAGTAGTGGCTTGCACCAACGCTCCTGCCCTAATTTTCAATATTTTT
TTTGTAGAGATAGGATCTCACTGTGTTACCCAAGCTAGACTTGAACCTCTGGCCTCAAGC
GATCCTTCCGCCTTGGCCTCCCAAAGTGTGGGATTACAGGCATTAGCTACCACACCTGG

CCAAGGCCCCAGGTTTCGACAGAAAGGGAGAGAAAACCTGCCAGAGATGCCATTTTCGGAGC
CACTCTGCTTGGCAGGGACCTGTGTTCCCCTCATGCAGGTTCATCCTTAGAGGGCTGCGG
TCTTATCTGGTTGTGCAAAAGTCCCACAACCTTTCTGGATTGATAGTTTGTGGTGAAATA
AACAATTTTAGTTTGTGGAGAATCTTTGTATACAAAATACAAATAAAACCTAAATCA
AAGAAACAGA

>BC010437

GAGGGGCGGAGGCGTCCCCGCTCCCGCTCGCTACTAGCCCGCGGGCCAGCGCCGCGTCC
 CGAGCCCCGGGAGCCATGGCTCTAAAAGGACAAGAAGATTATATTTATCTTTTCAAG
 GATTCAACACATCCAGTGGATTTTCTGGATGCATTTCAGAACATTTTACTTGGATGGATTA
 TTTACTGATATTACTCTTCAGTGTCTTCAGGCATAATTTTCCATTGTCACCGAGCCGTT
 TTAGCTGCTTGCAGCAATTATTTTAAGGCAATGTTTCACAGCTGACATGAAAGAAAAATTT
 AAAAATAAAATAAACTCTCTGGCATCCACCATGATATTCTGGAAGGCCCTTGTAATTTAT
 GCATACACTTCCCAAATTGAAATACTAAAAGAAATGTTCAAAGCCTGCTTGAGGCAGCG
 GATCTGCTACAGTTCTTTTCAGTAAAGAAGGCTTGTGAGCGGTTTTTGGTAAGGCACCTG
 GATATTGATAATTGTATTGGAATGCACTCCTTTGCGAATTTTCATGTGTGTCCAGAATA
 GAGAAGGAATCTCGAAGAATTCTATGTTCAAAGTTTAAGGAAGTGTGGCAACAAGAAGAA
 TTTCTGGAAATCAGCCTTGAAAAGTTTCTCTTTATCTTGTCCAGAAAGAATCTCAGTGT
 TGGAAAGAAGAAGCTATCATAGAGCCAGTTATTAAGTGGACTGCTCATGATGTAGAAAAT
 CGAATTGAATGCCTCTATAATCTACTGAGCTATATCAACATTGATATAGATCCAGTGTAC
 TTA AAAACAGCCTTAGGCCCTCAAAGAAGCTGCCTGCTCACCAGAAAATAAGATCCGCTCC
 CTAATATACAATGCCTTGAATCCCATGCAATAAAGAGATTTCCAGAGGTCCACAGCCACA
 ATGTATATAATTGGAGGCTATTACTGGCATCCTTTATCAGAGGTTTACATATGGGATCCT
 TTGACAAATGTTTGGATTTCAGGGAGCAGAAATACCAGATTATACCAGGGAGAGCTATGGT
 GTTACATGTTTAGGACCCAACATTTATGTAACCTGGGGGCTACAGGACGGATAACATAGAA
 GCTCTTGACACAGTGTGGATCTATAACAGTGAAAGTGATGAATGGACAGAAGGTTTGCCA
 ATGCTCAATGCCAGGTATTACCACTGTGCAGTCACCTTGGGTGGCTGTGTCTATGCTTTA
 GGTGGTTACAGAAAAGGGGCTCCAGCAGAAGAGGCTGAGTTCTATGATCCTTTAAAAGAG
 AAATGGATTCTATTGCAAACATGATTAAAGGTGTGGGAAATGCTACTGCCTGTGTCTTA
 CATGATGTTATCTACGTCAATTGGTGGCCACTGTGGCTACAGAGGAAGCTGCACCTATGAC
 AAAGTTTCAGAGCTACAATTCCGATATCAACGAATGGAGCCTCATCACCTCCAGTCCACAT
 CCAGAATATGGATTGTGCTCAGTTCGGTTTGAATAAAGCTCTATCTAGTCGGTGGACAA
 ACTACAATCACAGAATGCTATGACCCTGAACAAAATGAATGGAGAGAGATAGCTCCCATG
 ATGGAAAGGAGGATGGAGTGCCTGCGTGCATCATGAATGGATGTATTTATGTCACCTGGA
 GGATACTCCTACTCAAAGGGAACGTATCTTCAGAGCATTGAGAAATATGATCCAGATCTT
 AATAAGTGGGAAATAGTGGTAATCTTCCAGTGCCATGCGGTCTCATGGGTGTGTTTGT
 GTGTATAATGTCTAATTGAATCTGCAGAAATGACCAAGCAATCACTTTTTTGGAGTATAG
 TTTTATAAAAAAAGAATGCAGGGTTTGAAGTTCCTTACCTGATAATTGTGTCTGGCACAT
 GATAGGGGATCAGTAAATTGTAATTCCTAACCTACTGTACTCCCAAACATGGTGATTCA
 TGGTCAAGAAAAATCTTATATATATATATACACACATATATATGTGTTTCATATATATG
 TATACATATATGTGTATATATACGCATGTATGTATACATATATGTGTATATATACGCATG
 TATGTATGCATATGTGTGTATATATACGTATGTATGTATACATATGTGTATATATACGTA
 TGTATGTATACATATATGTGTATATATGCGTATGTATGTATACATATATGTGTATATATA
 CGTATGTATGTATACATATATGTGTATATATACGTATGTATGTATACATATATGTGTATA
 TATACGTATGTATGTATACATATATGTGTGTATATACGTGTGTATGTATACATATATGTG
 TATATATACGTGTGTATGTATACATATATGTGTATATATGCGTGTGTATGTATACATATA
 TGTGTATATATACGTGTGTATGTATACATATATGTGTATATATACGTGTGTATGTATACA
 TATATGTGTATATATGCGTGTGTATATATATACATATATACGTATATATGTATATATA
 TATACACAGTTGAATCAGTGGGATTAATACCTATAATCTCTGGTTTTCAAAGGTAATATG
 GAATATTTGACACTTGGTAAAAAGGTGAACCTTTGTAGTGAATCTTTTCTCTTGGTA
 GCATCAACACTGGGGATAAATCAGAACCATTCTGTGGAATGAAATGTTTCTCAAGAGCCT
 ATAATATAGTAGATAGTGCATATTAAGATGTCTGGCTGGGCATGGTGGCTCATGCCTGTA
 ATCCCAGCACTTTGGGAGGCTGAGGCGGGAGGATCACTTGAGCCTAGAAGTTGGAGACTA
 ACCTGGCGAGACCCTGTCTCAAAAAAAAAAAAAAAAAAAAA

>R15881

ACCCTTTTGTGACCACTGCATACCCCAAACCTTTTGGAAATCTGGGCTAACTGGCTGTG
 CCTACATCAACAGCACCCGTTGAACCCCGTGTGCTATGCTCTGTGCAACAAAACATTCAG
 AACCCACTTTCAAGATGCTGCTGCTGTGCCAGTGTGACAAAAAAGAGGCGCAAGCAGC
 AGTACCAGCAGAGACAGTCGGTCAATTTTCAACAAGCGCGCACCCGAGCAGGCCCTGTAGA
 ATGAGGTTGTATCAATAGCAGTGACAAAACGCACACATCAACCCACAGACCTTAGGAGGA

GGAAGGCGAGGGCGGGGTGACTTCTGGTGATGATAAAATGGTTTTATCACCCAGATGTG
AAAGAAGCTGCCTGTTTACTGATCCATTGAATAAACCCATTTTAATAGAAAAAGTCAATA
CCAATTACAGCAAAAAAAAAA

>AF191770

TATCTATGTAACAAATCGCAGCACAGGAGTCCCCTGGGCTCCCTCAGGCTCTGGTATGAC
ATATTTGAGCCATATAAATTCAGCTTCTCCTCTGGCATCTGTTAGCCGACTCACTTGCAA
CTCCACCTCAGCAGTGGTCTCTCAGTCCTCTCAAAGCAAGGAAAGAGTACTGTGTGCTGA
GAGACCATGGCAAAGAATCCTCCAGAGAATTGTGAAGACTGTCACATTCTAAATGCAGAA
GCTTTTAAATCCAAGAAAAATATGTAAATCACTTAAGATTTGTGGACTGGTGTGTTGGGTATC
CTGACCCCTAATCTAATTGTCTGTGTTTTGGGGGAGCAAGCACTTCTGGCCGGAGGTACCC
AAAAAAGCCTATGACATGGAGCACACTTTCTACAGCAGTGGAGAGAAGAAGAAGATTTAC
ATGGAAATTGATCCTGTGACCAGAACTGAAATATTTCAGAAGCGGAAATGGCACTGATGAA
ACATTGGAAGTACACGACTTTAAAAACGGATACACTGGCATCTACTTCGTGGGTCTTCAA
AAATGTTTTATCAAACTCAGATTAAAGTGATTCTGAATTTTCTGAACCAGAAGAGGAA
ATAGATGAGAATGAAGAAATTACCACAACCTTTCTTTGAACAGTCAGTGATTTGGGTCCCA
GCAGAAAAGCCTATTGAAAACCGAGATTTTCTTAAAAATTCCAAAATTCTGGAGATTTGT
GATAACGTGACCATGTATTGGATCAATCCCACTCTAATATCAGTTTCTGAGTTACAAGAC
TTTGAGGAGGAGGAGAGAAGATCTTCACTTTCTGCAACGAAAAAAGGGATTGAACAA
AATGAACAGTGGGTGGTCCCTCAAGTGAAAGTAGAGAAGACCCGTCACGCCAGACAAGCA
AGTGAGGAAGAAGCTTCCAATAAATGACTATACTGAAAATGGAATAGAATTTGATCCCATG
CTGGATGAGAGAGGTTATTGTTGTATTTACTGCCGTGAGGCAACCGCTATTGCCGCCGC
GTCTGTGAACCTTTACTAGGCTACTACCCATATCCATACTGCTACCAAGGAGGACGAGTC
ATCTGTGCTGTATCATGCTTGTAACTGGTGGGTGGCCCGCATGCTGGGGAGGGTCTAA
TAGGAGGTTTTGAGCTCAAATGCTTAACTGCTGGCAACATATAATAAATGCATGCTATTC
AATGAATTTCTGCCTATGAGGCATCTGGCCCTGGTAGCCAGCTCTCCAGAATTACTTGT
AGGTAATTCCTCTCTTCATGTTCTAATAAACTTCTACATTATCAAAAAA

5

>BC005364

GCGGATCGCTGCTCCCTCTCGCCATGGCGCAGGTGCTGATCGTGGGCGCCGGGATGACAG
GAAGCTTGTGCGCTGCGCTGCTGAGGAGGCAGACGTCCGGTCCCTTGTACCTTGCTGTGT
GGGACAAGGCTGACGACTCAGGGGAAGAATGACTACAGCCTGCAGTCCTCATAATCCTC
AGTGACACAGCTGACTTGGGTGCTCAGTACATCACCTGCACCTCCTCATTATGCCAAAAAAC
ACCAACGTTTTTATGATGAACTGTTAGCCTATGGCGTTTTGAGGCCTCTAAGCTCGCCTA
TTGAAGGAATGGTGATGAAAGAAGGAGACTGTAACCTTTGTGGCACCTCAAGGAATTTCTT
CAATTATTAAGCATTACTTGAAAGAATCAGGTGCAGAAGTCTACTTCAGACATCGTGTGA
CACAGATCAACCTAAGAGATGACAAATGGGAAGTATCCAAACAAACAGGCTCCCTTGAGC
AGTTTGATCTTATTGTTCTCACAATGCCAGTTCCTGAGATTCTGCAGCTTCAAGGTGACA
TCACCACCTTAATTAGTGAATGCCAAAGGCAGCAACTGGAGGCTGTGAGCTACTCCTCTC
GATATGCTCTGGGCCCTCTTTTATGAAGCTGGTACGAAGATTGATGTCCCTTGGGCTGGGC
AGTACATCACCAAGTAATCCCTGCATACGCTTCGTCTCCATTGATAATAAGAAGCGCAATA
TAGAGTCATCAGAAATTGGGCCTTCCCTCGTGATTACACCACTGTCCCATTTGGAGTTA
CATACTTGAACACAGCATTGAGGATGTGCAAGAGTTAGTCTTCCAGCAGCTGGAAAAACA
TTTTGCGGGTTTTGCCTCAGCCAATTGCTACCAAAATGCCAAAAATGGAGACATTACAGG
TTACAAATGCTGCTGCCAACTGTCTGGCCAAATGACTCTGCATCACAAACCTTTCTCTG
CATGTGGAGGGGATGGATTTACTCAGTCCAATTTGATGGCTGCATCACTTCTGCCCTAT
GTGTTCTGGAAGCTTTAAAGAATTATATTTAGTGCCTATATCCTTATTCTCTATATGTGT
ATTGGGTTTTTATTTTACAAATTTCTGTTATTGATTATTTTGTGTTTCTATTTTGCTAAG
AAAAATTACTGGAAAATTGTTCTTCACTTATTATCATTTTTTCTGTGGAGTATAAAATCA
ATTTTGTAATTTTGATAGTTACAACCCATGCTAGAATGGAAATTCCTCACACCTTGCACC
TTCCCTACTTTTCTGAATTGCTATGACTACTCCTTGTGGAGGAAAAGTGGTACTTAAAA
AATAACAAACGACTCTCTCAAAAAAATTACATTAAATCACAATAACAGTTTGTATGCCAA
AACTTGATTATCCTTATGAAAATTTCAATTCTGAATAAAGAATAATCACATTATCAAAG
CCCCATCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

10

>NM_001337

ACTCGTCTCTGGTAAAGTCTGAGCAGGACAGGGTGGCTGACTGGCAGATCCAGAGGTTCC
 CTTGGCAGTCCACGCCAGGCCTTCACCATGGATCAGTTCCCTGAATCAGTGACAGAAAAAC
 TTTGAGTACGATGATTTGGCTGAGGCCTGTTATATTGGGGACATCGTGGTCTTTGGGACT
 GTGTTCTGTCCATATTCTACTCCGTATCTTTGCCATTGGCCTGGTGGGAAATTTGTTG
 GTAGTGTTCGCCCTCACCAACAGCAAGAAGCCCAAGAGTGTACCGACATTTACCTCCTG

AACCTGGCCTTGTCTGATCTGCTGTTTGTAGCCACTTTGCCCTTCTGGACTCACTATTTG
 ATAAATGAAAAGGGCCTCCACAATGCCATGTGCAAATTCACCTACCGCCTTCTTCTTCATC
 GGCTTTTTTTGGAAGCATATTCTTCATCACCGTCATCAGCATTGATAGGTACCTGGCCATC
 GTCCCTGGCCGCCAACTCCATGAACAACCGGACCGTGCAGCATGGCGTCACCATCAGCCTA
 GGCGTCTGGGCAGCAGCCATTTTGGTGGCAGCACCCAGTTCATGTTACAAAGCAGAAA
 GAAAATGAATGCCTTGGTGAATAACCCGAGGTCTCCAGGAAATCTGGCCCGTGCCTCCGC
 AATGTGGAACAAATTTTCTTGGCTTCTACTCCCCCTGCTCATTATGAGTTATTGCTAC
 TTCAGAATCATCCAGACGCTGTTTCTCTGCAAGAACCACAAGAAAGCCAAAGCCATTAAA
 CTGATCCTTCTGGTGGTTCATCGTGTTCCTCTCTCTGGACACCCACAAACGTTATGATT
 TTCCTGGAGACGCTTAAGCTCTATGACTTCTTCCAGTTGTGACATGAGGAAGGATCTG
 AGGCTGGCCCTCAGTGTGACTGAGACGGTTCATTTAGCCATTGTTGCCTGAATCCTCTC
 ATCTATGCATTTGCTGGGGAGAAGTTCAGAAGATACCTTTACCACCTGTATGGGAAATGC
 CTGGCTGTCTGTGTGGGCGCTCAGTCCACGTTGATTTCTCCTCATCTGAATCACAAAGG
 AGCAGGCATGGAAGTGTCTGAGCAGCAATTTACTTACCACACGAGTGATGGAGATGCA
 TTGCTCCTTCTCTGAAGGGAATCCCAAAGCCTTGTGTCTACAGAGAACCTGGAGTTCTCTG
 AACCTGATGCTGACTAGTGAGGAAAGATTTTGTGTGTTATTTCTTACAGGCACAAAATGA
 TGGACCCCAATGCACACAAAACAACCTAGAGTGTTGTTGAGAATTGTGCTCAAAAATTTGA
 AGAATGAACAAATTTGAATCTTTGAATGACAAAGAGTAGACATTTCTCTTACTGCAAAATG
 TCATCAGAACTTTTTGGTTTGCAGATGACAAAAATTCAACTCAGACTAGTTTAGTTAAAT
 GAGGGTGGTGAATATTGTTTCATATTGTGGCACAAGCAAAAGGGTGTCTGAGCCCTCAAAG
 TGAGGGGAAACCAGGGCCTGAGCCAAGCTAGAATTCCCTCTCTCTGACTCTCAAATCTTT
 TAGTCATTATAGATCCCCAGACTTTACATGACACAGCTTTATCACCAGAGAGGGACTGA
 CACCCATGTTTTCTCTGGCCCCAAGGGAAAATTCACAGGGAAGTGCTCTGATAGGCCAAGT
 TTGTATCAGGTGCCATCCCTGGAAGGTGCTGTTATCCATGGGGAAGGGATATATAAGAT
 GGAAGCTTCCAGTCCAATCTCATGGAGAAGCAGAAAATACATATTTCCAAGAAGTTGGATG
 GGTGGGTACTATTCTGATTACACAAAACAATGCCACACATCACCCCTTACCATGTGCCTG
 ATCCAGCCTCTCCCCTGATTACACCAGCCTCGTCTTCATTAAGCCCTCTTCCATCATGTC
 CCCAAACCTGCAAGGGCTCCCCTGCTACTGCTATCGAGTCAAACTCAAATGCTTGGC
 TTCTCATACGTCCACCATGGGGTCTACCAATAGATTCCCCATTGCCTCCTCCTTCCCAA
 AGGACTCCACCCATCCTATCAGCCTGTCTCTTCCATATGACCTCATGCATCTCCACCTGC
 TCCCAGGCCAGTAAGGGAAATAGAAAAACCTGCCCCCAAATAAGAAGGGATGGATTCCA
 ACCCAACTCCAGTAGCTTGGGACAAATCAAGCTTCAGTTTCCTGGTCTGTAGAAGAGGG
 ATAAGGTACCTTTACATAGAGATCATCCTTTCCAGCATGAGGAACTAGCCACCAACTCT
 TGCAGGTCTCAACCCTTTTGTCTGCCTCTTAGACTTCTGCTTTCCACACCTGCACTGCTG
 TGCTGTGCCCAAGTTGTGGTGTGACAAAGCTTGGAAAGAGCCTGCAGGTGCCTTGGCCGC
 GTGCATAGCCCAGACACAGAAGAGGCTGGTTCTTACGATGGCACCCAGTGAGCACTCCCA
 AGTCTACAGAGTGATAGCCTTCCGTAACCCAACTCTCCTGGACTGCCTTGAATATCCCT
 CCCAGTCACCTTGTGCAAGCCCCTGCCATCTGGGAAAATACCCCATCATTATGCTACT
 GCCAACCTGGGGAGCCAGGGCTATGGGAGCAGCTTTTTTTTTCCCCCTAGAAACGTTTGG
 AACAAATGTAAACCTTTAAAGCTCGAAAACAATGTAAATAATGCTAAAGAAAAAGTCATCC
 AATCTAACCACATCAATATTGTCATTCTGTATTACCCCGTCCAGACCTTGTTCACACTC
 TCACATGTTTAGAGTTGCAATCGTAATGTACAGATGGTTTATAATCTGATTTGTTTCC
 TCTTAACGTTAGACCACAAATAGTGCTCGCTTTCTATGTAGTTTGGTAATTATCATTTTA
 GAAGACTCTACCAGACTGTGTATTCAATTGAAGTCAGATGTGGTAAGTGTAAATTTGCTGT
 GTATCTGATAGCTCTTTGGCAGTCTATATGTTGTATAATGAATGAGAGAATAAGTCATG
 TTCCTTCAAGATCATGTACCCCAATTTACTTGCCATTACTCAATTGATAAACATTTAACT
 TGTTTCCAATGTTTAGCAAAATACATATTTTATAGAACTTC

TGAACATATTCAGGCTGATTGGGGACGTGTCCACCTGGCGGCCATCGTCATCTTGATGG
TAGAGATCTGGAAGACGCGCTCCTGCGCCGGTATTTCTGGGAAAAGCCAGCTTCTGTCTG
CACTGGTCTTCACAACCTCGTGACCTGGATCTTTTCACTTCATTTATTTCAGTGTATCACA
CATCTATCAAGGTTATCTACGTTGCCTGCTCGTATGCCACAGTGTACCTGATCTACCTTA
AATTTAAGGCAACATCGGATGGAAATCATGATACCTTCCGAGTGGAGTTTCTGGTGGTCC
CTGTGGGAGGCCTCCTCATTTTTTAGTTAATCACGATTTCTCTCCTCTTGAGTACTCAAGG
GAAAGAAGCTCAGTTTGCCAGCATAAGTGCCAAAGACCATCGCCAGCATCTGTCCTTCAG
GGTGTTCCGACAGAATTCTTACCACAGCAAAGGCATAAGATGCTTGATACGGAAAATCAA
GAACTTAACTTTTTTGTTGCAGATAGTCATCAGTGGTTCTGTAAAAACGCAGAGGAAAAG
AGCCAGAAGGTTTCTGTTTAATGCATCTTGCCTTATCTTTTTTTATTACTGTGCACAAAG
ATTTTTTTACACAAACATCCTTAATGCTGTTTTTAATAAATTCAGTGTGTAGCTTCAAAAA
AA

>NM_024423

GGCAGGTCTCGCTCTCGGCACCCCTCCCGGCGCCCGCGTTCTCCTGGCCCTGCCCGGCATC
 CCGATGGCCGCGCTGGGCCCGGCGCTCCGTGCGCGGAGCCGTCTGCCTGCATCTGCTG
 CTGACCCCTCGTGATCTTCAGTCGTGATGGTGAAGCCTGCAAAAAGGTGATACTTAATGTA
 CCTTCTAAACTAGAGGCAGACAAAATAATTGGCAGAGTTAATTTGGAAGAGTGCTTCAGG
 TCTGCAGACCTCATCCGGTCAAGTGATCCTGATTTTCAGAGTTCTAAATGATGGGTCAGTG
 TACACAGCCAGGGCTGTTGCGCTGTCTGATAAGAAAAGATCATTTACCATATGGCTTTCT
 GACAAAAGGAAACAGACACAGAAAAGAGGTTACTGTGCTGCTAGAACATCAGAAGAAGGTA
 TCGAAGACAAGACACACTAGAGAAACTGTTCTCAGGCGTGCCAAGAGGAGATGGGCACCT
 ATTCCCTTGCTCTATGCAAGAGAATTCTTGGGCCCTTTCCCATTTGTTTCTTCAACAAGTT
 GAATCTGATGCAGCACAGAACTATACTGTCTTCTACTCAATAAGTGGACGTGGAGTTGAT
 AAAGAACCCTTTAAATTTGTTTTATATAGAAAGAGACACTGGAAATCTATTTTGCACCTCGG
 CCTGTGGATCGTGAAGAATATGATGTTTTTTGATTTGATTGCTTATGCGTCAACTGCAGAT
 GGATAFTCAGCAGATCTGCCCTCCCACTACCCATCAGGGTAGAGGATGAAAATGACAAC
 CACCCCTGTTTTTACAGAAGCAATTTATAATTTTGAAGTTTGGAAAGTAGTAGACCTGGT
 ACTACAGTGGGGTGGTTTTGTGCCACAGACAGAGATGAACCGGACACAATGCATACGCGC
 CTGAAATACAGCATTTTGCAGCAGACACCAAGGTACCTGGGCTCTTTCTGTGCATCCC
 AGCACAGGCGTAATCACCACAGTCTCTCATTATTTGGACAGAGAGGTTGTAGACAAGTAC
 TCAATTGATAATGAAAGTACAAGACATGGATGGCCAGTTTTTTGGATTGATAGGCACATCA
 ACTTGTATCATAACAGTAACAGATTCAAATGATAATGCACCCACTTTTCAGACAAAATGCT
 TATGAAGCATTTGTAGAGGAAAATGCATTCAATGTGGAAATCTTACGAATACCTATAGAA
 GATAAGGATTTAATTAACACTGCCAATTGGAGAGTCAATTTTACCATTTTAAAGGGAAT
 GAAAATGGACATTTCAAATCAGCACAGACAAAGAACTAATGAAGGTGTTCTTTCTGTT
 GTAAAGCCACTGAATTATGAAGAAAACCGTCAAGTGAACCTGGAAATGGAGTAAACAAT
 GAAGCGCCATTTGCTAGAGATATTCCAGAGTGAACGCTTGAACAGAGCCTTGGTTACA
 GTTCATGTGAGGGATCTGGATGAGGGGCTGAATGCACTCCTGCAGCCCAATATGTGCGG
 ATTAAGAAAACCTTAGCAGTGGGGTCAAAGATCAACGGCTATAAGGCATATGACCCCGAA
 AATAGAAATGGCAATGGTTTAAAGGTACAAAAATTCATGATCCTAAAGGTTGGATCACC
 ATTGATGAAATTTAGGGTCAATCATACTTCCAAAATCCTGGATAGGGAGGTTGAAACT
 CCCAAAATGAGTTGTATAATATTACAGTCTGGCAATAGACAAAGATGATAGATCATGT
 ACTGGAACACTTGCTGTGAACATTGAAGATGTAAATGATAATCCACCAGAAATACCTCAA
 GAATATGTAGTCATTTGCAACCAAAAATGGGGTATACCGACATTTTAGCTGTTGATCCT
 GATGAACCTGTCCATGGAGCTCCATTTTATTTTCAGTTTGCCCAATACTTCTCCAGAAATC
 AGTAGACTGTGGAGCCTCACCAAAGTTAATGATACAGCTGCCCCGTCTTTTCATATCAGAAA
 AATGCTGGATTTCAAGAATATACCATTCTTATTACTGTAAAAGACAGGGCCCGCCAAGCT
 GCAACAAAATTTATGAGAGTTAATCTGTGTGAATGTACTCATCCAACCTCAGTGTGCTGCG
 ACTTCAAGGAGTACAGGAGTAATACTTGGAAAATGGGCAATCCTTGCAATATTACTGGGT
 ATAGCACTGCTCTTTTCTGTATTGCTAACTTTAGTATGTGGAGTTTGTGGTCAACTAAA
 GGGAAACGTTTTCTGAAGATTTAGCACAGCAAAACTTAATTATATCAAAACAGAAAGCA
 CCTGGAGACGATAGAGTGTGCTCTGCCAATGGATTATGACCCAAACTACCAACAACCTCT
 AGCCAAAGTTTTTGTGGTACTATGGGATCAGGAATGAAAAATGGAGGGCAGGAAACCATT
 GAAATGATGAAAGGAGGAAACCAGACCTTGGAAATCCTGCGGGGGGCTGGGCATCATCAT
 ACCCTGGACTCCTGCAGGGGAGGACACACGGAGGTGGACAACCTGCAGATACACTTACTCG
 GAGTGGCACAGTTTTACTCAACCCCGTCTCGGTGAAGAATCCATTAGAGGACACACTGGT
 TAAAAATTAAACATAAAAGAAATTGCATCGATGTAATCAGAATGAAGACCGCATGCCATC
 CCAAGATTATGTCCTCACTTATAACTATGAGGGAAGAGGATCTCCAGCTGGTTCTGTGGG
 CTGCTGCAGTGAAGAGCAGGAAGAAGATGGCCTTGACTTTTTAAATAATTTGGAACCCAA
 ATTTATTACATTAGCAGAAGCATGCACAAAGAGATAATGTACAGTGCTACAATTAGGTC
 TTTGTGACACATTCTGGAGGTTTTCCAAAATAATATTGTAAAGTTCAATTTCAACATGTA
 TGTATATGATGATTTTTTTCTCAATTTTGAATTATGCTACTACCAATTTATATTTTTAA
 AGCCAGTTGTGCTTATCTTTTCCAAAAGTGAAAAATGTTAAACAGACAACTGGTAAA
 TCTCAAACCTCAGCACTGGAATTAAGGTCTCTAAAGCATCTGCTCTTTTTTTTTTTTACG
 GATATTTTAGTAATAAATATGCTGGATAAATATTAGTCCAACAATAGCTAAGTTATGCTA
 ATATCACATTATTATGTATTACTTTTAAAGTGATAGTTTAAAAATAAACAAGAAATATTG
 AGTATCACTATGTGAAGAAAGTTTTGGAAAGAAACAATGAAGACTGAATTAATTAATAA
 ATGTTGCAGCTCATAAAGAATTGGGACTACCCCTACTGCACTACCAAAATTCATTTGACT
 TTGGAGGCAAAATGTGTTGAAGTGCCCTATGAAGTAGCAATTTTCTATAGGAATATAGTT
 GGAAATAAATGTGTGTGTATATTATTATTAATCAATGCAATATTTAAATGAAATGAG
 AACAAAGAGGAAATGGTAAAAACTGAAATGAGGCTGGGGTATAGTTTGTCTTACAATA
 GAAAAAGAGAGAGCTTCTAGGCCTGGGCTCTTAAATGCTGCATTATACTGAGTCTAT

GAGGAAATAGTTCCTGTCCAATTTGTGTAATTTGTTTAAAATTGTAAATAAATTAACTT
 TTCTGGTTTCTGTGGGAAGGAAATAGGGAATCCAATGGAACAGTAGCTTTGCTTTGCAGT
 CTGTTTCAAGATTTCTGCATCCACAAGTTAGTAGCAAACCTGGGGAATACTCGCTGCAGCT
 GGGGTTCCCTGCTTTTTTGGTAGCAAGGGTCCAGAGATGAGGTGTTTTTTTCGGGGAGCTA
 ATAACAAAAACATTTTAAACTTACCTTTACTGAAGTTAAATCCTCTATTGCTGTTTTCTA
 TTCTCTCTTATAGTGACCAACATCTTTTAAATTTAGATCCAAATAACCATGTCTCTCCTAG
 AGTTTAGAGGCTAGAGGGAGCTGAGGGGAGGATCTTACTGAAAGCACCCCTGGGGAGATTG
 ATTGTCCTTAAACCTAAGCCCCACAACTTGACACCTGATCAGGTCTGGGAGCTACAAAA
 TTTCATTTTTCTCCTCACTGCCCTTCTTCTGAGTGGCATTTGGCCTGAATCAAGGAAAGCC
 AGGCCTTGTGGGCCCCCTTCTTTTGGCTTTCTGCTAAAAGCAACACCTCCAGCAGAGATT
 CCTTAAGTGACTCCAGGTTTCCACCATCCTTCAGCGTGAATTAATTTTAAATCAGTTTG
 CTTTCTCCAGAGAAATTTTAAATAATAGAAGAAATAGAAATTTTGAATGTATAAAAGAA
 AAAGATCAAGTTGTCATTTTGAACAGAGGGAACTTTGGGAGAAAGCAGCCCAAGTAGGT
 TATTTGTACAGTCAGAGGGCAACAGGAAGATGCAGGCCTTCAAGGGCAAGGAGAGGCCAC
 AAGGAATATGGGTGGGAGTAAAGCAACATCGTCTGCTTCATACTTTTTCTAGGCTTGG
 CACTGCCCTTTTCTTCTCAGGCCAATGGCAACTGCCATTTGAGTCCGGTGAGGGATCAG
 CCAACCTCTTCTCTATGGCTCACCTTATTTGGAGTGAGAAATCAAGGAGACAGAGCTGAC
 TGCATGATGAGTCTGAAGGCATTTGCAGGATGAGCCTGAACTGGTTGTGCAGAACAAACA
 AGGCATTCATGGGAATTGTTGTATTCTTCTGTCAGCCCTCCTTCTGGGCACATAAGAAGGT
 CTATGAATTAATGCCTATCTAAAATTCTGATTTATTCCTACATTTTCTGTTTTCTAATT
 TGACCCATAAATCTATGTGTTTTAGACTTAGACTTTTTTATTGCCCCCCCCCTTTTTTT
 TTGAGACGGAGTCTCGCTCTGACGCACAGGCTGGAGTGTCAGTGGCTCCGATCTCTGCTCA
 CTGAAAGCTCCGCTCCCGGGTTCATGCCATTCTCCTGCCTCAGCCTCCTGAGTAGCTGG
 GACTACAGGCGCCACCACCAGCGCCGGCTAATTTTTGTATTTTAAATAGAGACGGGGT
 TTCCTGTGTTAGCCAGGATGGTCTCGATCTCCTGACCTCGTGATCCGCTGCCTCGGCT
 TCCCAAAGTGCTGGGATTACAGGCATGACCCACCGCTCCCGGCCTTGTTTTCCGTTTAA
 GTCGCTCTTCTTTAATGTAATCATTTTGAACATGTGTGAAAGTTGATCATACGAATTGGA
 TCAATCTTGAATACTCAACCAAAAGACAGTCGAGAAGCCAGGGGGAGAAAGAACTCAGG
 GCACAAAATATTTGGTCTGAGAATGGAATTCTCTGTAAGCCTAGTTGCTGAAATTTCTGCT
 TGTAACCAGAAGCCAGTTTTATCTAACGGCTACTGAAACACCCACTGTGTTTTGCTCACT
 CCCACTCACCGATCAAAACCTGCTACCTCCCCAAGACTTTACTAGTGCCGATAAACTTTC
 TCAAAGAGCAACCAGTATCACTTCCCTGTTTATAAAACCTCTAACCATCTCTTTGTTCTT
 TGAACATGCTGAAAACCACCTGGTCTGCATGTATGCCCCGAATTTGTAATCTTTTCTCTC
 AAATGAAAATTTAATTTTAGGGATTCAATTTCTATATTTTACATATGTAGTATTATTATT
 TCCTTATATGTGTAAGGTGAAATTTATGGTATTTGAGTGTGCAAGAAAATATATTTTTAA
 AGCTTTCATTTTTCCCCAGTGAATGATTTAGAATTTTTTATGTAAATATACAGAATGTT
 TTTTCTTACTTTTATAAGGAAGCAGCTGTCTAAAATGCAGTGGGGTTTGTTTTGCAATGT
 TTTAAACAGAGTTTATGATTTGCTATTAAAGAAAGTTACTTTGCTTTTAAAGAACTTGG
 CTGCTTAAATAAGCAAAAATTGGATGCATAAAGTAATATTACAGATGTGGGGAGATGT
 AATAAAACAATATTAACCTTGGAAAAAATAAAAAAAAAAAAAA

>AA745593

GACTCAGNCTTCAGCCGCTCTCCTCCCCCTGGGCAACAGGACTCATCTGATGATGTGAG
 AAGAGTTTCAGAGGAGGGAGAAAAATCGTATTGCCGCCCAGAAAGAGCCGACAGAGGCAGAC
 ACAGAAGGCCGACACCCCTGCACCTGGAGAGCGAAGACCTGGAGAAAACAGAACGCGGCTCT
 ACGCAAGGAGATCAAGCAGCTCACAGAGGAACCTGAAGTACTTCACGTCCGGTGTGTAACAG
 CCACGAGCCCCGTGTGCTCGGTGCTGGCCGCCAGCACGCCCTCGCCCCCGAGGTGGTGTA
 CAGCGCCACGCATTCCACCAACCTCATGTGAGCTCCCCGCGCTTCCAGCCCTGAGCTTC
 CGATGCGGGGAGAGCAGAGCCTCGGGAGGGGCACACAGACTGTGGCAGAGCTGCGCCCAT
 CCCGCAGAGGCCCCGTGTCACCTGGAGACCCGGAGACAGAGGCCTGGACAAGGAGTGAAC
 ACCGGAACCTGTACGACTGGAAGGGCGTGAGGCCCTCCAGCAGTGCCGAGCGTTTCGAG
 GGGCGTGTGCTGGACCCACCACTGTGGGTTGTCAGGCCCAATGCAGAAGAGTATTAAGAA
 AGATGCTCAAGTCCATGGCACAGAGCAAGGCGGGCAGGGAACGGTTATTTTTCTAAATA
 AATGCTTTAAAGAAAAAATAAAAAAAAAAAAAA

5

>AI985118

ATGCAAGGNNNTAGGCAAAGATTGTTGACCCNGGAGATAGAGGTNNCAATGAGCCAGATCA
 TTCCATTGCATTCCAGCTTGGGCGACAGAATGAGACTCTGTCTCAAAATTA AAAANCAAA
 AAACCAAAANCAAATAGATGAAAAAGTAGACTGGAGACAAATAAAAGTGAGTTTCTAAAG
 GAAATTACAGTAATGCTGCATTAAACACTAAGCTCACTTAGGTCACTTTCTAGTGAGCT
 AACCGTAACAGAGAGCCTACAGGATACACGTGAGATAATGTCACGTGTAGAAGATCGTTG

TGAATTAAAGTTC AAAATTAAGACTTCTTAGATTATGATGTAGATTTTAGAGCTCCTTAA
 AACATAAAGCGAATCTTATAAATGTTCAATTCTAAAGTTATCCACTTGGAAAAATTAGC
 TTTTGGGACAATTTTAAAGAACTTTTGTGTAAAATGCAGCTCCATGTTTAGCATAATCTA
 AAAATAATTTCAAGCAATCCAGAATCTTCCAAGATGTTATTAAGCTTTAAACAAAGC
 AAAACAAAAGACCCCTTTTGTGCCTTATATGGGAAGACTAAAAAA

>AB038160

ACCGGGCACCGGACGGCTCGGGTACTTTCGTTCTTAATTAGGTCATGCCCGTGTGAGCCA
 GGAAAGGGCTGTGTTTATGGGAAGCCAGTAACACTGTGGCCTACTATCTCTCCGTGGTG
 CCATCTACATTTTGGGACTCGGGAATTATGAGGTAGAGGTGGAGGCGGAGCCGGATGTC
 AGAGGTCTGAAATAGTCACCATGGGGGAAAATGATCCGCTGCTGTTGAAGCCCCCTTC
 TCATTCCGATCGCTTTTGGCCTTGATGATTTGAAAATAAGTCTGTTGCACCAGATGCA
 GATGCTGTTGCTGCACAGATCTGTCACTGCTGCCATTGAAGTTTTC CAATCATCGTC
 ATTGGGATCATTGCATTGATATTAGCACTGGCCATTGGTCTGGGCATCCACTTCGACTGC
 TCAGGGAAGTACAGATGTCGCTCATCTTTAAGTGTATCGAGCTGATAGCTCGATGTGAC
 GGAGTCTCGGATTGCAAAGACGGGGAGGACGAGTACCGCTGTGTCCGGGTGGGTGGTCAG
 AATGCCGTGCTCCAGGTGTTACAGCTGCTTCGTGGAAGACCATGTGCTCCGATGACTGG
 AAGGGTCACTACGCAAATGTTGCCCTGTGCCAACTGGGTTTCCCAAGCTATGTGAGTTCA
 GATAACCTCAGAGTGAGCTCGCTGGAGGGGAGTTCCGGGAGGAGTTTGTGTCCATCGAT
 CACCTCTTGCCAGATGACAAGGTGACTGCATTACACCACTCAGTATATGTGAGGGAGGGA
 TGTGCCTCTGGCCACGTGGTTACCTTGCAGTGCACAGCCTGTGGTCATAGAAGGGGCTAC
 AGCTCACGCATCGTGGGTGGAAACATGTCCTTGCTCTCGCAGTGGCCCTGGCAGGCCAGC
 CTTCA GTTCCAGGGCTACCACCTGTGCGGGGGCTCTGTCACTACGCCCCCTGTGGATCATC
 ACTGCTGCACACTGTGTTTATGACTTGTACCTCCCCAAGTCATGGACCATCCAGGTGGGT
 CTAGTTTCCCTGTTGGACAATCCAGCCCCATCCCACTTGGTGGAGAAAGATTGTCTACCAC
 AGCAAGTACAAGCCAAAGAGGCTGGGCAATGACATCGCCCTTATGAAGCTGGCCGGGCCA
 CTCACGTTCAATGGTACATCTGGGTCTCTATGTGGTTCTGCAGCTCTTCCTTTGTTTCAA
 GAGGATTTGCAATTGCTCATTGAAGCACTTCTATGATGGCTGCTTTATAATCCTTGTCAG
 ATATTAATAATTCCAACCTCCTGATTCATGTTGGTGTGGCATCAGTTGATTATCTTTTCT
 CATTAAAATTGTGATGCTCCTAAAAA

5

>X69699

TTCAGAAGGAGGAGAGACACCGGGCCAGGGCACCCCTCGCGGGCGGGCGGACCCAAGCAG
 TGAGGGCCCTGCAGCCGGCCGGCCAGGGCAGCGCAGGCGCGGGCCGGACCTACGGGAGGA
 AGCCCCGAGCCCTCGGCGGGCTGCGAGCGACTCCCCGGCGATGCCTCACAACTCCATCAG
 ATCTGGCCATGGAGGGCTGAACCACTGGGAGGGGCCCTTTGTGAATGGCAGACCTCTGCC
 GGAAGTGGTCCGCCAGCGCATCGTAGACCTGGCCCCACCAGGGTGTAAAGCCCCTGCGACAT
 CTCTCGCCAGCTCCGCGTCAGCCATGGCTGCGTCAGCAAGATCCTTGGCAGGTACTACGA
 GACTGGCAGCATCCGGCCTGGAGTGATAGGGGGCTCCAAGCCCAAGGTGGCCACCCCCAA
 GGTGGTGGAGAAGATTGGGGACTACAAACGCCAGAACCCTACCATGTTTGCCTGGGAGAT
 CCGAGACCGGCTCCTGGCTGAGGGCGTCTGTGACAATGACACTGTGCCAGTGTGAGCTC
 CATTAATAGAATCATCCGGACCAAAGTGACGCAACCATTCAACCTCCCTATGGACAGCTG
 CGTGGCCACCAAGTCCCTGAGTCCCGACACACGCTGATCCCCAGCTCAGCTGTAACTCC
 CCCGAGTCACCCAGTCCGATTCCCTGGGCTCCACCTACTCCATCAATGGGCTCCTGGG
 CATCGCTCAGCCTGGCAGCGACAAGAGGAAAATGGATGACAGTGATCAGGATAGCTGCCG
 ACTAAGCATTGACTCACAGAGCAGCAGCAGCGGACCCCGAAAGCACCTTCGCACGGATGC
 CTTACGCCAGCACCACTCGAGCCGCTCGAGTGCCCATTTGAGCGGCAGCACTACCCAGA
 GGCTATGCTCCCGCAGCCACACCAAAGGCGAGCAGGGCCTCTACCCGCTGCCCTTGCT
 CAACAGCACCCCTGGACGACGGGAAGGCCACCCCTGACCCCTTCCAACAGCCACTGGGGCG
 CAACCTCTCGACTCACCAGACCTACCCCGTGGTGGCAGATCCTCACTCACCCCTTGGCCAT
 AAAGCAGGAAACCCCGAGGTGTCCAGTTCTAGCTCCACCCCTTGCTCTTTATCTAGCTC
 CGCCCTTTTGGATCTGCAGCAAGTCGGCTCCGGGGTCCCGCCCTTCAATGCCTTTCCCA
 TGCTGCCTCCGTGTACGGGCAGTTACGGGCCAGGCCCTCCTCTCAGGGCGAGAGATGGT
 GGGGCCACGCTGCCCGGATACCCACCCACATCCCCACCAGCGGACAGGGCAGCTATGC
 CTCCTCTGCCATCGCAGGCATGGTGGCAGGAAGTGAATACTCTGGCAATGCCTATGGCCA
 CACCCCTTACTCCTCTACAGCGAGGCTGGGGCTTCCCAACTCCAGCTTGCTGAGTTC
 CCCATATTATTACAGTTCCACATCAAGGCGGAGTGACCCGCCACCACTGCCACGGCCTT
 TGACCATCTGTAGTTGCCATGGGGACAGTGGGAGCGACTGAGCAACAGGAGGACTCAGCC
 TGGGACAGGCCCCAGAGAGTCACACAAAGGAATCTTTATTATTACATGAAAAATAACCAC
 AAGTCCAGCATTCGGGCACACTCCCTGTGTGGTTAATTTAATGAACCATGAAAGACAGGA
 TGACCTTGGACAAGGCCAACTGTCTCCAAGACTCCTTAATGAGGGGCAGGAGTCCCAG

GGAAAGAGAACCATGCCATGCTGAAAAAGACAAAATTGAAGAAGAAATGTAGCCCCAGCC
 GGTACCCCTCAAAGGAGAGAAGAAGCAATAGCCGAGGAACCTGGGGGGATGGCGAATGGT
 TCCTGCCCGGGCCCAAGGGTGACAGGGCACCTCCATGGCTCCATTATTAACACAACCTCT
 AGCAATTATGGACCATAAGCACTTCCCTCCAGCCCAAGTCACAGCCTGGTGGCCGAGGC
 TCTGCTCACCAGCCACCCAGGGAGTCACCTCCCTCAGCCTCCCGCCTGCCCCACACGGAG
 GCTCTGGCTGTCTCTTTCTCCACTCCATTGCTTGGCTCTTTCTACACCTCCCTCTTG
 GATGGGCTGAGGGCTGGAGCGAGTCCCTCAGAAATTCCACCAGGCTGTGAGCTGACCTCT
 TTTTCTCTGCTGCTGTGAAGGTATAGCACCAACCCAGGTCCCTCCTGCAGTGCGGCATCCCCCT
 TGGCAGCTGCCGTCAGCCAGGCCAGCCCCAGGGAGCTTAAACAGACATTCCACAGGGCC
 TGGGCCCCCTGGGAGGTGAGGTGTGGTGTGCGGCTTCACCCAGGGCAGAACAAAGGCAGAAT
 CGCAGGAAACCCGCTTCCCTTCCCTGACAGCTCCTGCCAAGCCAAATGTGCTTCTGCGAG
 CTCACGCCCACCAGCTACTGAAGGGACCCAAAGGCACCCCTGAAGCCAGCGATAGAGGGT
 CCCTCTCTGCTCCCCAGCAGCTCCTGCCCCAAGGCCTGACTGTATATACTGTAAATGAA
 ACTTTGTTTGGGTCAAGCTTCTTTCTTTCTAACCCCGAGACTTTGGCCTCTGAGTGAAAT
 GTCTCTCTTTGCCCTGTGGGGCTTCTCTCTCTGATGCTTCTTTCTTTTTTAAAGACAAC
 CTGCCATTACCACATGACTCAATAAACCATTTGCTCTTCAAAAAAAAAAAAAAAAAAAAAA
 AAAAAAAAAA

>AK025615

TGCTTCATAAAATTTACCTAAGCAAGTGGTCTTGCTTGCCCTCAAATCCAAGCAGTCTTGA
 ACACTTGGAGGCAATTAATGAGTATATCTTAGTCAAAAGAATTGTTGGAGCTTTTTATTA
 AAGCTGCAGTTTCAGTTCTGCTTTTGGGGAATTGTGCTATGAAAGCAGCTGCCAAAATAA
 GCTCATTATTTTCTTCAATCCCAGTCTGCTCAGTCACTATATTCTGTTTCCCTTTTTT
 TTTTCAAGTTGCATATTTGGTTTCCCCTTATGATTGGGAAAGATGAATTTTCAGCAGAA
 AACAGTGTGTTTCACTTTCAAAGAGTGATAGTTTCTAAACATTTAGAGCAATAAATAT
 TCATCAGAGGTACCAAGTAAGCCAGCAGAAGAGTTAAGGGTTAGAGAAATCCCTTATTTT
 ATGTCTTGACTCTAAAATGATCAAAGTACTTTTCTTGTAATGTGGATTCTTCTTATGC
 GGATATGCAAAAACCTTCAGTTATACGTAGTAATGCTAGCAGGTAATTTTAGTGACATTT
 TATAACAACTGTCACTTTGTGTTTGGCCACATGTAGAGTTTGTTCAGCTATTTTCCAGATAT
 CTCCCCACAAAAGGAGGCAAAGGGTACCAGCTTTTCAATGAGCATTACCTATTACTTGGC
 AAAGATGATGAAGACTCTATTAATAGTTCATTTGATAAATGTTGACATAACCAACAATAG
 AGATTAGGAAGTTAGTTTTAAGAAATCAATAGCATATAGACATTACCCTCATGGAGTTTG
 TATTCTACTACTTGAAGTATTGTAGCTATAAAAGCATAGTTAGATAGCTGAATAGTTAG
 ATCATAAGCAAAAGGAGGCCAGAACACATCTCTTATCAAGAAATCAATGAATAGTTTATCT
 CATTTTTAAAGCAACTTTATCCTTCTTTAATTCTTCTTCTTCTTAGTGCAAAACTACT
 TAATAAGGTTGGTGTGTTAGGTTAGTGTTCACACCATTCTCATCTGGTGTGAATTACCTT
 CTCTTTCTTACTATTTACTACCAACCTAGTACATGTGTTGACTGAATCTTTTCAAACA
 ATGTTGAGTTATCATGGTGCACCTAATAAATTAACACCACAGATTACAGCATCCTTGCTG
 ATTTTCTCAGCAAAGCCAGATTAGATGGAATAAAACAAAGAAAATGATCCTAGAGTGAAT
 TTTTCTAGAAAATATCTATTATGAACCATGCTGTTTAAAGTATTAGCTTGAAGGTGATGG
 ATCCAGCTATTTCAGAAAATAACTTTTCATATAACCATGATTTTGCACAGTATGAGGTCTTA
 AATGTGTGGAAGAGATAAAATTTTTTCTCTTCAGCATACATATATAAAAGACTTGTTCAGATG
 GTGGAAGAAAGTGATTTATTTTTTCTCTTCAGCATACATATATAAAAGACTTGTTCAGATG
 TTTAATTTGGGGAGGTTGATAATGAAACATATCAACAGAGTATAGTAGTTATAGTAGTGT
 TTGTGGGTAAATAATTTCTGGGGTCAGACATATATAAAACATATTTGCTTCAAATGATA
 AAGGCATGAAATCAGTCTTAAAAATTGAAATGGGGGTGATGGGGGAGAAAAAGAAGACA
 AATTTGAAGTGCCCTTTCAAATCTGCTGGATACAAGTATTGAAGTTTAAAGTCATCTTAT
 TCTGTCTGAAAGTGTATTTTTTCTTCTACAATAGACCAATCAACAAGACGTATAACTTG
 AGTTGCATGATGTTTCAGTTTATGTAATCTACTGTTGGGATGGTAAGAATTGATGTAGGCT
 GTGGTGTAGAATGAATTAATAATATAGTTTCACTGGCTTTTCTCTACATATCCACTATCA
 CAATGGCTAGGTTTCTGTGCTCACTGTTGGATTCTGGAGAAAAATTTAATGAAAGATG
 ATATCAGAGGAAGAATAAGTGGAGGTAGAGAAGAAAGGAGTGATAGAGGAGGGGAAAAAA
 ACAAACATATTTTTGTGTTATCCAAAGGAGCTTTTTTCTTATTCTGTCAAGCATTGAGA
 TCTTCTTCAGCTTTCAATGTAGTTGCTAAATACAAATAATGCTACTAGGTAGTGACTAAA
 TATAGCAAAACCTTCATCAGATATTAGAATTAGGTACACTATTGAGGTTATAATCTGAA
 GGTGTGTTACATAGAAACCACTTTAGATTATTATCAACTTGGGCTAGGCTTTATTTTAT
 AATAGCATAGTAAGTAATATCTATTGTGTCATTTCTTCAACCATTTTATTCTAAGATCCA
 TGAAGCTTCTTGAGGCCAAATAAAATAAAGTTTAGACAAGAAGTAGATTGTGACTTTT
 TTTCCCTTAGAGATACTATTTACTATCTCTATCTGATAGGTGGAAGGTTTACTGAATT
 GGAATTTGGTTGACTATTAGTTTTTAACTAAAATGTGCAATAACACATTGCAGTTTCTCTC
 AAAC TAGTTTCTTATGATCATTAACCTCATTCTCAGGGTTAAGAAAGGAATGTAAATTC

TGCCTCAATTTGTACTTCATCAATAAGTTTTTGAAGAGTGCAGATTTTTAGTCAGGTCTT
 AAAAATAAACTCACAAATCTGGATGCATTTCTAAATCTGCAAATGTTTCTGGGGTGAC
 TTAACAAGGAATAATCCCACAATATACCTAGCTACCTAATACATGGAGCTGGGGCTCAAC
 CCACTGTTTTTAAGGATTTGCGCTTACTTGTGGCTGAGGAAAAATAAGTAGTTCGAGGAA
 GTAGTTTTTAAATGTGAGCTTATAGATAGAAACAGAATATCAACTTAATTATGAAATTGT
 TAGAACCTGTTCTCTGTATCTGAATCTGATTGCAATTACTATTGTACTGATAGACTCCA
 GCCATTGCAAGTCTCAGATATCTTAGCTGTGTAGTGATTCTTGAAATCTTTTTAAGAAA
 AATTGAGTAGAAAGAAATAAACCTTTGTAAATGAGGCTTGGCTTTTGTGAAAGATCATC
 CGCAGGCTATGTTAAAGGATTTTGTCTCACTAAAAGTGTAATAATGGAAATGTGAAAA
 TATCGTAGGTAAAGGAACTACCTCATGCTCTGAAGGTTTTGTAGAAGCACAAATTAACA
 TCTAAATGGCTTTGTTACACCAGAGCCATCTGGTGTGAAGAACTCTATATTTGTATGTT
 GAGAGGGCATGGAATAATTGTATTTTGTCTGGCAATAGACACATTCTTTATTATTTGCAGA
 TTCTCTCATCAAAATCTGTAATTATGCACAGTTTCTGTTATCAATAAAACAAAGAATCCTG
 TTAACAAAAA

TGGCTCTCTCCTTCAAAGGNCCAGGCCCTGTCCCCCTTTCTCCCCGANTCCAACCCAG
CTCCCCGTGTGAAGAAAAAGTTAAAAAATTTGTTATTTATTTGCTTTTTTGCGTTGGGATG
GGTTCGTGTCCAGTCCCGGGGGTCTGATATGGCCATCACAGGCTGGGTGTTCCCAGCAGC
CCTGGCTTGGGGGGCTTGACGCCCTTCCCCCTGCCCCAGGCCATCATCTCCCCACCTCTCC
TCCCCTCTCCTCAGTTTTGCGGACTGCTTTTCATCTGAGTCACCATTTACTCCAAGCATG
TATCCAGACTTGTCACTGACTTTCTTCTGAGCAGGTGGCTAGAAAAAGAGGCTGTGG
GCAGGAAAGAAAGGCTCCTGTTTCTCATTGTGAGGCCAGCCTCTGGCTTTTCTGCCGTG
GATTCTCCCCCTGTCTTCTCCCCCTCAGCAATTCCTGCAAAGGGTTAAAAATTTAACTGGT
TTTTACTACTGATGACTTGATTAAAAAATAACAAAGATGCTGGATGCTAACTTGATAC
TAACCATCAGATTGTACAGTTTGGTTGTTGCTGTAAATATGGTAGCGTTTTGTTGTTGTT
GTTTTTTTCATGCCCATACTACTGAATAAACTAGTTCTGTGCGGGTAAAAAAAAAAAAA
AAAAAAAAA

>AL137761

CACAAAGAAAAAGAAATACCTGTAGAAGCGCATCGAAAGCTCCTGGAACAGAGTTGTGT
CTCATATTTGCAAAGATGCAGAAAAATAAACCCGGGACATCCAGCTTTCTTTTCTTTTC
TTCTTTGACTATTCTGAGAAGCTATGCGACTAGGAGCACATTTAGGTAAACACGTGGCT
TGAGTAGCCATAAGGCCACTCTTCCCTGTCTGTGACCCGCGCTGGGCCTTTAAGAGAT
ATTGGTGTGTTGAAAAGGGAGGAATCTGTTTGGCCCTCAGATATTTAGTTCAACTGCCTGCA
TTGCTTCCTATTTTGTGTCCAACCTCTGTAGTAGTTAGCACTGGCCTTACCAACATGTAA
AGAAATTTTCTTTACTGCCCATGAGTAGTTGGAGGCAAAGAGAAATTTTTAAAGCGCAG
AAAAAGGCCTGCAGGGAGATGGAATTTGTTCTGCCAGAGAAACGAGATGATAGCTGTATT
TAATAAAGTTACTGACCTCTTGTCAAATTTAAACCGCAAAGAGATGTTTCAAATGC
AGAGAATGTCAGAAAACAAACTACAGGGACCAGACCAGTATAATGTTTAGTTTTCATT
ATACTAACTTTTGTCTAGACTGGAGTTGATTCACTATTTTTTCTTTAACTCCTCAGGAAG
CAAACCTTCCCGATGATGAAGACTTCTTGAAGGATTTTCATGGGTGATTTGGGATCCCAGG
ACCATTGGCTAGTGTGCCTAGGTGACCACATGATTGCTGTTTTACCAGGAATGGCAGCAT
CCCATTGACAAAACAAGTGCTCTGAGAAGGTTTAAATACTACAGAGAATATGGGAACAC
AGACCTTGAAATTTAGCTGAGTTGTAACAGCTGAACTCCAAGAGGTGCTTCTCTTGTGTT
GAGGTGAAACTAGTGTGCTTCCAGAGGGCAGCTGGAAACCGTAAAGCTGTTTGGAAATC
TTTTTGACTGACTTGCTGACAAAGAGGTACTGTGATGCATTTTAAACAATATCTAAGTTGA
TTTTTTTTTAAATCAAGGAAAAATAAAACCAAGCATGAATGCTATGGTATGTGCCCTTT
TGACCATCCTGGGCTGATTAACATCATTTAAATCAAAGTAATCATAAAAAGGCATATTCT
ACTTCAATTATGTGGTCAAATAAGAGTAAACACACACTCACACATGCTGACCCCAATT
GCCAGAGCATTACTGCACTATAAATTACGGTTAATTCCCAAATTATACTACTGTTTATCT
TATTTAACAAGTCAGAAAGCACTTTTAAATAAATTTGAGGGCTACAAGGTCATTCTATTA
ATGTCATTCTCCATTCCGGTTGTAGGCATGTGGAAGTACCCATTAAGATAAGTTAGAG
TTTAAATACTGATAAACAAAACCTTTTATTGCAACTGGACAGTTTCTGGAGAGTTAGCGG
AAGAATCTTGGAGTTTCTTTGGTTCAGATGAATACAACATTTCACTTTTGCAGCACTATT
TAGAATGTACTCCATGGTTCTCTTGTTCCTCAACTTCCAAAAAGAACAGAAAACCTTTGGTT
TACACAGAACACGGGCATCTGAGGCAGGACCTCTTCCCTGCCCTTTGATCTGACTCACAC
CTCCACATATGACGTAATCAACCCAAATTTGACACCAATTCACTCTTTTCTGCAAAGGGC
ATATTTTGAAACAAGGGACAGCCTGAGGGCGGCTATAATGAGAATGTTTCATGGGGGTTAC
TGGGTCCCTAATTCTGAACTTGCTTATGACACCCAGAGTGAATAGATTAGATTTCAGAAC
CTTCTGAGAAATAACCCAAAGAAAATTTGTTACCCAGCCAATTCCTCGAAAGCTTAATAT

CAAAATATATCTTTTCAAGAAGAAAATCGTTAGAGAGAAGAATGTGGAGGGGAGAGAAAT
 GGGTTTCTCATTGATATGATATTTTGTAAACCATTTTCATTTTGAATTATTCAAGTTTGG
 TTAATATTGTATTCTTTTTTCGTAACTATTTTACCGTGAGAGTAGGTCATTGGGTACTT
 AGATATTTATTTTACACAGTTATTAGTCTTCAGATAGTTTTATTTTACTTCATATGATT
 TTAGTTTTTGTCAGTATAATTTTAAATCATGTTTTCTTGGTCATCTCTTTGTGTATATT
 GTGTAATTGGATTTTCATTGACTGCAAGTGGAGTGTGTTGCCACTCAATTCAGTACTCAGT
 ACTATGGTGACTTGTTTTCAAATAAGTCTCAGATACACATTTAGGGAGCCTTTGCTGGCC
 GAAATATAGACTCTGTTCAGGACAGCAGGTCCCTGATCTAAGAATTTTCCCAATGGTTGC
 TCTAAAAATGCTGCTATTTTGTCTGTTCACTGTATTGCACTTAGTTAAAAAGAAGATAATG
 TGAAAGATGAGAGCAGTTTTTTAAAGGATCTTTTCATATACCCAATTCCCTTATTTTCAG
 ATGTCCCATCAATTTTAGATATGAAAGCTTTAAGTAAAAGTGTGTATGCCTTTCTACTGT
 CAGAACAGGATGGATGCAGCCTGGGTCAGATTTATTTAAGATAAAAAATCATGCAGACTCA
 TCATTCATATCATAGGTGAAAAATGTAAAAACCAAATGGTTTCCACTAAAGCCACCAAGA
 TCTTTTAGAAATGTTTGCACCTTTGGTGGTGGCACAGGAAAAGAGAAGAATTCAGCTGGA
 GTGAATTCTAGAAGTAGATATCAGAAACGGGGCATGAAGAACAGGGGAAGTGGGTGGCAT
 CAGACTCCTAAAGAAGTGAGTTAATTTCCCTCCCTTCCATTTCAGATTTCATGCCACAGCT
 CCATATCTTGAGTATGTGTAAGAGGTGAGTTCCTTCTTCAGCCAGGGGCGGTGGCTCATG
 CCTTTAATCCCAATGCTTTGGGAGGCCAAGGTGGGAGGATCACTTGTGCCTTGGGGTTCA
 AGGTTGCAGTGAACCATGATTGCACCACTGCCTCCAGCCTGAGTGACAGAGCAAGACCC
 TGTCTCTAAAAATATATATAAAAAAGTAAAACTAAAGAACTTCTTGCCTAAACCTGAATTA
 CCGCAATTTGCTGAGTGACTTTGAGAAAAATCAGACTGTTTAGTTTCAGTCGGGATGAAAA
 GCTTGCATTTGCTTCCCAAGAATGGGCAATAGTGACGGCTGCAAGGTACTTTTATTTG
 TTCATGAAAGAACGACAATTTTCAAATGTAATTAAACATAATAGAATGTTTAAACTA
 CTGGGCACTGAACTGGAAGAAAAAGGAGGCTTTATTGAACATTCCCTTTTTCAGTTGG
 TTCAAAGTTTCAGCACTGTGGTTATCATTTGGTGATGCCAGAAAACATTAGTAGACTTAGAC
 AATTGCTATGGCAGTTTCTAAACAGAGCTTTTTCTATACACTATTTGCAACTGGAGTGCA
 ATATTGTATATTCTGTGTTAAAGAAATAAAGTATTTTTATCATTTATTAAAAA
 AAAAA

>AF038191

CCATCCAGAACGATGAGGCCGTGGCCCCGCTCATGAAGTACCTGGATGAGAAGCTGGCCC
 TGCTGAACGCCTCGCTGGTGAAGGGGAACCTGAGCAGGGTGCTGGAGGCCCTGTGGGAGC
 TACTCCTCCAGGCCATTCTGCAGGCGCTGGGTGCAAACCGTGACGTCTCTGCTGATTTCT
 ACAGCCGCTTCCATTTACGCTGGAGGCCCTGGTCAAGTTTTTCCACGCAGAGGGTCAGG
 GTTTGCCCCCTGGAGAGCCTGAGGGATGGAAGCTACAAGAGGCTGAAGGAGGAGCTGCGGC
 TGCACAAATGTTCCACCCGCGAGTGCATCGAGCAGTTCTACCTGGACAAGCTCAAACAGA
 GGACCCTGGAGCAGAACC GGTTTGGACGCCTGAGCGTCCGTTGCCATTACGAGGCCGCTG
 AGCAGCGGCTGGCCGTGGAGGTGCTGCACGCCGCGGACCTGCTCCCCCTGGATGCCAACG
 GCTTAAGTGACCCCTTTGTGATTGTGGAGCTGGGCCCCACCGCATCTCTTTCCACTGGTCC
 GCAGCCAGAGGACCCAGGTGAAGACCCGGACGCTGCACCCTGTATACGACGAACTCTTCT
 ACTTTTCCGTGCCTGCCGAGGCGTGCCGCCCGCGCGCGGCTGTGTGTTGTTACCGTCA
 TGGACCACGACTGGCTGTCCACCAACGACTTCGCTGGGGAGGCGGCCCTCGGCCCTAGGTG
 GCGTCACTGGTGTGCCCCGGCCCCAGGTGGGCGGGGTGCAAGGGCTGGGCAGCCTGTCA
 CCCTGCACCTGTGCCGGCCAGAGCCAGGTGAGATCTGCGCTGAGGAGGCTGGAAGGCC
 GCACCAGCAAGGAGGCGCAGGAGTTTCGTGAAGAACTCAAGGAGCTGGAGAAGTGATGG
 AGGCGGACCCCTGAGTCCATCAGCTGCCAGCCCCGGCCCTGGCCCCCACCCCAAGTTCCC
 TGAAGCATCCTCCAGCTCACTGTGGCCAGCTTTGTGCAACCAGGGCCACGGCGCCCCCTC
 CTGTGCTGTGACGTGTGTGCTGCTGGCTGGCCCCGCGGCGCCTACCGCCCTGGCCGTGTCT
 GTCTGGTGTGTGCTGTGAACCCCTGCACCCAACCCACATCTGGGTGGCCAACTTGGCAG
 GACTTGGCCAGCAGCTGCCCAGGACACAGTGCAGGCCAGAGCGGGCTTGACCACCTGGTG
 GGCTTCCCTGCCCGCTTCCTTGGGCTCCCCGGCCCTGGGTGGGCGGTGCGCAGCTGGTCT
 CCAGGGACTCAGTGAGTGGCTGTGCTCTCTGCACAACGGGCAATGTGCAGACGCATTTTT
 GGTAAATCACAGCTGGGGAGTGAAAAGGGTGCCACTGGCACCCTGGGTGGATGGTCCAGA
 GCCTCCACCCACAGAGGGGATGCAAAGGGCAGGTGAGTCAAGAACCGCATAGGTCTCCAG
 TCCCCACGGGGCTCCCAGGCCGGGGAAAAGGTTCCTCTGAGGTCACTCTGAGGCCAGGGAC
 GTCACCCAAGGCTGGTGGTCAGTGTGAAGGGCTCCGTGCCAACTGGTCAGCTGTCTTCA
 CGCACATATCCGTGGCCACCTGAGACCTGCTCCACGACCCTTCAGGCAGAGCCGAGAGT
 TCGCCCCAACCTTCCCCAGGCCAGTGTGAAAAACAGACTCACAAGGGGCTTCTTGGCC
 TGCAGCTTCATTTGCGAGAGCGCCGAGGCAGGACACAGAGCACAGCTGTGCTGGAAGTGT
 GGGGAGAACCCGACAGCTCAGTCTTCCAGCAGCCGCAAAGAGCCGAGGCTGCCAGGCC
 CATTTATGTCCCTCATGTCTCTAGATTTCTCGTCACCCAGCCTCAAAAATATATGTGTC

TGCAACCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>BC016340

GGGGGGGCTCCGTGACAGCCAACGCAGTGACCCCTCGCCCTTCCTTGGCAGCACATCATG
 CTTGTGTCAGCGGCAGATGTCTGTGATGGAAGAGGACCTGGAAGAATTCCAGCTCGCTCTG
 AAACACTACGTGGAGAGTGCTTCCTCCCAAAGTGGATGCTTGGCTATTTCTATACAGAAG
 CTTTCAAATGAATCTCGCTACATGATCTATGAGTTCTGGGAGAATAGTAGTGTATGGAAT
 AGCCACCTTCAGACAAATTATAGCAAGACATTCCAAAGAAGTAATGTGGATTCTTGGA
 ACTCCAGAACTCACATCTACAATGCTAGTTCTGCTTCGTGGTGGATCCTGAACAACCTAG
 ATGTTCTTAGACATTTCTTTATGGTTCCAAGTGCAAAACAGGTGTTCTTATCTAAAACG
 TCAATTAGAAAATTATCTGCGGTTGTTAATCTACTGTATATTTTGTGTTGGTATATTTAC
 TAAGTGCACTCTTTCAAAACCTTATTCTATAACTTTATCAATTCATGTGAATTTTAGCTCA
 ATTTTCAAAGTTCACATAATATTCTCAATATTTAATGCTAAATGCTTTGCTACATTGTAAC
 TCACCTAAAACCTTTTAGTGACAAAATCCTAATATGTGGAAAAAAGCATATGCATAAAGG
 AATAATATTGTGAAAAATGAATCTGTTATGATAAAGAAAAAATAAAGTGGAAACTTTTGA
 GTATTACTTCATAGGGCAGATTTTGTAACTGTCTGATATACTGTAAAGGGTTAAATCAGCG
 TTTTGTGATTTTAAAGTAACTGTGAGTGAAGTTTATTCTTCAACAATGTCTACTCCATCC
 CCAACCCAACTCACAGCCCTATGACTACTATCTTTGCATTAGTTAAAAAGTTAGTATATA
 GGCATCAAAACAACCTTGGCTGTAACTATAGAATCTCTATCCATGTATCAGGTTATAGAC
 TGGTTTTTCAAAGTGAACAATCCTGTGATAAGTTGGAGTACCATTAGTAATACAGCAA
 CATTGTGTCAATTATTAGCATCATAATTCTTTGTTATGTAAGTTAAATATATCAAGAAAG
 AAGAGACTGTTTGGAAAAATGTGGTTCAAGTTTTATGCTATATAGTTTTGGTATGCGATA
 CAGACAGCTAACTTTCTTATGAAAAATACATATTTGCATGTAAACAATGATTTCAAAT
 ACTTGAAAAATAAAATTTTAACCCAAATGAATACTAAGAAATATAAAACAAGCACAAAA
 TCTTAGGGAAGTCATAAAATAGTAGTGAAAGTATTAGACAGAAGACATCTGTTTTCGAAT
 TTCAACACTAGAATGACTAAACTATCTACCTATAGAACTATCTGTAGATAGTATACTAT
 CTACACTCTGCTCAACAAGCTCAGAAATTAATATTTTATAGTAATAAAATCTGTTCTGG
 TTATAAACCTTGCTAATGAAAATACAATACATATAAAAAATGTATAGCCATGTTATTTCT
 AGTATAAATTCCTTTGAACTATAAGTCTTTGAGGAAAAATATAAGGTAAATTTTCTG
 TTTTCCCCCTTTGAAAACTCAGGAAAAAGGAAGATTGAACTAATAAAATTTTATTTCT
 TTAAATATAAATTTGACCTAAATATTTTCTCAAATAATTCATGAAACAGCAACTTTTA
 CCAATACCTTTGTATACTCTCAGTTCTCATTGAGTATAAATAAAATTTTAAATCCTTTC
 ATAGTTCTATTAGAAATAAGTAGTAAATTTTGATATATTGTACATACACACGTGTGTGTG
 TGTGTGTGTGTGTGTGTGTATTTGTGTGCCTCTGGTCAACTCTAAGGATGACAGACACTG
 TGTAAACAACCTGGGTCAACTCTTTTAATTTATATACAAAGCAAAGAACAACATTAATG
 GAGATGCACAATGATTATTCAACAAGCTATATATATGTACAAAGGCAAACAGACACATA
 ACAGTCTCTGCAGACTGATTGTATATAGTAAGAAAAGATCAAAAGACTTTAAACCTAAA
 TGACTTTTGACATACAACTCTTCTTGAGAATGTTTGTGTAAATGGTTTCAAAAATACA
 AATTATAGCCAATCAAAACATTGCTTTGGTTGGTGCAATTAAGTATCCAACCTCAAAAAGC
 ATATCAAATATTTTGGGTACTAGGCAGTTTCCAAAGTAGCATGGTAGTATTACTTGTAA
 AAGGGTTCTGTTTTCATTAAACAGTACTAAGTGAAGGGATCTGCAGATTCCAAATTGGAA
 TAAGCTCTATCATATTCTGAAACAAGAATTAGAATGACTTGAGAACGGGCAAATAACAAA
 GCAAACCAATATAATTATATGGTCATTCTGACCCAGCTCTTATACAAATTATACATGTA
 TTTTGTGTATGTTTGTGAGAGTTGTATGTATGTGAATGTGTGTGAGTGTGTATTACAT
 ACACATATATACTGGAACCTATAGTAGAAAAGGAACTAGTAGGGCCAAAAAAGAA
 AAAAGAAAAAGAAAAAGAAAAAAGAAAAAAGTGGGACCTAAGTATAAATATCTCAT
 CCTAAAGTAAACAATAAGTTTATAGTTAACGAAGATTTTTTTCTATTTAAACCCCATTT
 TCCTAAAGAACAAAAAAGAAAAAAGAAAAAAGAAAAAAGTGGGACCTAAGTATAAATATCTCAT

>BC013282

GGCACGAGGGCAGGGGGAAGGGAAGTGCGGCTCGGTTCGGCGCGGGTGGAGGGGGCGTGAG
 GCCGCCCTACGGTGGCCGTGAGGGACGGCGCTACGGCTCCCACGCTAGGCCAAACGCCT
 CCGGCGGGCCGCGCCCGAGAGCCCCCTTCACCTGCAGGGCGACCCAGCCGCGACGCGTGA
 ACCACGCCCTCAGCCGCCTTGCCAGCGCCCCCAGCCGCGCGCCCCAGCACCATGCGGGCCG
 CCCTGCGCACGGAGCCCCGAGGGACAGGGGCACCCGAGGCCCCGCCCCCTAGCACCGCCG
 GCCGGCCCCGAGGTCCGGGACGCCGGCGCGCGCGGAGAGGGACACCGGGCCGACGCCTC
 CCCCAGGGTCAGCTGCGGGCTCCAGGCCCTAGGCGCCCATGACCCCTACGCCAACCGCC
 GCCTGGACACCGCCGCGCCACTGCGACCTAGCGCGCGCGCGCGGGGCCCAATGCCGG
 TCATGCCCATTCCGCGGGCGGGTGCGCTCCTTCCACGGCCCGCACACCACCTGCCTGCATG
 CGGCCTGCGGGCCCCGTGCGCGCTCCACCTGGCCCGCACCAAGTACAACAACCTTCGACG
 TGTACATCAAGACGCGCTGGCTGTACGGCTTCATCCGCTTCCTACTCTACTTTAGCTGCA

GCCTGTTCCTGCGGCGCTCTGGGGTGCGCTGGCCGCCCTCTTCTGCCTACAGTACCTGG
 GCGTTCGCGTCTGCTGCGCTTCCAGCGCAAGCTGTGCGGTGCTGCTGCTGCTGCTGGGCC
 GCCGGCGCGTGGAATTCGCGCTGGTGAACGAGCTGCTCGTCTATGGCATCCACGTACCA
 TGCTGCTGGTCCGGGGCGCTGGGGCTGGTGCTTCATGGTCTTCGTGGACATGTGAGGGCCGT
 GGGTGCGAGCTTGATGTATCGTCCCGGCCTGTGGCTGTGTTCTCTCCATGGGTGGGGTGG
 GCCAGCGCCTTCCCTTCGCCCATCCCCCAGGCAGTCCGTGCTGCCCGGCGCCACGGAGA
 GAAAAGAAAGGGCTGAGACTTCTGTGATGGGGGCGCGGACACCACCCCTAGGCTGGCTTC
 CTGGACCCACCCTCCCCGTATGCACTCTCAGGGGCGAGCGCCACCCTGCCGGTGGCTCCTG
 CTCACATGTCTTCGGTCTGACTGCGGGGTGGGCCCTCCGTTCGCGCTCTCTGTGGGCCT
 CTCTCCAGGACCACAGCTGCCAGGGAATTTAGACATCACCTGGGAGGCCCTGGACACA
 GAGGGCTGTGTGCCAGGAGCAATTCGGAGGGGGGCCCTCCTGGCTGCACAGCCCCCTC
 TGGCTGCCCTGGCCCCAGCCCCAGCCAACGGGACACGGAAGGCTCCCCCTGCTGACACAC
 CACACTGCCACAAAGCTGCTTACTCTGCCCTGGGCCGCGCTGAGGCCTGGCACTGCCGCG
 GACCACCTGTGTGTGTCATCCTGAGGGGTGTGTGGGTCTGAGTCCCCAGCCAGCCTT
 CAGGGTCCCCTTGGATTGTGTAGATGCAGTCTAGCGGGGGCCGGAGAAGGGCTCAGGTG
 GGAGGGGCTCAGCAGGCTCCAGCTCAGGGGCTGGCCTGGGGGGAACCTTGGGAGCCAG
 GGGCTGACTCCAGCAACACTGGCCTGTCTGCCTGTTCTGGGAGGGCTGTGAGGATGTCTT
 GCAGATGCTCTGGATTTCTGCGGAGGCACCTCCATTCTTTCTGGCTTTTTTTTGGGGGG
 AGGGCTTTGGGCCTCTTTCTTTGAGGGAACACCGTCAAAGAAAGCCTGGGAGATCGAGGC
 TTCAGTGAGCCAGGATGGAACGCGTGTCCAAGTGTCCGAGCAGGCGGCAGAGGCCTC
 AGTGCGGCAACACAGCCCCAGAGCCTGTGTGGCACCAGCAGCATCTTAGAGCCCCAGGT
 ATATGCTGAGATCTTATCTCACGCTGTCTCCAGTGTCTGGGGGGCCCAATGATGGCAC
 AGGGTCAGGTGGGCTGGAGGGGCGCAGATGCCTGTGTTACGGGAGGGTGGCCACCATGGG
 CCGAGGTCTCACCCAGGACCCCTTGCTCTGCTCCTCAGCCTTGCACTCACGGCAGCACTA
 TGGTGGACTGCCCATGGCCGTGTGACTTTGGGGGCAAGTGGGAGGGCGCCCTGAATAATG
 ATTGCAAGGACAACAGGCAGAGGCTACCCTAGAGCAGGACACAGGGTGTGGTACTGACAA
 CCCTAGTGTACCTCAAATCCATGTCCCACACTCTGGGCATGGGTGGGACTTGTGACCC
 TACCCTGTCAGGCGGACCAGTGGCCCAGGAGCCATGAGGACAGTTGTGTGCCACTGGAAG
 AGAAACTTTTTGAAAAACCCTAAATCAGGTAGAGAAAGCAAAAAATCTCTGGCCGTAAAC
 CGTGCTCTCTAATTTATCGGCAGCTTCTGTGGATGACCTCTGATGAGCCCGGGCTGCGTC
 CACGCCCTGGGCAGGTAGGCGGGAGCTTCCCTGCGTGGGCCCTCATTTCTTGCTGCAGAGA
 ATCTTTTGCACCTAAGTCAATGCTGTTTCTCAAAGAAGCTTTGTTTTTTGTTAACGTATTA
 CTCAGAGTCACCCAAGCCTCTTGGCTGAGGGTGAAGGTGGGACGGGAGGCGGGAGGGGGC
 TGGTGGTGCCGCTCGTGCGGTGTCAACGCTGCAGGGAGTTGTGGCACCTTGGTGCCCTCT
 GAGCACCTGGCCGCTGCTGTCCCCGGTGCCTGTGAAATTCGTGATGCCATGACCCACCT
 GCATTAAACCTATTTTTTTTAAATGTGTTAAAAA

>H09748

GNGGAAACACGGGCCAAACCCGTGANTTTGGTGCCCCCTTGTAACCTCANCCCCCTGCAAAN
 CCAAAGACCCCAATGGATTAAAGTTGNNTGGCATTGTACTGGCAAGGCAAAANATTTT
 TAANTACCTTTTCCTAATACTTATTGTATGAGCTTTGNTGTTTACTTGGAGGTTTGTGTC
 TTTTACTACAAGTTTGGAACATTTANTATTGCCTTGCTATTTGTGCTCTGTTTAAAGAAA
 CAGGCACCTTTTTTTTATTATGGATAAAATGTTGAGATGACAGGAGGTCATTTCAATATGG
 CTTAGTAAAAATTTTATTGTTCCCTTTATTCTCTGTACAAGATTTTGGGCCTCTTTTTTTC
 CTTAATGTCACAATGTTGAGTTCAGCATGTGTCTGTCCATTTCAATTTGTACGCTTGTTC
 AAACCAAGTTTGTCTGGTTTCAAGTTATAAAAAATAAATTGGACATTTAACTTGATCTCC
 AAAAAAAAAAAAAAAAAA

>BC001665

GGCACGAGGCAATCTGAGGAGCAGGAGGACCGGGGCGCCGGTGTCTGCGCCTCCTTCT
 CCTTGCTCTCACCTGCGCCTATTAGTCCACGCGCCTTCAAGGCCAGGGGCTACAGCCAG
 ACAGAGAGGGGACAGCAGAGGGAGAGAGAGCACCTGAGGATACAGAGCTGGCACTGGACT
 GCCTTTTTCACCCCCCAGGTGATGAGTGAGGTTGGAAGAACGGAAGATTTAAAAAGCAGCC
 GGGGCCTCCGTATTGAATGAAAGACCCAGTGCAAAGACATCACCATGAACACTAGCATTC
 CTTATCAGCAGAATCCTTACAATCCACGGGGCAGCTCCAATGTCATCCAGTGCTACCGCT
 GTGGAGACACCTGCAAAGGGGAAGTGGTCCGCGTGCACAACAACCACTTCCACATCAGAT
 GCTTCACCTGTCAAGTATGTGGCTGTGGCCTGGCCAGTCAGGCTTCTTCTTCAAGAACC
 AGGAGTACATCTGCACCCAGGACTACCAGCAACTCTATGGCACCCGCTGTGACAGCTGCC
 GGGACTTCATCACAGGCGAAGTCATCTCGGCCCTGGGCCGCACTTACCACCCCAAGTGCT
 TCGTGTGCAGCTTGTGCAGGAAGCCTTTCCCCATTTGGAGACAAGGTGACCTTCAGCGGTA
 AAGAATGTGTGTGCCAAACGTGCTCCCAGTCCATGGCCAGCAGTAAGCCCATCAAGATTC

GTGGACCAAGCCACTGTGCCGGGTGCAAGGAGGAGATCAAGCACGGCCAGTCACTCCTGG
 CTCTGGACAAGCAGTGGCACGTGCTGCTTCAAGTGCCAGACCTGCAGCGTCATCCTCA
 CCGGGGAGTATATCAGCAAGGATGGTGTTCATAGTGTGAGTCCGACTACCATGCCCAGT
 TTGGCATTTAAATGTGAGACTTGTGACCGATACATCAGTGGCAGAGTCTTGGAGGCAGGAG
 GGAAGCACTACCACCCAACCTGTGCCAGGTGTGTACGCTGCCACCAGATGTTTACCCGAAG
 GAGAGGAAATGTACCTCACAGGTTCCGAGGTTTGGCACCCCATCTGCAAACAGGCAGCCC
 GGGCAGAGAAGAAGTTAAAGCATAGACGGACATCTGAAACCTCCATCTCACCCCCTGGAT
 CCAGCATTTGGGTACCCAAACCGAGTCATCTGCGACATCTACGAGAACCTGGACCTCCGGC
 AGAGACGGGCTCCAGCCCAGGTACATAGACTCCCCACCTACAGCCGGCAGGGCATGT
 CCCCCACCTTCTCCCGCTCACCTCACCCTACTACCGCTCTGGTGATTTGTCTACAGCAA
 CCAAGAGCAAAACAAGTGAAGACATCAGCCAGACCTCCAAGTACAGTCCCATCTACTCGC
 CAGACCCCTACTATGCTTCGGAGTCTGAGTACTGGACCTACCATGGGTCCCCCAAAGTGC
 CCGAGCCAGAAGGTTCTCGTCTGGAGGAGAGGAGGATGATTTTGACCGCAGCATGCACA
 AGCTCCAAAGTGGAATTGGCCGGCTGATTCTGAAGGAAGAAATGAAGGCCCGGTTCGAGCT
 CCTATGCAGATCCCCTGGACCCCTCCCCGGAGCTCCACCAGCAGCCGGGAAGCCCTGCACA
 CAGCTGGCTATGAGATGTCCCTCAATGGCTCCCCCTCGGTTCGACTACCTGGCTGACAGTG
 ATCCTCTCATCTCCAAATCTGCCTCCCTGCCTGCCTACCGAAGAAATGGGCTGCACAGGA
 CACCCAGCGCAGACCTCTTCCACTACGACAGCATGAACGCAGTCAACTGGGGCATGCGAG
 AGTACAAGATCTACCCCTTATGAACTGCTGCTGGTGACTACAAGAGGAAGAAACCGACTGC
 CCAAGGATGTAGACAGGACCCGTTTAGAGGGAACTTTTGGAAGAGTGGCTGCTTATGAG
 ATTCCAAAATGAAGTGTGGCCAAACACCGCTCATGGCCATCCTGGATTTTCCCAGTGGCT
 TCCCTTCTGCTCGCCTCCCTGAACAGGGGAGAAAAGCTTAACCTCTCTTCTCCTCTCCAA
 ACCTTTCACCTTGAATGGGTAATGTTTGGTGGGGGCTGTTCTTCTTGGAGAAGCCTTGA
 GTCGGACCATTTTGAGATCATGGAGGAAGGATGAAGAAGTGAAAATGACAATAATGACTC
 TCAAGAGGCTGGCGATGTGACATGGCAAATGTAGAACTGACTTAAATTGAACAAACCTC
 ACTGAGCACCTCTGATGTTGAGCACCTGCTGAATACTGAGCACTGAATGGGGGAGGGGGA
 GGGGAGCACGGGTGAGTCAACCTGGGACTCGGTCTCAGGGATATGCCTACCAATAGCGG
 GTATCGTAAGGCATGTACCCAAACATAACGGATGTAAGGCAGAAAGTGATCGGAGAAGGA
 ATGAGAAAAGTGTGCGTGATGTTAATGAAAAGTCATATGCAGCTAGAGCAGACCCAGGAAA
 GCTTTCTGGAAGAGATTGCATCTGAGGAAATTCAGGAAGGATCTTTGTAGATTGGGGGGA
 GATTCTAAATTTGAAGGGGTGATGGGGTGAGGGGCCAGAGGGAAGTCTGCTGTGTTCTCAT
 GTAGGATGTCAGCCCTCCCTGCAACTTCTCTTTTGGCCAATGTCTTTTCACTTTTCTGTA
 CCCTTTAGAAATCATCCCAGCCAGACGCAATCATGGAAGTTGCCTTATTGTCACTGGTTA
 AGAACTTGGCGAGATTGAAGGGCTTTTGTATTGTTGTTGGATATTTTGTTCCTCATAA
 AAGCACATCATTTCAACCCTAAAAA .

>BC016451

GAAGAATTAGATACTTTTGAGTGGGCTTTGAAGAGCTGGTCTCAGTGTTCAAAACCTGT
GGTGGAGGTTTCCAGTACACTAAATATGGATGCCGTAGGAAAAGTGATAATAAAATGGTC
CATCGCAGCTTCTGTGAGGCCAACAAAAAGCCGAAACCTATTAGACGAATGTGCAATATT
CAAGAGTGTACACATCCACTCTGGGTAGCAGAAGAATGGGAACACTGCACCAAAACCTGT
GGAAGTCTGGCTATCAGCTTCGCACTGTACGCTGCCTTCAGCCACTCCTTGATGGCACC
AACCGCTCTGTGCACAGCAAATACTGCATGGGTGACCGTCCCGAGAGCCGCCGCCCTGT
AACAGAGTGCCCTGCCCTGCACAGTGGAAAACAGGACCCTGGAGTGAGTGTTCAGTGACC
TGCGGTGAAGGAACGGAGGTGAGGCAGGTCTCTGCAGGGCTGGGGACCCTGTGATGGT
GAAAAGCCTGAGTCGGTCAGAGCCTGTCAACTGCCTCCTTGTAATGATGAACCATGTTTG
GGAGACAAGTCCATATTCTGTCAAATGGAAGTGTGGCACGATACTGCTCCATACCAGGT
TATAACAAGTTATGTTGTGAGTCTGCAGCAAGCGCAGTAGCACCCCTGCCACCACCATAC
CTTCTAGAAGCTGCTGAAACTCATGATGATGTCTCTAACCCTAGTGACCTCCCTAGA
TCTCTAGTGATGCCTACATCTTTGGTTCCCTTATCATTCAGAGACCCCTGCAAAGAAGATG
TCTTTGAGTAGCATCTCTTCAGTGGGAGGTCCAAATGCATATGCTGCTTTTCAGGCCAAAC
AGTAAACCTGATGGTGCTAATTTACGCCAGAGGAGTGCTCAGCAAGCAGGAAGTAAGACT
GTGAGACTGGTCACCGTACCATCCTCCCCACCCACCAAGAGGGTCCACCTCAGTTCAGCT
TCACAAATGGCTGCTGCTTCTTTTGAGCCAGTGATTCAATAGGTGCTTCTTCTCAG
GCAAGAACCCTCAAAGAAAGATGGAAAGATCATTGACAACAGACGTCCGACAAGATCATCC
ACCTTAGAAAGATGAGAAAGTGAACCAAAAAGGCTAGAAACCAGAGGAAAACCTGGACAA
CCTCTCTCTCCCATGGTGCATATGCTTGTGTTAAAGTGGAAATCTCTATAGATCGTCAGC
TCATTTTATCTGTAATTGGAAGAACAGAAAGTGTCTGGCTCACTTTCTAGTTGCTTTTCATC
CTCCTTTTGTCTGCATTGACTCATTTACCAGAATTCATTGGAAGAAATCACCAGAAAT
ATTACAAAAGAAAAATATGTTGCTAAGATTGTGTTGGTCTGCTCTCTGAAGCAGAAAAGGG
ACTGGAACCAATTGTGCATATCAGCTGACTTTTTGTTGTTTGTAGAAAAGTTACAGTAAA

AATTA AAAAGAGATACCAATGGTTTACACTTTAACAAGAAATTTTGGATATGGAACAAAG
AATTCTTAGACTTGTATTCCTATTTATCTATATTAGAAATATTGTATGAGCAAATTTGCA
GCTGTTGTGTAAATACTGTATATTGCAAAAATCAGTATTATTTTAAGAGATGTGTTCTCA
AATGATTGTTTACTATATTACATTTCTGGATGTTCTAGGTGCCTGTCTGTTGATTTGCC
TTGTTTGGACATTTCTATAGGTTAATTTTCAAAGCAGAGTATTACAAAAGCAAGTTAGAAT
TACAGCTACTGACAATATAAAGGGTTTTGTTGAATCAACAATGTGATACGTAATTTATAG
AAAAAGAAAAGAAACACAAAAGCTATAGATATACAGATATCAGCTTACCTATTGCTTCT
ATACTTATAATTTAAAGGATTGGTGTCTTAGTACACTTGTGGTCACAGGGATCAACGAAT
AGTAAATAATGAACTCGTGCAAGACAAAACCTGAAACCCTCTTCCAGGACCTCAGTAGGC
ACCGTTGAGGTGTCTTTGTTTTGTGTGTGTGTTCTTTTTTAATTTTCGCATTGTTG
ACAGATACAAACAGTTATACTCAATGTACTGTAATAATCGCAAAGGAAAAAGTTTTGGGA
TAACCTATTTGTATGTTGGTAGCTGAGAAAAATATCATCAGTCTAGAATTGATATTTGAG
TATAGTAGAGCTTTGGGGCTTTGAAGGCAGGTTCAAGAAAGCATATGTCGATGGTTGAGA
TATTTATTTTCCATATGGTTCATGTTCAAATGTTCAACCCACAATGCATCTGACTGCAA
TAATGTGCTAATAATTTATGTCAAGTAGTCACCTTGCTCAGCAAGGCCAGAAATGCTCT
CTCCAGGGAGTAGATGTAAAGTACTTGTACATAGAATTCAGAACTGAAAGATATTTATTAA
AAGTTGATTTTTTTTTCTTGATAGTATTTTTATGTACTAAATATTTACACTAATATCAAT
TACATATTTTGGTAAACTAGAGAGACATAATTAGAGATGCATGCTTTGTTCTGTGCATAG
AGACCTTTAAGCAAACCTACTACAGCCAACTCAAAGCTAAAACCTGAACAAATTTGATGTT
ATGCAAACATCTTGCAATTTTAGTAGTTGATATTAAGTTGATGACTTGTTCCTTCAAG
GAAACATTAATTTGTATGGACTCAGCTAGCTGTTCAATGAAATTTGTGAATTAGAAACATT
TTTAAAAGTTTTTGAAAGAGATAAGTGCATCATGAATTACATGTACATGAGAGGAGATAG
TGATATCAGCATAATGATTTGAGGTGAGTACCTGAGCTGTCTAAAAATATATTATACAA
ACTAAAATGTAGATGAATTAACCTCTCAAAGCACAGAATGTGCAAGAACTTTTGCAATTTT
AATCGTTGTAAACTAACAGCTTAAACTATTGACTCTATACCTCTAAAGAAATGCTGCTAC
TTTGTGCAAGAACTTTGAAGGTCAAATTAGGCAAATTCAGATAGTAAAAACAATCCCTAA
GCCTTAAGTCTTTTTTTTTTCTTAAATTTCCCATAGAATAAAATTTCTCTAGTTTAC
TTGTGTGTGCATACATCTCATCCACAGGGGAAGATAAAGATGGTCACACAAACAGTTTCC
ATAAAGATGTACATATTCATTATACCTTGACCTTTGGGCTTTCTTTTCTACTAAGCTAA
AAATTCCTTTTTTATCAAAGGTGACACTACTGATGCTGTTTGTGTTGACTGAGAGCACGTAC
CAATAAAAATGTTAACAAAATATAAAAAA

TCCTGTGTTCTAGACCTCTGGAGGCTGCTGTGGGGACCACACTGATCCTGGAGAAAAGGG
 ATGGAGCTGAAAAAGATGGAATGCTTGCAGAGCATGACCTGAGGAGGGAGGAACGTGGTC
 AACTCACACCTGCCTCTTCCTGCAGCCTCACCTCTACCTGCCCCCATCATAAGGGCACTG
 AGCCCTTCCCAGGCTGGATACTAAGCACAAAGCCCATAGCACTGGGCTCTGATGGCTGCT
 CCACTGGGTTACAGAATCACAGCCCTCATGATCATTCTCAGTGAGGGCTCTGGATTGAGA
 GGGAGGCCCTGGGAGGAGAGAAGGGGGCAGAGTCTTCCCTACCAGGTTTCTACACCCCCG
 CCAGGCTGCCCATCAGGGCCCAGGGAGCCCCAGAGGACTTTATTCCGACCAAGCAGAGC
 TCACAGCTGGACAGGTGTTGTATATAGAGTGGAATCTCTTGATGCAGCTTCAAGAATAA
 ATTTTTCTTCTTTTCAAAAATGTATAAAAATCATTATACATAGCATTAAAGAAACATT
 TTTGAGAAGTACAAAACAAAAA

>AF301598

CGGGCGCCGAGGAGCGAGTGAGCTGGGAGCGAGGGGCGAAGGCGCGGAGAAGCCCGGCC
 GCCCGGTGGGCGGCAGAAGGCTCAGCCGAGGCGGCGGCGCCGACTCCGTTCCTCTCTCGG
 CCCGGATCCAGGCCTCCGGGTTCCAGGCGCTCACCTCCCTCTGACGCACTTTAAAGAGT
 CTCCCCCTTCCACCTCAGGGCGAGTAATAGCGACCAATCATCAAGCCATTTACCAGGCT
 TCGGAGGAAGCTGTTTATGTGATCCCCGCACTAATTAGGCTCATGAACTAACAAATCGTT
 TGCACAACCTGTGAAGAAGCGAACACTTCCATGGATTGTCCTTGGACTTAGGGCGCCCTG
 CCCGCCCTTTTGCAGAGGAGAAAAAACTTTTTTTTTTTTTTGCTCCCCGAGAACTTTCC
 CCCCTTCTCCTCCCTGCCTCTAACTCCGATCCCCCACGCCATCTCGCCAAAAA
 AAAAAAAAAAAGAAAAAAGAAAAAAGAAAAAATTACCCCAATCCACGCCT
 GCAAATTCTTCTGGAAGGATTTTCCCCCTCTCTTCAGGTTGGGCGCGTTTGGTGCAAGA
 TTCTCGGGATCCTCGGCTTTGCCTCTCCCTCTCCCTCCCCCTCCTTTCTTTCTTT
 CCTTTCCTTTCTTTCTTCTTCTTCCCTTCCCCCACCCCCACCCCCACCCCAACAAACGAG
 TCCCAATTCTCGTCCGTCTCTCGCCGCGGGCAGCGGGCGGCGGAGGCAGCGTGCGGCGGT
 CGCCAGGAGCTGGGAGCCAGGGCGCCCGCTCTCGGCGCAGCATGTTCCAGCCGGCGCC
 CAAGCGCTGCTTCACCATCGAGTCGCTGGTGGCCAAGGACAGTCCCTGCCCGCCTCGCG
 CTCCGAGGACCCCATCCGTCCCGCGGCACTCAGCTACGCTAACTCCAGCCCCATAAATCC

GTTCCTCAACGGCTTCCACTCGGCCGCCGCCGCCGCCGGTAGGGGCGTCTACTCCAA
 CCCGGACTTGGTGTTTCGCCGAGGCGGTCTCGCACCCGCCCAACCCCGCCGTGCCAGTGCA
 CCCGGTGCCGCCGCCGCACGCCCTGGCCGCCACCCCTACCCCTCCTCGCACTCGCCACA
 CCCCCTATTGCGCTCGCAGCAGCGGGATCCGTCCACCTTCTACCCCTGGCTCATCCACCG
 CTACCGATATCTGGGTCATCGCTTCCAAGGGAACGACACTAGCCCCGAGAGTTTCTTTT
 GCACAACGCGCTGGCCCGAAAGCCCAAGCGGATCCGAACCGCCTTCTCCCCGTCCCAGCT
 TCTAAGGCTGGAACACGCCTTTGAGAAGAATCACTACGTGGTGGGCGCCGAAAGGAAGCA
 GCTGGCACAACAGCCTCAGCCTCAGGAACTCAGGTAAAAGTATGGTTTCAGAACCGAAG
 AACAAAGTTCAAAGGCAGAAGCTGGAGGAAGAAGGCTCAGATTGCAACAAAAGAAAAA
 AGGGACGCACCATATTAACCGGTGGAGAATCGCCACCAAGCAGGCGAGTCCGGAGGAAAT
 AGACGTGACCTCAGATGATTAATAAACATAAACCTAACCCACAGAAACGGACAACATGGA
 GCAAAAGAGACAGGGAGAGGTGGAGAAGGAAAAAACCTACAAAACAAAAACAAACCGCA
 TACACGTTACCGAGAAAGGGAGAGGGAATCGGAGGGAGCAGCGGAATGCGGCGAAGACT
 CTGGACAGCGAGGGCACAGGGTCCCAACCGAGGCCGCGCAAGATGGCAGAGGATGGAG
 GCTCCTTCATCAACAAGCGACCCCTCGTCTAAAAGAGGCAGCTGAGTGAGAGACACAGAGAG
 AAGGAGAAAGAGGGAGGGAGAGAGAGAAAGAGAGAGAAAGAGAGAGAGAGAGAGAGAG
 AGAAAGCTGAACGTGCACTCTGACAAGGGGAGCTGTCAATCAAACACCAAACCGGGGAGA
 CAAGATGATTGGCAGGTATTCCGTTTATCAGTCCACTTAAAAAATGATGATGATGATA
 AAAACCACGACCCAACCAGGCACAGGACTTTTTTGTTTTTTGCCTTCGCTGTGTTTCCC
 CCCCATCTTTAAAAATAATTAGTAATAAAAAACAAAAATTCATATCTAGCCCCATCCCA
 CACCTGTTTCAAATCCTTGAAATGCATGTAGCAGTTGTTGGGCGAATGGTGTTTAAAGAC
 CGAAAATGAATTGTAATTTTCTTTTCTTTTAAAGACAGGTTCTGTGTGCTTTTATTTT
 GATTTTTTTTCCCAAGAAATGTGCAGTCTGTAAACACTTTTTTGATACCTTCTGATGTCAA
 AGTGATTGTGCAAGCTAAATGAAGTAGGCTCAGCGATAGTGGTCTCTTACAGAGAAACG
 GGGAGCAGGACGACGGGGGGGCTGGGGGTGGCGGGGAGGGTGCCCAAAAAAGAAATCAG
 GACTTGTACTGGGAAAAAAACCCCTAAATTAATTATATTTCTTGGACATTCCCTTTCCTA
 ACATCTGAGGCTTAAACCCCTGATGCAAACTTCTCCTTTCAGTGGTTGGAGAAATTGGC
 CGAGTTCAACCATTCAGTCAATGCCTATTCCAAACTTTAAATCTATCTATTGCAAAACC
 TGAAGGACTGTAGTTAGCGGGGATGATGTTAAGTGTGGCCAAGCGCACGGCGGCAAGTTT
 TCAAGCACTGAGTTTCTATTCCAAGATCATAGACTTACTAAAGAGAGTGACAAATGCTTC
 CTTAATGTCTTCTATACCAGAATGTAAATATTTTTGTGTTTTGTGTTAATTTGTTAGAAT
 TCTAACACACTATATACTTCCAAGAAGTATGTCAATGTCAATATTTTGTCAATAAAGATT
 TATCAATATGCCAAAAA

>Hs.77031_ARNm_1 gi|16741772|gb|BC016680.1|BC016680 clon de Homo sapiens MGC:21349
 IMAGE:4338754 poliA = 3

GTGGCGGCGGAGGCGGCGGAGGCCAGGGAGGAAGATGTCGTAATGAGCGATCCACAGACC
 AGCATGGCTGCCACTGCTGCTGTGAGTCCCAGTGACTACCTGCAGCCTGCCGCCTCCACC
 ACCCAGGACTCCCAGCCATCTCCCTTAGCCCTGCTTGCTGCAACATGTAGCAAAATTGGC
 CCTCCAGCAGTTGAAGCTGCTGTGACACCTCCTGCTCCCCCACAGCCACACCGCGGAAA
 CTTGTCCCTATCAAACCTGCCCTCTCCCTCTCAGCCCCGGCAAGAATAGCTTTGGAATC
 TTGTCTCCAAAGGAAATATACTTCAGATTGAGGGGTGCAAACTGAGCGCCTCCTATCCT
 GGAGGGCAGCTGGTGTTCGCTATCCAGAATCCCACCATGATCAACAAAGGGACCCGATCA
 AATGCCAATATCCAGTACCAGGCGGTCCCTCAGATTCAGGCAAGCAATTCCCAAACCATC
 CAAGTACAGCCCAATCTCACCACCCAGATCCAGATCATCCCTGGCACCAACCAAGCCATC
 ATCACCCTCACCCTGTCAGTCACAAGCCTGTCCCATCAAGCCAGCCCCCATCCAGAAG
 TCGAGTACGACCACCACCCCGTGCAGAGCGGGGCAATGTGGTGAAGTTGACAGGTGGG
 GGCGGCAATGTGACGCTCACTCTGCCCCTCAACAACCTCGTGAACGCCAGTGACACCGGG
 GCCCCTACTCAGCTCCTCACTGAAAGCCCCCAACCCCGCTGTCTAAGACTAACAAGAAA
 GCAAGGAAGAAGAGCCTTCCTGCCTCCAGCCCCCTGTGGCTGTGGCTGAGCAGGTGGAG
 ACGGTGCTGATCGAGACCACCGCGGACAACATCATCCAGGCAGGAAATAACCTGCTCATT
 GTTCAGAGCCCTGGTGGGGGCCAGCCAGCTGTGGTCCAGCAGGTCCAGGTGGTGGCCCCC
 AAGGCCGAGCAGCAGCAGGTGGTACAGATCCCCCAGCAGGCTCTGCGGGTGGTGCAGGCG
 GCATCTGCCACCCTCCCCACTGTACCCCAAGGCCCTCCAGAACTTTTCAGATCCAGGCA
 GCTGAGCCGACACCTACTCAGGTCTACATCCGCACGCCCTCCGGTGAGGTGCAGACAGTC
 CTTGTCCAGGACAGCCCCCAGCAACAGCTGCAGCCACCTCTAACACCACCTGTAGCAGC
 CCTGCATCCCGTGTCTCCCATCTGAGTGGGACCAGCAAAAAGCACTCAGCTGCAATTCTC
 CGAAAAGAGCGTCCCTTGCCAAAGATTGCCCCAGCCGGGAGCATCATCAGCCTGAATGCA
 GCCCAGTTGGCGGCAGCTGCCCAGGCAATGCAGACCATCAACATCAATGGTGTCCAGGTC

CAGGGCGTGCCCTGTCACCATCACCAACACAGGCGGGCAGCAGCAGCTGACAGTGCAGAAT
 GTTTCTGGGAACAACCTGACCATCAGTGGGCTGAGCCCCACCCAGATCCAGCTGCAAATG
 GAACAAGCCCTGGCCGAGAGACCCAGCCCGGGGAGAAGCGGCGCCGCATGGCCTGCACG
 TGTCCTCAACTGCAAGGATGGGGAGAAGAGGTCTGGAGAGCAGGGCAAGAAGAAGCAGTG
 TGCCACATCCCGACTGTGGCAAGACGTTCCGTAAGACGTCCTTGCTGCGTGCCCATGTG
 CGCCTGCACACTGGCGAGCGGCCCTTTGTCTGCAACTGGTTCTTCTGTGGGAAGAGGTTT
 ACACGGAGTGACGAGCTCCAACGGCATGCTCGCACCCACACAGGGGACAAACGCTTCGAG
 TGCGCCCACTGTGAGAAGCGCTTCATGAGGAGTGACCACCTCACCAAGCATTACAAGACC
 CACCTGGTCACGAAGAACTTGTAAGGCCAACTGCGGCGGGAGGCCCTGAAGATGCAGTCC
 CCCACCTGTGTCTCCCTGGGCCCCCTGGTGGAAAGGAGCCCTGTGGCTGCCCTTGGGCCTG
 CCCTCAGCCCCACTCCTGTTCTGCAACTGTCCCCACAGGAAGGGGCTCTGTTCCCTGTAT
 TGTCTCCTTCTGAAGCCCCCTTGGCTCTGCCTTGGCCCTTCCCCTCACCACGAGCTCCCG
 GCCTGCCCAGACTGTGGACACTGGCCGTGCCCAATGAGACGTTCTAAACCAGGACGCGTG
 GGAACCCCTATTTCCAAAGGAAAAACATGCATTTCACTCCGTGAGGAGCAAAGTGAGCC
 CCTACCCCCACCCCGATCCCCGCTCCCAACACTGCCGGAGTCGCGTCATGCCATGCCCC
 CTCTCCTGCACCTCCCTGGCCCTGCCGGCCACTGTGGACGCCCTGGGGCTTGGCACCCAC
 CTCTGGAGAACTCGGGGCCACCTCCACTCCATGTGCCCAGCCCCGCCACAACCTCTCCT
 CCAGCACATTCCAGCTCTATTTAAAAAGTAAAGACACCCACCGACTCCTGATCCCCCTCT
 TTTTCTATGGAGAACGTTGCCCTTATACTCTCTACTTCAGATGATGAACACTGTGTACTGT
 GTGTGCTTTAAAGAAGTTTATTTAATTGCTCCCTTCTTCCTTTCCTTGTATTACCTC
 CCTGATGCCTGCTTTCAGTTGAGGGTTGGGGGCAATGATGAGCATATGAATTTTTCTCA
 CTCTAGCAATTCCTTTTCTAAATGACACAGCATTAACTCAAATCTGGATTTCAGATAA
 CAGCACCTGCACATCCTGCACCTCCTCCCTCTCCCTCACCTCACCCTGCCCGGCCCAA
 GCTCTACTTGTGTACAGTGTATATTGTATAATAGACAATTGTGTCTACTACATGTTTAA
 AACACATTGCTTGTATTATTTTGAGGCTTTTAAATTAAACAAAAATCCAACTTTAAAAAA
 AAAAAAA

>Hs.77541_ARNm_1 gi|12804364|gb|BC003043.1|BC003043 clon de Homo sapiens MGC:4370 IMAGE:2822973
 polIA = 3

CCGCGCTCGGTGCCCCGCGCCCTCCCCGGGCCCCGCCATGGGCCTCACCGTGTCCGCGCT
 CTTTTCGCGGATCTTCGGGAAGAAGCAGATGCGGATTCTCATGGTTGGCTTGGATGCGGC
 TGGCAAGACCACAATCCTGTACAACTGAAGTTGGGGGAGATTGTCACCACCATCCCAAC
 CATAGGCTTCAATGTAGAAACAGTGGAAATATAAGAACATCTGTTTCACAGTCTGGGACGT
 GGGAGGCCAGGACAAGATTTCGGCCTCTGTGGCGGCACTACTCCAGAACACTCAGGGCCT
 CATCTTTGTGGTGGACAGTAATGACCGGGAGCGGGTCCAAGAATCTGCTGATGAACTCCA
 GAAGATGCTGCAGGAGGACGAGCTGCGGGATGCAGTGCTGCTGGTATTTGCCAACAAGCA
 GGACATGCCCCAACGCCATGCCCGTGAGCGAGCTGACTGACAAGCTGGGGCTACAGCACTT
 ACGCAGCCGCACGTGGTATGTCCAGGCCACCTGTGCCACCCAAGGCACAGGTCTGTACGA
 TGGTCTGGACTGGCTGTCCCACGAGCTGTCAAAGCGCTAACAGCCAGGGGCAGGCCCCCT
 GATGCCCCGAAGCTCCTGCGTGCATCCCCGGATGACCATACTCCCGGACTCCTCAGGCAG
 TGCCCCTTTCTCCCACTTTTCTCCCCCATAGCCACAGGCCTCTGCTCCTGCTCCTGCCT
 GCATGTTCTCTCTGTGTTGGAGCCTGGAGCCTTGCTCTCTGGGCACAGAGGGGTCCACT
 CTCCTGCCTGCTGGGACCTATGGAAGGGGCTTCTTGCCCAAGGCCCCCTCTTCCAGAGGA
 GGAGCAGGGATCTGGGTTTCTTTTTTTTTTCTGTTTGGGTGTACTCTAGGGGCCAGGT
 TGGGAGGGGGAAGGTGAGGGCTTCGGGTGGTGCTATAATGTGGCACTGGATCTTGAGTAA
 TAAATTTGCTGTGGTTTGAAAAAAAAAAAAAAAAAAAAA

>Hs.7001_ARNm_1 gi|6808256|emb|AL13727.1|HSM802274 ARNm de Homo sapiens; ADNc
 DKFZp434M0519 (del clon DKFZp434M0519); cds parcial poliA = 3

5

GTGGCGGTGGCTGCGGCGACGGCAGAGGCGAAGGGAGCCGGATCGCCGACCTGAGCGGGA
 GGCGGCGGTGGCGGCCATGGCGGCAGATGAGAGCGTTCCCCGCTGCTGTCTGAGCCCAT
 CGACGGTGGCGCGGGCGGCAACGGTTTAGTGGGGCCCGGCGGAGTGGGGCTGGGCCCGG
 GGGAGGCCTGACCCCTCCGCACCACCGTACGGAGCCGGTAAACATGCCCCGCCCCAGGG
 TAAGCCGGGGCGGGTCCGAGGTGCTCCCCGGGGTACTCTGAAAGCCGGGGAGGGGGCGGG
 ACCGAGGGCGGAGGCGGGTCCCAGTCGCCAGGTGCGGGACTGCTGCACCTGTGACTGGGC
 GAGGCTTCCTTCCCTCCGTAATCGCGACCACAGCCTAGGGACGGAAGGGGGTCTGAGCA
 ACCTGATAGAAGTGCCAATTATGAGAAGCCCTCCGAGCTTGGTCAGAGGGTTGAAGATCA
 GAAGGACTTCCCTACCACCGTGGAGCATCAGTGGGGGTGTAAGTGATCCCAGCCCTTCTA
 TTTGCTTCCTCTCCAGCATTTCCCCCGTTTCCCGAGGGGCATCCAGCCGTGTTGCCTGGG
 GAGGACCCACCCCCCTATTACCCCTTAACTAGCCCGGACAGTGGGAGTGCCCCATGATC
 ACCTGCCGAGTCTGCCAATCTCTCATCAACGTGGAAGGCAAGATGCATCAGCATGTAGTC
 AAATGTGGTGTCTGCAATGAAGCCACCGTGAGTTACACATATCTATGAAATGGGCCCTGT
 TTCCTGGATCCTCTTCTGATGTCTTGGTTCTAGACCTGACCTTCCGGCTATTAGCCAA

GTGCTTTTGATGATACCCAGGTTTCAGTTCCAGGTGTCTCACACAGCCATTTCCCCAGAA
 GCCACTCACCAAAGCTAATGTTCACTTTCTCTCACTTTTACACCTAGCCTAGTTCCCTATT
 TGCAAATCTCATGATATAGTCTTTCTTTTATTTCTCCTTCTGGTTAGCACCTTATTTTT
 CTGATCTCATAAAGTGTTTTTGGAGGGGAAGTGGAGGGGATTGGGATTAGAGGTTTGCTTG
 CTGATGACCCCTATTATTCTCTAGCCAATCAAGAATGCACCCCAAGGAAATATGTTT
 GATGCCCCCTGTAACCTGTCTCCTTATCTGCAAGTGACATCCCAACGGATTGCATGCCCTC
 GTCCCTACTGGTAAGAGGCATAAGGTGGGGAAGGGCCTAAGTGGGGAAGTGGAAAGTCAA
 AAAAGGATGAGCGTATACAGAGAATGTAAAGGTGAGAGAGCCTAGTGTTTATTAGGAGA
 AAAGGCTTTGAAGCATGTGCCTCAGGAATGTTATAGCTGTCTTTCTCGTTTCTCAATAAA
 AATATTGAGATGAAATGATGTCTGTTTCGGAGAATAGAGAGCCTTGGGGACTGGGTGTGTT
 ATCTTGAGGTGCGAGGGGAATTGGGGACCTGAAGTTTAAACAGTGCTCTTTCTTTCTCAA
 GGATTCTTGAGGGTATACAGTTGGGGACAGAGTATCTTAAGTACAGAGAAGTCGAGTGA
 CTTAATAGACAGGGAGTGGGGGATGTGGAACAGGGACTGTGAAGATTTTATAGGATTAATA
 ATTTTTCAAACACAAGTTTGAAAATACAAGTCTTTTTCTTTTGTATAGCAAAAGAATCAT
 CAACCTGGGGCCTGTGCATCCCGACCTCTGAGTCCAGAACCCCAACCCATGGGTGTGAG
 GGTATCTGTGGACATTGCAAGAATACTTTTCTGGTGAGGAAGGGGTATTGGGAAGGGGA
 GGGGAAGGAGACTAAGAGTCATTTTCGAGTATATTTCTTAGAGTAATGGTAATGACCCCT
 GAAAGGTCTGTCTATGGGAACATGTTCTGCATCCCCACCCCAAGGTTCTCATTTAGGGGA
 GACCTGTCTGTGCTATTATTTTTGTTTTCTTTCTCCATAGTGGACAGAGTTTACAGACC
 GCACTTTGGCACGTTGTCTCACTGCAGGAAAGTGTATCTATTGGGCGCAGATACCCAC
 GTAAGAGATGTATCTGCTGCTTCTTGCTTGGCTTGCTTTTGGCAGTCACTGCCACTGGCC
 TTGCCGTGAGTACCCTTGCCCCAACCTCTTTCACTCTGCAGCCTCATCTCCATAGGCTAA
 GATTTGGGAAACTGCTACCCCTAAAAAAGTGGAAGAACTTAGGGGACTAGTTTGTGTTT
 GTTTTAAGATATGGATGAGCTAAGGTGCAAGTGCGCTGATCAAACAGACTTTATTACTAC
 TACAAGAGTGAAAAACAGCCTTCTTCTCTGTAGGATGAGGATAGGACAGTGAAATTCT
 TAATTTAAGAGTTGCTATTTTTCAAACCTGGCTCAGTTGTGAGATATTAAGAAAACTGA
 GATACAGTGTGGGATGGGATGAGTATGTTACGCCCTAAGGGAAGGAAGCTGATCAGCTCTG
 CCTTTAAGAAGGTCCCTGAGGGTGGCTACATGTGGATAAGGAACAAGGACTGAAGCGTGA
 GTTATTACTGTTCTTAGAACTAATAGGAGGTAGTGGAGACCAACATTAACCCCATCTTTC
 TTTTCTTCTCCCTCCTTATCTTCATCAGTTTGGCACATGGAAGCATGCACGGCGATATGG
 AGGCATCTATGCAGCCTGGGCATTTGTATCCTGTTGGCTGTGCTGTGTTTGGGCCGGGC
 TCTTTATTGGGCCTGTATGAAGGTCAGCCACCTGTCCAGAACTTCTCCTGAGCCTGATG
 ACCCACAGACTGTGCCTGGCCCCCTCCCTGGTGGGGACAGTGACACTACGAAGGGAGCTGG
 GGTAGTTAAAGGCTCCCGGGGCTTCTAGAAGGAAGCCAAGCAGCTGCCTTCCCTTTCCCT
 GGGGAGAGGTAGGAAGGAACAGGCCCTCACTTAGGTTTGGAGGGGCAGATAAGAGCACT
 GCTGACCATCTGCTTTCTCCAAGGGTTGCTGTGTCTAGGGTGAAGTAGGCAAAACGTTG
 CCTTTAAACTGGGCCCTGAAGACGGTTCCAGCCTTGCTCTTCTGTGTGCTCCCTGAGA
 GCCATTCTGTCCCTTACACATTCCAGGGCAGGGTGGGGGTGGGTAGCCCTGGGGGTTC
 CCTCCCTCTGTGCACCATTAGGACTTTGCTGCTGCTATTGCACTTACACAGAGGTGGC
 TCTGGCCTCAGTACCCTCAGTCTCCTCTCCCACTTGTGTCTGTGGGGGTGGGGTCAG
 CCGCTGCTCTGTACAGAACCACAGGAAGTGTGTATATACTATTTAATGTGGGATAT
 GTTCCCTATTTCCTGTATTTCCCTTAATTCCTCCTCCCGACCTTTTTTACCCCCCAGTT
 GCAGTATTTAACTGGGCTGGGTAGGGTTGCTCAGTCTTTGGGGGAGGTTAGGGACTTATC
 CTGTGCTTGTAAATAAATAAGGTCATGACTCTAAAAA
 AAAAAA

>Hs.302144_ARNm_1 gi|11493400|gb|AF130047.1|AF130047 clon de Homo sapiens FLB3020 poliA = 0

CTGT CAGC ACGGG CCTGG CATGTAATTGGTCTGCACCCACTGGTGC ACTGAACTGCCAT
AACCTCAGGTTTTCTTTCTTGCTGATACCCCTGGGTCATGTTCTTTGGCAAATAACATGA
TTCATTATGAAGTAGAGTTCAGCAAAGGACAAGGATGAAAGTTGTCAATTTAGAGAACTG
CATT CAGACTTTCTTGCTTAGGTAAAGAGCAAGGTCTTCTCTCTTTTCAACTCATTTCT
AAAFTTAAACTGACGATGAGAATATGGATGATGTGTAGCTTCTCTCTCCCCACTGATTT
TTGGTT CAGGCTCTGGGTTTTTGGCAAGAACTTACAGATCTCACATTATTATTGGCCACCC
TTCTGCTTTAAGACCTGTCAGGGCTTGCTCTGAAATAAACTGGAAGCACTTCTGATTCCA
TCCTCACTGCTTTCTCTCTTACCGTCAGACAGCATTACTGTATAGCACTGAGTGAGGGG
CCCTGACACTGGAAGGTGGCAGGTGGGGCTGGCCGCCAGTGAGGTATCATATCATTTGTGT
GTGCTCATGTGTGCGTTGGGCTTGTTGTATCTGAGGCATGAACATTCCATATACACGGGT
TAAAGAGTTTTCTTCCCATACCGAAAGCATATATTCCGAGAGGACCAACTTATTCAGCA
TAGCCTTGTTCCCATAGTAGCCATCTATTCCCCACAGCCTCTACTTTAGGAAAGCTCC
CCGTCCCCATATGAAATCCAAACCAAAAAGATATATCACTTTTCAGCTCAATTATTCCAT
AATTACAAGATATTAGGCTAGTGGGCTCTTTATTGGTTGGGCTTATATTAATGTTATAT

GCTAGCCTTGTAATTGTGAGCTCCTCTATGGATGTTAATTTTAGTGAAACTCTATATTGA
AGAAAAGATGGGACTAAGGGGGAGACAGGAGGAGGAAAAGAAAGCAGAGACAGGCAAAGAA
TCATAGCCTGAAATTCACAGCAAGCATGGCTTATGAAGATCAAGTTATATTTTTGCTTC
ATGAATCATTGTGAGACAAATTAAGAACATATTGTTTCTTATTTATCTATTGTCAAGGAT
TCACTATCAGACACTAAGAATGAATCTTGATTTTCATAAGCTCTGTTGACACCATGGAGC
CACAGAGCATAAAACCTGTCATCTAATAAAGAAAGTGCAACATGGAACAGCAGGGAGTGGA
ATACCAGCACAACTCACAGCTGCTTCCTGTTCCCTCGTCCCTGTTTTCAGGAATGTTTCTT
AGCAGGAAGTTTTTTAATAGACCGAGAATTTGTTATATGTATTCTAAGAAAAGTTGTAGT
TGTAGATGCATTACTCTCCCAAATCTTAGAGATCAGGGATGATTATGTTCCATTTTTGTT
TGGTGAGTTCCCATCTTTGTATGTACCTCCTTGCTCCCGGCTGTCCTCCTCTCCTCTTCC
CTAGTGAGTGTTAATGAGTGTGTTAATGCCTAAACCATACTTGTTTTATGGACACTTCTAT
AATGGATTTCGTTGCATAATTTTCATGCAGTGATAGTGTTACTAGTTGGAATTCCTTGA
GGACTCTTAGCTGTCTGATGAATTCCTAGTAGAAATTTTGTGTTTGAATTCCTAAAGTT
GAAATATGAAAATTATATTTTTAATTTGATTC

>Hs.26510_ARNm_2 gi|11345385|gb|AF308803.1|AF308803 cromosoma 15 map 15q26 de Homo sapiens polIA
= 3

AGTTTTTCTGGTAGAAGGCGGGGTTCTCCTCGTACGCTGCGGAGTCTCTGCGGGGTGTAG
 ACCGGAATCCTGCTGACGGGCAGAGTGGATCAGGGAGGGAGGGTCGAGACACGGTGGCTG
 CAGGTCTGAGACAAGGCTGCTCCGAGGTAGTACTCTCTTGCTGGAGGTGGCCATTCAT
 TCCTGGAGTGCTGCTGAGGAGCGAGGGCCCATCTGGGGTCTCTGGAAGTCGGTGCCGAGG
 CCTGAAGGATAGCCCCCTTGCGCTTCCCTGGGCTGCGGCCGGCCTTCTCAGAACGAAGG
 GCGTCTTCCACCCCGCGCGCAGGTGACCGCTGCCATGGCTTTTCCCCATCGGCCGGAC
 GCCCCTGAGCTGCCTGACTTCTCCATGCTGAAGAGGCTGGCTCGAGACCAGCTCATCTAT
 CTGCTGGAGCAGCTTCTTGAAAAAAGGATTTATTATTGAGGCAGATCTCATGAGCCCT
 TTGGATCGAATTGCCAATGTCTCCATCCTGAAGCAACACGAAGTAGACAAGCTATACAAG
 GTGGAGAACAAGCCAGCCCTCAGCTCCAATGAACAATTGTGCTTCTTGGTCAGACCCCGC
 ATCAAGAATATGCGATACATTGCCAGTCTTGTCAATGCTGACAAATTGGCTGGCCGAAGT
 CGCAAATACAAAGTGATCTTCAGCCCTCAAAAGTTCTATGCGTGTGAGATGGTGCTTGAG
 GAAGAGGGAATCTATGGAGATGTGAGCTGTGATGAATGGGCCTTCTCTTTGCTGCCTCTT
 GATGTGGATCTGCTGAGCATGGAACACCAGAATTTTTTCAGGGATTACTTTCTGGAAGGA
 GATCAGCGTTGGATCAACACTGTAGCTCAGGCCCTTACACCTTCTCAGCACTCTCTATGGA
 CCTTTTCAAACCTGCTATGGAATTGGCAGGTGCGCCAAGATGGCATATGAATTGTGGAGG
 AACCTGGAGGAGGAGGAGGATGGCGAAACCAAGGGCCGAAGGCCAGAGATTGGACATATC
 TTTCTCTTGGACAGAGATGTGGACTTTGTGACAGCACTTTGCTCCCAAGTGCTTTATGAG
 GGCTTAGTAGATGACACCTTCCGCATCAAGTGTGGGAGTGTGACTTTGGCCCAAGATC
 ACATCTCTGACAAGAGCCTGAAGGTGCTACTCAATGCCGAGGACAAGGTGTTTAATGAG
 ATTCGGAACAGCACTTCTCCAATGTCTTTGGCTTCTTGAGCCAGAAGGCCGGAAGTGTG
 CAGGCCAGTATGATCGCCGAGAGGCATGGACATTAAGCAGATGAAGAATTTCTGTGTC
 CAGGAGCTCAAGGGCCTGAAACAGGAGCACCGCCTGCTGAGTCTCCATATTGGGGCCTGT
 GAATCCATCATGAAGAAGAAAACCAAGCAGGATTTCCAGGAGCTAATCAAGACTGAGCAT
 GCACTGCTAGAGGGGTTCAACATCCGGGAGAGCACAGCTACATTGAGGAACACATAGAC
 CGGCAGGTGTGCTATAGAAAGCCTGCGCCTCATGTGCTTTTGTCCATCACTGAGAAT
 GGTTTGATCCCCAAGGATTACCGATCTCTGAAAACACAGTATCTGCAGAGCTATGGCCCT
 GAGCACCTGCTAACCTTCTCCAATCTGCGAAGAGCTGGGCTCCTAACGGAGCAGGCCCCC
 GGGGACACCCTCACAGCCGTGGAGAGTAAAGTGAGCAAGCTGGTGACCGACAAGGCTGCA
 GGAAAGATTACTGATGCTTCTCAGTCTCTGCGCAAGAGGAGCAATTTCTGTCATCAGC
 AAAAGCTGAATTTGATCCCACTGTGGACGGCGAGTATGATCTGAAAGTGCCCCGAGAC
 ATGGCTTACGTCTTCTAGTGGTGCTTATGTGCCCTGAGCTGCCGAATCATTTGAGCAGGTG
 CTAGAGCGGCGAAGCTGGCAGGGCCTTGATGAGGTGGTACGGCTGCTCAACTGCAGTGAC
 TTTGCATTACAGATATGACTAAGGAAGACAAGGCTTCCAGTGAGTCCCTGCGCCTCATC
 TTGGTGGTGTCTTGGGTGGTTGTACATTCTCTGAGATCTCAGCCCTCCGGTTCCTGGGC
 AGAGAGAAAGGCTACAGGTTCAATTTCTTGACGACAGCAGTCACAAACAGCGCTCGCCTT
 ATGGAGGCCATGAGTGAGGTGAAAGCCTGATGTTTTTCCGGCCAGTGTGACATCTTCC
 CTGAACACATTCTCAGTGAGATGCAGGCATCTGGCACCCAGCTGCTATAACCAAGTGTG
 CACCAACTACCTGCTAAGAGCCGGGAGCATGGAACGTGTTGGGATTTAGAGAACATTATC
 TGAGAAAAGAGTTCACTTCTGCTCCAGGATATTTCTCTTTTCTGTTTATGAAGTACAA
 CCCATGCTGCTAAGATGCGAGCAGGAAGAGGCATCCTTTGCTAAATCCTGTTTGAATGTC
 ATTGTAAATAAAGCCTCTGCTCTCAGATGTAAAAA

>Hs.324709_ARNm_2_gij|12655026|gb|BC001361.1|BC001361 clon de Homo sapiens MGC:2474
 IMAGE:3050694 poliA = 2

5

GGCACGAGGGGTGCGCTGCGCCGTTTATTTGAAGACATCGTCCAGTTCTGACCATGG

ACTCGCAGCCATCGGCCCTTAGTTTCCATCCCCCTCTAGTGGGCCTTCGGGGGCTCTACTG
 ACGTCCCTCCTTCCCTTGGTACCGGGCCGGGGAAGTGTCTCGGGCGCGGGAGGTCCGC
 ATGCCCAGGCCTGGCCAGGGGAGATGACCGATCCGTGCTGGGGCTGACAGTCCCCATGG
 CGCCGCCCTCTGGCCCCGCTCCCTCCCCGGGACCCAAACGGGGCGGGATCCGAGTGGAGAA
 AGCCCCGGGGCCGTGAGCTTCGCCGACGTGGCCGTGTACTTCTCCCGGGAGGAGTGGGGCT
 GCCTGCGGCCCGCGCAGAGGGCCCTGTACCGGGACGTGATGCGGGAGACCTACGGCCACC
 TGGGCGCGCTCGGTGAGAGCCCCACCTGCTTGCCTGGGCCCTGCGCCTCCACAGGCCCTG
 CCGCGCCTCTGGGAGCTGCGTGTGGAGTTGGGGGCCCCGGGGCCGGGCAGGCGGCCCTCCT
 CGCAGCGTGGGGTTTGCCTTCTTCTCCCCAGGAGTCGGAGGCAGCAAGCCGGCGCTCAT
 CTCCTGGGTGGAGGAGAAGGCCGAACCTGTGGGATCCGGCTGCCAGGATCCGGAGGTGGC
 GAAGTGTCCGACAGAAGCGGACCCAGCAGATTCCAGAAACAAGGAAGAGGAAAGACAAAG
 GGAAGGACGGGAGCCCTGGAGAAGCCCGACCTGTGGCCGCCGGGTCTCCTGGGCTGAA
 GGCTCCCCAAGCCCCCTTTGCCGGGTGGAGCAGCTGTCCAAGGCCCGCGCCGGAGTCG
 CCCCCTTTTGGCCACCCCCCTGTCCCCGAGCTGACCAGCGTCACGGCTGCTACGT
 GTGCGGGAAGAGCTTCGCTTGGCGCTCCACACTGGTGGAGCACATTTACAGCCACAGGGG
 CGAGAAGCCCTTCCACTGCGCAGACTGCGGCAAGGGCTTCGGCCACGCTTCCCTCCCTGAG
 CAAACACCGGGCCATCCATCGTGGGGAGCGGCCCCACCGCTGTCCGAGTGTGGTCGGGC
 CTTTCATGCGCCGACGGCGCTGACTTCTCACCTGCGCGTTACACTGGCGAGAAGCCCTA
 CCGCTGCCCGCAGTGTGGCCGCTGCTTCGGCCTGAAGACCGGCATGGCCAAGCACCAATG
 GGTCCATCGGCCCGGGGGCGAGGGGCGTAGGGGCCGGCGCCCTGGGGGGCTGTCTGTGAC
 CCTGACTCCTGTCCGCGGGGACCTGGACCCGCTGTGGGCTTCCAGCTGTATCCAGAGAT
 ATTCCAGGAATGTGGGTGACGGCCTAAAAAGTGACCATCTAGACATTGTGGGCGGCCCGA
 GATGGGCTCAGGGGCCGAACCTCTGCAGCGGCCTGCAGGGAGGTCCAGAATCCACCGC
 AAGAGCTGGCCTGGGGTGGGACAGTCTGATCTTGGGCTCTCAGCAGCCTCTTCTGCCAG
 CACTTGTCTCCCCGCTGCCCTGGGCTCTCCAAGGCCCCCTTTGCTGAGGCAGGGCTGAGG
 TGAGAACCCCCAGACCTCCATACAGGGGAAGCAAAAGCTGTTTCTCCTCCAGAGATGCT
 AAGAGGATTGAGGTAGAGAAGAACCTTGTTFCTCTGTTGTCTTTTTCTTTTTACTTTTT
 TAATTTTTTGAGACGGAGTTTTGCTCTTGTGTCCTGAGGCTGGAGTGCAATGGTGGCATCT
 CGACTCACTGCAACTTCCACCTCCTGGAGTCAAGCGATTCTCCTGCCTCAGCCACCCAAG
 TAGCTGGAATTACAGGCACCTGCCACTATGCCCCGGCTAACTTTTTGTATTTTAGTAGAG
 ATGGGGTTTCACCATGTTGGCTAGGCTGGTCTCGAACTCCTGCCCTCAGGTGATCCACCC
 ACCTCTGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCTCACCTGGCCTTTCTT
 TTTTATTCCTTTGACCTTCCCACAAGACAATACCCATTGTCTGTTTTTTTGTATTATTTAT
 TTACTTTATTAAGACAGCATCTTGCTCCTCACCCAGGCTGGAATGCAGTGGTGTGAAGTGG
 GCTCACTGCAGCCTAGACCTGCTGGGCTCAAGGAATCCTCCTGCCCCAGCCTCTCAGATG
 GCTGTGACTACAGGTGGGCAACACTATGCCTGGTTAATTTTTTAAATTTTTTTCAGAGAT
 GGGGTTCCCACTATGTTGATCAGGCTGGTCTCAAACCTCCTCGGTTCAAGCAATTCGCCCA
 CCTTGGCCTCCCAAAGTGCTGGGATTACAGGGGAGCCACTGCACTGGCCTTCATTGTCTT
 TTTGCTGCACAACCTAAAAAACCAGTGACCCTGTATTGGAAAAAAAAAAAAAAAAAAAAA
 A

>Hs.65756_ARNm_3 gi|3641494|gb|AF035154.1|AF035154 cromosoma de Homo sapiens 16 map 16p13.3 poliA
= 3

GCCATGGCCGCGCCGCCCCGCGCCGCCCCCGGCCGCCCCCGGGCGCAGATGCCGCATCTG
 AGGAAGGTGCGAGGCGGATGGAGCGGGTGGTCTGTGAGCATGCAGGACCCCGACAGGGCG
 TGAAGATGCGGAGCCAGCGCCTGCTGGTCAACGTCAATCCCCACGCGGTGACAGGCAGCG
 ACGTCGTGCAGTGGTTGGCCCCAGAGTTCTGCGTCTCGGAGGAGGAGGCCCTGCACCTGG
 GCGCCGTCCTGGTGCAGCATGGCTACATCTACCCGCTGCGCGACCCCCGTAGCCTCATGC
 TCCGGCCAGACGAGACGCCCTACAGGTTCCAGACCCCGTACTTCTGGACAAGTACCCTGA
 GGCCGGCTGCAGAGCTGGACTATGCCATCTACCTGGCCAAGAAGAACATCCGAAAACGGG
 GGACCCTGGTGGATTATGAGAAGGACTGCTATGACCGGCTACACAAGAAGATCAACCACG
 CATGGGACCTGGTGTGATGCAGGCGAGGGAGCAGCTGAGGGCAGCCAAGCAGCGCAGCA
 AGGGGGACAGGCTGGTCATTGCGTGCCAGGAGCAGACCTACTGGCTGGTGAACAGGCCCC
 CGCCCGGGGCCCCCGATGTGCTGGAGCAGGGTCCAGGGCGGGGATCCTGCGCTGCCAGCC
 GTGTGCTCATGACCAAGAGTGCAGATTTCCATAAGCGGGAGATCGAGTACTTCAGGAAAG
 CGCTGGGCAGGACCCGAGTGAAGTCCCTCCGTCTGCCTTGAGGCGTACCTGAGTTTCTGCG
 GCCAGCGTGGACCCACGATCCCTCCTGTGCGGGTGCCTGCCAGCAATCCCTGGATCT
 CAGACAATGACGCCTACTGGGTCAATGAATGCCCCACGGTGGCTGCCCCACGAAGCTCC
 GTGTGGAGAGATGGGGCTTCACTTCCGGGAGCTCCTGGAGGACCCCGTGGGGCGGGCCC
 ACTTCATGGACTTTCTGGGAAAGGAGTTCAGTGGAGAAAACCTCAGCTTCTGGGAGGCAT
 GTGAGGAGCTTCGATATGGAGCGCAGGCCCAGGTCCCCACCCTGGTGGATGCCGTGTACG
 AGCAGTTCTGGCCCCCGGAGCTGCCCACTGGGTCAACATCGACAGCCGGACCATGGAGC

AGACCCTGGAGGGGCTGCGCCAGCCCCACCGCTATGTCCTGGATGACGCCCAGCTGCACA
 TATACATGCTCATGAAGAAGGACTCCTACCCAAGGTTCTGAAGTCTGACATGTACAAGG
 CCTCCTGGCAGAGGCTGGGATCCCCGCTGGAGATGAAGAGACGCGTGTTCCTGGTTTACGT
 GGAGGCCACGGCACTCGAGCCCCAGCCCTGCACTCCTTCCCACCCCTGTGGAGCCCCACAG
 CGGCTTGTGGCCCTGGGGGTGGAGATGGGGTGGCCTAGTGGACCTGGCCCCTCTGCCACT
 CTAGTCCCTGCAGCTCAACGTCTGCGTGAATGCAGCAGCCACCCCGTCTTGGCCCCAGG
 TCCTGGGGGCTGCTGAACCCAGCACCAGTGTCCCTTGTGCCAGGGGGGCCAGTCTTCT
 GTGGGGTGCACAGCCTCCCTCCCTCCAGCAAGCCCTCCCTGCCAGAAGGAATGGGTCCA
 GGTGTGGATTCCCAGGGAGGGGGTTCATTGGCTCAGCTTGGGTGAGGGCAGAGCCTGTTA
 CCTGAAGAGAGGTGAGACCAAGGCCACAGGGAGCTCCACCTTCTCTGGTCTTCAGTCCAG
 CACTGGGTGCCCATCCCCATCTCTAAAACAGTAAATCAGCCAGCGAATACCCGGAAGCA
 AGATGCACAGGCGGGCGGCTTCCCACACACCCGTCACAAGACGCGGACATGCAGGTCTCG
 GCGCGAGCTCTGCCCCGTCCAAGAGCCTCTCCGCTGTGCGCCAGTGTGAGCCTGGAAGAG
 GACCCAAGAGAGTGGCGTGCTGAGGCTGCCTCGAGGTCACTGCCTTCCGGAGCTGCGCCT
 ATTCCTCCCTCGCCAAACGCGTTCCAGAATTTGTCCACAGGTGCGCCGGCACCTGCTTTT
 CCACCTCGAGGCCGCGGCCTCCCCCGGATTTATAGACAACCTCTGACATTGTACCCCCAC
 TGACGAGGCCCGATTCCATAGGGTGGATCCTTGCCAGGCGTCCCTGATCCTCCCTGCCCA
 AGTCTTCTTCTGAGCTGGCCTTGCTCCCCATCCCCCAAGTGCCTCACCAGTCCCCCAG
 ACTGGGTGAAGGTACAGCTGGCTCCTTTCCGGGGTGCAGCTTCAACTCTCTCGGCGGTAG
 GGCGGTGCCATCCCCACCCATAGGGCTGGCTCACATCCAGTCACTCCCAACAGCGTCCAG
 CACACAAATAAAAGACCCTTGGGCCCTGGCTCTGAGAAAAAAA

>Hs.165743_ARNm_2 gij13543889|gb|BC006091.1|BC006091 clon de Homo sapiens MGC:12673
 IMAGE:3677524 poliA = 3

AGACTGCCGAGCAGCCTTGAGCCGTTGAGCAGCTGAACAGAGGCCATGCCGGGGCACTCC
 GAGGCCTGAGACGACCACGCCCTGTGCCGCTGAGGACCTTCATCAGGGCTCCGTCCACTTG
 GCCCCTTTGGCTGTCCAATCACACTCCAGTGTCAACCACTGGCACCCAGCAGCCAAGAGA
 GGTGTGGCGTGGCCCTGGGGACGCATGGCTGAGGCAGGAACAGGTGAGCCGTCCCCCAGC
 GTGGAGGGCGAACACGGGACGGAGTATGACACGCTGCCTTCCGACACAGTCTCCCTCAGT
 GACTCGGACTCTGACCTCAGCTTGCCCCGGTGGTGCTGAAGTGGAAGCACTGTCCCGATG
 GGGCTGCCTGGGGAGGAGGATTCAAGTTCCTGATGAGCCGCCCTCACCCCCGTGAGGCCTC
 CTCCCAGCCACGGTGCCAGCCATTCCATCTGAGAGGCATGAGCTCCACCTTCTCCCAGCGC
 AGCCGTGACATCTTTGACTGCCTGGAGGGGGCGGCCAGACGGGCTCCATCCTCTGTGGCC
 CACACCAGCATGAGTGACAACGGAGGCTTCAAGCGGCCCTAGCGCCCTCAGGCCGGTCT
 CCAGTGGAAGGCCTGGGCAGGGGCCATCGGAGCCCTGCCTCACCAAGGGTGCCTCCGGTC
 CCCGACTACGTGGCACACCCCGAGCGCTGGACCAAGTACAGCCTGGAAGATGTGACCGAG
 GTCAGCGAGCAGAGCAATCAGGCCACCGCCCTGGCCTTCTGGGCTCCAGAGCCTGGCT
 GCCCCCACTGACTGCGTGTCTCTTCAACCAGGATCCCTCCAGCTGTGGGGAGGGGAGG
 GTCATCTTCACCAAACAGTCCGAGGGGTGGAAGCCAGACACGAGAGGAAGAGGGTCTTG
 GGAAGGTGGGAGAGCCAGGCAGGGCGGCCCTTGGGAATCCTGCCACAGACAGGGGCGAG
 GGCCCTGTGGAGCTGGCCCATCTGGCCGGGCCCCGGGAGCCAGAGGCTGAGGAGTGGGGC
 AGCCCCCATGGAGGCCCTGCAGGAGGTGGAGGCACTGTGAGGGTCTGTCCACAGTGGGTCT
 GTGCCAGGTCTCCCGCCGGTGGAACTGTTGGCTTCCATGGCAGCAGGAAGCGGAGTCGA
 GACCACTTCCGGAACAAGAGCAGCAGCCCCGAGGACCCAGGTGCTGAGGTCTGAGAGGGA
 GATGGCCCAGCCTGACCCCACTGGCCACTGCCATCCTGCTGCCTTCCCAGTGGGGCTGGT
 CAGGGGGCAGCCTGGCCACTGCCCTAGCTGGAATGGGAGGAAGCCTGCAGGTGGCACCGGT
 GGCCCTGGCTGCAGTTCTGGGCAGCATCCTCCCAAGCAGAGACCTTGCTGAAGCTCCTGG
 GGTGTGGGGTGTGGGCTGGAAGCACTGGCTCCCTGGTAGGGACAATAAAGTTTGGGTC
 TTTCAAAC

REIVINDICACIONES

1. Un método para clasificar una muestra que contiene células como que contiene de células tumorales de un tipo de tejido, comprendiendo dicho método la determinación de los niveles de expresión de cinco a 49 secuencias transcritas de células en una muestra que contiene células obtenidas de un sujeto humano, y
5 comparar los niveles de expresión con los niveles de expresión de las mismas secuencias transcritas en una pluralidad de tipos de tejido tumoral conocidos, y
clasificar la muestra como que contiene células tumorales de un tipo de tejido de la pluralidad de tipos de tejido tumoral conocidos, en la que las cinco a 49 secuencias transcritas no se seleccionan basándose en sus valores de
10 correlación, o una clasificación basada en los valores de correlación, con los tipos de tejido tumoral.
2. El método de acuerdo con la reivindicación 1, en el que la proporción, de secuencias transcritas con respecto al número de tipos de tumor, usada en dicha clasificación es aproximadamente 5:2 o superior.
- 15 3. El método de la reivindicación 2, en el que dicha proporción es hasta aproximadamente 20:1.
4. El método de una cualquiera de las reivindicaciones 1-3, que comprende adicionalmente la determinación de los niveles de expresión de un exceso del número de secuencias transcritas, más allá del número basado en dicha proporción y el número de tipos de tumor.
- 20 5. El método de la reivindicación 4, en el que dichos niveles de expresión se determinan mediante el uso de una micromatriz.
6. El método de una cualquiera de las reivindicaciones 1-5, en el que dicha clasificación se realiza usando un
25 algoritmo de clasificación basado en la distancia.
7. El método de una cualquiera de las reivindicaciones 1-6, en el que dichas cinco a 49 secuencias transcritas comprenden dos o más seleccionadas del conjunto de genes que se muestran en el Apéndice 1 de la descripción.
- 30 8. El método de la reivindicación 7, en el que dichas cinco a 49 secuencias transcritas comprenden cinco o más seleccionadas del conjunto de genes que se muestran en el Apéndice 1 de la descripción.
9. El método de la reivindicación 6, en el que dicho algoritmo de clasificación basado en la distancia es el algoritmo vecinal más cercano a k (KNN) o algoritmo de máquina de vector de soporte (SVM).
- 35 10. El método de una cualquiera de las reivindicaciones 1-9, en el que dicha determinación comprende la amplificación de todas o parte de las secuencias transcritas, o transcripción inversa y marcado de ARN que corresponde a dichas secuencias transcritas.
- 40 11. El método de la reivindicación 10, en el que dicha amplificación comprende amplificación de ARN lineal o PCR cuantitativa.
12. El método de la reivindicación 10, en el que dicha amplificación es de las secuencias presentes dentro de 600
45 nucleótidos de los sitios de poliadenilación de los transcritos.
13. El método de la reivindicación 10, en el que dicha amplificación es amplificación de PCR cuantitativa de al menos 50 nucleótidos de los transcritos.
14. El método de una cualquiera de las reivindicaciones 1-9, en el que dichas secuencias transcritas se seleccionan
50 para que no sean redundantes.
15. El método de la reivindicación 14, que comprende adicionalmente la determinación de los niveles de expresión de un exceso del número de secuencias transcritas que son redundantes a las usadas para dicha clasificación.
- 55 16. El método de una cualquiera de las reivindicaciones 1-15, en el que dicha muestra es una muestra clínica de un paciente humano.
17. El método de la reivindicación 16, en el que dicha muestra es una muestra fijada.
- 60 18. El método de la reivindicación 17, en el que dicha muestra es una muestra embebida en parafina, fijada con a formalina (FFPE).
19. El método de la reivindicación 1, que comprende adicionalmente, antes de dicha determinación de los niveles de expresión de cinco a 49 secuencias transcritas,
65 diagnóstico de un sujeto humano como en necesidad de dicha determinación; o
recepción de una muestra que contiene células obtenida a partir de un sujeto humano; o

sección de una muestra que contiene células obtenida de un sujeto humano; o
aislamiento de células de una muestra que contiene células obtenidas de un sujeto humano; u
obtención de ARN de células de una muestra que contiene células obtenida de un sujeto humano.

- 5 20. El método de la reivindicación 1, en el que dicha pluralidad de tipos de tejido tumoral comprende tumor de
glándula adrenal, tumor de mama, tumor intestinal carcinoide, adenocarcinoma del cuello uterino, tumor del cuello
uterino de origen escamoso, tumor de vesícula biliar, tumor de ovario de origen en células germinales, células
transparentes de ovario, serosas de ovario, tumor del estroma gastrointestinal (GIST), leiomiosarcoma, tumor de
10 hígado, tumor de pulmón de células microcíticas, tumor de pulmón escamoso, adenocarcinoma de pulmón
macrocítico, meningioma, osteosarcoma, tumor de ovario de origen en células transparentes, tumor de ovarios de
origen en células serosas, tumor de piel de células basales, tumor de piel de células escamosas, liposarcoma de
tejido blando, histiocitoma fibroso maligno de tejido blando (MFH), tumor sinovial de sarcoma de tejido blando, tumor
de testículos por seminoma, tumor de testículos no seminoma, carcinoma folicular o papilar de la tiroides, y
15 carcinoma medular de la tiroides.

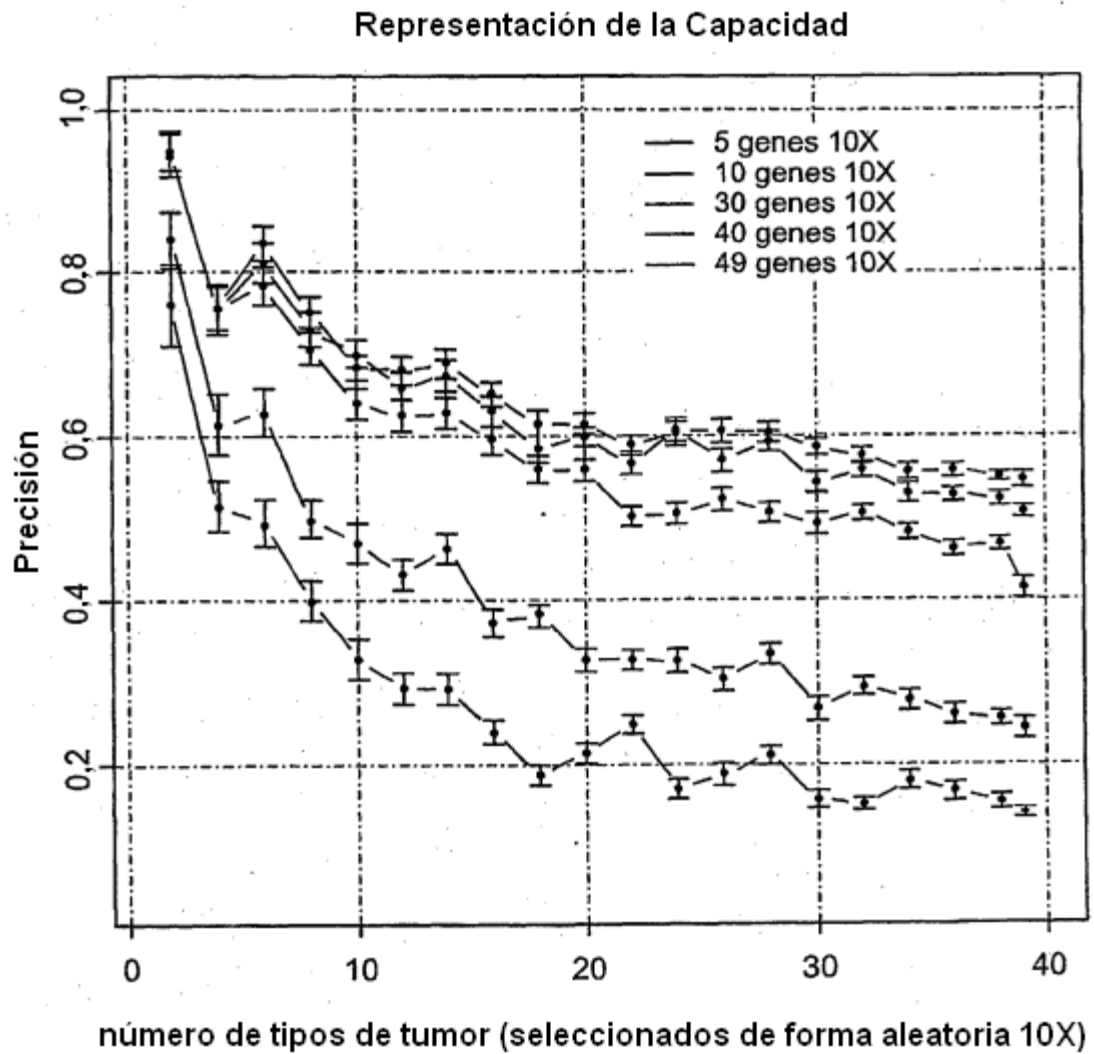


FIG. 1

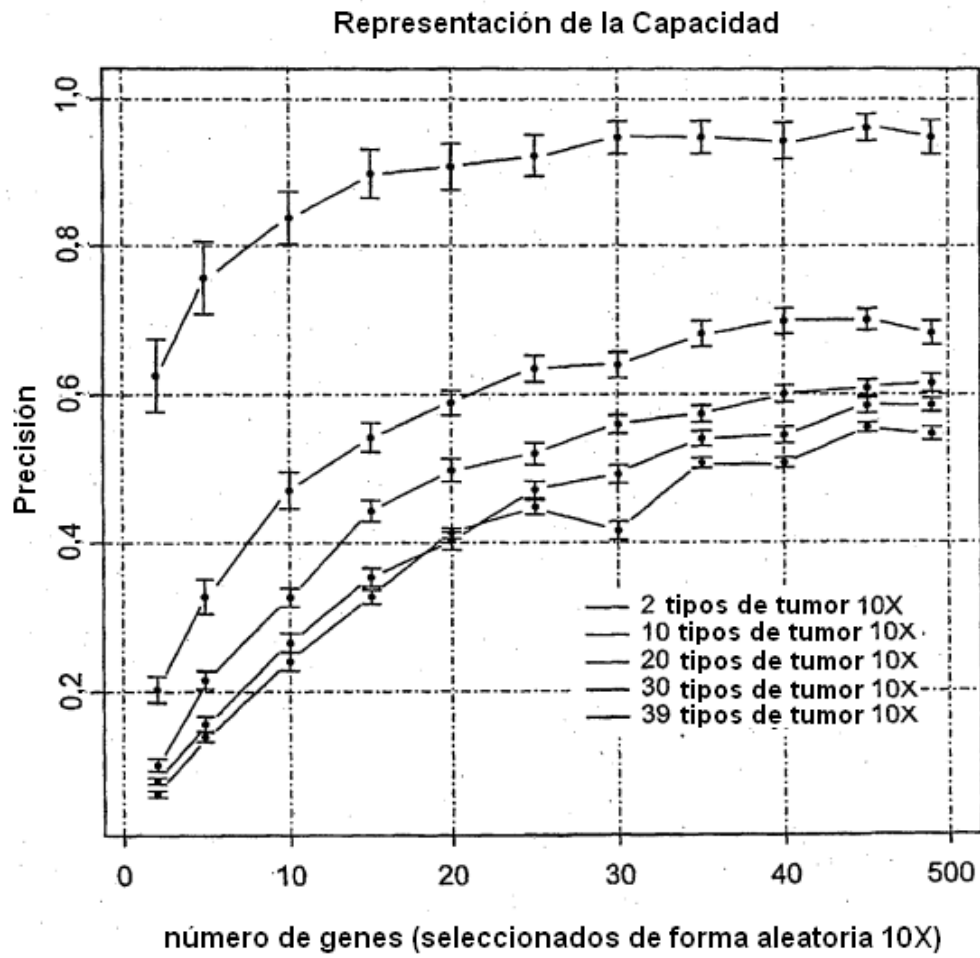


FIG. 2

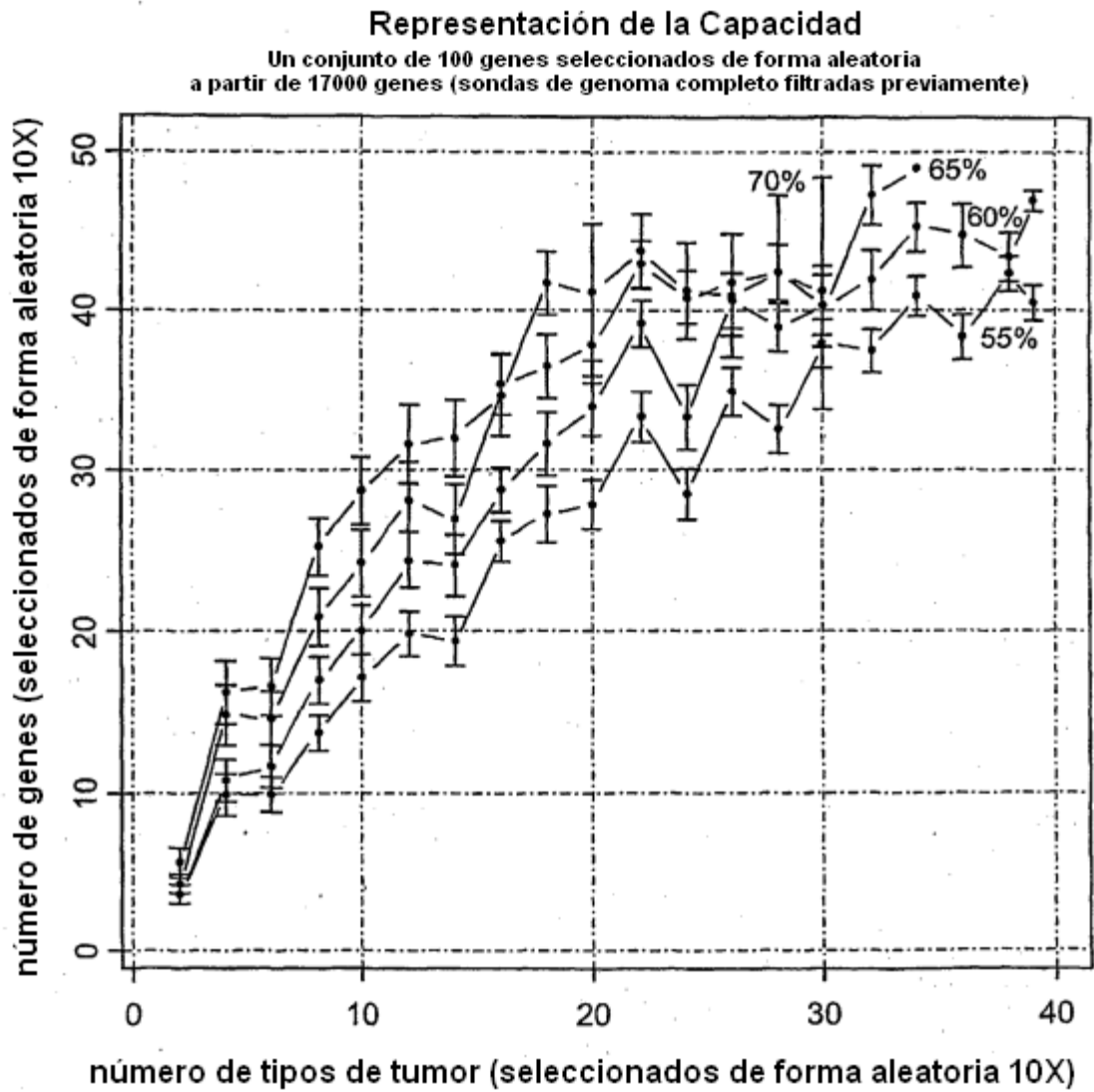


FIG. 3

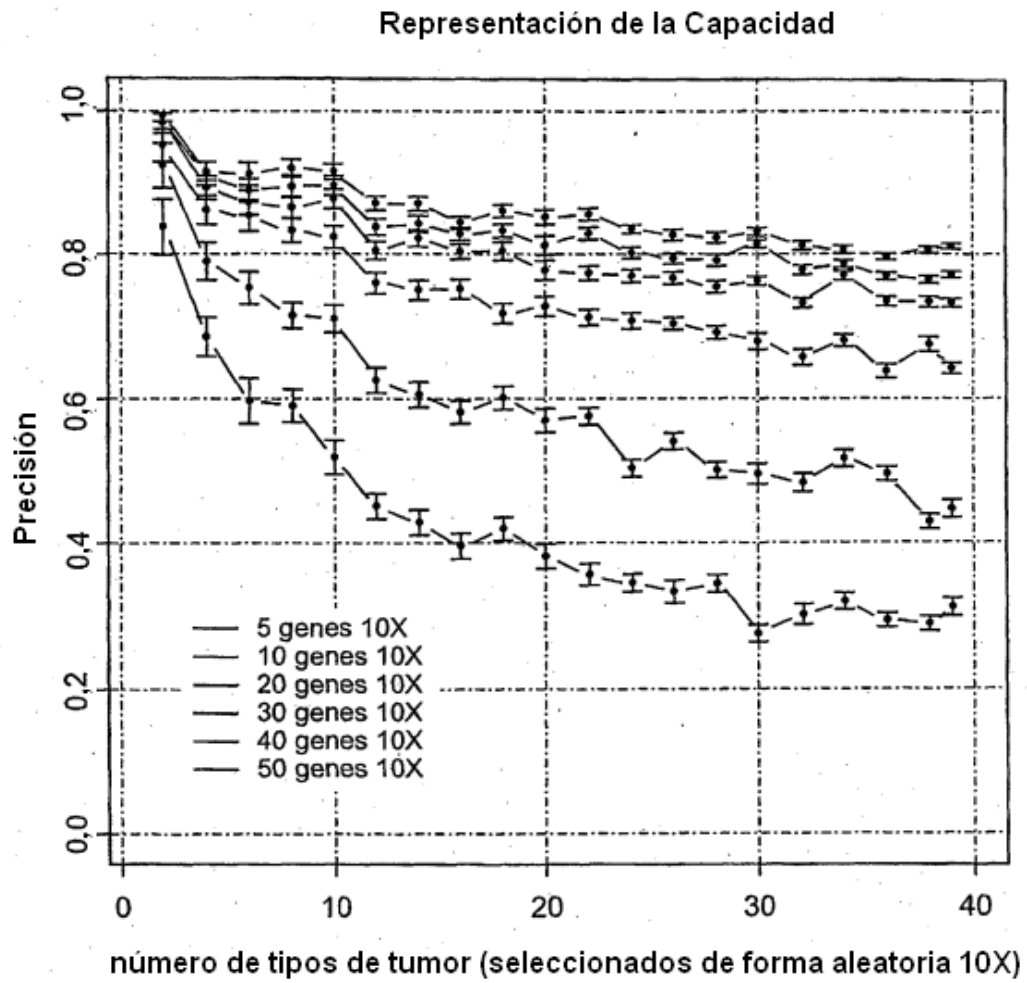


FIG. 4

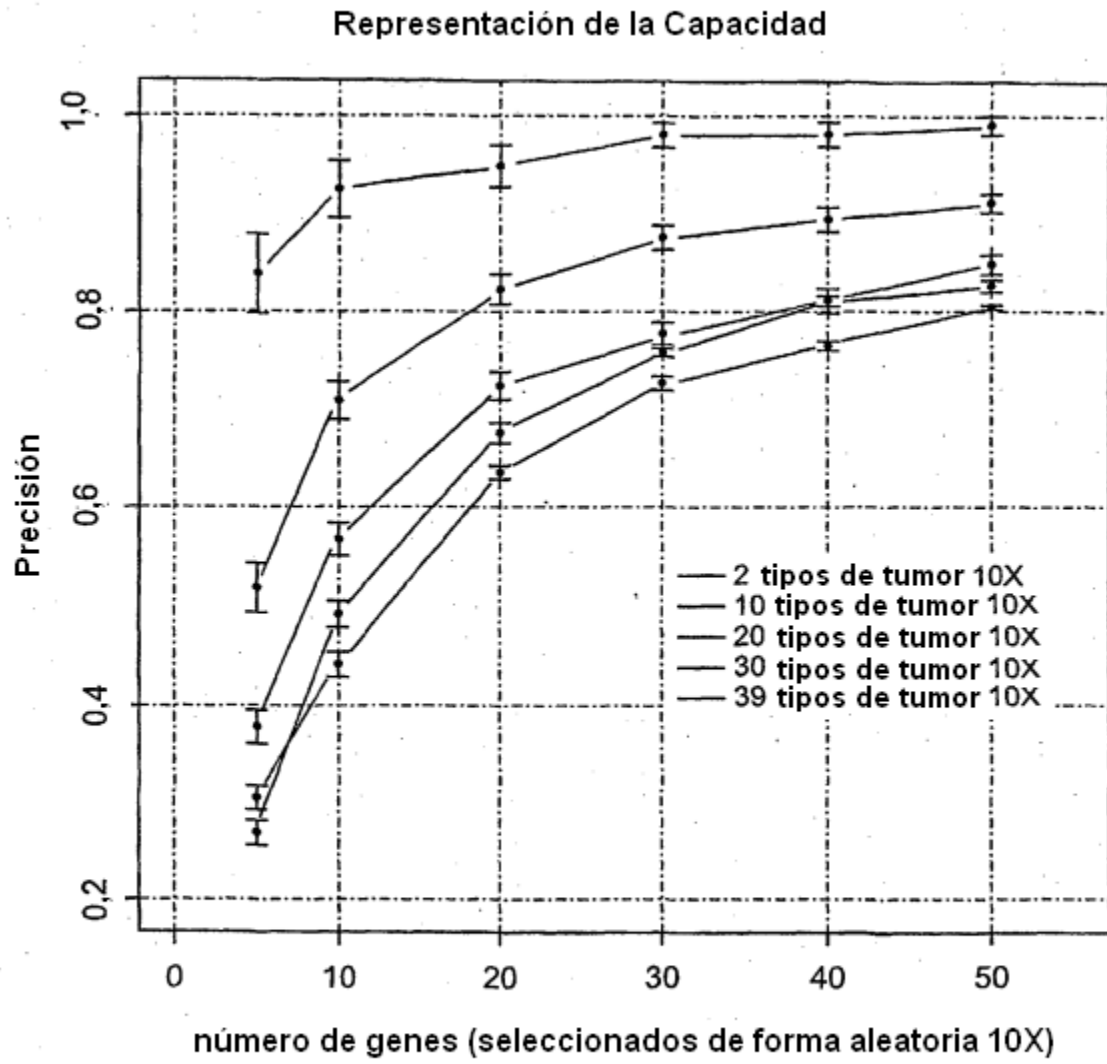


FIG. 5

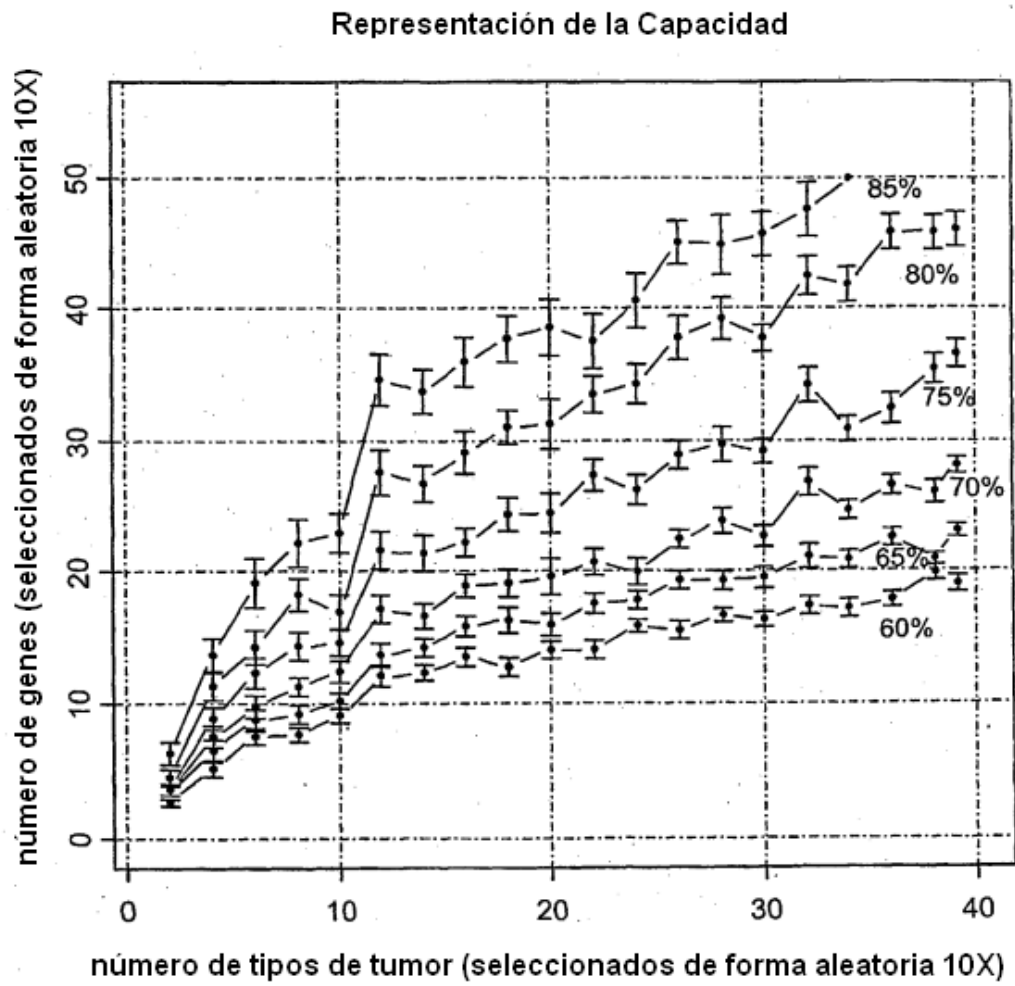


FIG. 6

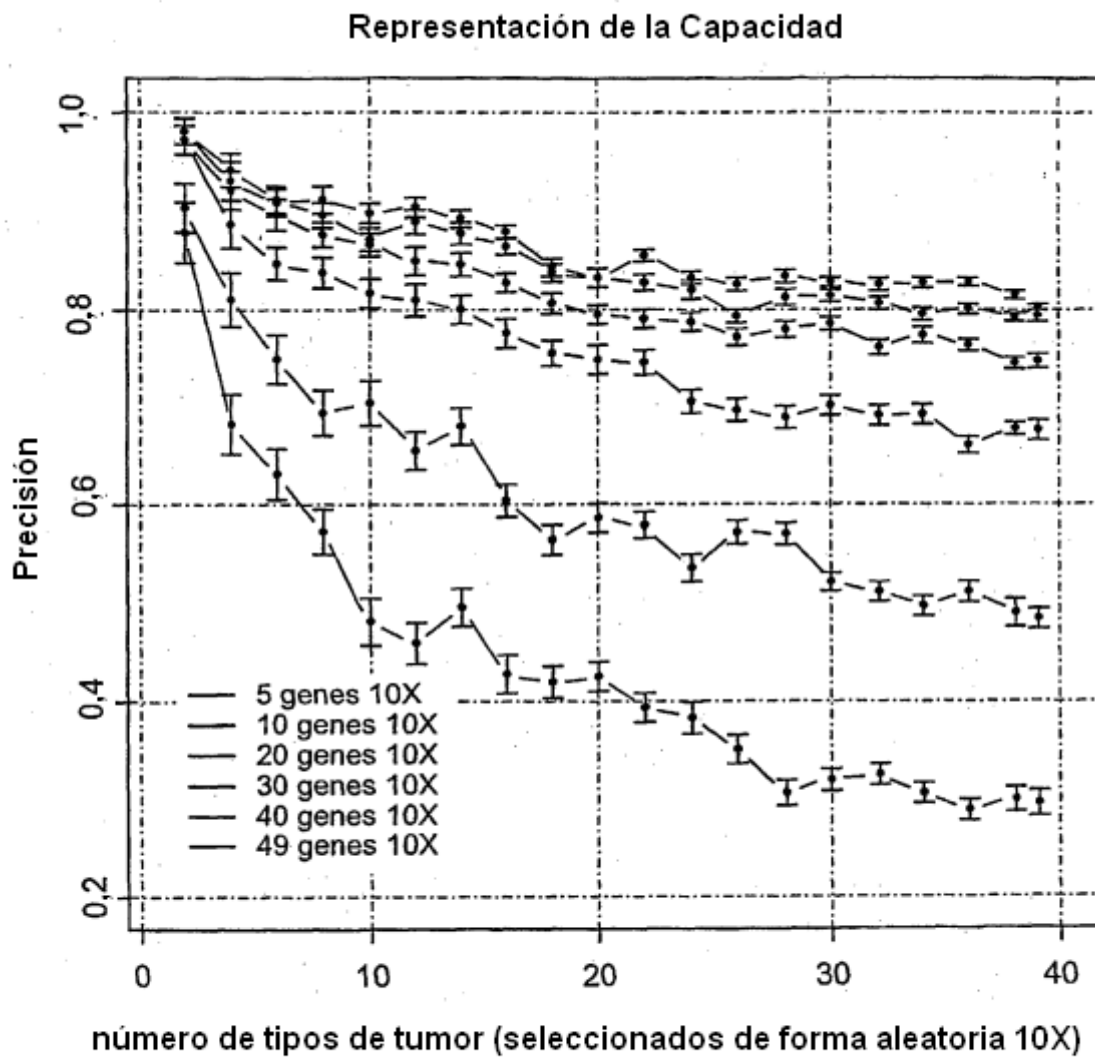


FIG. 7

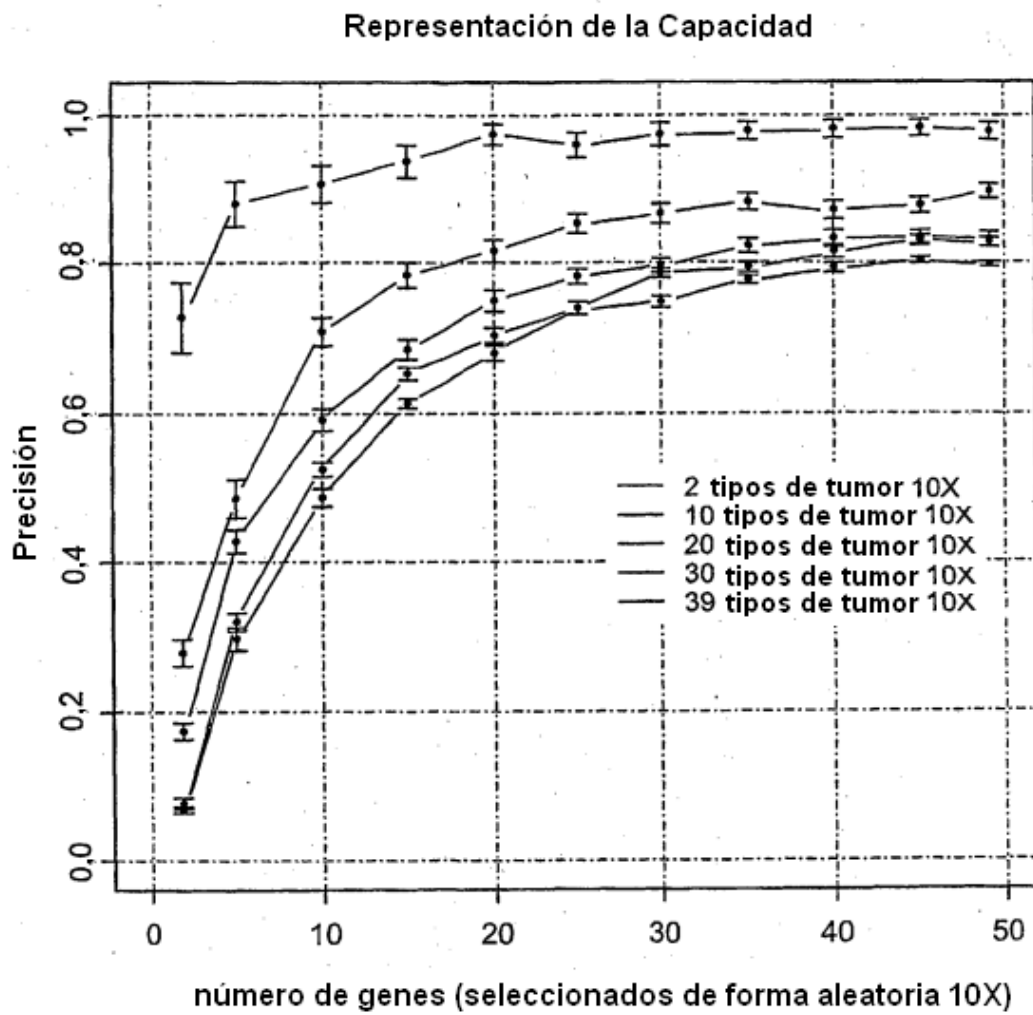


FIG. 8

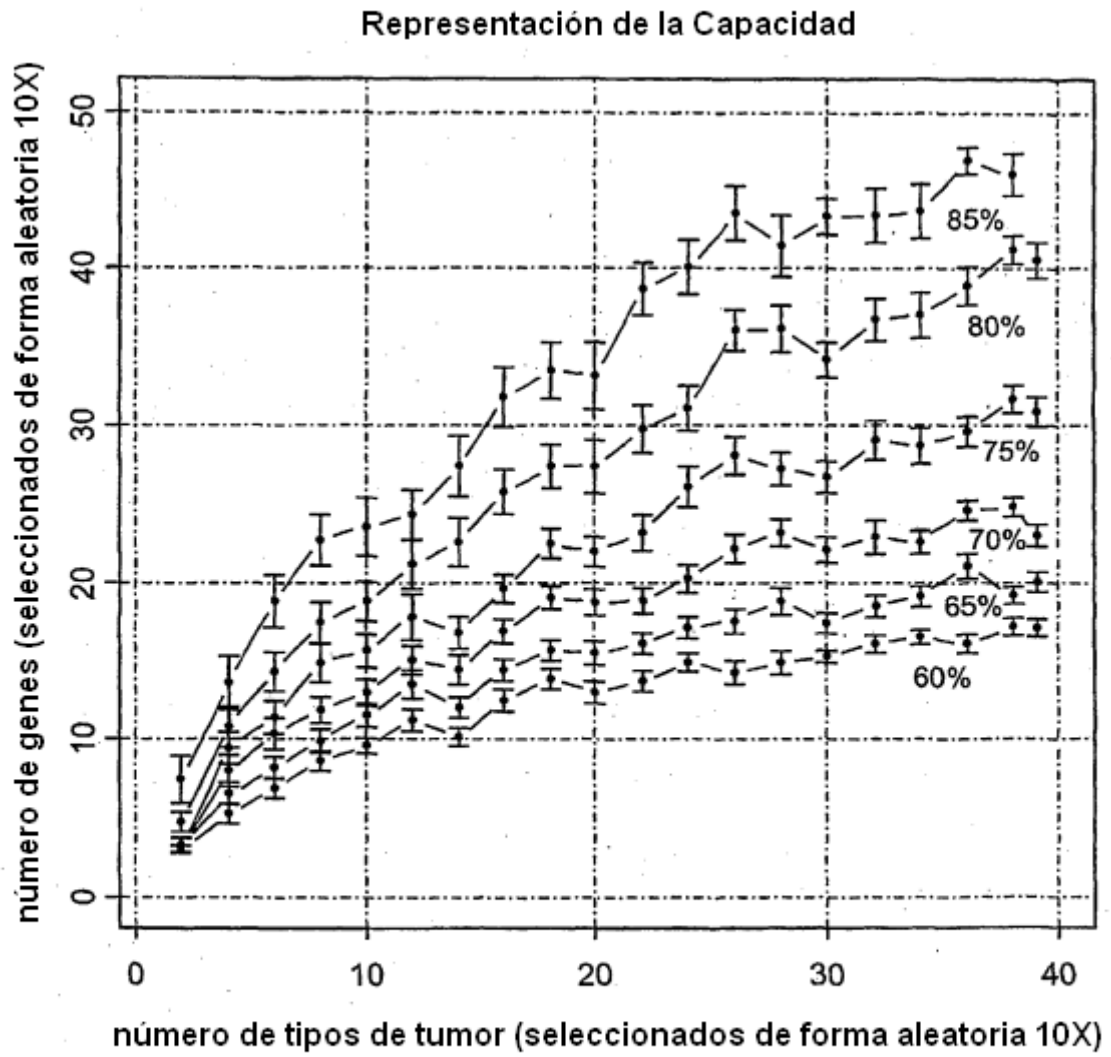


FIG. 9

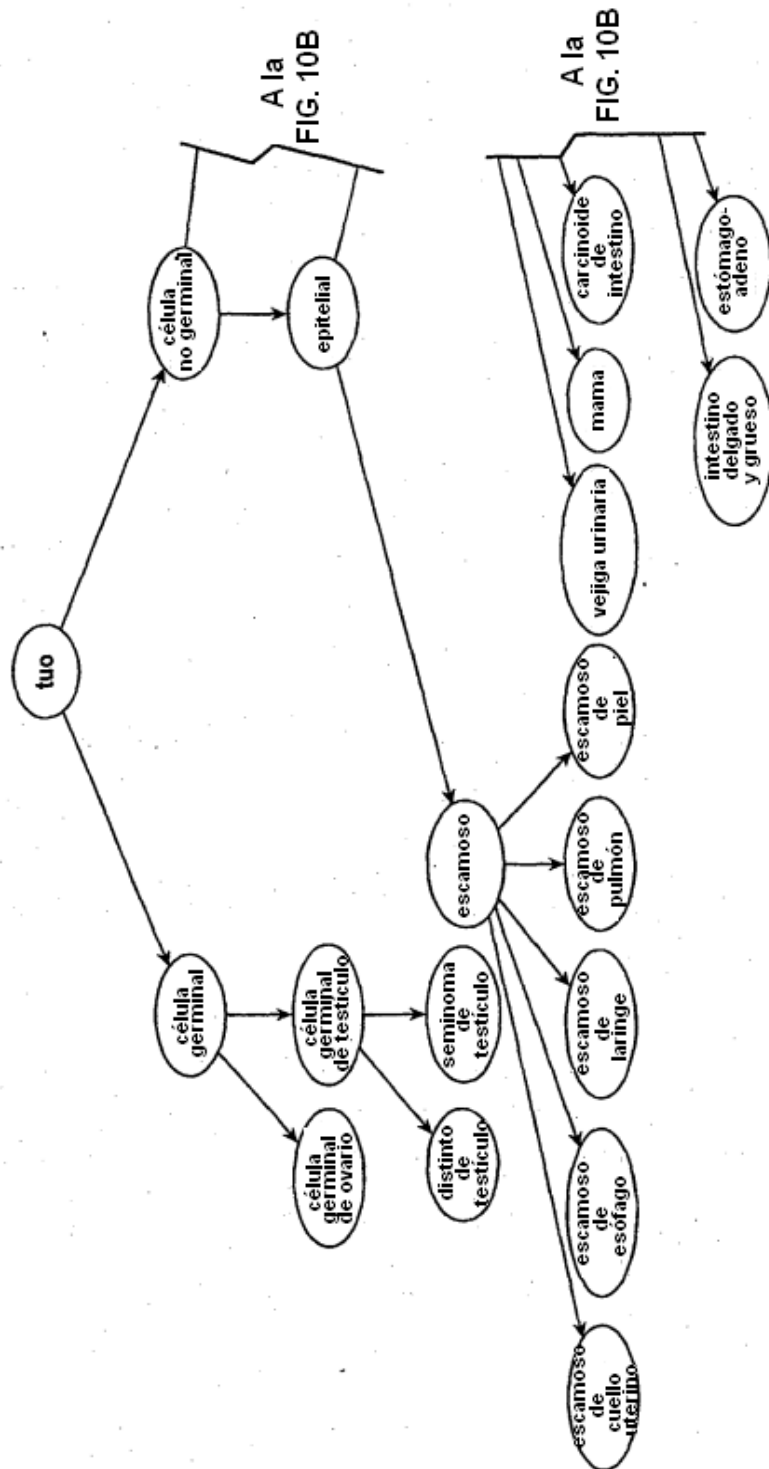


FIG. 10A

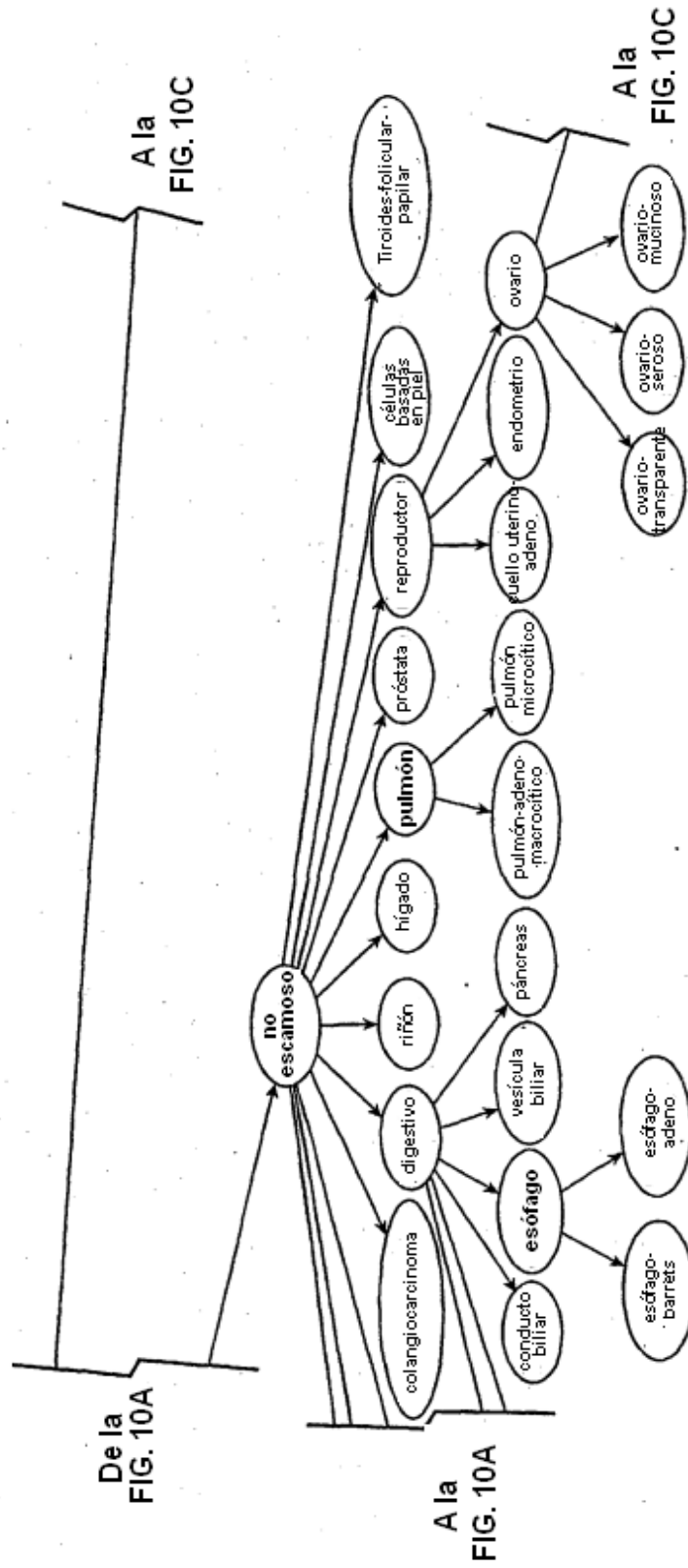
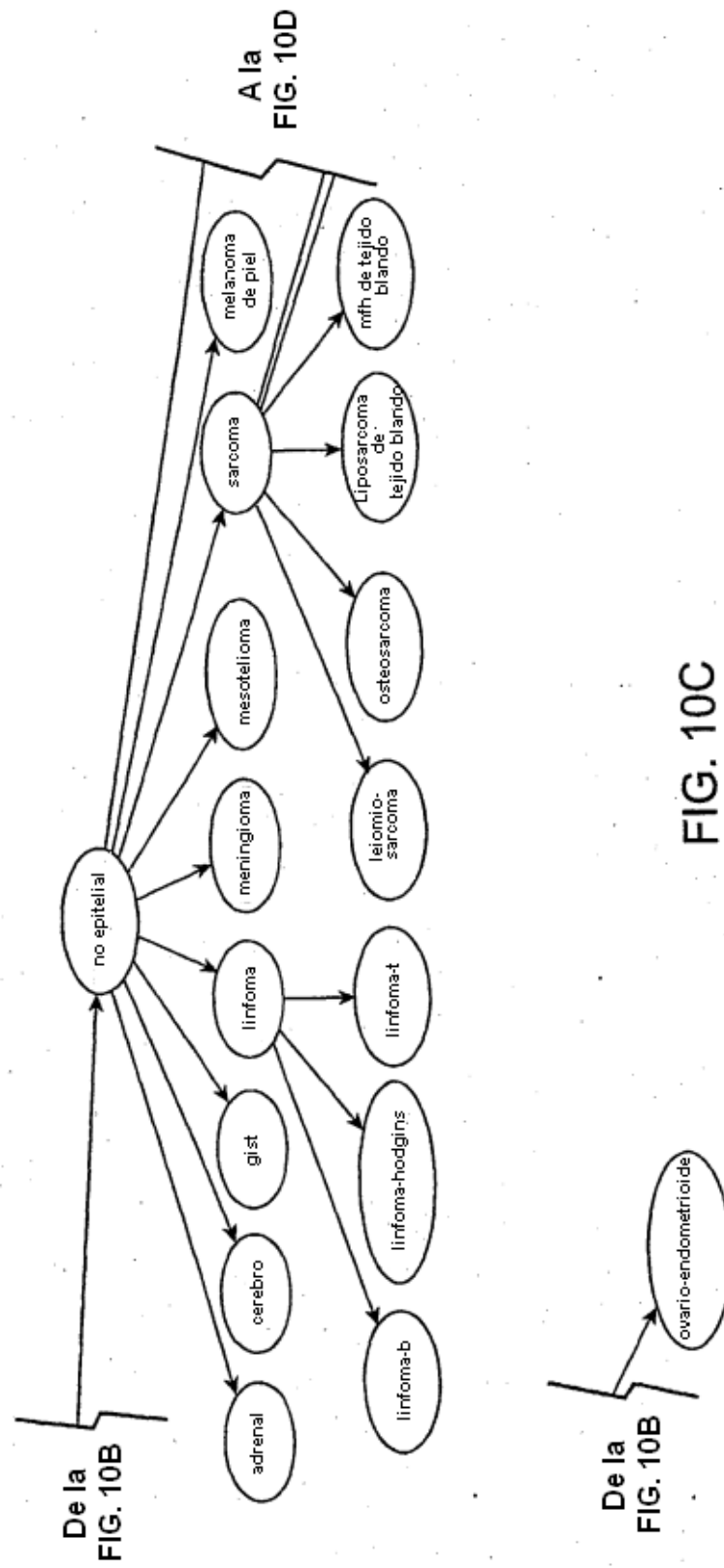


FIG. 10B



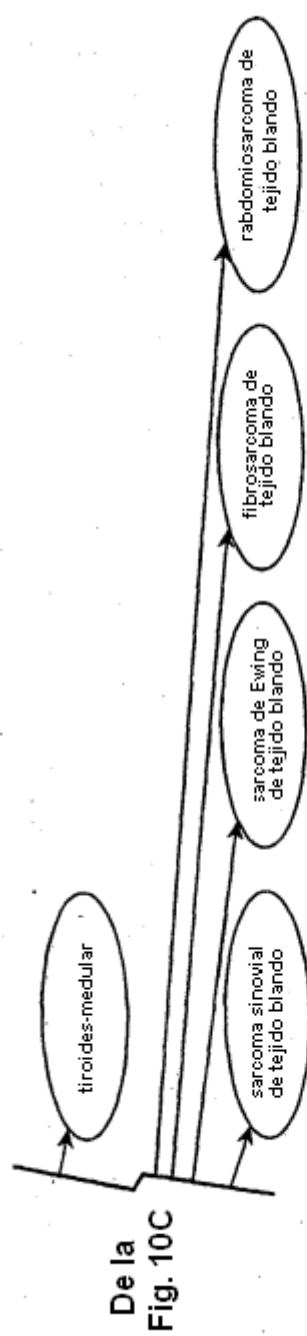


FIG. 10D