(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 553 435**

(51) Int. CI.:

C12Q 1/68 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **20.07.2007 E 11173357 (2)**

(97) Fecha y número de publicación de la concesión europea: **19.08.2015 EP 2468885**

(54) Título: **Infidelidad de la transcripción, su detección y sus usos**

(30) Prioridad:

20.07.2006 EP 06291176

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
09.12.2015

(73) Titular/es:

GENCLIS (100.0%)
15 Rue du Bois de la Champelle
54500 Vandoeuvre-les-Nancy, FR

(72) Inventor/es:

BIHAIN, BERNARD

(74) Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

DESCRIPCIÓN

Infidelidad de la transcripción, su detección y sus usos

La presente invención se refiere a péptidos y composiciones para usos terapéuticos, de diagnóstico, farmacogenómicos o para diseño de fármacos. Tal como se describirá, la invención resulta particularmente adecuada para detectar, controlar o tratar trastornos de células proliferativas.

Introducción a la invención

Los ácidos nucleicos ADN y ARN son macromoléculas que actúan para almacenar información genética en el núcleo celular (ADN) y para tranferir información genómica hacia el citoplasma después de un proceso denominado transcripción, que genera un ARN mensajero para producir una proteína. Ambos son polímeros lineales compuestos de monómeros denominados nucleótidos. El ADN y el ARN representan cada uno combinaciones de cuatro tipos de nucleótidos. Todos los nucleótidos tienen una estructura común: un grupo fosfato unido mediante un enlace fosfodiéster a una pentosa que, a su vez, está unida a una base orgánica: adenina, citosina, timina y guanina (A, C, T, G) para el ADN, y adenina, citosina, uracilo y guanina (A, C, U, G) para el ARN. En el ARN, la pentosa es ribosa, y en el ADN, es desoxirribosa. La molécula de ADN está organizada como una doble cadena de nucleótidos ordenados dispuestos en forma de una doble hélice. Las moléculas de ARN son polinucleótidos monocatenarios que principalmente producen 4 tipos de moléculas: 1) ARN mensajero (ARNm), que son nucleótidos monocatenarios que se transcriben desde el ADN y se traducen en proteínas en el citoplasma, 2) ARN de transferencia (ARNt), que adopta una estructura tridimensional bien definida y se une a aminoácidos (AA) específicos que son transferidos a cadenas polipeptídicas para formar proteínas con secuencias de AA específicas en un proceso denominado traducción, 3) ARN ribosómico (ARNr), que son moléculas más grandes que, junto con proteínas específicas, constituyen el ribosoma, una estructura que permite el ensamblaje de AA para producir proteínas siguiendo la información proporcionada por la secuencia especificada por los codones (un codón = 3 nucleótidos) presentes en un orden concreto sobre el ARNm, y 4) ARN corto regulador, denominado ARN no codificador (ARNnc). La secuencia de ARNm entonces se transforma en un lenguaje diferente, es decir, el lenguaje de AA, mediante un proceso denominado traducción.

En la presente, los inventores demuestran que, por primera vez, la fidelidad de la transcripción, es decir, la transferencia de información del ADN al ARNm, se reduce notablemente en células patológicas, en particular en células del cáncer. Esta falta de fidelidad en la transcripción en las células del cáncer es un fenómeno general que afecta a todos los cánceres y a la mayoría de los genes ensayados en los ejemplos ofrecidos, así como a la mayoría de las transcripciones presentes en las bases de datos disponibles, y que tiene varias implicaciones inmediatas e importantes. En particular, el descubrimiento de la infidelidad de la transcripción permite mejorar las estrategias proteómicas y transcriptómicas que se emplean en la actualidad para la investigación de trastornos patológicos. El descubrimiento de la infidelidad de la transcripción también permite una mejor clasificación de las enfermedades con respecto a su gravedad, y mejora la predicción del efecto terapéutico y el diseño de nuevos métodos para seleccionar fármacos con respecto a su eficacia basándose en su capacidad para corregir la falta o una modificación en la fidelidad de la transcripción. La presente invención, por tanto, tiene implicaciones y utilidades muy importantes en el diagnóstico y en el desarrollo de fármacos, así como en el tratamiento, la detección y el control de pacientes y en la eficacia de fármacos.

Para comprender el impacto del descubrimiento descrito a continuación, es importante reconocer que, en la actualidad, se cree que es necesaria una absoluta fidelidad en la transcripción del ADN al ARN para la función celular normal. Sin embargo, en la actualidad se sigue planteando una cuestión importante aún no resuelta. En efecto, la secuenciación y la anotación del genoma humano condujo a la idea de que están codificados 30-40 mil genes. Este cálculo sigue siendo mucho menor que el número de proteínas que pueden identificarse en la actualidad mediante métodos proteómicos: se han identificado hasta 300 mil proteínas. Para reconciliar estas diferencias, se ha propuesto que un único gen pueda producir varios ARNm diferentes que codifiquen diferentes proteínas mediante un proceso denominado corte y empalme alternativo. El corte y empalme alternativo elimina diferentes partes de un ARN premensajero (ARNpm), conduciendo con ello a la eliminación de diferentes elementos que corresponden a intrones específicos, y produce diferentes ARNm maduros. El análisis de la base de datos RefSeq indica que contiene 11259 transcripciones que corresponden a 6946 genes. Por tanto, el corte y empalme alternativo puede aumentar la heterogeneidad de las transcripciones en 76%. El mecanismo que describen en la presente los inventores es diferente del corte y empalme alternativo, e induce una heterogeneidad de proteínas mucho mayor. En efecto, los inventores demuestran la aparición de modificaciones no aleatorias en la secuencia del ARNm que producen cambios en los AA codificados, la introducción de codones de fin prematuros que producen isoformas de proteínas más cortas, y la alteración de los codones de fin naturales que implica la introducción de nuevas secuencias codificadoras que producen isoformas de proteínas alargadas. La introducción de huecos e inserciones modifica así el marco de lectura del ARNm y crea de ese modo secuencias de proteínas insospechadas. Los datos descritos en esta invención demuestran que este fenómeno de infidelidad de la transcripción está presente en células normales pero está notablemente potenciado en células patológicas, tales como células derivadas de un cáncer. Por tanto, se propone que la infidelidad de la transcripción (IT) contribuye a la diversificación de la información transferida desde el ADN al ARN. Los resultados de los investigadores también establecen que la infidelidad de la transcripción no se produce de modo aleatorio, sino que sigue unas reglas específicas en las que

resulta importante el contexto que rodea a b0, las bases afectadas por el acontecimiento de IT.

Es posible imaginarse otro mecanismo para explicar que la heterogeneidad del ARN no aparece a nivel genómico sino a nivel de ARN: la edición de ARN. Sin embargo, la edición de ARN no puede explicar dos tipos del acontecimiento de IT, la introducción de huecos y la inserción. La edición de ARN se caracteriza por unos cambios de bases postranscripcionales en el ARNm y ARNt en eucariotas. La inmensa mayoría de los acontecimientos de edición de sustitución conocidos consisten en conversiones de C a U, o de A a I (leído como G) (Gott, J.M. y Emeson, R.B. (2000), *Annu. Rev. Genet.*, **34**, 499-531; Maas, S. y Rich, A. (2000), *Bioessays*, **22**, 790-802; Niswender, C.M. (1998), *Cell Mol. Life Sci.*, **54**, 946-964). Diferentes trabajos han demostrado que estas conversiones surgen a través de mecanismos de desaminación, catalizados por adenosina y citidina desaminasas. Otro acontecimiento de edición de sustitución es la converción de “U a C”. Aunque la teoría de la reversibilidad microscópica dicta que la reacción de la citidina desaminasa puede ir hacia atrás para generar esta conversión, también se ha propuesto que una actividad de tipo CTP sintetasa puede ser responsable. Se ha demostrado en varios ejemplos que la edición de ARN puede ser específica de células de cáncer frente a células normales. Además, la tasa de este fenómeno puede verse afectada en condiciones de cáncer.

En el conjunto con cáncer, a nivel de ADNc, los inventores observaron cambios de 5,7% de C $\rightarrow$ T, 9,2% de T $\rightarrow$ C, y 4,7% de A $\rightarrow$ G. Por tanto, la edición de ARNm puede justificar no más que 20% de las sustituciones de una sola base descritas en la presente. En efecto, las sustituciones de bases más comunes, A $\rightarrow$ C (24,6%) y T $\rightarrow$ G (16,8%), representan cambios en la familia de bases que, en la actualidad, no pueden explicarse por los procesos de edición del ARN enzimáticos humanos conocidos.

Además, se espera que la infidelidad de la transcripción provoque deleciones y/o inserciones de bases a través de un mecanismo (o mecanismos) diferente de la edición del ARNm. Además, un estudio reciente ha demostrado que una serie de registros de dbSNP (“Single Nucleotide Polymorphism dababase”, base de datos de polimorfismos de nucleótidos individuales) son, de hecho, sitios de edición (Eisenberg, E. et al. (2005), *Nucleic Acids Res.*, **33** (14), 4612-4617). En la estrategia de los inventores, no se han considerado todos los SNP de dbSNP, por tanto se excluyen los SNP conocidos o los SNP falsos que corresponden a sitios de edición.

Por tanto, el mecanismo para explicar la heterogeneidad del ARNm en el cáncer observada es la infidelidad de la transcripción.

Otro estudio (Ruiz et al., 2004, Clinical cancer Research 10 (8), 2560-2567) se refiere a la identificación de linfocitos T colaboradores determinantes de antígeno carcinoembriónico, y que se predice que son potenciales enlaces a moléculas HLA-DR. Sin embargo, este documento no describe ninguna secuencia ni péptido sintético, como reivindican los presentes inventores.

Sumario de la presente invención

La presente invención se refiere a péptidos sintéticos, composiciones y usos de los mismos para fines terapéuticos y de diagnóstico, por ejemplo, para detectar y/o tratar enfermedades de células proliferativas El objeto de esta invención se refiere al uso de un péptido sintético como se define en las reivindicaciones.

La invención también describe el uso de un compuesto que altera (por ejemplo, reduce o aumenta) la tasa de la infidelidad de la transcripción de un gen de mamífero para la fabricación de una composición farmacéutica para su uso en un método de tratamiento de un ser humano o de un animal, en particular para tratar enfermedades tales como trastornos de células proliferativas incluyendo, pero no limitándose a, cánceres, enfermedades inmunes, enfermedades inflamatorias o envejecimiento.

La invención también describe métodos y productos (tales como sondas, cebadores, anticuerpos o sus derivados) para detectar o medir (el nivel de) la infidelidad de la transcripción en una muestra, así como los correspondientes kits.

La invención describe además métodos para identificar secuencias de infidelidad de la transcripción en proteínas o en ácidos nucleicos, así como su uso, por ejemplo, como marcadores, inmunógenos y/o para generar ligandos específicos para ellos.

A este respecto, la invención describe un método para identificar y/o producir biomarcadores, comprendiendo dicho método identificar, en una muestra procedente de un sujeto, la presencia de un sitio o sitios de infidelidad de la transcripción en una proteína o un ácido nucleico diana y, opcionalmente, determinar la secuencia de dicho sitio o sitios de infidelidad de la transcripción.

La invención también describe un método para identificar y/o producir un ligando específico para un rasgo o un trastorno patológico, comprendiendo dicho método identificar, en una muestra procedente de un sujeto que tiene dicho rasgo o trastorno patológico, la presencia de al menos un sitio de infidelidad de la transcripción en una o varias proteínas o ácidos nucleicos diana, opcionalmente determinando la secuencia de dicho al menos un sitio de infidelidad de la transcripción, y produciendo un ligando que se une de modo específico a dicha al menos una proteína (o dominio) o ácido nucleico creado mediante la infidelidad de la transcripción.

La invención resulta particularmente adecuada para identificar biomarcadores de trastornos de células proliferativas, tales como cánceres, enfermedades inmunes, enfermedades inflamatorias o envejecimiento. Es particularmente útil para producir ligandos que son específicos de tales trastornos en mamíferos, en particular ligandos que pueden detectar la presencia o gravedad de un trastorno proliferativo en un sujeto.

Esta invención también describe un ligando que se une de modo específico a una proteína (o dominio) o a un ácido nucleico creado mediante la infidelidad de la transcripción. El ligando puede ser un anticuerpo (o cualquier fragmento (tal como Fab, Fab', CDR, etc.) o un derivado de éste, tal como se describirá a continuación), que se une de modo específico a una proteína (o dominio) creado mediante la infidelidad de la transcripción. El ligando también puede ser un ácido nucleico que se une de modo específico con un ácido nucleico creado mediante la infidelidad de la transcripción (por ejemplo, una sonda, un cebador, un ARNi (ARN de interferencia), etc.).

Esta invención también describe un péptido que comprende un dominio de una proteína creada mediante la infidelidad de la transcripción, en particular de una proteína de mamífero, más preferiblemente de una proteína humana. El péptido es generalmente un péptido sintético, es decir, un péptido que se ha preparado de modo artificial, por ejemplo, mediante síntesis química, síntesis y/o extensión de aminoácidos, digestión de proteínas, ensamblaje de péptidos, expresión recombinante, etc. El péptido comprende generalmente la secuencia de un fragmento C-terminal de dicha proteína. El péptido comprende preferiblemente menos de 100, 80, 75, 70, 65, 60, 50, 45, 40, 35, 30, 25 o incluso 20 aminoácidos (aunque, en otras realizaciones, la longitud del péptido puede ser mayor). La proteína puede ser una proteína de la superficie celular (por ejemplo, un receptor, etc.), una proteína segregada (por ejemplo, una proteína plasmática, etc.) o una proteína intracelular.

Otro aspecto de esta invención reside en el uso de un péptido sintético como se define anteriormente con una longitud de 100 aminoácidos o menos y, en particular, que comprende una secuencia seleccionada entre las SEQ ID NOs:1 a 5, 7 a 13, 15 a 18 y 20 a 32 para detectar o controlar trastornos de células proliferativas.

La invención también describe una composición de vacuna que comprende un péptido que comprende una parte de una proteína creada mediante la infidelidad de la transcripción. Un objetivo adicional de la invención reside en una composición de vacuna que comprende un péptido sintético, como se define anteriormente, con una longitud de 100 aminoácidos o menos, que comprende una secuencia seleccionada de SEQ ID NOs: 1 a 5, 7 a 13, 15 a 18 y 20 a 32, y opcionalmente un vehículo, excipiente y/o adyuvante adecuado.

Esta invención también se describe un método para producir un anticuerpo, comprendiendo dicho método inmunizar a un mamífero no humano con un péptido creado mediante la infidelidad de la transcripción. Un objetivo adicional de esta invención reside en un método para producir un anticuerpo, comprendiendo el método inmunizar a un mamífero no humano con un péptido sintético con una longitud de 100 aminoácidos o menos, que comprende una secuencia seleccionada de SEQ ID NOs: 1 a 5, 7 a 13, 15 a 18 y 20 a 32, como se define anteriormente, y recuperando anticuerpos que se unen a dicho péptido, o las correspondientes células productoras de anticuerpos. Opcionalmente, pueden producir derivados del anticuerpo.

La invención también describe un método para producir un anticuerpo, comprendiendo dicho método (i) identificar un dominio de una proteína creada mediante la infidelidad de la transcripción, y (ii) producir un anticuerpo que se une de modo específico a ese dominio. Opcionalmente, pueden producirse derivados del anticuerpo.

Esta invención también describe un anticuerpo que se une de modo específico a una parte de una proteína creada mediante la infidelidad de la transcripción, o un derivado de dicho anticuerpo que tiene sustancialmente la misma especificidad de antígeno. En particular, la invención reside en un anticuerpo, o derivado del mismo, que se une específicamente a un péptido sintético con una longitud de 100 aminoácidos o menos, que comprende una secuencia seleccionada de SEQ ID NOs: 1 a 5, 7 a 13, 15 a 18 y 20 a 32. El anticuerpo puede ser policlonal o monoclonal. El término derivado incluye cualquier fragmento (tal como Fab, Fab', CDR, etc.) u otro derivado, tal como anticuerpos monocatenarios, anticuerpos bifuncionales, anticuerpos humanizados, anticuerpos humanos, anticuerpos quiméricos, etc.

Otro aspecto de esta invención se refiere a un anticuerpo o su derivado, según se definió anteriormente, que está conjugado con una molécula. La molécula puede ser un fármaco, un marcador, una molécula tóxica, un radioisótopo, etc.

La invención también describe al uso de un anticuerpo o su derivado, tal como se definió anteriormente, para detectar o cuantificar (por ejemplo, *in vitro*) la infidelidad de la transcripción de un gen.

La invención también se refiere al uso de un anticuerpo conjugado o su derivado, tal como se definió anteriormente, como medicamento o reactivo de diagnóstico.

La invención también describe un dispositivo o producto que comprende, inmovilizado sobre un soporte, un reactivo que se une de modo específico con una proteína o un ácido nucleico creado mediante la infidelidad de la transcripción. El reactivo puede ser, por ejemplo, una sonda, o un anticuerpo o su derivado.

La invención también describe un método para provocar o inducir o estimular la infidelidad de la transcripción en una

célula, tejido u organismo. Este método puede comprender, generalmente, la introducción de un sitio de infidelidad de la transcripción en secuencias genómicas normales, por ejemplo, mediante un vector de terapia génica. Esta modificación puede dar como resultado la destrucción de una célula o un tejido. En particular, es posible crear un marco de lectura abierto en cualquier secuencia génica que de cómo resultado la producción de compuestos o proteínas tóxicas celulares. También es posible provocar la expresión, en una célula enferma (o diana), de un biomarcador específico que pueda fijarse como objetivo para moléculas tóxicas o terapéuticas. Utilizando esta estrategia, es por tanto posible provocar la muerte celular o una terapia cuando la infidelidad de la transcripción aparezca o exceda una tasa concreta.

La invención también describe el uso de un ácido nucleico que contiene un sitio o sitios de infidelidad de la transcripción para la fabricación de una composición farmacéutica para su uso en un método de tratamiento de un ser humano o animal. El gen puede ser cualquier gen, incluyendo un gen de mamífero o un gen procedente de un agente patógeno, tal como un gen vírico, un gen bacteriano, etc. A este respecto, la invención se refiere a un método para tratar una enfermedad provocada por un patógeno, comprendiendo dicho método provocar o estimular la infidelidad de la transcripción de un gen codificado por dicho patógeno.

En una realización particular, la invención se refiere a una molécula de ácido nucleico sintético que codifica un péptido de 100 aminoácidos o menos, que comprende una secuencia seleccionada de SEQ ID NOs: 1 a 5, 7 a 13, 15 a 18 y 20 a 32 para detectar o controlar trastornos de células proliferativas.

La invención también describe métodos para producir polipéptidos recombinantes *in vitro* con una infidelidad de la transcripción reducida. Estos métodos permiten una reducción en la microheterogeneidad en polipéptidos recombinantes. El método comprende utilizar células hospedantes que comprenden un ácido nucleico recombinante con una utilización de codones adaptada, para reducir la aparición de la infidelidad de la transcripción. El método también puede incluir el uso de cualquier compuesto o tratamiento que reduzca la aparición de la infidelidad de la transcripción. El método puede utilizarse en células hospedantes procariotas o eucariotas, por ejemplo, en cepas de *E. coli* o CHO.

La invención puede utilizarse en cualquier sujeto mamífero, en particular cualquier sujeto humano, para detectar, controlar o tratar una diversidad de trastornos patológicos asociados con trastornos de células proliferativas (por ejemplo, cánceres) y/o para producir, diseñar o seleccionar fármacos terapéuticamente activos.

Leyendas de las figuras

Figura 1: Principio de construcción y secuenciación de bancos de ADNc.

Figura 2: (a-q) Secuencias de referencia de ARNm utilizadas para el análisis. Para evitar distorsionar BLAST, la cola de poliA de las secuencias de referencia de ARNm se eliminó sistemáticamente. La figura 2r proporciona una lista de genes ensayados.

Figura 3: Archivos de salida MegaBLAST típicos.

Figura 4: Porcentaje de desviación de la secuencia de nucleótidos en cualquier posición para cada gen estudiado.

Figura 5: TPT1, VIM = número de EST y análisis del ensayo de proporción.

Figura 6: Variación en los EST antes y después de la aplicación secuencial de filtros electrónicos. Efectos de la eliminación de líneas celulares y Clip 400.

Figura 7: Contexto del ADN: a) Efecto de la composición de bases del ARNpm sobre la heterogeneidad de las bases b0. b) Composición de la base de sustitución para cada base sustituida. c) Reparto de bases afectadas y de sustitución dentro de ensayos C>N estadísticamente significativos. d) Efecto de la composición de bases del ARNpm sobre la base de sustitución que corresponde a la repetición de b-1 o b+1.

Figura 8: Variante de ARNm virtual y proteína para VIM.

Figura 9: Impacto en la codificación.

Figura 10: Proteínas de interés = secuencias de aminoácidos en las que el codón de fin se ve estadísticamente afectado.

Figura 11: DHPLC = principio y límites de la detección.

Figura 12: DHPLC = cebadores y productos de la PCR esperados.

Figura 13: Lista de 60 proteínas estudiadas.

Figura 14: Producción de anticuerpos.

Figura 15: Resultados de DHPLC.

Figura 16: Sensibildad de Sanger

Figura 17: Detección de PSP de APOAII y APOCII en plasma de pacientes cancerosos.

Figura 18: PSP potencial en proteínas plasmáticas.

Figura 19: Evaluación de las predicciones de sustituciones en la secuencia del ARNm.

Descripción detallada de la invención

Definiciones

Infidelidad de la transcripción

La expresión infidelidad de la transcripción indica un nuevo mecanismo mediante el cual se producen varias moléculas de ARN diferenciadas en una célula a partir de una única secuencia génica. Este mecanismo recién identificado afecta potencialmente a muchos genes, no es aleatorio, y sigue una reglas concretas, tal como se describirá a continuación. Tal como se muestra en los ejemplos, la infidelidad de la transcripción puede introducir sustituciones, deleciones e inserciones en moléculas de ARN, creando con ello una diversidad de proteínas a partir de un único gen. La infidelidad de la transcripción también puede afectar a secuencias de ARN no codificadoras, modulando con ello sus funciones. La medición, la modulación o la localización de la infidelidad de la transcripción, por tanto, representa una nueva estrategia para detectar o tratar trastornos, así como para el desarrollo de fármacos.

Sitio de infidelidad de la transcripción

La expresión sitio de infidelidad de la transcripción indica una secuencia y/o una posición afectada por la infidelidad de la transcripción. Puede ser un dominio de ácido nucleico o de aminoácido que contenga al menos una modificación generada como resultado de la infidelidad de la transcripción. Dicha modificación puede ser el resultado, por ejemplo, de una conmutación en el marco de lectura (inserción y/o deleción), de la introducción o supresión de codones de fin, de la introducción de nuevos codones de inicio, de una sustitución o sustituciones de nucleótidos (que impliquen o no cambios en AA), etc. Un sitio de infidelidad de la transcripción puede comprender una o varias variaciones de la secuencia que sean el resultado de la infidelidad de la transcripción. Un sitio de infidelidad de la transcripción puede comprender, por ejemplo, un resto nucleótido o aminoácido modificado hasta, por ejemplo, 150 restos nucleótidos o aminoácidos modificados, o incluso más. El sitio de infidelidad de la transcripción generalmente se diferencia de la secuencia que resulta de una transcripción fiel en al menos una modificación de nucleótido o de aminoácido (por ejemplo, sustitución, deleción, inserción, inversión, etc.). Una proteína o un dominio creado mediante la infidelidad de la transcripción (proteína-IT) generalmente comprende de 1 a 50 aminoácidos (o incluso más), con al menos un resto aminoácidos modificado. Una secuencia de infidelidad de la transcripción de un ácido nucleico comprende generalmente de 1 a 150 nucleótidos (o incluso más), con al menos un resto nucleótido modificado.

Reglas e identificación de la infidelidad de la transcripción

Basándose en técnicas específicas de detección y/o reglas concretas de la infidelidad de la transcripción según se define en la presente solicitud, ahora es posible predecir e identificar sitios de infidelidad de la transcripción en cualquier gen o proteína. Estos sitios también pueden obtenerse mediante el alineamiento de secuencias disponibles para un ARN concreto y la identificación de las sustituciones de bases. En última instancia pueden validarse mediante diversas técnicas, por ejemplo, utilizando ligandos específicos para una proteína-IT predicha.

Un método particular para identificar un sitio de infidelidad de la transcripción comprende:

- proporcionar la secuencia de un gen, molécula de ARN o ADNc concretos, o una porción de éstos, e

- identificar, dentro de dicha secuencia, la presencia de un cambio en algún nucleótido (por ejemplo, sustitución, deleción, inserción, inversión) que resulte de la infidelidad de la transcripción, siguiendo las reglas de la infidelidad de la transcripción, por ejemplo, tal como se analiza en (iii) a continuación.

Otro método particular para identificar un sitio de infidelidad de la transcripción comprende:

- proporcionar la secuencia de una proteína concreta, o una porción de ésta, e

- identificar, dentro de dicha secuencia, la presencia de un cambio en algún aminoácido (por ejemplo, sustitución, deleción, inserción, inversión) que resulte de la infidelidad de la transcripción, siguiendo las reglas de la infidelidad de la transcripción, por ejemplo, tal como se analiza en (iii) a continuación.

La presencia de un cambio resultante de la infidelidad de la transcripción en una molécula puede identificarse en un proceso en tres etapas, que comprende:

(i) la identificación de sitios de infidelidad de la transcripción con una máquina de aprendizaje que se basa en la

discriminacion cuadrática de fáctores de análisis factorial de correspondencia múltiple (MCFA),

(ii) la identificación de la categoría de los sitios de infidelidad de la transcripción (b-1 o b+1 u otros) con el mismo método, y

(iii) la predicción de la base de sustitución utilizando las siguiente reglas de la infidelidad de la transcripción:

· Para cualquier base presente como un singletón, es decir, cualquier base que no esté precedida ni seguida por ella misma, entonces es muy probable que la base sustituida sea idéntica a la base que precede a la que está sustituida (en caso de la categoría b-1), o idéntica a la base que sigue a la que está sustituida (en el caso de la categoría b+1). Por ejemplo, CAT será CTT (regla b-1); ATG será AGG (regla b+1);

· Cuando las sustituciones se producen dentro de tres A consecutivas, la sustitución de la segunda A preferentemente es a C;

· Cuando las sustituciones se producen dentro de tres T consecutivas, la sustitución de la segunda T preferentemente es a C, después a A, después a G;

· Los tramos de C y G raramente están sustituidos en la segunda, pero si se produce, la sustitución de la segunda C preferentemente es a A, y la sustitución de la segunda G preferentemente es a C;

· Para otros casos, la base de sustitución es preferentemente C.

Otros métodos para identificar sitios de infidelidad de la transcripción dentro de la secuencia de cualquier proteína o ácido nucleido diana comprenden, por ejemplo, la comparación de marcadores de secuencia expresados ("expressed sequence tag", EST) existentes, la amplificación directa de secuencias identificadas y la secuenciación de productos de la amplificación con un método de secuenciación específico (Sanger, químico, pirosecuenciación). Esta estrategia produce secuencias variantes que implican cambios en AA, modificaciones de la longitud de la proteína, etc. Entonces pueden diseñarse oligonucleótidos que se aparean de modo selectivo con ARNm o ADNc que portan sitios de infidelidad de la transcripción, permitiendo la amplificación selectiva de nucleótidos con sitios de infidelidad de la transcripción a partir de un agrupación de secuencias. También pueden producirse anticuerpos que se dirijan a una proteína-IT.

Como alternativa, es posible identificar un sitio de infidelidad de la transcripción mediante el análisis de la secuencia o mediante reglas bioinformáticas. La secuencia resultante puede clonarse en cualquier vector de transfección. También es posible diseñar una construcción que comprenda dicha secuencia de infidelidad de la transcripción dentro de marco con un gen indicador, que será transcrito y traducido sólo si se produce la infidelidad de la transcripción. El gen indicador puede ser una enzima o una proteína fluorescente o cualquier gen que haga que la célula diana sea sensible, o resistente a la exposición a una toxina.

Un método detallado que permite la identificación de los sitios de infidelidad de la transcripción (o sustituciones) incluye el método TDG según se describe a continuación.

Una técnica descrita por Pan y Weissman y Liu et al. (PNAS, 2002, 99 (14), 9346-9351; y Anal. Biochem., 1 de septiembre, 2006, 356 (1):117-124) puede adaptarse para detectar un sitio o sitios de infidelidad de la transcripción en el ARN. Esta técnica se basa en el uso de una enzima, la timina ADN glicosilasa (TDG), que es capaz de enriquecer por separado dúplex de ADN que contienen desapareamientos en mezcla complejas. Estas enzimas reconocen de modo específico los desapareamientos de nucleótidos y generan sitios abásicos (el enlace entre la desoxirribosa y una de las bases del ADN está roto). Entonces, la TDG puede unirse de modo reversible a estos sitios abásicos y, por tanto, puede utilizarse para la purificación por afinidad de los fragmentos de ADN que contienen desapareamientos. En un experimento típico, el ARN se extrae de 2 tipos celulares (normal y canceroso) y se somete a una transcripción inversa para fabricar ADNc bicatenario. Después de una desnaturalización térmica y una lenta renaturalización de cada muestra, se genera un ADNc con uno o más desapareamientos y puede separarse de los dúplex perfectos. Estos ADNc entonces pueden analizarse mediante secuenciación directa, o compararse mediante DHPLC (véase a continuación).

Otras técnicas que pueden adaptarse para identificar sitios de infidelidad de la transcripción o mutaciones se describen, por ejemplo, en los documentos US 6.329.147, US 4.979.330, WO 02/077286 o US 6.120.992.

Generalmente, los métodos para identificar sitios de infidelidad de la transcripción comprenden también una etapa de validar la secuencia del sitio de infidelidad de la transcripción produciendo una molécula que comprende la secuencia identificada, generando un ligando que se une de modo específico a dicha molécula, y verificando en una muestra biológica la presencia de un antígeno que es reconocido de modo específico por dicho ligando.

Una secuencia de infidelidad de la transcripción típica de un ácido nucleico es una secuencia con una longitud entre 1 y 150 nucleótidos que comprende al menos una sustitución, deleción o inserción de bases provocada por la infidelidad de la transcripción, preferiblemente siguiendo una regla según se describe en el ejemplo 5. La secuencia de infidelidad de la transcripción puede ser mayor.

Un dominio típico de una proteína-IT es una secuencia de aminoácidos con una longitud entre 1 y 50 aminoácidos que comprende al menos una sustitución de un aminoácido (o más cambios debidos a la modificación en el marco de lectura como resultado de una deleción o inserción de bases) provocada por la infidelidad de la transcripción, preferiblemente siguiendo una regla según se describe en el ejemplo 5. El dominio de una proteína-IT puede ser mayor. Los ejemplos de proteínas-IT se proporcionan en la sección experimental, por ejemplo, en las figuras 13 y 14.

Detectar o medir la infidelidad de la transcripción

Esta invención describe un descubrimiento inesperado de que se produce una alta tasa de anomalías de secuencias naturales durante o al poco tiempo después de la transcripción del ADN a ARN. El proceso puede estar presente en células normales, pero aumenta mucho en células patológicas, por ejemplo, células del cáncer. El descubrimiento de que la transcripción del ADN en ARN está introduciendo variaciones en la secuencia siguiendo unas reglas que son diferentes de las definidas por las complementariedades de bases una a una permite el diseño lógico de nuevos reactivos (por ejemplo, sondas, cebadores, anticuerpos, aptámeros, etc.) que sean específicos para el ácido nucleico o la proteína creada mediante la infidelidad de la transcripción. Estos reactivos por tanto permiten detectar o medir la infidelidad de la transcripción, y discriminar entre condiciones normales o enfermas, así como el transporte dirigido de moléculas a células que muestren una infidelidad de la transcripción.

Además, la invención también describe el diseño lógico de reactivos (por ejemplo, sondas, cebadores, anticuerpos, aptámeros) que sean predictivos de la gravedad de una enfermedad (por ejemplo, cáncer). En efecto, se espera que una mayor tasa de infidelidad de la transcripción se correlacione con la gravedad de la enfermedad, y esta tasa aumenta progresivamente a medida que más y más productos génicos se ven afectados. La medición directa de la (in)fidelidad de la transcripción en una célula o células enfermas utilizando los métodos descritos en la presente o cualquier otra tecnología permite detectar células en diferentes etapas de avance de la enfermedad en cualquier tejido concreto. La medición precisa de estas variaciones en la expresión génica (transcripciones y proteínas) también mejora la capacidad para evaluar la eficacia de fármacos con respecto a la gravedad de la enfermedad. En la actualidad, diversas técnicas de micromatrices permiten la medición de los cambios en la expresión génica. Sin embargo, la fiabilidad y, de modo más importante, la reproducibilidad de estos datos en la actualidad no es suficiente. Los inventores ahora pueden postular que una gran parte de la variabilidad de estos experimentos transcriptómicos es provocada, en parte, por la infidelidad de la transcripción según ha sido descubierta por los inventores. Por tanto, el descubrimiento de este fenómeno común permite el diseño de reactivos de expresión génica que se vean mínimamente afectados por la infidelidad de la transcripción, o que directamente reflejen la infidelidad de la transcripción que se produce en secuencias específicas siguiendo las reglas descritas en la presente solicitud.

También es posible e incluso probable que la infidelidad de la transcripción esté modulada por la tasa de transcripción. Los inventores especulan que un aumento en la expresión de un gen concreto en un trastorno patológico aumentará la IT y así aumentará la heterogeneidad de proteínas.

La infidelidad de la transcripción puede detectarse o medirse utilizando una serie de técnicas conocidas per se en la técnica, que pueden adaptarse. En particular, la infidelidad de la transcripción puede medirse utilizando reactivos específicos para ácidos nucleicos o proteínas creadas mediante la infidelidad de la transcripción, tales como sondas, cebadores o anticuerpos específicos, mediante electroforesis, análisis de migración en gel, espectrometría, etc., que permitan la detección entre varias formas de una proteína o de un ácido nucleico. La tasa de la infidelidad de la transcripción puede determinarse evaluando el número de genes en una células que están sometidos a la infidelidad de la transcripción y/o el número de sitios de infidelidad de la transcripción generados para un gen concreto. Esta tasa puede determinarse comparando el nivel de ácidos nucleicos o proteínas creados mediante la infidelidad de la transcripción en una muestra de ensayo, con el obtenido en una muestra de referencia.

Detección de la infidelidad de la transcripción a nivel de ácidos nucleicos

Tal como se analizó anteriormente, la infidelidad de la transcripción introduce una variación o variaciones en la secuencia de moléculas de ARN. La detección o la medición de la infidelidad de la transcripción por tanto puede realizarse detectando la presencia o la cantidad (absoluta o relativa) de dicha variación o variaciones en la secuencia de los ARN codificados por uno o varios genes, o en una célula entera o tejido o muestra. La detección de la infidelidad de la transcripción en ácidos nucleicos puede realizarse mediante diversas técnicas, tales como hibridación, amplificación, formación de heterodúplex, etc.

Casi todas las tecnologías utilizadas para la identificación de nucleótidos presentes en diversos tipos de tejidos se basa en un proceso denominado hibridación. La hibridación resulta crítica para tecnologías, tales como micromatrices, utilizadas para identificar polimorfismos y para buscar variaciones en la expresión génica. Esta técnica se aplica para medir el nivel de expresión génica. La hibridación de sondas oligonucleotídicas también se emplea para retirar secuencias de genes que se expresan con abundancia para permitir un mejor estudio de los mensajes de baja abundancia; este procedimiento se denomina hibridación sustractiva. La hibridación también es la primera etapa de la reacción de amplificación de genes que se emplea habitualmente en experimentos de PCR, en la forma directa o después de la aplicación de la transcriptasa inversa para convertir la secuencia del mensaje desde una secuencia de tipo ARN a una secuencia de tipo ADN.

El mecanismo básico que subyace a la hibridación es que la secuencia o secuencias de nucleótidos monocaternarias pueden unirse entre sí con la condición de que sus respectivas secuencias sean complementarias. Tanto la longitud como la ordenación específica de cada una de las 4 bases define la eficacia de la hibridación y la especificidad de la secuencia. Esto implica que cada base específica se une de manera no covalente a su base complementaria: A a T, y C a G, y viceversa. Esta unión no covalente en un apareamiento de secuencias está condicionada por la ordenación de la base. Así, en la práctica general, una secuencia definida de bases se une a una única secuencia complementaria. Esta reacción de unión específica proporciona la base para la hibridación que permite la identificación de cualquier secuencia complementaria en cualquier mezcla concreta de nucleótidos. La eficacia de la hibridación viene determinada por el número de bases que, en el orden apropiado, se aparean entre sí (se tolera un cierto grado de desapareamientos) y también por las condiciones del experimento, tales como la rigurosidad del tampón de unión, el contenido relativo en CG frente a TA (CG y TA están unidos mediante 3 y 2 enlaces de hidrógeno, respectivamente) y por la temperatura de fusión, es decir, la temperatura necesaria para permitir la disociación del 50% del ADN bicatenario.

La detección de la infidelidad de la transcripción empleando la hibridación generalmente comprende colocar una muestra en contacto con una sonda de ácido nucleico que sea específica para una secuencia de infidelidad de la transcripción, y detectar la presencia (o cantidad) de los híbridos formados, siendo dicha presencia o cantidad una indicación directa de la presencia o la tasa de la infidelidad de la transcripción. En una realización preferida, el método emplea un conjunto de sondas de ácidos nucleicos que son específicas para secuencias de infidelidad de la transcripción diferenciadas de uno o varios genes, respectivamente. Pueden prepararse sondas de ácidos nucleicos específicas para una secuencia de infidelidad de la transcripción para cualquier gen, empleando la información de la secuencia y las reglas de sustitución, inserción o deleción de nucleótidos según se describe en la presente solicitud (véase, por ejemplo, el ejemplo 5).

Una “sonda” de ácido nucleico se refiere a un ácido nucleico u un oligonucleótido que tiene una secuencia polinucleotídica que es capaz de una hibridación selectiva con una secuencia de infidelidad de la transcripción o su complemento, y que es adecuada para detectar la presencia (o su cantidad) en una muestra que contiene dicha secuencia o complemento. Las sondas preferiblemente son perfectamente complementarias con una secuencia de infidelidad de la transcripción, aunque se tolera un cierto grado de desapareamiento. Las sondas comprenden generalmente ácidos nucleicos monocatenarios con una longitud de 8 a 1500 nucleótidos, por ejemplo entre 10 y 1000, más preferiblemente entre 10 y 800, generalmente entre 20 y 700. Debe entenderse que también pueden emplearse sondas más largas. Por ejemplo, una sonda puede ser una molécula de ácido nucleico monocatenario de 8 a 400 nucleótidos de longitud, que puede hibridarse de modo específico con una secuencia de infidelidad de la transcripción.

Típicamente una sonda de ácido nucleico se hibrida de modo selectivo con una región de una molécula de ácido nucleico que no contiene una secuencia de infidelidad de la transcripción.

La selectividad, cuando se emplea para indicar la hibridación de ácidos nucleicos, indica que la hibridación de la sonda con la secuencia diana es diferente de la hibridación de dicha sonda con otra secuencia. A este respecto, aunque no sea necesaria una complementariedad perfecta entre la sonda y la secuencia diana, será lo suficientemente elevada como para permitir una hibridación selectiva.

Las sondas típicamente comprenden una secuencia que es complementaria con una secuencia de infidelidad de la transcripción en una molécula de ARN (o la correspondiente molécula de ADNc) que codifica una proteína de la superficie celular o una proteína segregada. Ejemplos de sondas se seleccionan de sondas que tienen una secuencia complementaria con una secuencia de infidelidad de la transcripción, o codifican una secuencia peptídica de una cualquiera de SEQ ID NO:1 a 32.

La secuencia de las sondas puede modificarse, por ejemplo, químicamente, por ejemplo para aumentar la estabilidad de los híbridos (por ejemplo, grupos intercalantes o nucleótidos modificados, tales como 2’-alcoxirribonucleótidos) o para marcar la sonda. Los ejemplos típicos de marcadores incluyen, sin limitación, radiactividad, fluorescencia, luminiscencia, marcaje enzimático y similares. La sonda puede hibridarse con el ácido nucleico diana en disolución, suspensión, o unido a un soporte sólido tal como, sin limitación, una esfera, una columna, una placa, un sustrato (para producir chips o matrices de ácidos nucleicos), etc.

La infidelidad de la transcripción se puede detectar o medir mediante amplificación selectiva utilizando cebadores específicos. La detección de la infidelidad de la transcripción mediante amplificación selectiva generalmente comprende colocar una muestra en contacto con un cebador de ácido nucleico que amplifica de modo específico una secuencia de infidelidad de la transcripción, y detectar la presencia (o la cantidad) de los productos de la amplificación formados, siendo dicha presencia o cantidad una indicación directa de la presencia o la tasa de la infidelidad de la transcripción. En una realización preferida, el método utiliza un conjunto de cebadores de ácidos nucleicos que permiten una amplificación específica de una secuencia de infidelidad de la transcripción diferenciada de uno o varios genes, respectivamente. La amplificación puede realizarse según diversas técnicas conocidas per se en la técnica tales como, sin limitación, la reacción en cadena de la polimerasa (PCR), la reacción en cadena de la ligasa (LCR), la amplificación mediada por la transcripción (TMA), la amplificación por desplazamiento de hebra (SDA), y la amplificación basada en la secuencia del ácido nucleico (NASBA).

Los cebadores de ácidos nucleicos específicas para una secuencia de infidelidad de la transcripción pueden prepararse para cualquier gen utilizando la información de la secuencia y las reglas de sustitución, inserción o deleción de nucleótidos según se describe en la presente solicitud (véase, por ejemplo, el ejemplo 5).

El término “cebador” indica un ácido nucleico o un oligonucleótido que tiene una secuencia polinucleotídica que es capaz de una hibridación selectiva con una secuencia de infidelidad de la transcripción o su complemento, o con una región de un ácido nucleico que flanquea un sitio de infidelidad de la transcripción, y que es adecuado para amplificar todo o parte de dicho sitio de infidelidad de la transcripción en una muestra que contenga dicha secuencia o complemento. Típicamente los cebadores se pueden seleccionar entre moléculas de ácidos nucleicos monocatenarios con una longitud de aproximadamente 5 a 60 nucleótidos, más preferiblemente una longitud de aproximadamente 8 a aproximadamente 50 nucleótidos, aún más preferiblemente una longitud de aproximadamente 10 a 40, 35, 30 ó 25 nucleótidos. Se prefiere una complementariedad perfecta, para asegurar una alta especificidad. Sin embargo, pueden tolerarse ciertos desapareamientos, tal como se analizó anteriormente para las sondas.

El término “flanquea” indica que la región debe localizarse a una distancia del sitio de infidelidad de la transcripción que sea compatible con las actividades polimerasa convencionales, por ejemplo, no más de 250 pb, preferiblemente no más de 200, 150, 100 o, más preferiblemente, 50 pb cadena arriba de dicho sitio.

Esta invención también describe al menos un par de cebadores de ácidos nucleicos, en el que dicho par de cebadores comprende un cebador directo y un cebador inverso, y en el que dichos cebadores directo e inverso permiten la amplificación selectiva de una secuencia de infidelidad de la transcripción, o de la secuencia exactamente complementaria, por ejemplo, en la figura 12.

La infidelidad de la transcripción se puede medir mediante una cromatografía líquida de alta resolución desnaturalizante (DHPLC). El principio de este método se describe, por ejemplo, en la figura 11. Básicamente, los productos de la amplificación se desnaturalizan y se reasocian. Durante la reasociación se forman homodúplex y heterodúplex, como resultado de la presencia de secuencias de infidelidad de la transcripción. La mezcla entonces se analiza mediante DHPLC. Puesto que las estructuras de ADN heterodúplex y homodúplex son diferentes en la temperatura del análisis se eluyen de modo diferente y su cantidad (relativa) puede evaluarse.

Como verificación de que las secuencias de infidelidad de la transcripción existen en células cancerosas con una frecuencia más alta que en células normales, la infidelidad de la transcripción predicha de genes seleccionados (ENO1, GAPDH, TMSB4X) se amplifica mediante RT-PCR utilizando los oligonucleótidos indicados (véase el ejemplo 6). Se verificaron las variaciones de homo- y heterodúplex (figura 15).

Otra técnica adecuada para detectar ácidos nucleicos puede utilizarse o adaptarse para detectar, cuantificar o controlar la infidelidad de la transcripción.

Detección de la infidelidad de la transcripción a nivel de proteínas

Debido a que la infidelidad de la transcripción conduce a cambios en la secuencia de proteínas, la presencia o el nivel de infidelidad de la transcripción también puede medirse detectando la presencia o la cantidad de proteínas-IT.

Pueden utilizarse y/o adaptarse diversas técnicas conocidas per se en la técnica para medir la infidelidad de la transcripción en proteínas. En particular, puesto que la infidelidad de la transcripción provoca modificaciones en las proteínas que conducen a cambios directos en su comportamiento, estos cambios puede detectarse, por ejemplo, mediante una electroforesis en gel bidimensional y espectrometría de masas, o mediante ionización de desorción por láser inducida en superficie. Como verificación de que una proteína-IT está presente en plasma humano, los inventores muestran el perfil de espectrometría de masas de un péptido localizado después del codón de fin canónico de ApoAII. Los datos de la espectrometría de masas muestran que la arginina está sustituida en el codón de fin canónico. Además, puesto que la infidelidad de la transcripción provoca la aparición de isoformas de proteínas (celulares y plasmáticas) más largas y más cortas (como resultado del codón de fin prematuro debido a la sustitución de bases, o una conmutación en el marco de lectura abierto en el caso de deleciones o inserciones), dichas isoformas pueden detectarse o cuantificarse utilizando sus ligandos específicos y/o estrategias de secuenciación de proteínas. A este respecto, los inventores demuestran que los cambios en la longitud de la proteína y de la secuencia primaria de AA se producen predominantemente en el dominio carboxi-terminal de la proteína. Por consiguiente, las estrategias de secuenciación deben dirigirse principalmente al extremo C-terminal de las proteínas, una región previamente insospechada (en efecto, los métodos de secuenciación de proteínas directos que se emplean en la actualidad sólo pueden realizarse comenzando en el AA NH_2-terminal).

La detección de la infidelidad de la transcripción utilizando ligandos específicos generalmente comprende colocar una muestra en contacto con un ligando que sea específico para un dominio de una proteína-IT, y detectar la presencia (o cantidad del complejo formado, siendo dicha presencia o cantidad una indicación directa de la presencia o la tasa de la infidelidad de la transcripción. En una realización preferida, el método emplea un conjunto de ligandos que son específicos para dominios diferenciados de una o varias proteínas-IT, respectivamente. Los ligandos específicos para estos dominios pueden prepararse para cualquier proteína utilizando la información de la secuencia y las reglas de sustitución de aminoácidos según se describe en la presente solicitud (véase, por ejemplo, el ejemplo 5). El ligando puede utilizarse en forma soluble, o puede revestirse sobre una superficie o un soporte.

La invención también describe ligandos que se unen de modo selectivo a un dominio de una proteína-IT. Pueden contemplarse diferentes tipos de ligandos, tales como anticuerpos específicos, moléculas sintéticas, aptámeros, péptidos y similares.

El ligando puede ser un anticuerpo, o su fragmento o derivado. Por consiguiente, un aspecto concreto de esta invención reside en un anticuerpo que se une de modo específico a un dominio de una proteína-IT.

Dentro del contexto de esta invención, un anticuerpo indica un anticuerpo policlonal, un anticuerpo monoclonal, así como sus fragmentos o derivados que tienen sustancialmente la misma especificidad de antígeno. Los fragmentos incluyen, por ejemplo, Fab, Fab'2, regiones CDR, etc. Los derivados incluyen anticuerpos monocatenarios, anticuerpos humanizados, anticuerpos humanos, anticuerpos polifuncionales, etc.

Los anticuerpos contra dominios de proteínas-IT pueden producirse mediante procedimientos conocidos en general en la técnica. Por ejemplo, pueden producirse anticuerpos policlonales inyectando el dominio de la proteína-IT (por ejemplo, como una proteína o un péptido) solo o acoplado con una proteína adecuada en un animal no humano. Después de un periodo apropiado se sangra al animal, se recupera el suero y se purifica mediante técnicas conocidas en la técnica (Paul, W.E., "Fundamental Immunology", 2ª ed., Raven Press, NY, p. 176, 1989; Harlow et al., "Antibodies: A Laboratory Manual", CSH Press, 1988; Ward et al., Nature, 341 (1989), 544).

Pueden producirse péptidos con la misma secuencia que las proteínas-IT mediante procedimientos conocidos en general en la técnica. Estos péptidos pueden acoplarse a un soporte adecuado y utilizarse para la detección de autoanticuerpos presentes en muestras biológicas (por ejemplo, fluidos corporales).

Pueden prepararse anticuerpos monoclonales contra dominios de proteínas-IT, por ejemplo, mediante la técnica de Kohler-Millstein (2) (Kohler-Millstein, Galfre, G., y Milstein, C., Methods Enz., 73, p. 1 (1981)) que implica la fusión de un linfocito B inmunológico con células de mieloma. Por ejemplo, un inmunógeno, según se describió anteriormente, puede inyectarse en un mamífero no humano, según se ha descrito. Posteriormente se retira el bazo y se realiza la fusión con el mieloma según una diversidad de métodos. Las células de hibridoma resultantes entonces pueden ensayarse para la secreción de anticuerpos contra dominios de proteínas-IT.

Un anticuerpo "selectivo" para un dominio concreto de proteínas-IT indica un anticuerpo cuya unión a dicho dominio (o a su fragmento que contiene un epitopo) puede discriminarse de modo fiable de la unión no específica (es decir, de la unión a otro antígeno, en particular a la proteína nativa que contiene dicho dominio). Los anticuerpos selectivos para los dominios de las proteínas-IT permiten la detección de la presencia de proteínas que contienen dichos dominios en una muestra.

Diagnóstico

La presente invención describe la realización de ensayos de detección o de diagnóstico que pueden emplearse, entre otras cosas, para detectar la presencia, la ausencia, la predisposición, el riesgo o la gravedad de una enfermedad a partir de una muestra derivada de un sujeto. Debe considerarse que el término "diagnóstico" incluye métodos de farmacogenómica, prognóstico, etc.

La invención describe un método para detectar *in vitro* o *ex vivo* la presencia, la ausencia, la predisposición, el riesgo o la gravedad de enfermedades en una muestra biológica, preferiblemente una muestra biológica humana, que comprende colocar dicha muestra en contacto con un ligando que se une de modo específico a un sitio de infidelidad de la transcripción, y determinar la formación de un complejo.

Esta invención también describe un método para detectar la presencia, la ausencia, la predisposición, el riesgo o la gravedad de cánceres en un sujeto, comprendiendo dicho método colocar *in vitro* o *ex vivo* una muestra procedente de un sujeto en contacto con un ligando que se une de modo específico a un sitio de infidelidad de la transcripción expresado por las células del cáncer, y determinar la formación de un complejo.

Esta invención también describe un método para evaluar la respuesta o la sensibilidad de un sujeto a un tratamiento de un cáncer, comprendiendo dicho método detectar la presencia o la tasa de la infidelidad de la transcripción en una muestra procedente del sujeto en diferentes momentos antes y durante el desarrollo del tratamiento.

Esta invención también describe un método para determinar la eficacia de un tratamiento de una enfermedad cancerosa, comprendiendo dicho método (i) proporcionar una muestra de tejido procedente del sujeto durante o después de dicho tratamiento, (ii) determinar la presencia y/o la tasa de la infidelidad de la transcripción en dicha muestra, y (iii) comparar dicha presencia y/o tasa con la cantidad de infidelidad de la transcripción en una muestra de referencia procedente de dicho sujeto tomada antes o en una etapa más temprana del tratamiento.

La presencia (o aumento) de la infidelidad de la transcripción en una muestra es indicativa de la presencia, la predisposición o la etapa de avance de una enfermedad cancerosa. Por tanto, la invención describe el diseño de la intervención terapéutica apropiada, que será más eficaz y personalizada. Además, esta determinación a nivel presintomático permite aplicar un régimen preventivo.

Los métodos de diagnóstico descritos en la presente invención pueden realizarse *in vitro, ex vivo* o *in vivo*, preferiblemente *in vitro* o *ex vivo*. La muestra puede ser cualquier muestra biológica derivada de un sujeto, que contenga ácidos nucleicos o polipéptidos, según sea apropiado. Los ejemplos de dichas muestras incluyen fluidos corporales, tejidos, muestras de células, órganos, biopsias, etc. Las muestras más preferidas son sangre, plasma, suero, saliva, orina, fluido seminal y similares. La muestra puede tratarse antes de realizar el método, para rendir o mejorar la disponiblidad de los ácidos nucleicos o los polipéptidos para el ensayo. Los tratamientos pueden incluir, por ejemplo, uno o más de los siguientes: lisis celular (por ejemplo, mecánica, física, química, etc.), centrifugación, extracción, cromatografía en columna, y similares.

Un método descrito en la presente invención consiste en determinar la presencia en muestras humanas de anticuerpos dirigidos contra nuevos péptidos producidos mediante la infidelidad de la transcripción, y generar una estimulación inmunológica que conduzca a la producción de anticuerpos. Un segundo método de esta invención se dirige a la detección de células que portan estructuras inmunológicas dirigidas contra los péptidos de infidelidad de la transcripción.

Diseños de fármacos y terapia

Tal como se analizó anteriormente, la invención describe el diseño (o la selección) de nuevos fármacos evaluando la capacidad de una molécula candidata para modular la infidelidad de la transcripción. Estos métodos incluyen ensayos de unión y/o ensayos (de actividad) funcionales, y pueden realizarse *in vitro* (por ejemplo, en sistemas celulares o en ensayos no celulares), en animales, etc.

Esta invención describe un método para seleccionar, caracterizar, buscar u optimizar un compuesto biológicamente activo, comprendiendo dicho método determinar *in vitro* si un compuesto de ensayo modula la infidelidad de la transcripción. La modulación de la infidelidad de la transcripción puede evaluarse con respecto a un gen o una proteína concretos, o con respecto a un conjunto predefinido de genes o proteínas, o de modo global.

La presente invención también describe un método para seleccionar, caracterizar, buscar u optimizar un compuesto biológicamente activo, comprendiendo dicho método colocar *in vitro* un compuesto de ensayo en contacto con un gen, y determinar la capacidad de dicho compuesto de ensayo para modular la producción, a partir de dicho gen, de moléculas de ARN que contengan sitios de infidelidad de la transcripción.

La invención también describe un método para seleccionar, caracterizar, buscar y/u optimizar un compuesto biológicamente activo, comprendiendo dicho método poner en contacto un compuesto de ensayo con una célula, y determinar, en dicha célula, si el compuesto de ensayo modula la infidelidad de la transcripción.

Los anteriores ensayos de selección pueden realizarse en cualquier dispositivo adecuado, tal como placas, tubos, platos, matraces, etc. Generalmente, el ensayo se realiza en placas de microvaloración de múltiples pocillos. Por ejemplo pueden ensayarse varios compuestos en paralelo. Además, el compuesto de ensayo puede tener diversos orígenes, naturaleza y composición. Puede ser una sustancia orgánica o inorgánica, tal como un lípido, un péptido, un polipéptido, un ácido nucleico, una molécula pequeña, aislados o mezclados con otras sustancias. Los compuestos pueden ser todo o parte de un banco combinatorio de compuestos, por ejemplo.

Los compuestos que modulan la infidelidad de la transcripción pueden tener muchas utilidades, tales como utilidades terapéuticas, para reducir o aumentar la infidelidad de la transcripción en una célula. Dichos compuestos pueden utilizarse para tratar (o prevenir) enfermedades provocadas o asociadas con una infidelidad de la transcripción anómala, tales como enfermedades proliferativas (por ejemplo, cánceres), enfermedades inmunológicas, envejecimiento, enfermedades inflamatorias, etc.

Otros compuestos descritos en la presente invención son compuestos que se dirigen a la infidelidad de la transcripción, es decir, compuestos que se unen de modo selectivo a sitios de infidelidad de la transcripción. Estos compuestos (por ejemplo, anticuerpos, ARNi, etc.) pueden utilizarse como reactivos de diagnóstico, o como agentes terapéuticos, solos o conjugados con un marcador.

Los compuestos que modulan o se dirigen a la infidelidad de la transcripción que se describen en la presente invención pueden administrarse mediante cualquier vía adecuada, que incluye la vía oral, o mediante administración sistémica, por inyecciones intravenosas, intraarteriales, intracerebrales o intratecales. La dosificación puede variar dentro de amplios límites, y deberá ajustarse a las necesidades del individuo en cada caso concreto, dependiendo de diversos factores conocidos por los expertos en la técnica. Puede utilizarse cualquier forma de dosificación farmacéuticamente aceptable conocida en la técnica, tal como cualquier disolución, suspensión, polvo, gel, etc., incluyendo disoluciones isotónicas, disoluciones tamponadas y disoluciones salinas, etc. Los compuestos pueden administrarse por sí solos, pero en general se administran con un vehículo farmacéutico, con respecto a la práctica farmacéutica convencional (tal como se describe en Remington's Pharmaceutical Sciences, Mack Publishing). Los anticuerpos pueden administrarse según métodos y protocolos conocidos per se en la técnica, que se emplean en la actualidad en terapias y ensayos humanos.

Efecto de la infidelidad de la transcripción sobre la función celular normal

En general se acepta que las proteínas son responsables de la mayoría de las funciones celulares. Sin embargo, también es evidente que el ARN conocido como ARN no codificador también es una molécula funcional en sí misma. Estas transcripciones, cuya importante ha sido subestimada, estarían presentes en muchos organismos y regulan muchas funciones celulares. Entre éstos, se encuentran ARNr, ARNt, ARNtm (ARN transmensajero), pero también otros ARN no codificadores (ARNsno, ARNsn, etc.) que intervienen en las modificaciones postranscripcionales del ARN, en el corte y empalme u otras funciones celulares. Las funciones del ARN y de las proteínas pueden verse afectadas por la falta de infidelidad de la transcripción descrita en esta solicitud. La secuencia de ADN se transcribe en ARNm en el núcleo por la acción de una proteína denominada ARN polimerasa II que reconoce la secuencia de la molécula de ADN y sintetiza un polímero monocatenario de nucleótidos que es complementario con el molde de ADN de origen. El orden y el tipo de base del ARN en cualquier posición concreta viene determinado por el orden y el tipo de base presente en la hebra de ADN que actúa como molde. El extremo 5' del ARN transcrito, que se corresponde con la primera base del ARNm, se une a 7-metilguanilato unido al nucleótido inidical, constituyendo así 5' CAP que protege al ARN de la degradación. Esta modificación se produce antes de que se haya completado la transcripción. El procesamiento en el extremo 3' de la transcripción primaria implica la ruptura por una endonucleasa para producir un grupo 3'-hidroxilo libre al cual se añade una cadena de restos adenina por una enzima denominada poli(A) polimerasa. La cola de poli(A) resultante contiene entre 100-250 nucleótidos. La etapa final del procesamiento es el corte y empalme, que consiste en la escisión de la transcripción primaria del elemento que corresponde a las secuencias intrónicas, seguido del acoplamiento de los exones. Este proceso es controlado por numerosas proteínas y ARNnc, parte de los cuales se une al ARN. Es posible que esta unión sea una etapa que se vea afectada por la falta de infidelidad de la transcripción. En efecto, todas estas proteínas están codificadas por un ARN específico; por tanto, la función de estas proteínas puede verse potencialmente afectada por la falta de infidelidad de la transcripción. El proceso, muy complejo, de la maduración conduce a la producción de ARNm que será exportado desde el núcleo para que se produzca la traducción. Es importante reconocer que, en esta etapa, no todos los tripletes de nucleótidos que están presentes en el ARNm maduro serán traducidos a AA. El ARN mensajero maduro contiene regiones no codificadoras denominadas regiones 5' y 3' no traducidas que tienen importancia funcional para determinar la estabilidad global del ARNm y otros papeles en la traducción. Para trasladar de forma eficaz la información contenida en el ADN a la secuencia de AA apropiada de la proteína, se requiere una fidelidad absoluta de la transcripción. En efecto, cualquier error que se produzca durante la transcripción o durante los procesos de maduración dará como resultado potencialmente la introducción de cambios en la estabilidad del ARNm y, en último término, una variación en la secuencia primaria de AA de la proteína. Estas variaciones en la secuencia de la proteína pueden entonces exacerbar el fenómeno de la infidelidad de la transcripción mediante la alteración de la secuencia de la proteína implicada en el inicio de la transcripción, en la propia transcripción, en la formación del casquete 5' del ARN, en la poliadenilación 3', en el corte y empalme del ARN y/o en la exportación del ARN. Por tanto, la demostración descrita en esta invención de que la infidelidad de la transcripción afecta a un gran número de genes aislados a partir de células del cáncer abre la posibilidad de que el fenómeno pueda exacerbarse, conduciendo a un aumento progresivo en la gravedad de la enfermedad. La variación en las secuencias de ARN introducida por la infidelidad de la transcripción puede tener consecuencias inmediatas sobre la función celular. En efecto, la introducción de bases en la transcripción primaria del ARN que no son complementarias con las del molde de ADN tiene consecuencias potenciales inmediatas sobre la secuencia primaria de AA de la proteína resultante. Cuando el cambio afecta a las primeras 2 bases del codón, generalmente ejerce un impacto directo sobre la secuencia de AA. Las variaciones que afectan a la tercera base del codón tendrán un impacto menor sobre la secuencia de AA de la proteína, porque el codigo genético está degenerado. Por tanto, los cambios en la base localizada en la tercera posición del codón pueden no influir directamento a la secuencia de AA de la proteína. Basándose en los datos descritos a continuación, los inventores creen que la infidelidad de la transcripción es un fenómeno que afecta a todas las bases independientemente de su posición en el codón y, por tanto, impacta sobre la secuencia primaria de AA de la proteína. Los cambios en la proteína pueden ser neutros o provocar una modificación de la función de la proteína, tanto un aumento como la pérdida de actividad. El fenómeno de la infidelidad de la transcripción se observa predominantemente después de codificar y completar al menos las primeras 400-500 bases del ARNm maduro; el extremo 5' del ARNm se ve relativamente menos afectado. Se cree que estarán presentes fragmentos más cortos, y también más largos, de la proteína en las células del cáncer, así como en el plasma de los pacientes con cáncer. Estas isoformas más cortas y más largas pueden deducirse directamente de las reglas de la infidelidad de la transcripción descritas a continuación. Esto permite la producción de métodos diseñados de forma lógica para la identificación de marcadores específicos del cáncer o de la gravedad de la enfermedad inmunológica. Debido a que la tasa de transcripción de la mayoría de los genes está controlada para adaptarse a las necesidades celulares, se propone que la infidelidad de la transcripción provoca cambios en la expresión génica directamente debidos a una excesiva función de la proteína o a la falta de función de ésta. Por tanto, el fenómeno de la infidelidad de la transcripción ejerce un profundo efecto sobre la capacidad celular para realizar su tarea. La identificación de este defecto como una característica común a la mayoría de los genes aislados de todos los tipos de células del cáncer proporciona, por tanto, un fundamento para la búsqueda de nuevos métodos que permitan la medición cuantitativa de la tasa de la infidelidad de la transcripción en cualquier célula concreta. Este ensayo de selección permite el ensayo de nuevos fármacos capaces de limitar la infidelidad de la transcripción, evitando así el avance de la enfermedad y restableciendo la función celular normal.

Infidelidad de la transcripción y patología

La presente invención describe que la infidelidad de la transcripción conduce a importantes cambios en la secuencia

de las proteínas. La función de cualquier proteína concreta viene determinada por su estructura tridimensional que depende directamente de su secuencia de AA. Las proteínas con un AA variante, un truncamiento de la proteína o un alargamiento de la proteína pueden producir una modificación profunda de la actividad de la proteína o permanecer neutras. Los ejemplos de proteínas modificadas que inhiben significativamente la función de sus homólogos normales se han descritos. La infidelidad de la transcripción es un fenómeno que afecta a un gran número de genes, produciendo con ello un gran número de proteínas. Debido a que varias de estas proteínas participan en mantener estable la estructura del ADN y participan en la reparación del ADN, una función defectuosa de estas proteínas puede producir una reparación defectuosa del ADN y dar como resultado una capacidad significativamente disminuida de las células para reparar con éxito el ADN dañado. Debido a que la fidelidad de la transcripción en las células normales es debida a la actividad de varios complejos de proteínas que controlan la transcripción, es bastante probable que unas pequeñas alteraciones iniciales de estas proteínas puedan exacerbar progresivamente el fenómeno. Esto producirá, en último término, una mayor tasa de infidelidad de la transcripción que, por consiguiente, suprimirá el estado de diferenciación celular y conducirá a formas cada vez más graves de la enfermedad. La demostración de que la infidelidad de la transcripción realmente se produce en las células del cáncer y que conduce a una amplia diversificación de las proteínas codificadas por cualquier gen concreto abre un área nueva para la selección de nuevas sondas de diagnóstico y para el diseño de nuevas dianas terapéuticas.

La presente invención describe un nuevo fenómeno que contribuye a la diversificación de la información presente en el ADN y que sigue una reglas específicas. Este proceso de infidelidad de la transcripción no aleatoria se ve muy exacerbado en células del cáncer, pero también se produce en células normales siguiendo las mismas reglas. Este mecanismo introduce bases nuevas en el ARNm, provocando así cambios profundos en el mensaje que será traducido a nivel de ribosomas. Es probable que el fenómeno sea bastante general, porque está presente en la mayoría de los genes ensayados. La consecuencia inmediata de la infidelidad de la transcripción es que un único gen puede producir muchas más proteínas de las sospechadas. Por tanto, el descubrimiento de los inventores tiene el potencial de explicar en parte la relativa discrepancia entre el número limitado de genes presentes en el genoma y el gran número de proteínas presentes en las muestras biológicas. Los inventores han demostrado que la infidelidad de la transcripción es un proceso no aleatorio y está gobernado por reglas específicas, algunas de las cuales ya se han descrito en la presente. Algunas será reveladas por una mayor caracterización del proceso utilizando métodos de bioinformática y biológicos. Los inventores se han centrado en la presente en la implicación de la infidelidad de la transcripción en el cáncer. Sin embargo, creen que el proceso es general y que puede contribuir a generar la diversidad de proteínas. Esta diversificación de las proteínas puede ejercer una influencia significativa sobre la capacidad del sistema inmunológico para reconocer un patrón de proteínas específico. A este respecto, se cree que la infidelidad de la transcripción desempeña un papel en la patogénesis de las enfermedades inmunológicas. Los inventores han considerado en la presente los acontecimientos de infidelidad de la transcripción que conducen a la sustitución de una sola base. Se han centrado en las sustituciones de una sola base porque no es probable que este mecanismo provoque una desestabilización significativa de la estructura del ARNm. Por tanto, las sustituciones de una sola base no deberían interferir con la traducción, sino ser reconocidas como un nuevo mensaje auténtico. A continuación se verá que también se espera que la infidelidad de la transcripción provoque deleciones y/o inserciones de bases. El mismo método descrito anteriormente detectará mecanismos alternativos de infidelidad de la transcripción.

El nuevo proceso descrito en esta solicitud tiene varias aplicaciones prácticas inmediatas.

Optimización de los experimentos proteómicos para identificar nuevos biomarcadores específicos de enfermedades y para diseñar nuevos anticuerpos que se dirigirán a proteínas de la enfermedad específicas

Basándose en las indicaciones de la presente solicitud, ahora es posible predecir los cambios que pueden producirse en cualquier secuencia de proteína como resultado de la infidelidad de la transcripción y, por tanto, predecir los cambios en las propiedades fisicoquímicas, tales como las masas moleculares aparentes y los puntos isoeléctricos. Esto puede producir unos patrones de proteínas modificados en geles bidimensionales y en espectrometría de masas. Debido a que la infidelidad de la transcripción afecta a la mayoría de los genes, ahora es posible centrarse en un conjunto limitado de proteínas y caracterizar sus cambios para identificar biomarcadores específicos de enfermedades (por ejemplo, cánceres). Los algoritmos de la infidelidad de la transcripción pueden acelerar el proceso de descubrimiento de biomarcadores. Como alternativa, es posible diseñar anticuerpos específicos dirigidos contra dominios de proteínas-IT, es decir, los dominios de las proteínas que contienen AA modificados, así como nuevos AA adicionales que extienden la proteína nativa generada mediante la infidelidad de la transcripción. Los anticuerpos dirigidos contra estos dominios serán específicos para el variante de la proteín y, por tanto, estarán dirigidos hacia proteínas específicas de enfermedades (por ejemplo, cáncer) presentes en un fluido corporal, en la superficie celular, o en el interior de las células del cáncer. Estos anticuerpos se emplearán para detectar proteínas características de un trastorno patológico en fluidos corporales, para detectar células enfermas circulando en el plasma, para identificar células enfermas en muestras histológicas, para evaluar la gravedad de una enfermedad, y también para dirigir medicaciones hacia células enfermas específicas. Esto puede lograrse mediante la transferencia génica de secuencias propensas a la infidelidad de la transcripción mediante un vector de terapia génica, por ejemplo, adenovirus, liposomas, etc. Como alternativa, los inventores han diseñado péptidos específicos que tienen la misma secuencia que un dominio que se produce como consecuencia de la IT. Éstos pueden utilizarse para detectar autoanticuerpos específicos y naturales dirigidos contra los dominios de proteínas-IT presentes, por ejemplo, en fluidos corporales.

La infidelidad de la transcripción se produce de modo no aleatorio y con frecuencia afecta a los codones de fin del ARNm en una gran proporción de los genes ensayados. Además, puesto que es posible calcular que hasta 6% de cualquier población concreta de ARNm contiene estas secuencias adicionales, por tanto es posible detectar, en una célula cancerosa concreta, una población específica de proteínas que muestre una nueva secuencia en el extremo carboxi-terminal. La localización de estas secuencias, por tanto, no resultaba concebible porque ningún proceso expuesto, excepto las enfermedades genéticas raras que afectan al ADN, ha sido capaz de introducir la lectura dentro del marco de secuencias que siguen inmediatamente a los codones de fin canónicos. Previamente no se sospechaba la existencia de estas secuencias ocultas. El descubrimiento de la infidelidad de la transcripción que conduce a la sustitución de una sola base que se produce en los codones de fin canónicos revela la existencia de dichas secuencias codificadoras y también demuestra que su presencia aumenta en células del cáncer porque la infidelidad de la transcripción está potenciada en estas células. Por tanto, es posible identificar y localizar estas secuencias que normalmente están escondidas para diseñar nuevos fármacos específicos del cáncer.

Una deleción o inserción de una sola base en relación con la IT que se produce en las secuencias codificadoras conduce a una conmutación en el marco de lectura. Como convergencia, las secuencias en la proteína correspondiente se modifican (un cambio en la secuencia de AA y en la longitud de la proteína). Por tanto, se puede detectar, en un célula cancerosa concreta, una población específica de proteínas que muestren una secuencia nueva. Por tanto, es posible identificar y localizar estas secuencias que normalmente están escondidas para diseñar nuevos fármacos específicos del cáncer.

Debido a la existencia de secuencias de proteínas nuevas o modificadas, y considerando que estas alteraciones están presentes en las proteínas presentes sobre la superficie celular, se propone utilizar estas secuencias para vacunar contra secuencias específicas de enfermedades. Estas vacunaciones se utilizarán para activar respuestas inmunológicas en pacientes diagnosticados con cualquier forma concreta de enfermedad (por ejemplo, cáncer) o para iniciar respuestas inmunológicas preventivas en pacientes con mayor riesgo de desarrollar enfermedades específicas (por ejemplo, cáncer) debido a una predisposición genética o debido a una mayor exposición a riesgos o toxinas ambientales.

Por tanto, esta invención describe un método para detectar la presencia o la etapa de una enfermedad en un sujeto, comprendiendo dicho método evaluar (*in vitro* o *ex vivo*) la presencia o la tasa de la infidelidad de la transcripción en una muestra procedente de dicho sujeto, siendo dicha presencia o tasa una indicación de la presencia o etapa de una enfermedad en dicho sujeto. La muestra puede ser cualquier tejido, célula, fluido, biopsia, etc., que contenga ácidos nucleicos y/o proteínas. La muestra puede tratarse antes de la reacción, es decir, mediante dilución, concentración, lisis, etc. El método comprende generalmente evaluar la infidelidad de la transcripción de una serie de genes o proteínas diferenciados, tales como proteínas plasmáticas, proteínas de la superficie celular, etc.

Esta invención también describe un método para tratar a un sujeto que lo necesite, el método comprende administrar a dicho sujeto una cantidad eficaz de un compuesto que altera (p. ej., reduce o aumenta) la tasa de la infidelidad de la transcripción de un gen de mamífero.

Esta invención también describe el uso de un compuesto que altera (por ejemplo, reduce o aumenta) la tasa de la infidelidad de la transcripción de un gen de mamífero para la fabricación de una composición farmacéutica para su uso en un método de tratamiento de un ser humano o animal, en particular para tratar un trastorno de células proliferativas, tal como cánceres y enfermedades inmunes.

La invención también describe un método para evaluar la eficacia de un fármaco o de un candidato a fármaco, comprendiendo dicho método una etapa de evaluar si dicho fármaco altera la tasa de la infidelidad de la transcripción de un gen de mamífero, de forma que dicha alteración es una indicación de la eficacia del fármaco.

La invención también describe métodos y productos (tales como sondas, cebadores, anticuerpos o sus derivados) para detectar o medir (el nivel de) la infidelidad de la transcripción en una muestra, así como los correspondientes kits.

La invención también describe un método para identificar y/o producir biomarcadores, comprendiendo dicho método identificar, en una muestra procedente de un sujeto, la presencia de un sitio o sitios de infidelidad de la transcripción en una proteína, ARN o gen diana y, opcionalmente, determinar la secuencia de dicho sitio o sitios de infidelidad de la transcripción. Por ejemplo, la proteína diana es una proteína de la superficie celular o una proteína segregada, en particular una proteína de la superficie celular o una proteína plasmática.

La invención describe un método para producir o identificar ligandos específicos para un rasgo o un trastorno patológico, comprendiendo dicho método identificar, en una muestra procedente de un sujeto que tiene dicho rasgo o trastorno patológico, la presencia de un sitio o sitios de infidelidad de la transcripción en una o varias proteínas, ARN o genes diana, opcionalmente determinando la secuencia de dicho sitio o sitios de infidelidad de la transcripción, y produciendo un ligando o ligandos que se unen de modo específico a dicho sitio o sitios de infidelidad de la transcripción.

La invención es particularmente adecuada para identificar biomarcadores de trastornos de células proliferativas, tales como cánceres, enfermedades inmunes, inflamación o envejecimiento. Es particularmente útil para producir

ligandos que son específicos para tales trastornos en mamíferos, en particular ligandos que pueden detectar la presencia o gravedad de un trastorno de células proliferativas en un sujeto.

En particular, esta invención describe un péptido (sintético) que comprende un sitio de infidelidad de la transcripción de una proteína, en particular de una proteína de mamífero, más preferiblemente de una proteína humana. El péptido generalmente comprende un fragmento interno de la proteína, o la secuencia de un fragmento C-terminal de dicha proteína. El péptido preferiblemente comprende menos de 100, 80, 75, 70, 65, 60, 50, 45, 40, 35, 30, 25 o incluso 20 aminoácidos. La proteína puede ser una proteína de la superficie celular (por ejemplo, un receptor, etc.), una proteína segregada (por ejemplo, una proteína plasmática), o una proteína intracelular. Los ejemplos de dichas proteínas plasmáticas se proporcionan, por ejemplo, en la figura 13, e incluyen apolipoproteínas (por ejemplo, AI, AII, CI, CII, CIII, D, E), componentes del complemento (por ejemplo, C1s, C3, C7), proteína reactiva C, inhibidores de serpina peptidasa, fibrinógeno (por ejemplo, FGA1, FGA2), plasminógeno, transferrina, transtiretina, etc. Los ejemplos de receptores de la superficie celular incluyen, por ejemplo, receptores de citoquinas y receptores de hormonas, etc. En una realización específica, la invención se refiere a un péptido sintético que comprende un sitio de infidelidad de la transcripción de una proteína plasmática humana o un receptor de la superficie celular abundantes, tal como se listó anteriormente.

Los ejemplos específicos de dichos péptidos se describen en las figuras 10, 14 y 18. De modo más específico, los ejemplos de péptidos sintéticos de la presente invención comprenden toda o un fragmento de las siguientes secuencias de aminoácidos:

QMWQLFWIYHLSS (TPT1) (SEQ ID NO: 1)KLHTLSAAIYYQQE (VIM) (SEQ ID NO: 2)
DFLSNK (RPS6) (SEQ ID NO: 3)MYTVEFSVHKNN (RPL7A) (SEQ ID NO: 4)NGSLGDMSDLCT (RPS4X) (SEQ ID NO: 5)ASG (FTH1) (SEQ ID NO: 6)
EPSEPSDF (FTL) (SEQ ID NO: 7)
APSIFPTLPAKPGTKQPRSPVTALSLHMLLMVSSAPSCGLIQTVSSFTVYIFTL (TPI1) (SEQ ID NO: 8)ARHGRDEEVWHRKHSHHFVQAWAWVGGLVCWPRKCHMRSTLISSLDSLLPVIPHRTEAEWVV VMFDRRH (AHSG) (SEQ ID NO: 9)
RSKAYSSVFLFRWCKANTLSKKHKFL (ALB) (SEQ ID NO: 10)
GLDSTRALENEMTV (APCS) (SEQ ID NO: 11)
GARRRPPSRCSE (APOA1) (SEQ ID NO: 12)
SVQTIVFQPQLASRTPTGQS (APOA2) (SEQ ID NO: 13)
IVFQPQLASRTPTGQS (APOA2) (SEQ ID NO: 14)QPDPPSVDKGRVPYSPDPPGSD (APOC2) (SEQ ID NO: 15)
DPPSVDKGRVPYSPD (APOCII) (SEQ ID NO: 16)
DLNTPSPPAYPSCELLGSCNLQGCPCRLLKRDSILSALLPHLMPGPPPGMLASQ (APOC3) (SEQ ID NO: 17)
PGSTGRLHPLHVTSASLSPTPPPPHKDKPINHDKGS (APOD) (SEQ ID NO: 18)
TPKPAAMRPHATPCLLPPRSLQRETLSPPQPSSWGGP (APOE) (SEQ ID NO: 19)
EARVGGNVGSQTQ (AZGP1) (SEQ ID NO: 20)
PSVLHTARGPRMPRPPLAPAGREPDHLPC (CHI3L1) (SEQ ID NO: 21)
DVDVAFAPTGASESSSPQDELQPPRESSARHQVTRPQPPGPQLRPASPRSGSCTLTLDSAAHGK NRIAPACN (CLU) (SEQ ID NO: 22)
NVIPLKRKMNNTLN (HRG) (SEQ ID NO: 23)
TPAARLMWSSNMPYFAQKTAKDMTSSWLQPRFIFLFVVN (IGFBP3) (SEQ ID NO: 24)
GWGVFLLNPMAGGHAPTIISWEERQSWEIDGSHSSLLSLLCLWATLPTPLLSQ (INHA) (SEQ ID NO: 25)
TGPTHHSPSPSISTWCLVPVHSVNKKP (KLK11) (SEQ ID NO: 26)
WTPEPLLQPLSHPLPPAHPLGQQRL (PKM2) (SEQ ID NO: 27)
LDGRQSDALTHLEAGTWVGI (PLG) (SEQ ID NO: 28)
SLPSSSGALSKELGMQAGCLGLWAQPGPCAPSGHGMCGPVCLSLEGDSDSLCSSHMHRGPWTLQ SGGSWAS (SERPINA3) (SEQ ID NO: 29)
NLRGRAATKVKMGTQMIHEFALVSLAQWCANHVCLHSSVLPCVLNKK (TF) (SEQ ID NO: 30)
GPAPPRPAPAGPAPPRPAPAALPMGAVFKDTRAPSPPGAPLKMERGLRISVSLGACLGSPSLTF PHSHSLSLPLCLLLPVCTIPLPGIKAQGTSGEHYCS (TGFB1) (SEQ ID NO: 31)
GTSPPVDLKDEGWDFM (TTR) (SEQ ID NO: 32)

Otro aspecto de esta invención se refiere al uso de un péptido con una longitud de 100 aminoácidos o menos, que comprende una secuencia seleccionada de SEQ ID NOs: 1 a 5, 7 a 13, 15 a 18 y 20 a 32 para detectar o controlar trastornos de células proliferativas.

Un aspecto adicional de esta invención se refiere al uso de un péptido, tal como se describió anteriormente, como inmunógeno. La invención también se refiere a una composición de vacuna que comprende dicho péptido, tal como se definió anteriormente, y opcionalmente un vehículo, excipiente y/o adyuvante adecuado.

La invención también describe un dispositivo o producto que comprende, inmovilizado sobre un soporte, un reactivo que se une de modo específico con sitio de infidelidad de la transcripción. El reactivo puede ser, por ejemplo, una sonda, o un anticuerpo o su derivado.

La invención puede utilizarse en cualquier sujeto mamífero, en particular cualquier sujeto humano, para detectar, controlar o tratar una diversidad de trastornos patológicos asociados con trastornos de células proliferativas (p. ej., cánceres) y/o para producir, diseñar o seleccionar fármacos terapéuticamente activos.

Producción y utilización de agentes que se dirigen a sitios de infidelidad de la transcripción

Utilizando las indicaciones de la presente invención, es posible producir agentes que puedan dirigirse a sitios de infidelidad de la transcripción, por ejemplo, que se unan a proteínas o ácidos nucleicos que contengan un sitio de infidelidad de la transcripción, o a células o tejidos que contengan o expresen estas proteínas o ácidos nucleicos. El agente de transporte dirigido puede ser un anticuerpo (o su derivado), una sonda, un cebador, un aptámero, etc., tal como se describió anteriormente.

Estos agentes de transporte dirigido pueden utilizarse, por ejemplo, como agentes de diagnóstico, para detectar, controlar, etc. la infidelidad de la transcripción en una muestra, tejido, sujeto, etc. Para este fin, el agente puede acoplarse a un resto marcador, tal como un radiomarcador, una enzima, un marcador fluorescente, un tinte luminiscente, etc.

Los agentes de transporte dirigido también pueden utilizarse como molécula terapéutica, por ejemplo, para tratar una célula o tejido u organismo que muestre una infidelidad de la transcripción anómala. El agente puede utilizarse como tal, o preferiblemente puede acoplarse a una molécula activa, tal como una molécula tóxica, un fármaco, etc.

Esta invención describe un método para producir un agente que se dirige a la infidelidad de la transcripción, comprendiendo dicho método (i) identificar un sitio de infidelidad de la transcripción de una proteína o un ácido nucleico, y (ii) producir un agente que se una de modo selectivo a dicho sitio. El sitio de infidelidad de la transcripción puede identificarse, por ejemplo, mediante el alineamiento de secuencias disponibles para una molécula de ARN concreta y la identificación de las variaciones de la secuencia, en particular en el extremo 3'. Los sitios de infidelidad de la transcripción también pueden identificarse aplicando las reglas de la infidelidad de la transcripción, según se describe en la presente solicitud (véase, por ejemplo, el ejemplo 5) o reglas nuevas que se describan después, a cualquier secuencia génica. Entonces puede producirse una molécula de péptido o de ácido nucleico que comprenda dicho sitio identificado utilizando métodos convencionales. La invención describe la etapa (i) del método que comprende identificar un sitio de infidelidad de la transcripción de una proteína segregada o de la superficie celular (por ejemplo, un receptor, etc.), o de un ácido nucleico. En efecto, las proteínas segregadas y las proteínas de la superficie celular pueden ser localizadas con facilidad utilizando un agente de transporte dirigido que se pone en contacto con la célula o se administra a un sujeto. Los ejemplos de dominios de dichas proteínas-IT se describen en los ejemplos.

Producción y utilización de una vacuna que provoca o estimula o inhibe una respuesta inmunológica contra un dominio de una proteína-IT

Utilizando las indicaciones de la presente invención es posible producir una composición de vacuna que provoca o inhibe una respuesta inmunológica contra proteínas-IT, o contra células o tejidos que contengan o que expresen dichas proteínas-IT o los correspondientes ácidos nucleicos.

Generalmente, la composición de vacuna comprende, como inmunógeno, una molécula que comprende la secuencia de un sitio de infidelidad de la transcripción. La composición de vacuna puede comprender cualquier vehículo, excipiente o adyuvante farmacéuticamente aceptable. La composición de vacuna puede utilizarse para generar una respuesta inmunológica contra una célula o tejido enfermo que exprese proteínas-IT, que de como resultado la destrucción de dicha célula o tejido por el sistema inmunológico. Como alternativa, la composición de vacuna puede utilizarse para inducir una tolerancia contra los sitios de infidelidad de la transcripción implicados en enfermedades autoinmunológicas.

La invención también describe un método para producir un inmunógeno que provoca, estimula o inhibe una respuesta inmunológica contra un dominio de proteínas-IT, o contra una célula o tejido que exprese dicha proteína, comprendiendo dicho método (i) identificar un sitio de infidelidad de la transcripción de una proteína o de un ácido nucleico, y (ii) producir un péptido que comprenda dicho sitio o su variante. El sitio de infidelidad de la transcripción puede identificarse, por ejemplo, mediante el alineamiento de las secuencias disponibles para una molécula de ARN

concreta, y la identificación de las variaciones en las secuencias, en particular en el extremo 3'. El sitio de infidelidad de la transcripción también puede identificarse aplicando las reglas de la infidelidad de la transcripción, tal como se describe en la presente solicitud (véase, por ejemplo, el ejemplo 5) o reglas nuevas que se describan después, a cualquier secuencia génica. Entonces puede producirse una molécula de péptido que comprenda dicho sitio identificado utilizando métodos convencionales. En una realización preferida, la invención se refiere a un método para producir un anticuerpo, que comprende inmunizar un mamífero no humano con un péptido de acuerdo con la invención y recuperar anticuerpos que se unen a dicho péptido o las correspondientes células que productoras de anticuerpos.

Reducción de la infidelidad de la transcripción en sistemas de producción de proteínas recombinantes

Esta invención también describe un método para prevenir o reducir la infidelidad de la transcripción que puede producirse en sistemas de producción recombinantes, en particular en sistemas de producción de proteínas terapéuticas. En efecto, dicha infidelidad de la transcripción puede disminuir el rendimiento de producción del sistema, o dar como resultado la producción de mezclas de proteínas no caracterizadas que pueden mostrar diversos perfiles de actividad. Por tanto, se recomienda y resulta muy valioso poder reducir la aparicion o la tasa de la infidelidad de la transcripción en sistemas de producción recombinantes, tales como sistemas de producción bacterianos (si las reglas de la infidelidad de la transcripción son las mismas para sistemas procariotas), sistemas de producción eucariotas (por ejemplo, levaduras, células de mamífero, etc.). Esta reducción puede lograrse, por ejemplo adaptando la secuencia de la molécula de ácido nucleico codificadora y/o inhibiendo las moléculas de ARN generadas mediante la infidelidad de la transcripción. Como alternativa, las proteínas que contienen sitios de infidelidad de la transcripción pueden retirarse del medio.

Reducción de la sensibilidad a la infidelidad de la transcripción para un gen

Esta invención también describe un método para prevenir o reducir la IT que puede producirse en un gen. En efecto, dicha IT puede afectar a la expresión o a la actividad de una proteína, o dar como resultado la producción de mezclas de proteínas no caracterizadas que pueden mostrar diversos perfiles de actividad. Por tanto, se recomienda y resulta muy valioso poder reducir la aparicion o la tasa de la infidelidad de la transcripción para dicho gen. Esta reducción puede lograrse, por ejemplo, modificando la secuencia del gen (terapia génica) y/o inhibiendo las moléculas de ARN generadas mediante la infidelidad de la transcripción. Como alternativa, las proteínas que contienen sitios de infidelidad de la transcripción pueden degradarse de modo específico (por ejemplo, pueden localizarse de modo específico para su degradación).

Medición de la infidelidad de la transcripción a nivel de ARN

La transcriptómica es la ciencia que mide la variación a nivel de ARN (preferiblemente ARNm) que aparece en una diversidad de trastornos patológicos, de los cuales el cáncer es el más frecuente. La transcriptómica se basa en la hibridación de ADNc con un conjunto específico de oligonucleótidos que identifican subconjuntos predefinidos de genes. La eficacia de la hibridación depende de las secuencias específicas de cualquier ARN concreto que tendrá su transcripción inversa en el ADNc. La introducción de una variación insospechada en la secuencia de un ARN concreto disminuirá la eficacia de la hibridación y, por tanto, provocará una variación en la señal percibida por los chips transcriptómicos. El descubrimiento de que la infidelidad de la transcripción provoca alteraciones en la secuencia de bases del ARN tiene dos consecuencias inmediatas: en primer lugar, permite la optimización de los chips transcriptómicos para minimizar la consecuencia del desapareamiento de bases debido a la infidelidad de la transcripción y, por tanto, mejora la precisión del experimento transcriptómico. En efecto, la limitación actual de los experimentos transcriptómicos es su falta de reproducibilidad entre estudios. Los inventores ahora creen que una parte importante de dicha variabilidad está provocada por la infidelidad de la transcripción. La comprensión de las reglas de la infidelidad de la transcripción, tal como describió anteriormente, permite ahora el diseño de chips de micromatrices que miden de modo específico la tasa de la infidelidad de la transcripción en cualquier mezcla concreta de ADNc obtenido a partir de células normales o enfermas. El control de la tasa de la infidelidad de la transcripción a nivel de ARN permite una selección de alta capacidad de procesamiento de la tasa relativa de infidelidad de la transcripción que se produce en una célula concreta. Esto proporciona una información esencial para determinar la gravedad de una enfermedad, para definir las estrategias terapéuticas, y para clasificar a las células enfermas según su perfil de sensibilización farmacológico. La misma selección también permite ensayar la eficacia de nuevos fármacos, por ejemplo, en la terapia del cáncer.

Otras aplicaciones de la infidelidad de la transcripción

La infidelidad de la transcripción es un mecanismo natural recién descubierto que añade diversificación al código genético bien establecido. Los inventores han observado que la infidelidad de la transcripción se produce, aunque a tasas bajas, incluso en tejidos normales. Los inventores proponen que la infidelidad de la transcripción contribuye a explicar la relativamente baja abundancia de genes identificados en el genoma y la mucha mayor diversidad de proteínas presentes en las células vivas de tejidos y fluidos biológicos.

Los inventores proponen que la infidelidad de la transcripción tiene una función específica a nivel inmunológico. También proponen que el sistema inmunológico depende, al menos en parte, en que una tasa concreta de

infidelidad de la transcripción afecte preferentemente a genes específicos para identificar al propio individuo. Por tanto, las anomalías en la tasa de infidelidad de la transcripción que se producen por razones aún desconocidas podrían contribuir a activar una respuesta inmunológica específica y contribuir a la patogénesis de enfermedades autoinmunológicas y alégicas. Las diferencias interindividuales en la infidelidad de la transcripción también podrían estar condicionadas por los polimorfismos de genes del sistema inmunológico, determinando por tanto la tasa de infidelidad de la transcripción que se produce en las células en circulación del sistema inmunológico, por ejemplo, linfocitos T o B, y también podrían contribuir a la evaluación de la idoneidad del donante y receptor de injertos, y para determinar la gravedad de las enfermedades de injerto frente al receptor.

Los inventores también postulan que la infidelidad de la transcripción se produce a una tasa mayor durante el proceso normal del envejecimiento. Esto crearía unas condiciones de reducción progresiva de la actuación que afectarían a todas las enzimas en general, y de forma más específica a las proteínas que están implicadas en la replicación celular, la reparación del ADN y el mantenimiento de la tasa normal de fidelidad de la transcripción. El concepto descrito anteriormente, por tanto, podría redirigir las investigaciones sobre los mecanismos del envejecimiento. Los métodos bioquímicos, proteómicos, transcriptómicos y otros métodos que permiten seleccionar secuencias de ARNm y ADNc, y la fidelidad a las especificadas por el ADN, podrían utilizarse, por tanto, para cuantificar la tasa de infidelidad de la transcripción que se produce en un individuo concreto. El ensayo de nuevas moléculas que reduzcan la tasa de infidelidad de la transcripción, por tanto, podría someterse a una selección biológica y conducir al desarrollo de fármacos que reduzcan la velocidad de envejecimiento.

Otros aspectos y ventajas de la invención se describen en la siguiente sección experimental.

Sección experimental

Ejemplo 1. Principio de construcción y secuenciación de un banco de ADNc típico (véase la figura 1)

La primera etapa para preparar un banco de ADN complementario (ADNc) es aislar el ARNm maduro del tipo de célula o tejido de interés. Debido a su cola de poli(A), resulta sencillo obtener una mezcla de todo el ARNm celular mediante hibridación con oligo dT complementarios unidos de forma covalente a una matriz. El ARNm unido entonces se eluye con un tampón de salinidad baja. La cola de poli(A) del ARNm entonces se deja hibridar con oligo dT en presencia de una transcriptasa inversa, una enzima que sintetiza una hebra de ADN complementaria a partir del molde de ARNm. Esto produce nucleótidos bicatenarios que contienen el molde de ARNm original y su secuencia de ADN complementaria. Después se obtiene ADN monocatenario retirando la hebra de ARN mediante un tratamiento alcalino o mediante la acción de la ARNasa H. Entonces se añade una serie de dG al extremo 3' del ADN monocatenario mediante la acción de una enzima denominada terminal transferasa, una ADN polimerasa que no requiere un molde, sino que añade desoxioligonucleótidos al extremo 3' libre de cada hebra de ADNc. El oligo dG se deja hibridar con oligo dC, que actúa como cebador para sintetizar, mediante la ADN polimerasa, una hebra de ADN complementaria con la hebra de ADNc original. Estas reacciones producen una molécula de ADN bicatenario completa que se corresponde con las moléculas de ARNm que se encuentran en la preparación original. Cada una de estas moléculas de ADN bicatenario se denominan habitualmente ADNc, y cada una contiene una doble cadena de oligo dC-oligo dG en un extremo, y una región de doble cadena de oligo dT-oligo dA en el otro extremo. Este ADN entonces se protege mediante metilación en los sitios de restricción. Entonces se acoplan conectores de restricción cortos en ambos extremos. Son segmentos de ADN sintético bicatenario que contienen el sitio de reconocimiento para una enzima de restricción concreta. El acoplamiento es realizado por la ADN ligasa del bacteriófago T4 que puede unir moléculas de ADN bicatenario con "extremos romos". El ADN con extremos romos bicatenario resultante con un sitio de restricción en cada extremidad entonces se trata con una enzima de restricción que crea extremos pegajosos. La etapa final en la construcción de bancos de ADNc es el acoplamiento de la doble cadena rota por enzimas de restricción con un plásmido específico que se transfecta en una bacteria. Las bacterias recombinantes entonces se cultivan para producir un banco de plásmidos, en presencia de antibióticos que se corresponden con la resistencia a antibióticos específica del plásmido. Cada clon porta un ADNc derivado de una única parte del ARNm. Cada uno de estos clones entonces se aisla y se secuencia utilizando métodos de secuenciación clásicos. Un ensayo típico de secuenciación comienza en el sitio de inserción y produce secuencias de 400 a 800 pares de bases para cada clon. Esta secuencia actúa como molde para comenzar la segunda ejecución de la secuenciación. Este avance conduce a la secuenciación progresiva del inserto del plásmido completo. Los resultados de la secuenciación de numerosos ADNc denominados EST se han depositado en varias bases de datos públicas.

Ejemplo 2. Anotaciones de las bases de datos

Las bases de datos de EST contienen información de la secuencia que se corresponde con la secuencia de ADNc obtenida en los bancos de ADNc y, por tanto, se corresponde fundamentalmente con la secuencia del ARNm individual presente en cualquier momento concreto en el tejido que se utilizó para producir estos bancos. La calidad de estas secuencias se ha puesto en cuestión por varios motivos. En primer lugar, tal como se analizó anteriormente, el proceso para producir bancos de ADNc inicialmente se basaba en gran medida en la presencia de una cola de poli(A) en el extremo 3' del ARNm eucariota. En segundo lugar, los ARNm son moléculas bastante frágiles que son digeridas con facilidad por unas nucleasas muy abundantes, denominadas ARNasas. En tercer lugar, mientras se estaban construyendo y secuenciando estos bancos, no se prestó mucha atención a la calidad del

material original utilizado ni a su conservación. Debido a esto, se han utilizado secuencias de EST para anotar información genética, es decir, para determinar si un segmento de ADN genómico identificado y totalmente secuenciado codifica algún ARNm específico. En este contexto, las secuencias de EST fueron útiles para identificar la secuencia codificadora genómica. Sin embargo, no se prestó atención a la información que portaba la propia secuencia EST. En efecto, la secuencia genómica de ADN se considera mucho más fiable, con potentes argumentos técnicos que apoyan esta opinión. Los inventores especulan que la diversidad incluida en las secuencias EST podría contener información importante desde el punto de vista biológico, analítico o clínico. En efecto, las bases de datos de EST fueron producidas por una serie de investigadores que utilizaron diversos métodos: esto condujo a los inventores a especular que cada sesgo metodológico debería contribuir a un nivel de ruido de fondo con un cierto número de errores. Sin embargo, si existiesen diferencias en los errores debidas a la fuente del material utilizado para generar el banco, entonces la diferencia en la tasa de error estaría directamente relacionada con la fuente subyacente.

Para ensayar esto, los inventores recuperaron las bases de datos ***est_human*** disponibles en el sitio de NCBI ftp. Se seleccionaron estas bases de datos porque estas secuencias no estaba anotadas ni curadas con herramientas humanas ni bioinformáticas.

Los inventores utilizaron un sistema de identificación del banco para determinar si un EST se obtuvo de un tejido canceroso o normal. Cada banco se denominó ***normal*** o ***canceroso***. Mediante el emparejamiento del número de registro de cada EST con el identificador del banco correspondiente se clasificaron 2,6 millones de EST como obtenidos de tejidos cancerosos, y 2,8 millones de EST como obtenidos de tejidos normales.

Para obtener las alineaciones de los EST, los inventores utilizaron el programa informático disponible para el público megaBLAST 2.2.13 (Basic Local Alignment Search Tool, Zheng Zhang, A greedy algorithm for nucleotide sequence alignment search, J. Comput. Biol., 2000) y se especificaron los siguientes parámetros de selección:

-d est_human: base de datos para la selección (todos EST humanos de dbEST)

-i sequences.fasta: formato de archivo FASTA que contiene la secuencia de referencia que se corresponde con genes expresados con abundancia (figura 2)

-D 2: salida de BLAST tradicional

-q -2: puntuación de -2 para un desapareamiento de nucleótidos

-r 1: puntuación de 1 para un único apareamiento de nucleótidos

-b 100000: número máximo de secuencias para las cuales se indica el alineamiento

-p 90: porcentaje de coincidencia mínimo entre EST y la secuencia de referencia

-S 3: toma en cuentra EST que se aparean con la secuencia de referencia

-W 16: longitud del mejor apareamiento perfecto para comenzar con la extensión del alineamiento

-e 10: valor esperado (valor E), número de alineamientos que se espera encontrar simplemente por azar para una secuencia concreta y una base de datos concreta, según el modelo estocástico de Karlin y Altschul (1990)

-F F: no se filtra la secuencia de referencia

-o NonFiltre_test_90.out: archivo de salida de informe de megaBLAST

-v 0: número máximo de secuencias que muestran una línea de descripción

-R T (T de “true”, verdadero): indica la información logarítmica al final de la salida de megaBLAST

-I T: muestra los GI en la definición de línea [T/F]

La línea de comando para el BLAST es:

megablast -d est_human -i C:/SeqRef/TPT1/sequences.fasta

-o C:/blast/NonFiltre_test_90.out -R T -D 2 -I T -q -2 -r 1 -v 0 -b 100000 -p 90,00 -S 3 -W 16 -e 10,0 -F F

La figura 2(a-q) muestra las 17 secuencias utilizadas para el análisis; para evitar distorsionar el BLAST se eliminó sistemáticamente la cola de poliA de las secuencias de referencia de ARNm. La figura 2r proporciona una lista de genes que se utilizaron para ensayar la aparición de variaciones en la secuencia.

Ejemplo 3. Identificación de las variaciones en la secuencia entre EST procedentes de orígenes normales y cancerosos

Se seleccionaron 17 genes basándose en su gran representación en la base de datos. Cada secuencia EST entonces se alineó frente a su secuencia de referencia de ARNm curada (RefSeq) procedente de NCBI utilizando el MegaBLAST. Esto crea una matriz en la que cualquier base concreta se define por el número de EST que se aparean en esta posición. Entonces los inventores midieron la proporción de EST que se desvían de RefSeq en cualquier posición concreta. La comparación de las secuencias de EST alineadas según el tejido de origen condujo a la identificación de las variaciones en las secuencias que producen en cada posición de las bases en el grupo de cáncer o normal. La figura 4 proporciona una representación gráfica de estas variaciones que se producen en los conjuntos normales y cancerosos para los 17 genes seleccionados. El examen visual de los datos reveló que la variación en la secuencia se producía con más frecuencia en el conjunto de cáncer, y también que el fenómeno parecía predominar más sitios del ARNm específicos. La mayoría de las observaciones se localizaron al menos 400-500 bases después del comienzo de la secuencia del ARNm, por tanto predominantemente en la parte 3' del gen. Este número tan alto de variaciones no puede ser explicado por los SNP. En la gráfica se muestran los SNP putativos (□) y los SNP biológicamente validados (O); estos SNP se obtuvieron en la base de datos NCBI (dbSNP, construida 126, septiembre, 2006) (Sherry, S.T. et al. (2001), Nucleic Acids Res., 29, 308-311). Los SNP putativos y biológicamente validados que conducen a variaciones en EST (n = 442) se excluyeron de posteriores análisis (por tanto, también se excluyeron los sitios de edición del ARN (SNP falsos)).

La siguiente etapa de este análisis fue ensayar la significancia estadística de las diferencias en la variación en la secuencia que se producen entre EST de cáncer y normales. Para cada posición, los inventores compararon la proporción de la base de RefSeq con la de las tres otras bases entre los grupos normal y con cáncer utilizando un ensayo de proporción bilateral. Este ensayo se aplicó sistemáticamente siempre que se cumpliesen las siguientes condiciones: $n > 70$, y $(n_{i.} * n_{.j})/n > 5$, $i = 1, 2$, $j = 1, 2$ (siendo n = el número de EST de cáncer y normales para un gen, $n_{1.}$ = el número de EST de cáncer, $n_{2.}$ = el número de EST normales, $n_{.1}$ = el número de EST que tienen la RefSeq, $n_{.2}$ = el número de EST que tienen una sustitución). Se dice que un ensayo estadístico es positivo al nivel umbral de 5% cuando el correspondiente valor de P es menor que 0,05; en este caso, la hipótesis nula (es decir, ambas proporciones son iguales) se rechaza.

Además, se consideraron los siguientes dos ensayos de proporción unilateral para precisar en qué conjunto la variabilidad era mayor. El primero permitió concluir que las variabilidades eran diferentes en ambos grupos cuando el ensayo estadístico es positivo, y entonces se midió en esta caso si la variabilidad era estadísticamente mayor en el conjunto del cáncer. Por el contrario, el segundo ensayo verificó la hipótesis de que la variabilidad era significativamente mayor en el conjunto normal. Ambos ensayos unilaterales se realizaron cuando se cumplieron las mismas condiciones que los ensayos bilaterales.

La figura 3 proporciona una descripción de los resultados de BLAST típicos. Tal como se muestra en las figuras 5a y 5b, para el gen representativo TPT1, el número de EST en cualquier posición concreta sobre el gen es similar en los grupos de cáncer y normales. La figura 5c muestra que 489 de 830 ensayos de proporción posibles cumplían los criterios asignados. Los ensayos de proporción que eran estadísticamente significativos al nivel de 5% se muestran en la figura 5d. Puede calcularse un error estimado resultante de los múltiples ensayos, definido por el estimador basado en la localización (C. Dalmasso, P. Broet (2005), Journal de la Société FranÇaise de Statistique, tomo 146, n1-2, 2005). Los 26 ensayos de proporción positivos son debidos a las sustituciones de bases que aparecen en el tejido normal (N>C) en la figura 5d (n = 26; LBE = 33). Esto contrasta con la acumulación de ensayos de proporción estadísticamente significativos que aparecen debido a las variaciones en la secuencia en el grupo de cáncer (C>N) (n = 145; LBE = 15). Se realizó un análisis similar para un segundo gen, VIM (figura 5e-5h). De nuevo, el número de EST en cualquier posición concreta de un gen es superponible en los grupos de cáncer y normales. 752 de los 1847 ensayos de proporcion cumplían los criterios. De nuevo, se observó una gran diferencia en el número de ensayos de proporción positivos: n = 78 variaciones son debidas al grupo normal (LBE = 50), n = 269 variaciones son debidas al grupo de cáncer (LBE = 24). Se repitió el mismo análisis en 17 genes que están presentes en abundancia en la base de datos de EST. La figura 5i es un resumen de los resultados estadísticos obtenidos. De los 17 genes ensayados, 15 muestran mayores variaciones en la secuencia en los EST obtenidos de tejidos con cáncer (figura 5j). En esta etapa, el análisis estadístico cubre sólo del 9% al 91% de las posiciones de cada gen (una media de 32%) debido a las restricciones del ensayo estadístico.

La conclusión de esta primera ronda de análisis es que los EST del mismo gen fueron diferentes cuando se comparan según la fuente del tejido (normal frente a canceroso) de la cual se extrajo el ARNm. Suponiendo que la tasa de errores técnicos que produce la variación en la secuencia de EST no es diferente según los orígenes normal o canceroso del tejido, y considerando el hecho de que 15 de 17 de los genes ensayados mostraron variaciones en la secuencia debidas a la fuente del tejido, los inventores proponen que estas diferencias son el resultado directo del estado de las células (normal frente a canceroso) a partir de las cuales se produjeron los EST.

Se realizaron experimentos similares para estudiar las inserciones y las deleciones. Los resultados se obtuvieron después de aplicar el filtro F3 (véase el ejemplo 4). Se emplearon condiciones más rigurosas para evitar acontecimientos no biológicos (errores de la secuenciación, alineamientos erróneos de MegaBLAST…): se consideró un hueco o una inserción sólo si no había otras modificaciones presentes en una ventana -10/+10. Para cada posición, los inventores compararon la proporción de ventanas con deleción (o inserción) con otras ventanas entre los grupos normales y de cáncer utilizando un ensayo de proporción bilateral como se describió anteriormente. La figura 5k es un resumen de los resultados estadísticos obtenidos. De los 17 genes ensayados, 14 (o 10 para la

inserción) muestran mayores variaciones en la secuencia (deleciones e inserciones) en los EST obtenidos de tejidos cancerosos (figura 5l).

Ejemplo 4. Verificación del exceso de variación en el conjunto de cáncer C>N frente al conjunto normal N>C

Se procesaron bancos que se habían originado de muestras de cáncer o muestras normales fundamentalmente de la misma manera. Por tanto, se espera que aparezcan errores aleatorios de la construcción del banco o la secuenciación del clon en la misma proporción en ambos conjuntos y que no justifiquen las diferencias observadas entre los grupos de EST normales y de cáncer. El análisis matemático es coherente con esta interpretación (Brulliard M. et al., PNAS, 1 de mayo, 2007, vol. 104, nº 18, 7522-7527, véase el material suplementario, figura 7).

Después se realizó una búsqueda para eliminar otras fuentes de heterogeneidad de los EST aplicando secuencialmente procedimientos de filtrado basados en el siguiente fundamento. Los requerimientos iniciales fueron que los EST se alinearan con RefSeq con 100% de coincidencia en al menos 16 bases consecutivas, y con $\geq$90% de coincidencia en al menos 50 bases. Tal como se muestra en la figura 6b, esto produce, cuando se comparan los conjuntos de cáncer y normales para los 17 genes, 2281 y 725 diferencias estadísticamente significativas C>N y N>C, respectivamente (columna F1), y diferenciadas de los SNP putativos o biológicamente validados. El segundo filtro (F2) requiere que cada EST se alinee con RefSeq continuamente en más del 70% de su longitud. El tercer filtro (F3) retira los EST con una secuencia más relacionada con los parálogos y los pseudogenes que con la RefSeq auténtica. El cuarto filtro (F4) deleciona del análisis las primeras y últimas 50 bases de cada alineamiento de EST. Se empleó este filtro para eliminar los desapareamientos en los bordes 3' y 5' de los EST que pueden ser creados por el programa MegaBLAST para optimizar aún más los alineamientos y/o que resultan de la acumulación de errores en las extremidades de EST alineadas. El último filtro (F5) normaliza las longitudes de secuencia de EST para cada gen para eliminar cualquier diferencia en la longitud entre los conjuntos normales y de cáncer. En efecto, los inventores advirtieron que, después de aplicar los primeros 4 filtros, la longitud de los EST de cáncer media fue mayor que la de los EST normales (640 $\pm$ 248, y 554 $\pm$ 229 bases, respectivamente). Sin embargo, también observaron una heterogeneidad de los EST de cáncer significativamente mayor en 5 genes cuando la longitud media de los EST no era diferente entre los conjuntos normales y de cáncer (TPT1, VIM, HSPA8, LDHA, CALM2).

El número de ensayos C>N estadísticamente significativos siguió siendo mayor que el número de ensayos N>C, pero la proporción C/N disminuyó de 3,15 (F1) a 2,05 (F5) (figura 6c).

Para determinar aún más que la heterogeneidad de los EST de cáncer no era debida a la acumulacion de errores al final del ensayo de secuenciación, se tomó en cuenta sólo la información proporcionada después de las primeras 50 bases y no más larga de 450 bases en cualquier ensayo de secuenciación. Después de este filtrado drástico, los ensayos C>N estadísticamente significativos fueron 455 (LBE = 92), y los ensayos N>C fueron 292 (LBE = 119) para los 17 genes. Por tanto, puede considerarse de modo razonable que las variaciones en la secuencia que provocan la mayor heterogeneidad en los EST del grupo de cáncer son un reflejo directo de la heterogeneidad del ARNm de las células cancerosas (figura 6d).

Después los inventores verificaron de forma independiente que una mayor heterogeneidad de EST estadísticamente significativa persistía en el conjunto de cáncer después de eliminar los EST producidos a partir de células cancerosas cultivadas. Después de este filtrado, los ensayos C>N estadísticamente significativos fueron 1009 (LBE = 117), y los ensayos N>C fueron 445 (LBE = 193) para los 17 genes (figura 6e).

Ejemplo 5. La ruptura del código de sustitución de bases se produce debido a la infidelidad de la transcripción en células cancerosas

El siguiente objetivo de los inventores fue determinar si el fenómeno descrito de la sustitución de bases debido a la infidelidad de la transcripción es un fenómeno aleatorio o sigue reglas específicas. Para lograr esto, se centraron en las variaciones de EST en las que C>N es estadísticamente significativo. Para evitar sesgos que podrían ser introducidos por los procedimientos de filtrado, se emplearon todos los datos no filtrados disponibles. La primera indicación de que la infidelidad de la transcripción no es un proceso aleatorio fue que las sustituciones de bases raramente aparecen en la región 5' de los genes ensayados. Para todos los genes ensayados, muy raramente se observó la aparición del fenómeno de sustitución de bases en las primeras 400-500 bases del ARNm maduro. La segunda observación que indica que la infidelidad de la transcripción no es aleatoria surge de la observación de que existe una diferencia en la composición de bases en el molde de ADN genómico observado cuando se comparan secuencias cadena arriba y cadena debajo de los acontecimientos de sustitución (heterogéneos, H, n = 2281) con aquellas en que no se detectaron acontecimientos de sustitución (no heterogéneos, NH, n = 12273). El criterio para los sitios NH fueron unas variaciones en el conjunto de cáncer menores que 0,5% y no estadísticamente diferentes de las variaciones del conjunto normal (figura 7a). En este análisis, se denomina a la base que sufre la sustitución como b0, las bases localizadas en el extremo 5' del ARNpm se denominan b-n, y las bases localizadas en el extremo 3' se denominan b+n. Para mayor claridad, en este análisis los inventores hacen referencia a la composición de bases del ARNpm: esto se corresponde en el ADN con la hebra que no es el molde (la hebra no transcrita por ARNP). Los datos demuestran que no todas las 4 bases son igualmente susceptibles a la variación: b0 = A (33%) $\approx$ T (32%) >> C (21%) >> G (14%). Además, las composiciones de las 4 bases cadena arriba y las 3 bases cadena abajo del sitio del acontecimiento fueron estadísticamente diferentes (resultados del análisis de chi-

cuadrado) de las de los sitios sin una variación de EST significativa (figura 7a, la composición de bases se expresa en porcentaje, y el gris claro muestra las bases enriquecidas; el gris oscuro muestra las bases insuficientes). De modo específico, los sitios en los que se producen variaciones fueron precedidos y seguidos con más frecuencia por $A \geq G > T \approx C$.

Por tanto, la aparición de heterogeneidad en las EST de cáncer no es aleatoria, sino que está determinada en primer lugar por la naturaleza de la base que está siendo sustituida y en segundo lugar por la naturaleza de las bases que preceden y siguen inmediatamente al acontecimiento.

Después los inventores se preguntaron si la base de sustitución se selecciona de modo aleatorio. Resulta evidente a partir de la figura 7b que no es el caso. La significancia estadística de la diferencia en las proporciones se calculó con la hipótesis nula de que la base de sustitución se selecciona de modo aleatorio (ensayo de ajuste de las tres bases de sustitución a la distribución uniforme). A fue sustituida preferentemente por C ($p = 2{,}8 \times 10^{-125}$), T por G ($p = 5{,}7 \times 10^{-29}$), y G por A ($p = 2{,}2 \times 10^{-32}$). La sustitución de C mostró una distribución más uniforme, con una ligera insuficiencia de T ($p = 0{,}007$).

Después se buscaron las causas subyacentes a la sustitución de bases preferente. Para lograr esto, los inventores distinguieron entre dos conjuntos de acontecimientos informativos y no informativos.

Los acontecimientos informativos son situaciones en las que la base sustituida es diferente de la base precedente (b-1) o de la base siguiente (b+1) (n = 1676) (figura 7c). Los acontecimientos no informativos son situaciones que no cumplen estos criterios. Cuando se analizan acontecimientos informativos aparecen dos casos: la base sustituida es reemplazada por b-1 o b+1 (79%) o por otra base, diferente de b-1 y b+1 (21%). 1) En el primer subconjunto, la base de sustitución fue idéntica a b-1 (n = 799, figura 7c, panel B) o b+1 (n = 530, figura 7c, panel C). Cuando la base de sustitución fue b-1, entonces b0 = A (36%) > C (30%) >> T (21%) >> G (13%) (figura 7d, panel A). A fue sustituida preferentemente por C (71% de los casos). Cuando la base de sustitución fue b+1, entonces b0 = T (47%) >> A (21%) > C (19%) >> G (13%) (figura 7a, panel B). T fue sustituida preferentemente por G (71%). De manera interesante, la importancia estadística de las bases circundantes también fue diferente (figura 7d, panel A, y figura 7d, panel B). Para las sustituciones b-1, el patrón de influencia relativa de la composición de bases fue b0 > b-1 > b-2 > b+1 > b-3 > b+2. Para las sustituciones b+1, la influencia relativa de las bases circundantes sigue el patrón de b0 > b+1 > b-1 > b+2 > b-3 > b-2. 2) En el segundo subconjunto de acontecimientos informativos, la base de sustitución no se correponde con b-1 ni b+1 (n = 347, figura 7c, panel D). Las bases afectadas tuvieron el siguiente orden: A (47%) > T (29%) > C (14%) > G (10%). A fue reemplazada más habitualmente por C (91% de los casos), T por C (50%) y A (42%), C por G (46%), y G por C (73%). Por tanto, cuando la base de sustitución no se corresponde con b-1 ni b+1, la base de sustitución no se selecciona de modo aleatorio, sino que es C en muchísimas ocasiones.

Después los inventores consideraron el conjunto de acontecimientos no informativos, es decir, las situaciones en las que 1) b-1 = b+1, y en las que 2) b-1 = b0 = b+1 (figura 7c). Cuando b-1 y b+1 son idénticos pero diferentes de b0 (n = 339, figura 7c, panel E), las bases sustituidas tuvieron el siguiente orden: T (34,8%) > G (23,6%) > C (21,2%) < A (20,4%), y siguieron el mismo patrón de preferencia que en la figura 7b: $T \rightarrow G$, $G \rightarrow A$, $A \rightarrow C$. Las sustituciones que se produce en la base central de repetición de tres bases idénticas (n = 266, figura 7c, panel F) se observaron en el siguiente orden: A (46,2%) > T (36,9%) > G (10,5%) > C (6,4%). En este caso, los acontecimientos de sustitución más comunes fueron: $A \rightarrow C$, y $T \rightarrow C$ y A. Las raras sustituciones de GGG fueron reemplazadas, de modo más habitual, por GCG y CCC por CAC (figura 7c).

Por tanto, cuando las sustituciones aparecen dentro de tres bases idénticas consecutivas, y cuando las sustituciones no se corresponden con b-1 ni b+1, entonces C es la base de sustitución más habitual (figura 7c). Cuando la base de sustitución se corresponde con b-1, la sustitución más habitual es $A \rightarrow C$; cuando la base de sustitución se corresponde con b+1, la sustitución más habitual es $T \rightarrow G$.

Por tanto, puede concluirse que ni la base que se sustituye ni la base de sustitución se seleccionan de modo aleatorio. Ambos fenómenos siguen patrones predecibles definidos por la composición de la base que se sustituye y la de las bases localizadas cadena arriba y cadena debajo de este acontecimiento.

Pueden definirse algoritmos específicos para identificar con precisión la composición del motivo que determina la aparición de la sustitución de bases en cualquier secuencia concreta.

Después, los inventores separaron el conjunto C>N en dos conjuntos en los que N es estable (no se observa desviación en la misma posición en el conjunto normal) o no lo es. La desviación se considera sólo si excede de un cierto umbral definido como el porcentaje medio de desviación en el conjunto normal. La figura 7e demuestra que el conjunto C>N porta la firma -4/+3 en el que N se desvía. La longitud del contexto es menos importante en el otro conjunto. La figura 7f muestra una acumulación en el intervalo 500-1000 que es más importante en el conjunto C>N en el que N se desvía, que en el conjunto C>N en el que N es estable. En éste último, parece que aparecieron acontecimientos más tempranos (intervalo 0-500). Las bases de sustitución para los sitios C>N fueron similares entre los dos conjuntos (figura 7g).

En el caso de deleciones o inserciones, los resultados demuestran que las bases omitidas (398 casos C>N) tuvieron

el siguiente orden: C (46,0%) > T (37,9%) >> A (9,3%) > G (6,8%) (figura 5m), y que las bases insertadas (225 casos C>N) tuvieron el siguiente orden: G (36,0%) > C (30,2%) > A (21,8%) > T (12,0%) (figura 5o). Los inventores observaron que los acontecimientos de deleción o inserción a menudo se producen en bases idénticas conservadas. Un programa específico diseñado para analizar estos acontecimientos muestra efectivamente que 94,7% de las deleciones y 81% de las inserciones se producen en tramos de longitudes variables. Los tramos pueden ser bipolares y asimétricos (por ejemplo, AAATAA) o no (por ejemplo, CCCGG) o repeticiones de nucleótidos individuales (por ejemplo, AAA) (véanse los motivos en la figura 5m-5p). En el caso de las deleciones, las bases omitidas son idénticas a la base b-1 o b+1 (tramo) en 84% de los casos. La mayoría de los motivos son dobletes o tripletes de C > T > A ≈ G. Se observan resultados similares en el caso de las inserciones. De hecho, los motivos de los tramos no parecen ser diferentes entre C>N y N>C. La especificidad de C>N o N>C surgiría de otra información alrededor de los tramos. Otros análisis en más genes confirmarán estas observaciones.

Tomando en cuenta los 17 genes, los inventores observaron una heterogeneidad de EST con una tasa de 10 por 100 bases (figura 5j y figura 8a-c). Esta tasa es mayor que cualquier tasa de mutación descrita producida en el ADN genómico. Como referencia, se puede calcular que los polimorfismos de un solo nucleótido se producen aleatoriamente en el genoma una vez cada 300 bases. La tasa de polimorfismos de un solo nucleótido que afectan al ADN transcrito es mucho menor: se calcula que aparece una vez cada 3000 bases. Es evidente que las mutaciones en el ADN se producen con más frecuencia en el cáncer. Sin embargo, los trabajos en profundidad de secuenciación de ADN de cáncer de mama y colon que incluían 14 de los 17 genes utilizados en este estudio, condujeron a una tasa de mutación somática estimada de 3,1 mutaciones por 10^6 bases (Sjoblom, T. et al. (2006), Science, 314, 268-274). Por tanto, la infidelidad de la transcripción que se describe en esta invención se produce con una tasa mucho mayor que las mutaciones que afectan al ADN. De manera más importante, la mayoría de las sustituciones de bases en el ADN tienen consecuencias limitadas a nivel de proteínas, porque menos del 10% del ADN genómico se transcribe a ARNm. En contraste, la sustitución de bases debido a la infidelidad de la transcripción a menudo tiene consecuencias directas sobre la función de la proteína. En efecto, 1179 de 2281 sustituciones descritas en la presente (1548 CDS - 369 sustituciones silenciosas) condujeron a unas sustituciones de bases con un impacto inmediato sobre la secuencia primaria de AA de la proteína (figura 8d). De manera más importante, se observaron unas sustituciones de bases significativas que afectan al codón de fin en 9 de los 17 genes ensayados. Sin embargo, antes del concepto de la infidelidad de la transcripción no se había propuesto que las proteínas humanas contuviesen secuencias codificadoras adicionales codificadas por secuencias de ARN consideradas hasta la fecha “regiones no traducidas”. Los inventores ahora demuestran que la sustitución de bases que se produce en los codones de fin naturales debido a la infidelidad de la transcripción revela nuevas regiones codificadoras que codifican AA específicos. Estas nuevas regiones codificadoras están en fase con el marco de lectura abierto nativo. El codón de fin natural se transforma en un codón codificador. El siguiente triplete de bases entonces se lee como un AA y la traducción avanza con una nueva región codificadora hasta que se alcanza un nuevo codón de fin. Los inventores han verificado que éste es el caso en 8 de los 17 genes que mostraron infidelidad de la transcripción. Estos 8 contenían codones de fin alternativos dentro del marco con la correspondiente RefSeq (GAPDH no contiene un codón de fin alternativo). Esto tiene una consecuencia inmediata porque, en cada caso, se crean nuevas secuencias codificadoras de 14, 7, 13, 15, 13, 4, 9, 55 AA en TPT1, RPS6, RPL7A, VIM, RPS4X, FTH1, FTL y TPI1, respectivamente (figura 10). La adición de estos AA tiene el potencial de crear motivos que serán muy potenciados en el cáncer; estos motivos darán o no como resultado una nueva función de las proteínas. La predicción de este acontecimiento conduce al posible desarrollo de herramientas útiles que puedan ser utilizadas en el diagnóstico, con objetivos terapéuticos u otros objetivos. La predicción de este acontecimiento también conduce al posible desarrollo de anticuerpos específicos que reconocerán secuencias específicas del cáncer en el extremo carboxi-terminal de la proteína. Ningún método analítico actual es capaz de unaa secuenciación de proteínas directa en el extremo carboxi-terminal. Sin embargo, es posible romper las proteínas de forma enzimática y secuenciar los productos de la ruptura desde su extremo NH_2 terminal. También es posible analizar el contenido de AA de los péptidos generados mediante proteolisis utilizando espectrometría de masas. Los inventores además han demostrado que estos codones de fin alternativos también se ven afectados por la infidelidad de la transcripción (7/9 genes tiene el segundo codón de fin alternativo afectado). El mismo fenómeno descrito anteriormente puede expandir aún más la lectura a un nuevo conjunto de secuencias. La anotación de todas las secuencias de proteínas utilizando el método de los inventores revelará varias secuencias de ARNm codificadoras insospechadas que serán transcritas con más o menos eficacia dependiendo de la utilización de los codones, así como de la capacidad de la maquinaria de traducción para traducir correctamente o no las sustituciones de las bases. De hecho, las sustituciones de las bases pueden conducir a cambios en la estructura del ARNm, que pueden modificar la velocidad de lectura ribosómica. No obstante, los inventores suponen que las sustituciones de bases no implicarán cambios en la estructura del ARN. Basándose en la aparición de alteraciones en los codones de fin, calculan, en los genes afectados, que hasta 4% del ARNm en los tejidos cancerosos contiene estas regiones codificadoras adicionales.

Un programa específico basado en varios filtros puede utilizarse para anotar todas las secuencis de proteínas para la presencia de un péptido de post-fin (PSP) putativo. Después de recuperar la secuencia nucleica que se corresponde con las proteínas estudiadas, el programa busca la presencia o la ausencia de una secuencia nucleica en fase después del fin canónico, con otro fin en fase (la posibilidad de pasar por encima de uno o más codones de fin en el caso de que la infidelidad de la transcripción afecte a estos codones de fin alternativos debe tomarse en cuenta). Puede fijarse una longitud mínima (por ejemplo, sólo las secuencias que codifiquen más de 12 aminoácidos

pero, en último término, la longitud del péptido puede ser menor) según un criterio mínimo para la predicción de un patrón antigénico potencial. Entonces los PSP antigénicos se almacenan (figura 18). Se emplea otra etapa para validar los candidatos con la anotación mediante la formación de la máquina de aprendizaje del fin canónico: en efecto, puede determinarse la probabilidad de que una o más bases del codón de fin puedan ser sustituidas mediante la infidelidad de la transcripción. También es posible analizar el contenido en AA de los péptidos generados mediante proteolisis utilizando espectrometría de masas.

Los inventores también han descubierto que, además de crear condiciones que permiten la traducción de nuevas secuencias de proteínas, la infidelidad de la transcripción puede introducir codones de fin prematuros en el ARNm. Se identificaron 24 nuevos codones de fin que aparecían dentro del marco de lectura abierto canónico en 13 de los 17 genes. Esto indica que la infidelidad de la transcripción puede dar como resultado la producción de proteínas más cortas que carecen de dominios específicos. Estas proteínas truncadas pueden producir un aumento o la pérdida de la función. Es probable que la estructura tridimensional de la proteína se vea afectada y, por tanto, se crean nuevas entidades que pueden ser reconocidas por el sistema inmunológico.

Por último, la infidelidad de la transcripción en los 17 genes ensayados revela que 50% de todas las sustituciones C>N identificadas conducen a cambios en los AA. 17% se corresponden con la sustitución de AA por otros AA de la misma familia, y 33% se corresponden con sustituciones de AA por una clase diferente de AA. Por tanto, la infidelidad de la transcripción es capaz de generar proteínas con nuevas secuencias de AA con funciones o actividad potencialmente modificadas. La predicción de la infidelidad de la transcripción utilizando las reglas descritas anteriormente permitirá la predicción lógica de cambios en el comportamiento de proteínas y el resultado de experimentos proteómicos.

Ejemplo 6. Validaciones biológicas

Las validaciones biológicas se realizan a dos niveles: ARNm y proteínas. En primer lugar, se detectarán las sustituciones en el ARNm de los 17 genes de interés en tejidos cancerosos humanos. Los inventores utilizaron DHPLC (cromatografía líquida de alta resolución desnaturalizante), que es un método cromatográfico a gran escala para detectar mutaciones en las secuencias. El principio del experimento se describe en la figura 11a-c. Primero los inventores desarrollaron un método para ensayar el umbral de DHPLC transgenómico para estimar el porcentaje de ADN mutado en la muestra que es suficiente para permitir la formación y la detección de heterodúplex. En efecto, se emplearon fragmentos de PCR de 300 pb con 1 y 3 sustituciones de bases. Se prepararon diferentes proporciones de estos fragmentos: 0%, 2,5%, 5%, 7,5%, 10%, 20% o 50%. La DHPLC permite distinguir el ADN normal y el ADN normal más mutado en cuanto la muestra contenga del 2,5% al 5% de ADN mutado (figura 11d).

Estos resultados indican que los inventores fueron capaces de distinguir el ARNm de tejidos normales y cancerosos para los genes de interés. Se empleó ARNm extraído de tejidos adyacentes normales y cancerosos (Biochain Inc.) para ensayar tres genes: GAPDH, ENO1 y TMSB4X. Como la DHPLC se aplica al ADN, las muestras de ARNm se convierten en ADNc utilizando transcriptasa inversa. Los inventores elegieron regiones que muestran sustituciones mucho más significativas en EST procedentes de tejidos de cáncer que EST procedentes de tejidos normales. Los cebadores utilizados para la amplificación se muestran en la figura 12.

En un primer conjunto de experimentos, varios ADNc de tejido adyacente normal y canceroso (hígado, riñón, mama y colon) se amplificaron mediante PCR con dichos cebadores y se inyectaron en el sistema DHPLC. La temperatura de la estufa se seleccionó con el programa informático Navigator (Transgenomic). Se obtuvieron perfiles para los genes descritos anteriormente, y un experimento representativo se presenta en la figura 15 para los genes ENO1, GAPDH y TMSB4X. Tal como se muestra en las figuras 15a a 15c, los perfiles del cáncer son claramente diferentes de los perfiles normales para los genes GAPDH y ENO1. No se observaron diferencias para el gen TMSB4X, tal como se esperaba (tiene muchos menos sitios de infidelidad de la transcripción). Las inyecciones del mismo producto de PCR y de otros 2 productos de PCR se realizaron por triplicado (figura 15b y 15c), y los perfiles fueron muy reproducibles en el mismo experimento. Sin embargo, la naturaleza de la diferencia no puede deducirse de este experimento.

Por consiguiente, se realiza la continuación de esta validación biológica basada en ARNm. De hecho, la secuenciación de los productos de PCR procedentes de tejidos cancerosos puede permitir una detección precisa de las mutaciones más abundantes.

El método de secuenciación de Sanger clásico de ARNm amplificado con PCR de transcripción inversa no detecta los variantes de secuencia que se producen a unas tasas inferiores a 15-30% en una posición específica (figura 16). En efecto, la lectura automática de nucleótidos mutados obtenida mediante la secuenciación de mezclas de productos de la amplificación de secuencias conocidas mutadas y no mutadas se obtiene en mezclas del 50-50% y se detectan picos secundarios del 15%. La pirosecuenciación y la PCR de emulsión son métodos más sensibles que permiten la detección de la heterogeneidad del ADNc, y por tanto es posible el análisis de los productos de RT-PCR obtenidos de células cancerosas y normales (Thomas, R.K. et al. (2006), Nat. Med., 12, 852-855).

Un gen afectado por la infidelidad de la transcripción se clona por completo, o se clona un fragmento que comprende y que no comprende su sitio de infidelidad de la transcripción. La construcción con y sin el codón de fin canónico se

acopla dentro del marco con el gen que codifica la resistencia a la neomicina. Se transfectan células cancerosas y normales con esta construcción quimérica, primero de forma transitoria y después estable. Los inventores predicen que la infidelidad de la transcripción conducirá a cambiar el codón de fin canónico, permitiendo con ello la traducción del gen de neomicina y creando una célula resistente a la neomicina. También predicen que las células cancerosas que sean resistentes a la neomicina crecerán con más rapidez, y que la forma de estas células será significativamente diferente de la de las células normales. Además, predicen que estas células serán más invasivas y pueden compararse con las células de una etapa más tardía del avance del cáncer. Por tanto, esta técnica puede utilizarse para determinar la fase cancerosa de diferentes células procedentes de un paciente individual.

La prueba final de que la resistencia a la neomicina aparece como resultado de la infidelidad de la transcripción se obtendrá secuenciando la construcción insertada amplificada a partir del ADN genómico y demostrando que la información genómica permanece inalterada. Los inventores también verificarán que el ARNm de la célula resistente a la neomicina contiene un codón de fin mutado debido a la infidelidad de la transcripción.

Una técnica que permite la detección de sitios de infidelidad de la transcripción es la construcción de bancos de ADNc que consisten en clonar ADNc obtenido mediante transcripción inversa a partir del ARN extraído de tejidos de pacientes cancerosos y normales. Cada ADNc, amplificado o no a partir de un gen específico, se clona en plásmidos diferentes que se tranforman en *Escherichia coli*. Se escogen diferentes colonias de *E. coli* y los plásmidos se secuencian. El número de clones que deben secuenciarse depende del porcentaje de sustituciones en el fragmento de ADNc clonado. Un análisis estadístico proporcionará la variación en la secuencia en los sitios de infidelidad de la transcripción. Esta técnica puede mejorarse. En efecto, después de la transcripción inversa y de la amplificación con cebadores específicos, el ADNc puede clonarse en un plásmido que porte un gen indicador, por ejemplo, el gen *lacZ*. El ADNc y el gen indicador pueden clonarse en fase. Cuando aparece una sustitución en la secuencia del codón de fin del ARNm, el gen indicador se expresa. Cuando las células de *E. coli* se transforman con esta construcción, las colonias que portan el plásmido en el que el ADNc está mutado en el codón de fin se seleccionan con la expresión del gen indicador. Después de esto, el plásmido puede secuenciarse para verificar si la sustitución del codón de fin está presente.

Se emplea otra técnica basada en la PCR a tiempo real con cebadores específicos para detectar los sitios de infidelidad de la transcripción. Estos cebadores se diseñan para que se apareen con un ADNc con una variación o variaciones en la secuencia basadas en un análisis estadístico. Se diseña un segundo cebador para que se aparee con el ADNc sin la variación en la secuencia (referencia). El número de variaciones en la secuencia del cebador es muy importante, y los inventores han determinado que, en un experimento de ensayo con una variación conocida en la secuencia, son necesarias dos mutaciones fuera de la posición que se estudia para obtener una señal fluorescente específica. Cuando el cebador es complementario con el ADNc sin una sustitución en la secuencia (referencia), la señal fluorescente se detecta tras un número específico de ciclos de amplificación. El mismo ADNc amplificado con el cebador que porta la sustitución en la secuencia conduce a una señal de fluorescencia que aparece con un número grande de ciclos de amplificación. La diferencia entre el número de ciclos de amplificación entre los dos cebadores es una estimación directa de la presencia de un ADNc con una sustitución en la secuencia. Además, los inventores han verificado que el método es lo suficientemente sensible como para detectar 1% de la secuencia mutada en una mezcla de secuencias conocidas.

En un experimento típico, un ARN extraído de tejidos de pacientes cancerosos y normales se somete a transcripción inversa y se amplifica con cada uno de dichos cebadores específicos. Entonces se mide la diferencia entre los ciclos de amplificación con ambos cebadores.

Por último, los inventores se centraron en las nuevas proteínas inducidas cuando el codón de fin natural se ve afectado. Los inventores demostraron que el codón de fin se ve significativamente afectado en 9 de los 17 genes de interés. Esto conduce a poblaciones específicas de proteínas diferenciadas con una nueva secuencia en el extremo carboxi-terminal. Los inventores han calculado que los tejidos cancerosos con genes afectados contienen 4% de proteínas que son más largas que las normales.

A la vista de esta hipótesis, se analizaron 60 proteínas plasmáticas abundantes (figura 13) y se buscaron posibles PSP. Se descubrieron 22 genes para los cuales un PSP era lo suficientemente largo y antigénico (figura 18). Después se buscaron secuencias de proteínas más largas putativas inducidas (figura 18b-c). Después se buscaron nuevos péptidos putativos en el plasma obtenido de individuos normales y cancerosos (figura 18). Se seleccionaron 3 de estas 22 proteínas interesantes: APOAI, APOAII y APOCII. Basándose en los anteriores análisis, se espera encontrar proteínas más largas en el plasma de los pacientes cancerosos (13, 16 y 17 AA más largas). Se predice que estas secuencias peptídicas de infidelidad de la transcripción serán inmunogénicas, y los anticuerpos dirigidos contra estas nuevas secuencias representan ligandos específicos para medir la infidelidad de la transcripción (figura 14). Puesto que el análisis de Kyte-Doolittle indica que estas secuencias no son hidrófobas, se espera que estas tres nuevas proteínas se segreguen hacia la circulación (véase a continuación).

Identificación del péptido de post-fin (PSP) de ApoII y ApoCII debido a la sustitución en el codón de fin canónico

Los PSP que resultan de las sustituciones de bases en el codón de fin canónico pueden identificarse y caracterizarse de la siguiente manera. Se preparan anticuerpos policlonales de conejo que reconocen una porción o

porciones inmunogénicas del PSP en cuestión. Estos anticuerpos antipéptido pueden comprobarse mediante un análisis de la transferencia de puntos utilizando el péptido purificado para verificar que, en efecto, reconocen el PSP. Entonces puede realizarse un análisis de la transferencia Western con muestras de plasma obtenidas de pacientes cancerosos utilizando anticuerpos dirigidos contra el PSP. La figura 17a (panel derecho) demuestra que el anticuerpo anti-PSP de ApoAII reconoce una banda en las transferencias Western realizadas con muestras de plasma obtenidas de pacientes con cáncer de próstata, que no se observa cuando se emplea suero preinmunológico de conejo como control negativo (figura 17a, panel izquierdo). La banda de PSP de ApoAII también se observa en pacientes con cáncer en la fase metastásica (figura 17b, panel derecho). Esta banda tiene una masa molecular ligeramente mayor (11,4 kDa) comparada con la de la forma de monómero nativa de la ApoII (figura 17a, panel central, figura 17b, panel izquierdo). Esta masa molecular se corresponde con la predicha basada en el secuencia peptídica adicional. También pueden prepararse geles bidimensionales para caracterizar aún más esta banda.

Pueden realizarse experimentos de cromatografía de afinidad para aislar la forma de PSP de ApoAII utilizando el anticuerpo anti-PSP (figura 17e). El anticuerpo anti-PSP se inmoviliza sobre esferas de matriz y esta columna se incuba en presencia de plasma o HDL deslipidado, y después se lava secuencialmente para eliminar las proteínas unidas no específicamente y, por último, se eluye con detergente o reactivos caotrópicos. La fracción eluida se analiza mediante transferencia Western utilizando los anticuerpos anti-PSP y los anticuerpos comerciales anti-ApoAII. Dos bandas a 9 kDa y 11,4 kDa son reconocidas por el anticuerpo comercial anti-ApoAII, mientras que sólo la banda de 11,4 kDa es reconocida por el anticuerpo anti-PSP. Por tanto, la banda de 9 kDa se corresponde con la forma nativa de ApoAII y la de 11,4 kDa se corresponde con la forma de PSP de ApoAII. La presencia de ApoAII nativa en la fracción eluida sugiere que la proteína PSP de ApoAII puede formar dímeros con la proteína ApoAII nativa bajo condiciones no reductoras nativas. En efecto, la ApoAII existe normalmente en el plasma como un dímero de dos monómeros unidos mediante un único puente disulfuro.

La ApoAII se localiza principalmente en la fracción de HDL del plasma, que puede aislarse mediante ultracentrifugación secuencial. Unos análisis de la transferencia Western demuestran que la PSP de ApoAII no se detecta en la fracción de d < 1,07 g/ml que contiene VLDL y LDL, sino en la fracción de d > 1,07 g/ml que contiene HDL y proteínas plasmáticas (figura 17c). Después de más etapas de ultracentrifugación para purificar y lavar el HDL (d = 1,07-1,21 g/ml), la transferencia Western revela que la PSP de ApoAII permanece asociada con la fracción HDL en lugar de con la fracción d > 1,21 g/ml, también denominada suero deficiente en lipoproteínas (LPDS). Una cantidad correspondiente del plasma demuestra, en la transferencia Western, que esto no es debido a una dilución de la fracción de d > 1,21 g/ml durante las etapas de purificación. La separación de las lipoproteínas del plasma mediante una filtración en gel utilizando Superose 6B (Amersham, GE Healthcare) también demuestra que la PSP de ApoAII eluye de una manera similar a la de la ApoAII asociada con HDL.

La asociación de PSP de ApoAII con HDL permite la purificación de la forma de PSP de ApoAII de una manera similar a la de la ApoAII. En primer lugar, el HDL se deslipida para eliminar todos los lípidos. El sedimento de proteínas exento de lípidos resultante entonces se resuspende en tampón Tris 10 mM que contiene urea o guanidina en presencia o en ausencia de un agente reductor, y se aplica a una columna de filtración en gel (por ejemplo, Superdex 200, Amersham, GE Healthcare). Un análisis de la transferencia Western (figura 17d) demuestra que la forma de PSP de ApoAII sigue estando presente en HDL deslipidado. Se logró una mayor purificación mediante una electroforesis preparativa (por ejemplo, DE52). La forma de PSP de ApoAII fue rastreada mediante un análisis de la transferencia Western. La forma de PSP de ApoAII purificada entonces se rompió de forma enzimática (tripsina), y los péptidos resultantes se analizan en MS-MS para la secuenciación completa de los AA. Los resultados demuestran que el codón de fin canónico está reemplazado preferentemente por arginina, con una sustitución (U a C) o (U a A), seguido de serina, valina, ácido glutámico, treonina, isoleucina, valina, fenilalanina, ácido glutámico, prolina, ácido glutámico, leucina, alanina, serina, arginina. Esta es la secuencia exacta de aminoácidos que se predice que aparecerá tras pasar por encima del codón de fin canónico de ApoAII. Por tanto, el codón de fin canónico UGA de ApoAII se sustituye por AGA o CGA que conduce a una arginina. Una hipótesis aún inexplorada es que UGA se convierte en GGA, que conduce a la glicina, pero la detección de este variante mediante espectrometría de masas todavía se encuentra fuera de los límites tecnológicos actuales.

Otro ejemplo se ilustra con el PSP de ApoCII. La figura 17f muestra un experimento similar al de la figura 17a, con la excepción de que los análisis de la transferencia Western se realizan con anticuerpos comerciales anti-ApoCII y con anticuerpos anti-PSP de ApoCII. Puede seguirse un procedimiento similar al del PSP de ApoAII para ApoCII. Sin embargo, ApoCII es menos abundante en el plasma. Por tanto, se requiere una cantidad mucho mayor de plasma que contenga PSP de ApoCII para obtener una cantidad suficiente para el análisis con MS-MS.

Conclusiones

Los inventores describen en la presente un nuevo mecanismo que conduce a unas sustituciones de bases que se producen en su mayor parte en el extremo 3' de la secuencia codificadora y la región no traducida del ARNm. Estas sustituciones de bases conducen a cambios en la secuencia de AA de proteínas debido a cambios en la identidad de los AA, la introducción de codones de fin prematuros, así como la modificación de los codones de fin naturales que resultan en la introducción de nuevas regiones codificadoras. Este fenómeno de infidelidad de la transcripción también puede afectar al mundo del ARNnc y alterar la regulación mediada por estos ARN. Se produce en la mayoría de los genes con una tasa que excede cualquier fenómeno descrito que conduzca a una mutación en el

ADN. La infidelidad de la transcripción aumenta muchísimo en células cancerosas. La infidelidad de la transcripción proporciona un nuevo paradigma para comprender la patología del cáncer, la gravedad de la enfermedad y el avance de la enfermedad. Tiene implicaciones importantes no sólo en el diseño de nuevos experimentos transcriptómicos y proteómicos, sino también para el desarrollo de diagnósticos y productos terapéuticos específicos.

LISTA DE SECUENCIAS

<110> TRANSMEDI SA

<120> Infidelidad de la transcripción, su detección y sus usos

<130> B478EPPC

<160> 32

<170> Patent In versión 3.3

<210> 1
<211> 13
<212> PRT
<213> Homo sapiens

<400> 1
Gln Met Trp Gln Leu Phe Trp Ile Tyr His Leu Ser Ser
1 5 10

<210> 2
<211> 14
<212> PRT
<213> Homo sapiens

<400> 2
Lys Leu His Thr Leu Ser Ala Ala Ile Tyr Tyr Gln Gln Glu
1 5 10

<210> 3
<211> 6
<212> PRT
<213> Homo sapiens

<400> 3
Asp Phe Leu Ser Asn Lys
1 5

<210> 4
<211> 12
<212> PRT
<213> Homo sapiens

<400> 4
Met Tyr Thr Val Glu Phe Ser Val His Lys Asn Asn
1 5 10

<210> 5
<211> 12
<212> PRT
<213> Homo sapiens

<400> 5
Asn Gly Ser Leu Gly Asp Met Ser Asp Leu Cys Thr
1 5 10

<210> 6
<211> 3
<212> PRT
<213> Homo sapiens

<400> 6

Ala Ser Gly
1

<210> 7
<211> 8
<212> PRT
<213> Homo sapiens

<400> 7
Glu Pro Ser Glu Pro Ser Asp Phe
1 5

<210> 8
<211> 54
<212> PRT
<213> Homo sapiens

<400> 8
Ala Pro Ser Ile Phe Pro Thr Leu Pro Ala Lys Pro Gly Thr Lys Gln
1 5 10 15

Pro Arg Ser Pro Val Thr Ala Leu Ser Leu His Met Leu Leu Met Val
20 25 30

Ser Ser Ala Pro Ser Cys Gly Leu Ile Gln Thr Val Ser Ser Phe Thr
35 40 45

Val Tyr Ile Phe Thr Leu
50

<210> 9
<211> 69
<212> PRT
<213> Homo sapiens

<400> 9
Ala Arg His Gly Arg Asp Glu Glu Val Trp His Arg Lys His Ser His

1 5 10 15

His Phe Val Gln Ala Trp Ala Trp Val Gly Gly Leu Val Cys Trp Pro
20 25 30

Arg Lys Cys His Met Arg Ser Thr Leu Ile Ser Ser Leu Asp Ser Leu
35 40 45

Leu Pro Val Ile Pro His Arg Thr Glu Ala Glu Trp Val Val Val Met
50 55 60

Phe Asp Arg Arg His
65

<210> 10
<211> 26
<212> PRT
<213> Homo sapiens

<400> 10

Arg Ser Lys Ala Tyr Ser Ser Val Phe Leu Phe Arg Trp Cys Lys Ala
1 5 10 15

Asn Thr Leu Ser Lys Lys His Lys Phe Leu
20 25

<210> 11
<211> 14
<212> PRT
<213> Homo sapiens

<400> 11
Gly Leu Asp Ser Thr Arg Ala Leu Glu Asn Glu Met Thr Val
1 5 10

<210> 12
<211> 12
<212> PRT
<213> Homo sapiens

<400> 12
Gly Ala Arg Arg Arg Pro Pro Ser Arg Cys Ser Glu
1 5 10

<210> 13
<211> 20
<212> PRT
<213> Homo sapiens

<400> 13
Ser Val Gln Thr Ile Val Phe Gln Pro Gln Leu Ala Ser Arg Thr Pro
1 5 10 15

Thr Gly Gln Ser
20

<210> 14
<211> 16
<212> PRT
<213> Homo sapiens

<400> 14
Ile Val Phe Gln Pro Gln Leu Ala Ser Arg Thr Pro Thr Gly Gln Ser
1 5 10 15

<210> 15
<211> 22
<212> PRT
<213> Homo sapiens

<400> 15
Gln Pro Asp Pro Pro Ser Val Asp Lys Gly Arg Val Pro Tyr Ser Pro
1 5 10 15

Asp Pro Pro Gly Ser Asp
20

<210> 16
<211> 15
<212> PRT
<213> Homo sapiens

<400> 16

Asp Pro Pro Ser Val Asp Lys Gly Arg Val Pro Tyr Ser Pro Asp
1 5 10 15

<210> 17
<211> 54
<212> PRT
<213> Homo sapiens

<400> 17
Asp Leu Asn Thr Pro Ser Pro Pro Ala Tyr Pro Ser Cys Glu Leu Leu
1 5 10 15

Gly Ser Cys Asn Leu Gln Gly Cys Pro Cys Arg Leu Leu Lys Arg Asp
20 25 30

Ser Ile Leu Ser Ala Leu Leu Pro His Leu Met Pro Gly Pro Pro Pro
35 40 45

Gly Met Leu Ala Ser Gln
50

<210> 18
<211> 36
<212> PRT
<213> Homo sapiens

<400> 18
Pro Gly Ser Thr Gly Arg Leu His Pro Leu His Val Thr Ser Ala Ser
1 5 10 15

Leu Ser Pro Thr Pro Pro Pro Pro His Lys Asp Lys Pro Ile Asn His
20 25 30

Asp Lys Gly Ser
35

<210> 19
<211> 37
<212> PRT
<213> Homo sapiens

<400> 19
Thr Pro Lys Pro Ala Ala Met Arg Pro His Ala Thr Pro Cys Leu Leu
1 5 10 15

Pro Pro Arg Ser Leu Gln Arg Glu Thr Leu Ser Pro Pro Gln Pro Ser
20 25 30

Ser Trp Gly Gly Pro
35

<210> 20
<211> 13
<212> PRT
<213> Homo sapiens

<400> 20

Glu Ala Arg Val Gly Gly Asn Val Gly Ser Gln Thr Gln
1 5 10

<210> 21
<211> 29
<212> PRT
<213> Homo sapiens

<400> 21
Pro Ser Val Leu His Thr Ala Arg Gly Pro Arg Met Pro Arg Pro Pro
1 5 10 15

Leu Ala Pro Ala Gly Arg Glu Pro Asp His Leu Pro Cys
20 25

<210> 22
<211> 72
<212> PRT
<213> Homo sapiens

<400> 22
Asp Val Asp Val Ala Phe Ala Pro Thr Gly Ala Ser Glu Ser Ser Ser
1 5 10 15

Pro Gln Asp Glu Leu Gln Pro Pro Arg Glu Ser Ser Ala Arg His Gln
20 25 30

Val Thr Arg Pro Gln Pro Pro Gly Pro Gln Leu Arg Pro Ala Ser Pro
35 40 45

Arg Ser Gly Ser Cys Thr Leu Thr Leu Asp Ser Ala Ala His Gly Lys
50 55 60

Asn Arg Ile Ala Pro Ala Cys Asn
65 70

<210> 23
<211> 14
<212> PRT
<213> Homo sapiens

<400> 23
Asn Val Ile Pro Leu Lys Arg Lys Met Asn Asn Thr Leu Asn
1 5 10

<210> 24
<211> 39
<212> PRT
<213> Homo sapiens

<400> 24

Thr Pro Ala Ala Arg Leu Met Trp Ser Ser Asn Met Pro Tyr Phe Ala

1 5 10 15

Gln Lys Thr Ala Lys Asp Met Thr Ser Ser Trp Leu Gln Pro Arg Phe
20 25 30

Ile Phe Leu Phe Val Val Asn
35

<210> 25
<211> 53
<212> PRT
<213> Homo sapiens

<400> 25
Gly Trp Gly Val Phe Leu Leu Asn Pro Met Ala Gly Gly His Ala Pro
1 5 10 15

Thr Ile Ile Ser Trp Glu Glu Arg Gln Ser Trp Glu Ile Asp Gly Ser
20 25 30

His Ser Ser Leu Leu Ser Leu Leu Cys Leu Trp Ala Thr Leu Pro Thr
35 40 45

Pro Leu Leu Ser Gln
50

<210> 26
<211> 27
<212> PRT
<213> Homo sapiens

<400> 26
Thr Gly Pro Thr His His Ser Pro Ser Pro Ser Ile Ser Thr Trp Cys
1 5 10 15

Leu Val Pro Val His Ser Val Asn Lys Lys Pro
20 25

<210> 27
<211> 25
<212> PRT
<213> Homo sapiens

<400> 27
Trp Thr Pro Glu Pro Leu Leu Gln Pro Leu Ser His Pro Leu Pro Pro
1 5 10 15

Ala His Pro Leu Gly Gln Gln Arg Leu

20 25

<210> 28
<211> 20
<212> PRT
<213> Homo sapiens

<400> 28

```
Leu Asp Gly Arg Gln Ser Asp Ala Leu Thr His Leu Glu Ala Gly Thr
1               5                  10                  15

Trp Val Gly Ile
            20
```

<210> 29
<211> 71
<212> PRT
<213> Homo sapiens

<400> 29

```
   Ser Leu Pro Ser Ser Ser Gly Ala Leu Ser Lys Glu Leu Gly Met Gln
   1               5                  10                  15

   Ala Gly Cys Leu Gly Leu Trp Ala Gln Pro Gly Pro Cys Ala Pro Ser
               20                  25                  30

   Gly His Gly Met Cys Gly Pro Val Cys Leu Ser Leu Glu Gly Asp Ser
           35                  40                  45

   Asp Ser Leu Cys Ser Ser His Met His Arg Gly Pro Trp Thr Leu Gln
       50                  55                  60

   Ser Gly Gly Ser Trp Ala Ser
   65                  70
```

<210> 30
<211> 48
<212> PRT
<213> Homo sapiens

<400> 30

```
  Asn Leu Arg Gly Arg Ala Ala Thr Lys Val Lys Met Gly Thr Gln Met
  1               5                  10                  15

  Ile His Glu Phe Ala Leu Val Ser Leu Ala Gln Val Val Cys Ala Asn
              20                  25                  30

  His Val Cys Leu His Ser Ser Val Leu Pro Cys Val Leu Asn Lys Lys
          35                  40                  45
```

<210> 31
<211> 100
<212> PRT
<213> Homo sapiens

<400> 31

```
Gly Pro Ala Pro Pro Arg Pro Ala Pro Ala Gly Pro Ala Pro Pro Arg
1               5                  10                  15

Pro Ala Pro Ala Ala Leu Pro Met Gly Ala Val Phe Lys Asp Thr Arg
            20                  25                  30

Ala Pro Ser Pro Pro Gly Ala Pro Leu Lys Met Glu Arg Gly Leu Arg
        35                  40                  45

Ile Ser Val Ser Leu Gly Ala Cys Leu Gly Ser Pro Ser Leu Thr Phe
    50                  55                  60

Pro His Ser His Ser Leu Ser Leu Pro Leu Cys Leu Leu Leu Pro Val
65                  70                  75                  80

Cys Thr Ile Pro Leu Pro Gly Ile Lys Ala Gln Gly Thr Ser Gly Glu
                85                  90                  95

His Tyr Cys Ser
            100
```

<210> 32
<211> 16
<212> PRT
<213> Homo sapiens

<400> 32

```
Gly Thr Ser Pro Pro Val Asp Leu Lys Asp Glu Gly Trp Asp Phe Met
1               5                  10                  15
```

REIVINDICACIONES

1. Uso un péptido sintético con una longitud de 100 aminoácidos o menos, que comprende una secuencia seleccionada de SEQ ID NOs: 1 a 5, 7 a 13, 15 a 18 y 20 a 32 para detectar o controlar trastornos de células proliferativas.

2. El uso de un péptido de la reivindicación 1, para detectar o controlar trastornos de células proliferativas, donde el péptido es inmunogénico.

3. Una composición de vacuna que comprende un péptido de la reivindicación 1, y opcionalmente un vehículo, excipiente y/o adyuvante adecuado.

4. Una molécula de ácido nucleico sintética que codifica un péptido de 100 aminoácidos o menos que comprende una secuencia seleccionada de SEQ ID NOs: 1 a 5, 7 a 13, 15 a 18 y 20 a 32 para detectar o controlar trastornos de células proliferativas.

5. Un método para producir un anticuerpo, comprendiendo el método inmunizar a un mamífero no humano con un péptido de la reivindicación 1 y recuperar anticuerpos que se unen a dicho péptido, o las correspondientes células productoras de anticuerpos.

6. Un anticuerpo, o derivado del mismo, que se une específicamente a un péptido de la reivindicación 1.

7. Un anticuerpo, o derivado del mismo, según la reivindicación 6, donde dicho anticuerpo es el conjugado de una molécula, tal como un fármaco, un marcador, una molécula tóxica o un radioisótopo.

8. El anticuerpo conjugado, o derivado del mismo, de la reivindicación 7, para usar como medicamento o reactivo de diagnóstico.

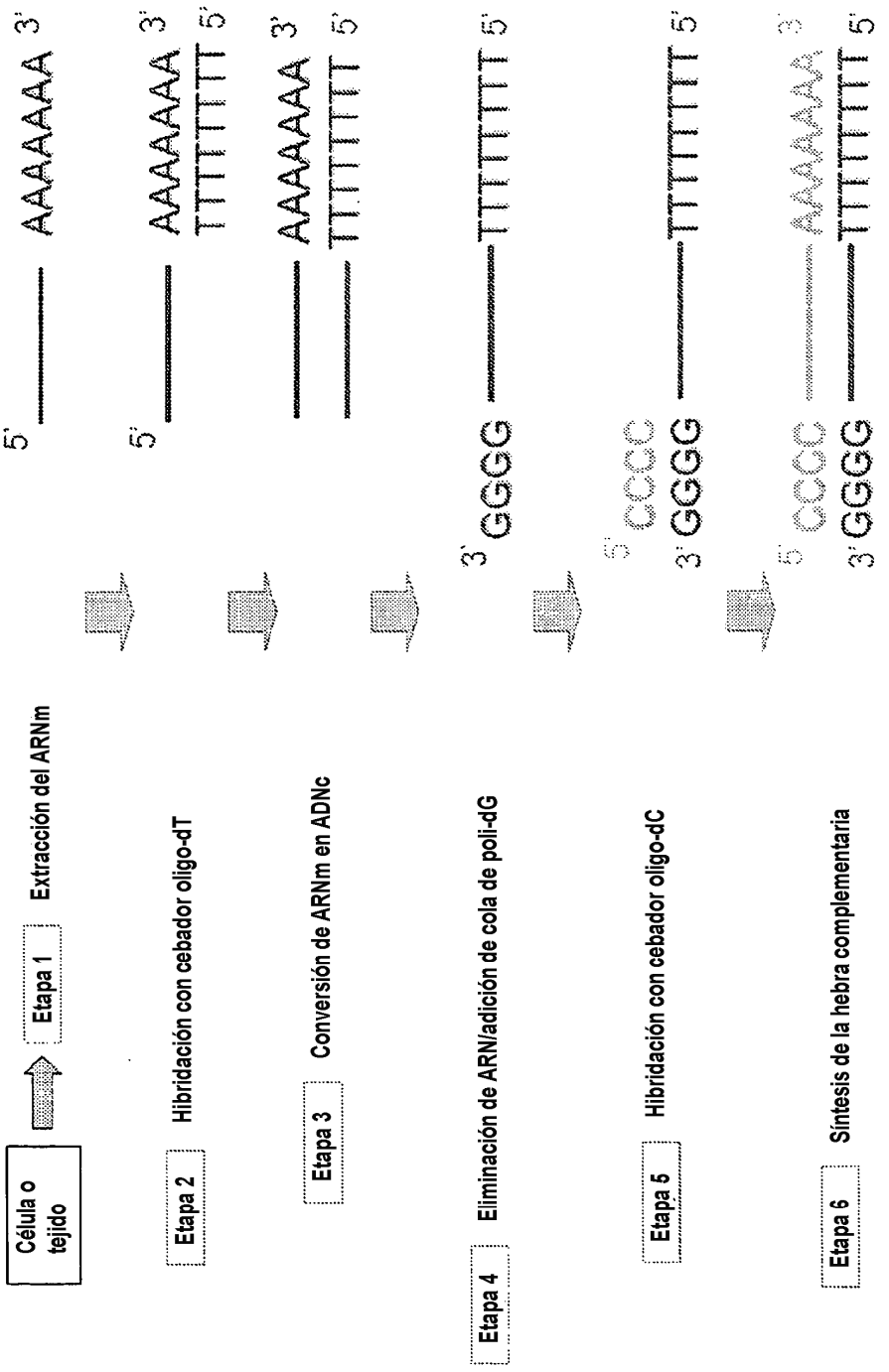

Célula o tejido
Etapa 1
Extracción del ARNm
Etapa 2
Hibridación con cebador oligo-dT
Etapa 3
Conversión de ARNm en ADNc
Etapa 4
Eliminación de ARN/adición de cola de poli-dG
Etapa 5
Hibridación con cebador oligo-dC
Etapa 6
Síntesis de la hebra complementaria
5' AAAAAAA 3'
5' AAAAAAA 3'
TTTTTTT 5'
AAAAAAA 3'
TTTTTTT 5'
3' GGGG TTTTTTT 5'
5' CCCC
3' GGGG TTTTTTT 5'
5' CCCC AAAAAAA 3'
3' GGGG TTTTTTT 5'

Figura 1a

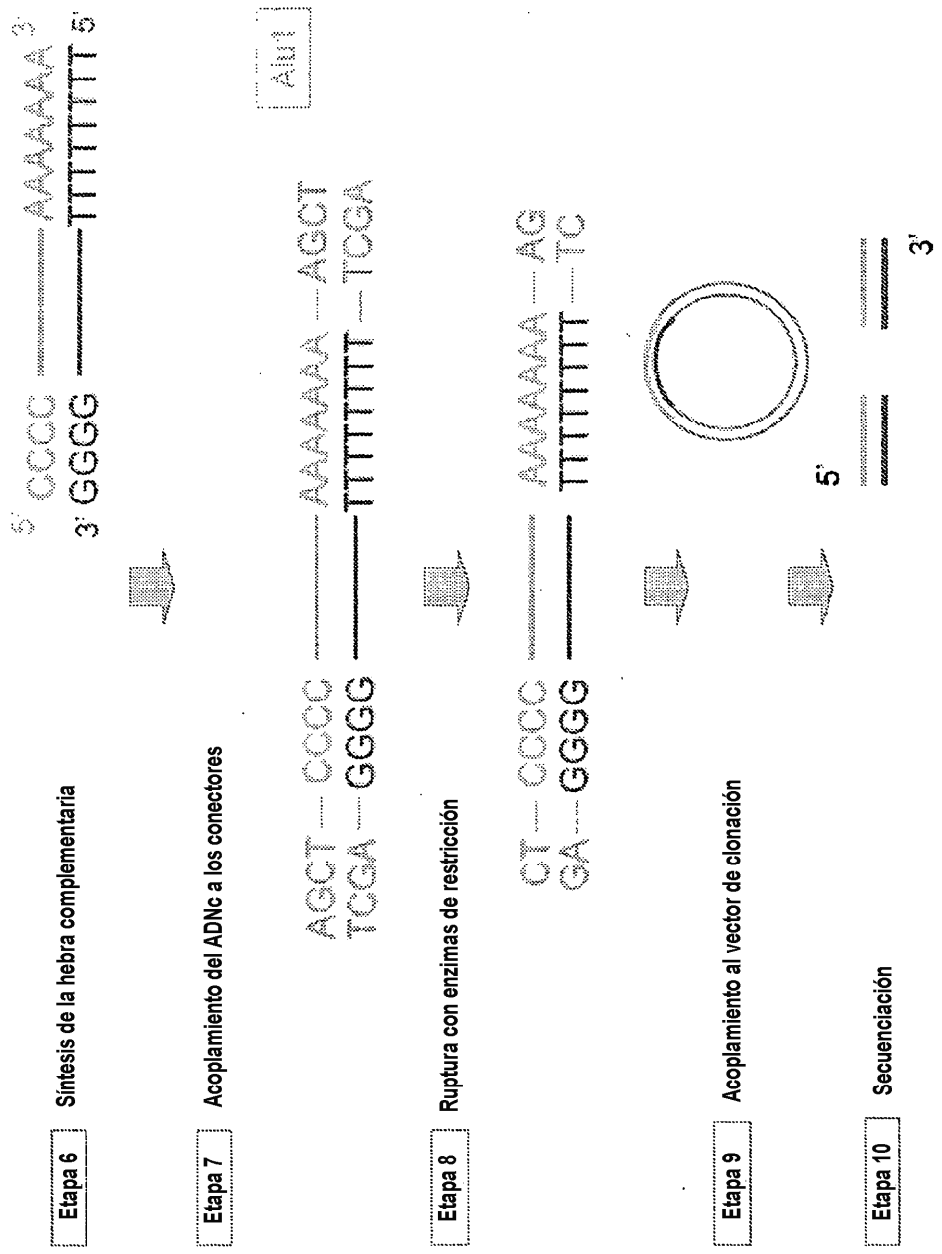

Etapa 6
Síntesis de la hebra complementaria
5' CCCC ——— AAAAAAA 3'
3' GGGG ——— TTTTTTTT 5'
Etapa 7
Acoplamiento del ADNc a los conectores
AscI
AGCT — CCCC ——— AAAAAAA — AGCT
TCGA — GGGG ——— TTTTTTTT — TCGA
Etapa 8
Ruptura con enzimas de restricción
CT — CCCC ——— AAAAAAA — AG
GA — GGGG ——— TTTTTTTT — TC
Etapa 9
Acoplamiento al vector de clonación
Etapa 10
Secuenciación
5'
3'

Figura 1b

>gi|16507965|ref|NM_001428.2| Enolasa 1 de *Homo sapiens*, (alfa) (ENO1), ARNm

TAGCTAGGCAGGAAGTCGGCGCGGGCGGCGCGGACAGTATCTGTGGGTACCCGGAGCACGGAGATCTCGCCGGCTTTACGTTCACCTCGG
TGTCTGCAGCACCCTCCGCTTCCTCTCCTAGGCGACGAGACCCAGTGGCTAGAAGTTCACCATGTCTATTCTCAAGATCCATGCCAGGGA
GATCTTTGACTCTCGCGGGAATCCCACTGTTGAGGTTGATCTCTTCACCTCAAAAGGTCTCTTCAGAGCTGCTGTGCCCAGTGGTGCTTC
AACTGGTATCTATGAGGCCCTAGAGCTCCGGGACAATGATAAGACTCGCTATATGGGGAAGGGTGTCTCAAAGGCTGTTGAGCACATCAA
TAAAACTATTGCGCCTGCCCTGGTTAGCAAGAAACTGAACGTCACAGAACAAGAGAAGATTGACAAACTGATGATCGAGATGGATGGAAC
AGAAAATAAATCTAAGTTTGGTGCGAACGCCATTCTGGGGGTGTCCCTTGCCGTCTGCAAAGCTGGTGCCGTTGAGAAGGGGGTCCCCCT
GTACCGCCACATCGCTGACTTGGCTGGCAACTCTGAAGTCATCCTGCCAGTCCCGGCGTTCAATGTCATCAATGGCGGTTCTCATGCTGG
CAACAAGCTGGCCATGCAGGAGTTCATGATCCTCCCAGTCGGTGCAGCAAACTTCAGGGAAGCCATGCGCATTGGAGCAGAGGTTTACCA
CAACCTGAAGAATGTCATCAAGGAGAAATATGGGAAAGATGCCACCAATGTGGGGGATGAAGGCGGGTTTGCTCCCAACATCCTGGAGAA
TAAAGAAGGCCTGGAGCTGCTGAAGACTGCTATTGGGAAAGCTGGCTACACTGATAAGGTGGTCATCGGCATGGACGTAGCGGCCTCCGA
GTTCTTCAGGTCTGGGAAGTATGACCTGGACTTCAAGTCTCCCGATGACCCCAGCAGGTACATCTCGCCTGACCAGCTGGCTGACCTGTA
CAAGTCCTTCATCAAGGACTACCCAGTGGTGTCTATCGAAGATCCCTTTGACCAGGATGACTGGGGAGCTTGGCAGAAGTTCACAGCCAG
TGCAGGAATCCAGGTAGTGGGGGATGATCTCACAGTGACCAACCCAAAGAGGATCGCCAAGGCCGTGAACGAGAAGTCCTGCAACTGCCT
CCTGCTCAAAGTCAACCAGATTGGCTCCGTGACCGAGTCTCTTCAGGCGTGCAAGCTGGCCCAGGCCAATGGTTGGGGCGTCATGGTGTC
TCATCGTTCGGGGGAGACTGAAGATACCTTCATCGCTGACCTGGTTGTGGGGCTGTGCACTGGGCAGATCAAGACTGGTGCCCCTTGCCG
ATCTGAGCGCTTGGCCAAGTACAACCAGCTCCTCAGAATTGAAGAGGAGCTGGGCAGCAAGGCTAAGTTTGCCGGCAGGAACTTCAGAAA
CCCCTTGGCCAAGTAAGCTGTGGGCAGGCAAGCCCTTCGGTCACCTGTTGGCTACACAGACCCCTCCCCTCGTGTCAGCTCAGGCAGCTC
GAGGCCCCCGACCAACACTTGCAGGGGTCCCTGCTAGTTAGCGCCCCACCGCCGTGGAGTTCGTACCGCTTCCTTAGAACTTCTACAGAA
GCCAAGCTCCCTGGAGCCCTGTTGGCAGCTCTAGCTTTGCAGTCGTGTAATTGGCCCAAGTCATTGTTTTTCTCGCCTCACTTTCCACCA
AGTGTCTAGAGTCATGTGAGCCTCGTGTCATCTCCGGGGTGGCCACAGGCTAGATCCCCGGTGGTTTTGTGCTCAAAATAAAAAGCCTCA
GTGACCCATGAG

Figura 2a

>gi|56682958|ref|NM_002032.2| Ferritina de *Homo sapiens*, polipéptido pesado 1 (FTH1), ARNm

ATAAGAGACCACAAGCGACCCGCAGGGCCAGACGTTCTTCGCCGAGAGTCGTCGGGGTTTCCTGCTTCAACAGTGCTTGGACGGAAC
CCGGCGCTCGTTCCCCACCCCGGCCGGCCGCCCATAGCCAGCCCTCCGTCACCTCTTCACCGCACCCTCGGACTGCCCCAAGGCCCC
CGCCGCCGCTCCAGCGCCGCGCAGCCACCGCCGCCGCCGCCGCCTCTCCTTAGTCGCCGCCATGACGACCGCGTCCACCTCGCAGGT
GCGCCAGAACTACCACCAGGACTCAGAGGCCGCCATCAACCGCCAGATCAACCTGGAGCTCTACGCCTCCTACGTTTACCTGTCCAT
GTCTTACTACTTTGACCGCGATGATGTGGCTTTGAAGAACTTTGCCAAATACTTTCTTCACCAATCTCATGAGGAGAGGGAACATGC
TGAGAAACTGATGAAGCTGCAGAACCAACGAGGTGGCCGAATCTTCCTTCAGGATATCAAGAAACCAGACTGTGATGACTGGGAGAG
CGGGCTGAATGCAATGGAGTGTGCATTACATTTGGAAAAAAATGTGAATCAGTCACTACTGGAACTGCACAAACTGGCCACTGACAA
AAATGACCCCCATTTGTGTGACTTCATTGAGACACATTACCTGAATGAGCAGGTGAAAGCCATCAAAGAATTGGGTGACCACGTGAC
CAACTTGCGCAAGATGGGAGCGCCCGAATCTGGCTTGGCGGAATATCTCTTTGACAAGCACACCCTGGGAGACAGTGATAATGAAAG
CTAAGCCTCGGGCTAATTTCCCCATAGCCGTGGGGTGACTTCCCTGGTCACCAAGGCAGTGCATGCATGTTGGGGTTTCCTTTACCT
TTTCTATAAGTTGTACCAAAACATCCACTTAAGTTCTTTGATTTGTACCATTCCTTCAAATAAAGAAATTTGGTACCCAGGTGTTGT
CTTTGAGGTCTTGGGATGAATCAGAAATCTATCCAGGCTATCTTCCAGATTCCTTAAGTGCCGTTGTTCAGTTCTAATCACACTAAT
CAAAAAGAAACGAGTATTTGTATTTATTAAACTCATTAGTTTGGGCAGTATACTAAGGTGTGGCTGTCTTGGATTCAGATAGAACTA
AGGGTTCCCGACTCTGAATCCAGAGTCTGAGTTAAATGTTCCAATGGTTCAGTCTAGCTTTCACAGTTTTTATGAATAAAAGGCAT
TAAAGGCTGA

Figura 2b

>gi|56682960|ref|NM_000146.3| Ferritina de *Homo sapiens*, polipéptido ligero (FTL), ARNm

GCAGTTCGGCGGTCCCGCGGGTCTGTCTCTTGCTTCAACAGTGTTTGGACGGAACAGATCCGGGGACTCTCTTCCAGCCTCCGACCGC
CCTCCGATTTCCTCTCCGCTTGCAACCTCCGGGACCATCTTCTCGGCCATCTCCTGCTTCTGGGACCTGCCAGCACCGTTTTTGTGGT
TAGCTCCTTCTTGCCAACCAACCATGAGCTCCCAGATTCGTCAGAATTATTCCACCGACGTGGAGGCAGCCGTCAACAGCCTGGTCAA
TTTGTACCTGCAGGCCTCCTACACCTACCTCTCTCTGGGCTTCTATTTCGACCGCGATGATGTGGCTCTGGAAGGCGTGAGCCACTTC
TTCCGCGAATTGGCCGAGGAGAAGCGCGAGGGCTACGAGCGTCTCCTGAAGATGCAAAACCAGCGTGGCGGCCGCGCTCTCTTCCAGG
ACATCAAGAAGCCAGCTGAAGATGAGTGGGGTAAAACCCCAGACGCCATGAAAGCTGCCATGGCCCTGGAGAAAAAGCTGAACCAGGC
CCTTTTGGATCTTCATGCCCTGGGTTCTGCCCGCACGGACCCCCATCTCTGTGACTTCCTGGAGACTCACTTCCTAGATGAGGAAGTG
AAGCTTATCAAGAAGATGGGTGACCACCTGACCAACCTCCACAGGCTGGGTGGCCCGGAGGCTGGGCTGGGCGAGTATCTCTTCGAAA
GGCTCACTCTCAAGCACGACTAAGAGCCTTCTGAGCCCAGCGACTTCTGAAGGCCCCTTGCAAAGTAATAGGGCTTCTGCCTAAGCC
TCTCCCTCCAGCCAATAGGCAGCTTTCTTAACTATCCTAACAAGCCTTGGACCAAATGGAAATAAAGCTTTTTGATGC

Figura 2c

ES 2 553 435 T3

>gi|83641890|ref|NM_002046.3| Gliceraldehído-3-fosfato deshidrogenasa de *Homo sapiens* (GAPDH), ARNm

AAATTGAGCCCGCAGCCTCCCGCTTCGCTCTCTGCTCCTCCTGTTCGACAGTCAGCCGCATCTTCTTTTGCGTCGCCAGCCGAGCCACA
TCGCTCAGACACCATGGGGAAGGTGAAGGTCGGAGTCAACGGATTTGGTCGTATTGGGCGCCTGGTCACCAGGGCTGCTTTTAACTCTG
GTAAAGTGGATATTGTTGCCATCAATGACCCCTTCATTGACCTCAACTACATGGTTTACATGTTCCAATATGATTCCACCCATGGCAAA
TTCCATGGCACCGTCAAGGCTGAGAACGGGAAGCTTGTCATCAATGGAAATCCCATCACCATCTTCCAGGAGCGAGATCCCTCCAAAAT
CAAGTGGGGCGATGCTGGCGCTGAGTACGTCGTGGAGTCCACTGGCGTCTTCACCACCATGGAGAAGGCTGGGGCTCATTTGCAGGGG
GAGCCAAAAGGGTCATCATCTCTGCCCCCTCTGCTGATGCCCCCATGTTCGTCATGGGTGTGAACCATGAGAAGTATGACAACAGCCTC
AAGATCATCAGCAATGCCTCCTGCACCACCAACTGCTTAGCACCCCTGGCCAAGGTCATCCATGACAACTTTGGTATCGTGGAAGGACT
CATGACCACAGTCCATGCCATCACTGCCACCCAGAAGACTGTGGATGGCCCCTCCGGGAAACTGTGGCGTGATGGCCGCGGGGCTCTCC
AGAACATCATCCCTGCCTCTACTGGCGCTGCCAAGGCTGTGGGCAAGGTCATCCCTGAGCTGAACGGGAAGCTCACTGGCATGGCCTTC
CGTGTCCCCACTGCCAACGTGTCAGTGGTGGACCTGACCTGCCGTCTAGAAAAACCTGCCAAATATGATGACATCAAGAAGGTGGTGAA
GCAGGCGTCGGAGGGCCCCTCAAGGGCATCCTGGGCTACACTGAGCACCAGGTGGTCTCCTCTGACTTCAACAGCGACACCCACTCCT
CCACCTTTGACGCTGGGGCTGGCATTGCCCTCAACGACCACTTTGTCAAGCTCATTTCCTGGTATGACAACGAATTTGGCTACAGCAAC
AGGGTGGTGGACCTCATGGCCCACATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAAGAGAGAG
ACCCTCACTGCTGGGGAGTCCCTGCCACACTCAGTCCCCCACCACACTGAATCTCCCCTCCTCACAGTTGCCATGTAGACCCCTTGAAG
AGGGGAGGGGCCTAGGGAGCCGCACCTTGTCATGTACCATCAATAAAGTACCCTGTGCTCAACC

Figura 2d

ES 2 553 435 T3

>gi|24234684|ref|NM_006597.3| Proteína 8 de choque térmico de 70 kDa de *Homo sapiens* (HSPA8), variante de transcripción 1, ARNm

CTCATTGAACTCGCCTGCAGCTCTTGGGTTTTTTGTGGCTTCCTTCGTTATTGGAGCCAGGCCTACACCCCAGCAACCATGTCCAAGGGAC
CTGCAGTTGGTATTGATCTTGGCACCACCTACTCTTGTGTGGGTGTTTTCCAGCACGGAAAAGTCGAGATAATTGCCAATGATCAGGGAAA
CCGAACCACTCCAAGCTATGTCGCCTTTACGGACACTGAACGGTTGATCGGTGATGCCGCAAAGAATCAAGTTGCAATGAACCCCACCAAC
ACAGTTTTTGATGCCAAACGTCTGATTGGACGCAGATTTGATGATGCTGTTGTCCAGTCTGATATGAAACATTGGCCCTTTATGGTGGTGA
ATGATGCTGGCAGGCCCAAGGTCCAAGTAGAATACAAGGGAGAGACCAAAAGCTTCTATCCAGAGGAGGTGTCTTCTATGGTTCTGACAAA
GATGAAGGAAATTGCAGAAGCCTACCTTGGGAAGACTGTTACCAATGCTGTGGTCACAGTGCCAGCTTACTTTAATGACTCTCAGCGTCAG
GCTACCAAAGATGCTGGAACTATTGCTGGTCTCAATGTACTTAGAATTATTAATGAGCCAACTGCTGCTGCTATTGCTTACGGCTTAGACA
AAAAGGTTGGAGCAGAAAGAAACGTGCTCATCTTTGACCTGGGAGGTGGCACTTTTGATGTGTCAATCCTCACTATTGAGGATGGAATCTT
TGAGGTCAAGTCTACAGCTGGAGACACCCACTTGGGTGGAGAAGATTTTGACAACCGAATGGTCAACCATTTTATTGCTGAGTTTAAGCGC
AAGCATAAGAAGGACATCAGTGAGAACAAGAGAGCTGTAAGACGCCTCCGTACTGCTTGTGAACGTGCTAAGCGTACCCTCTCTTCCAGCA
CCCAGGCCAGTATTGAGATCGATTCTCTCTATGAAGGAATCGACTTCTATACCTCCATTACCCGTGCCCGATTTGAAGAACTGAATGCTGA
CCTGTTCCGTGGCACCCTGGACCCAGTAGAGAAAGCCCTTCGAGATGCCAAACTAGACAAGTCACAGATTCATGATATTGTCCTGGTTGGT
GGTTCTACTCGTATCCCCAAGATTCAGAAGCTTCTCCAAGACTTCTTCAATGGAAAAGAACTGAATAAGAGCATCAACCCTGATGAAGCTG
TTGCTTATGGTGCAGCTGTCCAGGCAGCCATCTTGTCTGGAGACAAGTCTGAGAATGTTCAAGATTTGCTGCTCTTGGATGTCACTCCTCT
TTCCCTTGGTATTGAAACTGCTGGTGGAGTCATGACTGTCCTCATCAAGCGTAATACCACCATTCCTACCAAGCAGACACAGACCTTCACT
ACCTATTCTGACAACCAGCCTGGTGTGCTTATTCAGGTTTATGAAGGCGAGCGTGCCATGACAAAGGATAACAACCTGCTTGGCAAGTTTG
AACTCACAGGCATACCTCCTGCACCCCGAGGTGTTCCTCAGATTGAAGTCACTTTTGACATTGATGCCAATGGTATACTCAATGTCTCTGC
TGTGGACAAGAGTACGGGAAAAGAGAACAAGATTACTATCACTAATGACAAGGGCCGTTTGAGCAAGGAAGACATTGAACGTATGGTCCAG
GAAGCTGAGAAGTACAAAGCTGAAGATGAGAAGCAGAGGGACAAGGTGTCATCCAAGAATTCACTTGAGTCCTATGCCTTCAACATGAAAG
CAACTGTTGAAGATGAGAAACTTCAAGGCAAGATTAACGATGAGGACAAACAGAAGATTCTGGACAAGTGTAATGAAATTATCAACTGGCT
TGATAAGAATCAGACTGCTGAGAAGGAAGAATTTGAACATCAACAGAAAGAGCTGGAGAAAGTTTGCAACCCCATCATCACCAAGCTGTAC
CAGAGTGCAGGAGGCATGCCAGGAGGAATGCCTGGGGATTTCCTGGTGGTGGAGCTCCTCCCTCTGGTGGTGCTTCCTCAGGGCCCACCA
TTGAAGAGGTTGATTAAGCCAACCAAGTGTAGATGTAGCATTGTTCCACACATTTAAAACATTTGAAGGACCTAAATTCGTAGCAAATTCT
GTGGCAGTTTTAAAAAGTTAAGCTGCTATAGTAAGTTACTGGGCATTCTCAATACTTGAATATGGAACATATGCACAGGGGAAGGAAATAA
CATTGCACTTTATAAACACTGTATTGTAAGTGGAAAATGCAATGTCTTAAATAAAACTATTTAAAATTGGCACCAT

Figura 2e

>gi|18390348|ref|NM_000972.2| Proteína ribosómica L7a de *Homo sapiens* (RPL7A), ARNm

TTTCCTTTCTCTCTCCTCCCGCCGCCCAAGATGCCGAAAGGAAAGAAGGCCAAGGGAAAGAAGGTGGCTCCGGCCCCAGCTGTCGTGAA
GAAGCAGGAGGCTAAGAAAGTGGTGAATCCCCTGTTTGAGAAAAGGCCTAAGAATTTTGGCATTGGACAGGACATCCAGCCCAAAAGAG
ACCTCACCCGCTTTGTGAAATGGCCCCGCTATATCAGGTTGCAGCGGCAGAGAGCCATCCTCTATAAGCGGCTGAAAGTGCCTCCTGCG
ATTAACCAGTTCACCCAGGCCCTGGACCGCCAAACAGCTACTCAGCTGCTTAAGCTGGCCCACAAGTACAGACCAGAGACAAAGCAAGA
GAAGAAGCAGAGACTGTTGGCCCGGGCCGAGAAGAAGGCTGCTGGCAAAGGGGACGTCCCAACGAAGAGACCACCTGTCCTTCGAGCAG
GAGTTAACACCGTCACCACCTTGGTGGAGAACAAGAAAGCTCAGCTGGTGGTGATTGCACACGACGTGGATCCCATCGAGCTGGTTGTC
TTCTTGCCTGCCCTGTGTCGTAAAATGGGGGTCCCTTACTGCATTATCAAGGGAAAGGCAAGACTGGGACGTCTAGTCCACAGGAAGAC
CTGCACCACTGTCGCCTTCACACAGGTGAACTCGGAAGACAAAGGCGCTTTGGCTAAGCTGGTGGAAGCTATCAGGACCAATTACAATG
ACAGATACGATGAGATCCGCCGTCACTGGGGTGGCAATGTCCTGGGTCCTAAGTCTGTGGCTCGTATCGCCAAGCTCGAAAAGGCAAAG
GCTAAAGAACTTGCCACTAAACTGGGTTAAATGTACACTGTTGAGTTTTCTGTACATAAAAATAATTGAAATAATACAAATTTTCCTTC

Figura 2f

>gi|39812410|ref|NM_001007.3| Proteína ribosómica S4 unida a X de *Homo sapiens* (RPS4X), ARNm

GGGCGGAGCAGCTGAAAATCCGGCGCGCGCAGTCTCCAGCCCCAATTTCTACGCGCACCGGAAGACGGAGGTCCTCTTTCCTTGCCT
AACGCAGCCATGGCTCGTGGTCCCAAGAAGCATCTGAAGCGGGTGGCAGCTCCAAAGCATTGGATGCTGGATAAATTGACCGGTGTG
TTTGCTCCTCGTCCATCCACCGGTCCCCACAAGTTGAGAGAGTGTCTCCCCCTCATCATTTTCCTGAGGAACAGACTTAAGTATGCC
CTGACAGGAGATGAAGTAAAGAAGATTTGCATGCAGCGGTTCATTAAAATCGATGGCAAGGTCCGAACTGATATAACCTACCCTGCT
GGATTCATGGATGTCATCAGCATTGACAAGACGGGAGAGAATTTCCGTCTGATCTATGACACCAAGGGTCGCTTTGCTGTACATCGT
ATTACACCTGAGGAGGCCAAGTACAAGTTGTGCAAAGTGAGAAAGATCTTTGTGGGCACAAAAGGAATCCCTCATCTGGTGACTCAT
GATGCCCGCACCATCCGCTACCCCGATCCCCTCATCAAGGTGAATGATACCATTCAGATTGATTTGGAGACTGGCAAGATTACTGAT
TTCATCAAGTTCGACACTGGTAACCTGTGTATGGTGACTGGAGGTGCTAACCTAGGAAGAATTGGTGTGATCACCAACAGAGAGAGG
CACCCTGGATCTTTTGACGTGGTTCACGTGAAAGATGCCAATGGCAACAGCTTTGCCACTCGACTTTCCAACATTTTTGTTATTGGC
AAGGGCAACAAACCATGGATTTCTCTTCCCCGAGGAAAGGGTATCCGCCTCACCATTGCTGAAGAGAGACAAAAGACTGGCGGCC
AAACAGAGCAGTGGGTGAAATGGGTCCCTGGGTGACATGTCAGATCTTTGTACGTAATTAAAAATATTGTGGCAGGATTAATAGC

Figura 2g

>gi|17158043|ref|NM_001010.2| Proteína ribosómica S6 de *Homo sapiens* (RPS6), ARNm

CCTCTTTTCCGTGGCGCCTCGGAGGCGTTCAGCTGCTTCAAGATGAAGCTGAACATCTCCTTCCCAGCCACTGGCTGCCAGAAACTCA
TTGAAGTGGACGATGAACGCAAACTTCGTACTTTCTATGAGAAGCGTATGGCCACAGAAGTTGCTGCTGACGCTCTGGGTGAAGAATG
GAAGGGTTATGTGGTCCGAATCAGTGGTGGGAACGACAAACAAGGTTTCCCCATGAAGCAGGGTGTCTTGACCCATGGCCGTGTCCGC
CTGCTACTGAGTAAGGGGCATTCCTGTTACAGACCAAGGAGAACTGGAGAAAGAAAGAGAAAATCAGTTCGTGGTTGCATTGTGGATG
CAAATCTGAGCGTTCTCAACTTGGTTATTGTAAAAAAAGGAGAGAAGGATATTCCTGGACTGACTGATACTACAGTGCCTCGCCGCCT
GGGCCCCAAAAGAGCTAGCAGAATCCGCAAACTTTTCAATCTCTCTAAAGAAGATGATGTCCGCCAGTATGTTGTAAGAAAGCCCTTA
AATAAAGAAGGTAAGAAACCTAGGACCAAAGCACCCAAGATTCAGCGTCTTGTTACTCCACGTGTCCTGCAGCACAAACGGCGGCGTA
TTGCTCTGAAGAAGCAGCGTACCAAGAAAAATAAAGAAGAGGCTGCAGAATATGCTAAACTTTTGGCCAAGAGAATGAAGGAGGCTAA
GGAGAAGCGCCAGGAACAAATTGCGAAGAGACGCAGACTTTCCTCTCTGCGAGCTTCTACTTCTAAGTCTGAATCCAGTCAGAAATAA
GATTTTTTGAGTAACAAATAAATAAGATCAGACTCTG

Figura 2h

>gi|34328943|ref|NM_021109.2| **Timosina de *Homo sapiens*, beta 4, unida a X (TMSB4X), ARNm**

ACAACTCGGTGGTGGCCACTGCGCAGACCAGACTTCGCTCGTACTCGTGCGCCTCGCTTCGCTTTTCCTCCGCAACCATGTCTGACAAACCC
GATATGGCTGAGATCGAGAAATTCGATAAGTCGAAACTGAAGAAGACAGAGACGCAAGAGAAAAATCCACTGCCTTCCAAAGAAACGATTGA
ACAGGAGAAGCAAGCAGGCGAATCGTAATGAGGCGTGCGCCGCCAATATGCACTGTACATTCCACAAGCATTGCCTTCTTATTTTACTTCTT
TTAGCTGTTTAACTTTGTAAGATGCAAAGAGGTTGGATCAAGTTTAAATGACTGTGCTGCCCCTTTCACATCAAAGAACTACTGACAACGAA
GGCCGCGCCTGCCTTTCCCATCTGTCTATCTATCTGGCTGGCAGGGAAGGAAAGAACTTGCATGTTGGTGAAGGAAGAAGTGGGGTGGAAGA
AGTGGGGTGGGACGACAGTGAAATCTAGAGTAAAACCAAGCTGGCCCAAGGTGTCCTGCAGGCTGTAATGCAGTTTAATCAGAGTGCCATTT
TTTTTTTTGTTCAAATGATTTTAATTATTGGAATGCACAATTTTTTTAATATGCAAATAAAAAGTTTAAAAACTT

Figura 2i

>gi|4507668|ref|NM_003295.1| Proteína tumoral de *Homo sapiens*, traduccionalmente controlada 1 (TPT1), ARNm

CCCCCCCGAGCCGCCGCTCGCCGCTCCGAGTTTCAGGCTCGTGCTAAGCTAGCGCCGTCGTCGTCCTCCCTTCAGT

CGCCTATCATGATTATCTTACCGGGACCTCATCAGCGATGAGATGTTCTCCGACATCTACAAGATCGCGGACGGGTT

GTGCCTGGAGGTGGAGGGAAGATGGTCAGTAGGAGACAAGGAAGGTAACATTGATGACTCGCTCATTGGTGGAAATGCCTCCGCTGAGG

CCCCGAGGGCCGAAGGTACCGAAAGCACAGTAATCACTGGTCGATATTGTCATGAACCATCACCTGCAGGAAACAAGTTTCACAAA

AGAAGCCTACAAGAAGTACATCAAAGATTACATGAAATCAATCAAAGGGAAACTTGAAGAACAGAGACCAGAAAGAGTAAAAACCTTT

TATGACAGGGCCTGCAGAACAAATCCAAGCACATCCTTGCTAATTTCAAAAACTACCAGTTCTTTATTGGTGAAAACATGAATCCAGA

TGGCATGGTTGCTCTATTGGACTACCGTGAGGATGGTGTGACCCCATATATGATTTTCTTTAAGGATGGTTTAGAAATGGAAAAATG

TTAACAAATGTGGCAATTATTTTGGATCTATCACCTGTCATCATAACTGGCTTCTGCTTTGTCATCCACACAACCAGGACTTAAGA

CAAATGGGACTGATGTCATCTTGAGCTCTTCATTTATTTTGACTGTGATTTATTGGAGTGGAGGCATTGTTTTTAAGAAAAACATG

TCATGTAGGTTGTCTAAAAATAAAATGCATTTAAACTCATTTGAGAG

Figura 2j

>gi|62414288|ref|NM_003380.2| Vimentina de *Homo sapiens* (VIM), ARNm

GTCCCCGCGCCAGAGACGCAGCCGCGCTCCCACCACCCACACCCACCGCGCCCTCGTTCGCCTCTTCTCCGGGAGCCAGTCCGCGCC
ACCGCCGCCGCCCAGGCCATGCCCACCCTCCGCAGCCATGTCCACCAGGTCCGTGTCCTCGTCCTCCTACCGCAGGATGTTCGGCGG
CCCGGGCACCGCGAGCCGGCCGAGCTCCAGCCGGAGCTACGTGACTACGTCCACCCGCACCTACAGCCTGGGCAGCGCGCTGCGCCC
CAGCACCAGCCGCAGCCTCTACGCCTCGTCCCCGGGCGGCGTGTATGCCACGCGCTCCTCTGCCGTGCGCCTGCGGAGCAGCGTGCC
CGGGGTGCGGCTCCTGCAGGACTCGGTGGACTTCTCGCTGGCCGACGCCATCAACACCGAGTTCAAGAACACCCGCACCAACGAGAA
GGTGGAGCTGCAGGAGCTGAATGACCGCTTCGCCAACTACATCGACAAGGTGCGCTTCCTGGAGCAGCAGAATAAGATCCTGCTGGC
CGAGCTCGAGCAGCTCAAGGGCCAAGGCAAGTCGCGCCTGGGGGACCTCTACGAGGAGGAGATGCGGGAGCTGCGCCGGCAGGTGGA
CCAGCTAACCAACGACAAAGCCCGCGTCGAGGTGGAGCGCGACAACCTGGCCGAGGACATCATGCGCCTCCGGGAGAAATTGCAGGA
GGAGATGCTTCAGAGAGAGGAAGCCGAAAACACCCTGCAATCTTTCAGACAGGATGTTGACAATGCGTCTCTGGCACGTCTTGACCT
TGAACGCAAAGTGGAATCTTTGCAAGAAGAGATTGCCTTTTTGAAGAAACTCCACGAAGAGGAAATCCAGGAGCTGCAGGCTCAGAT
TCAGGAACACCATGTCCAAATCCATGTGCATGTTTCCAAGCCTGACGTCACCGCCTCCCCTCCGTCACCTACCTCACCAATATCAAAG
TGTGGCTGCCAAGAACCTGCAGGAGGCAGAAGAATGGTACAAATCCAAGTTTGCTGACCTCTCTGAGGCTGCCAACCGGAACAATGA
CGCCCTGCGCCAGGCAAAGCAGGAGTCCACTGAGTACCGGAGACAGGTGCAGTCCCTCACCTGTGAAGTGGATGCCCTTAAAGGAAC
CAATGAGTCCCTGGAACGCCAGATGCGTGAAATGGAAGAGAACTTTGCCGTTGAAGCTGCTAACTACCAAGACACTATTGGCCGCCT
GCAGGATGAGATTCAGAATATGAAGGAGGAAATGGCTCGTCACCTTCGTGAATACCAAGACCTGCTCAATGTTAAGATGGCCCTTGA
CATTGAGATTGCCACCTACAGGAAGCTGCTGGAAGGCGAGGAGAGCAGGATTTCTCTGCCTCTTCCAAACTTTTCCTCCCTGAACCT
GAGGGAAACTAATCTGGATTCACTCCCTCTGGTTGATACCCACTCAAAAAGGACACTTCTGATTAAGACGGTTGAAACTAGAGATGG
ACAGGTTATCAACGAAACTTCTCAGCATCACGATGACCTTGAATAAAAATTGCACACACTCAGTGCAGCAATATATTACCAGCAAGA
ATAAAAAAGAAATCCATATCTTAAAGAAACAGCTTTCAAGTGCCTTTCTGCAGTTTTTCAGGAGCGCAAGATAGATTTGGAATAGGA
ATAAGCTCTAGTTCTTAACAACCGACACTCCTACAAGATTTAGAAAAAAGTTTACAACATAATCTAGTTTACAGAAAAATCTTGTGC
TAGAATACTTTTTAAAAGGTATTTTGAATACCATTAAAACTGCTTTTTTTTTTCCAGCAAGTATCCAACCAACTTGGTTCTGCTTCA
ATAAATCTTTGGAAAAACTC

Figura 2k

>gi|8392890|ref|NM_000477.3| Albúmina de *Homo sapiens* (ALB), ARNm

AGCTTTTCTCTTCTGTCAACCCCACACGCCTTTGGCACAATGAAGTGGGTAACCTTTATTTCCCTTCTTTTTCTCTTTAGCTCGGCTTATT
CCAGGGGTGTGTTTCGTCGAGATGCACACAAGAGTGAGGTTGCTCATCGGTTTAAAGATTTGGGAGAAGAAAATTTCAAAGCCTTGGTGTT
GATTGCCTTTGCTCAGTATCTTCAGCAGTGTCCATTTGAAGATCATGTAAAATTAGTGAATGAAGTAACTGAATTTGCAAAAACATGTGTT
GCTGATGAGTCAGCTGAAAATTGTGACAAATCACTTCATACCCTTTTTGGAGACAAATTATGCACAGTTGCAACTCTTCGTGAAACCTATG
GTGAAATGGCTGACTGCTGTGCAAAACAAGAACCTGAGAGAAATGAATGCTTCTTGCAACACAAAGATGACAACCCAAACCTCCCCCGATT
GGTGAGACCAGAGGTTGATGTGATGTGCACTGCTTTTCATGACAATGAAGAGACATTTTTGAAAAAATACTTATATGAAATTGCCAGAAGA
CATCCTTACTTTTATGCCCCGGAACTCCTTTTCTTTGCTAAAAGGTATAAAGCTGCTTTTACAGAATGTTGCCAAGCTGCTGATAAAGCTG
CCTGCCTGTTGCCAAAGCTCGATGAACTTCGGGATGAAGGGAAGGCTTCGTCTGCCAAACAGAGACTCAAGTGTGCCAGTCTCCAAAAATT
TGGAGAAAGAGCTTTCAAAGCATGGGCAGTAGCTCGCCTGAGCCAGAGATTTCCCAAAGCTGAGTTTGCAGAAGTTTCCAAGTTAGTGACA
GATCTTACCAAAGTCCACACGGAATGCTGCCATGGAGATCTGCTTGAATGTGCTGATGACAGGGCGGACCTTGCCAAGTATATCTGTGAAA
ATCAAGATTCGATCTCCAGTAAACTGAAGGAATGCTGTGAAAAACCTCTGTTGGAAAAATCCCACTGCATTGCCGAAGTGGAAAATGATGA
GATGCCTGCTGACTTGCCTTCATTAGCTGCTGATTTTGTTGAAAGTAAGGATGTTTGCAAAAACTATGCTGAGGCAAAGGATGTCTTCCTG
GGCATGTTTTTGTATGAATATGCAAGAAGGCATCCTGATTACTCTGTCGTGCTGCTGCTGAGACTTGCCAAGACATATGAAACCACTCTAG
AGAAGTGCTGTGCCGCTGCAGATCCTCATGAATGCTATGCCAAAGTGTTCGATGAATTTAAACCTCTTGTGGAAGAGCCTCAGAATTTAAT
CAAACAAAATTGTGAGCTTTTTGAGCAGCTTGGAGAGTACAAATTCCAGAATGCGCTATTAGTTCGTTACACCAAGAAAGTACCCCAAGTG
TCAACTCCAACTCTTGTAGAGGTCTCAAGAAACCTAGGAAAAGTGGGCAGCAAATGTTGTAAACATCCTGAAGCAAAAAGAATGCCCTGTG
CAGAAGACTATCTATCCGTGGTCCTGAACCAGTTATGTGTGTTGCATGAGAAAACGCCAGTAAGTGACAGAGTCACCAAATGCTGCACAGA
ATCCTTGGTGAACAGGCGACCATGCTTTTCAGCTCTGGAAGTCGATGAAACATACGTTCCCAAAGAGTTTAATGCTGAAACATTCACCTTC
CATGCAGATATATGCACACTTTCTGAGAAGGAGAGACAAATCAAGAAACAAACTGCACTTGTTGAGCTCGTGAAACACAAGCCCAAGGCAA
CAAAAGAGCAACTGAAAGCTGTTATGGATGATTTCGCAGCTTTTGTAGAGAAGTGCTGCAAGGCTGACGATAAGGAGACCTGCTTTGCCGA
GGAGGGTAAAAAACTTGTTGCTGCAAGTCAAGCTGCCTTAGGCTTATAACATCTACATTTAAAAGCATCTCAGCCTACCATGAGAATAAGA
GAAAGAAAATGAAGATCAAAAGCTTATTCATCTGTTTTCTTTTTCGTTGGTGTAAAGCCAACACCCTGTCTAAAAAACATAAATTTCTTTA
ATCATTTTGCCTCTTTTCTCTGTGCTTCAATTAATAAAAAATGGAAAGAATCTAATAGAGTGGTACAGCACTGTTATTTTTCAAAGATGTG
TTGCTATCCTGAAAATTCTGTAGGTTCTGTGGAAGTTCCAGTGTTCTCTCTTATTCCACTTCGGTAGAGGATTTCTAGTTTCTGTGGGCTA
ATTAAATAAATCACTAATACTCTTCTAAGTT

Figura 2l

>gi|34577108|ref|NM_000034.2| Aldolasa A de *Homo sapiens*, fructosa-bisfosfato (ALDOA), variante de transcripción 1, ARNm

CCTAGCTTGGCGCGAATCCGTGAATTGCCCGCGGCCGAGGGTCCAGCTCCCGGACTGACTGGCTCTGCCCTTCCCCATGAGCCTCCT
CTAGCCCGTGGAATCCAACCCCGGCTCCTGTCAGCAGCCCTCCCCTGCTCCCCATCCCTGCCATCGCTCTTCATGCTGTGTGGCATCTA
TTTGTTGCTGCTGGGTCTAGTCCCTGCTGACTAGGAACTGCCTGCTGCCTGCTGCGGCTGCTGGCGGAGGTAGCTCCCCCTGCAGGAAGCAA
GGTTCCTCCGGGCCCCAGACTGCTGCTGGACTGTGCAGAAGCCTGCAACTTTCCTCTGCCTAGCCCGCCAACTTCCTGGATGCTTGCT
GCCCCCAGCCCACCAGAGCTGACTGGGCAGTCTCGCTGCCCCGCTGCTGCCCACTCTGCGACTGTGCCTGTACGTGCCAGCTCCCCGACTG
CCAGAGCCTCAACTGTCTCTGCTTCGAGATCAAGCTCCGATGAGGACCCAGGGCCCCTGCCCCTCTGGGGAGCGGCCAGCCCCCAGGCCCA
TGTGCCCTCGTCCCTGAGAGGCCTTCCCCACGCCACTGGAACCACAGATGGCCTGCCGAGCACCCAGCCTGGGAACTGGAAGTGGCAGC
GCAGGGCCTGGCTCCCTGCAGGGCAGGACTCTTGGCCGGCTGGACGGCAGCTCCTCTGGAGGGCCAGAAAAGAGAGGGGCTAGTGCTCGGG
CAGGTGCCTGGCTTCCCTCCCCTCCACACGTCAACGATTCTATTTGAAGTTGGGCAGGTGGCGCTGCTCACCACACAGCAAGTGTT
ATAGGAGGAGTCTGGCCCTTGAGTACCGGGTACGCAGGGGTGCCTCAACCACACTCCGTCCACGGACTCTCCGTTATTTAGGAGGTCCCT
GGCCAAAGATTTATTTCTTGACAACCAAGGGCCTCCGTCTGAATTTCCAAGGAAGAATTCCTCTGAAGCACCGGAACTTGCTACTACC
AGCACCATGCCCTACCAATATCCAGCACTGACCCGGAGCAGAAGGAGCTGTCTGACATCGCTCACCGCATCGTGGCACCTGGCAAGG
GCATCCTGGCTGCAGATGACTCCACTGGGAGCATTGCCAAGCGGCTGCAGTCCATTGGCACCGAGAACACCGGAGGAGAACCGGCTGCTTCTA
CCGCCAGCTGCTGCTGACAGCTGACGACCGCGTGAACCCCTGCATTGGGGGTGTCATCCTCTTCCATGAGACACTCTACGACAAGGCGGAT
GATGGCGTCCCTTCCCCCAAGTTATCAAATCCAAGGGCGGTGTTTGTGGCATCAAGGTCGTGGTCCCCCTGGCAGGGACAA
ATGGCGACTACCACCCAAGGGTTGGATGGGCTGTCTGAGCGCTGTGCCCAGTACAAGAAGGACTGAGTTCGCCAAGTGGCGTTG
TGTGCTGAAGATTGGGGAACACACCCCCTCAGCCCTCGCCATCATGGAAAATGCCAATGTTCTGGCCCGTTATGCCAGTATCTGCCAGCAG
AATGGCATTGTGCCCATCGTGGAGCCTGAGATCCTCCCTGATGGGGACCATGACTGTAAGCGCCTGCCAGTATGTGACCGACAAGGTGCTGG
CTGCTGTCTACAAGGCTCTGAGTGACCACATCACTTGCAAGGCACCTTGCTGAAGCCCAACATGGTCACCCCAGGCCATGCTTGCAC
TCAGAAGTTTCTCATGAGGAGATTGCCATGGCGACCGTCACAGCGCTGCGCCGCCTGTCACTGGGATCACCTTCCTG
TCTGGAGGCCAGAGTGAGGAGGAGCGTCCATCAATGCCATTAACAAGTGCCCCCTGCCTGAAGCCCTGGGCCCTGACGTTCTCCT
ACGGCCGAGCCCTGCAGGCCTCTGCCCTGAAGGCCTGGGGCGGAAGAAGGAGAAGCCTGAAGGCTGCGCAGGAGGAGTATGTCAAGCGAGC
CCTGGCCAACAGCCTTGCCGTGTCAAGGAAAGTACACTCCGAGTGGTCAGGCTGGGCCTGCTGGCAGGAGTCCCTCTTGTCTCTAACCAC
GCCTATTAAGCCGAGTGTTCCCAGCTGCCCCCAACACTCCAGCCCTGCCCCTCCCACTCTTGAAGAGGAGCCGCCTCCTCCTGGGGCT
CCAGGCTGGCTTGCCCGCGCTCTTTCTTCCCTCGTGACAGTGGTGTGGTGCTGAATGCTAAGTCCATCACCCTTTCCGGCACA
CTGCCAAATAAACAGCTATTTAAGGGGG

Figura 2m

>gi|50345983|ref|NM_001001937.1| ATP sintasa de *Homo sapiens*, transportador de H+, complejo F1 mitocondrial, subunidad alfa 1, músculo cardíaco (ATP5A1), gen nuclear que codifica la proteína mitocondrial, variante de transcripción 1, ARNm

TCTGGCATTGCAAGCCTCGCTTCGTTGCCACTTCCCAGCTCTTCCCGCCTTCCGCGGGTATAATCAACACTACGAGAGATAGAGCCGCCTAG
AACCAGTCCGGGAGGCTGCGGCTGCAGAAGTACCGCCTGCGGAGTAACTGCAAAGATGCTGTCCGTGCGCGTTGCTGCGGCCGTGGTCCGCG
CCCTTCCTCGGCGGGCCGGACTGGTCTCCAGAAATGCTTTGGGTTCATCTTTCATTGCTGCAAGGAACTTCCATGCCTCTAACACTCATCT
TCAAAAGACTGGGACTGCTGAGATGTCCTCTATTCTTGAAGAGCGTATTCTTGGAGCTGATACCTCTGTTGATCTTGAAGAAACTGGGCGT
GTCTTAAGTATTGGTGATGGTATTGCCCGCGTACATGGGCTGAGGAATGTTCAAGCAGAAGAAATGGTAGAGTTTTCTTCAGGCTTAAAGG
GTATGTCCTTGAACTTGGAACCTGACAATGTTGGTGTTGTCGTGTTTGGAAATGATAAACTAATTAAGGAAGGAGATATAGTGAAGAGGAC
AGGAGCCATTGTGGACGTTCCAGTTGGTGAGGAGCTGTTGGGTCGTGTAGTTGATGCCCTTGGTAATGCTATTGATGGAAAGGGTCCAATT
GGTTCCAAGACGCGTAGGCGAGTTGGTCTGAAAGCCCCCGGTATCATTCCTCGAATTTCAGTGCGGGAACCAATGCAGACTGGCATTAAGG
CTGTGGATAGCTTGGTGCCAATTGGTCGTGGTCAGCGTGAACTGATTATTGGTGACCGACAGACTGGGAAAACCTCAATTGCIATTGACAC
AATCATTAACCAGAAACGTTTCAATGATGGATCTGATGAAAAGAAGAAGCTGTACTGTATTTATGTTGCTATTGGTCAAAAGAGATCCACT
GTTGCCCAGTTGGTGAAGAGACTTACAGATGCAGATGCCATGAAGTACACCATTGTGGTGTCGGCTACGGCCTCGGATGCTGCCCCACTTC
AGTACCTGGCTCCTTACTCTGGCTGTTCCATGGGAGAGTATTTTAGAGACAATGGCAAACATGCTTTGATCATCTATGACGACTTATCCAA
AGAGGCTGTTGCTTACCGTCAGATGTCTCTGTTGCTCCGCCGACCCCCTGGTCGTGAGGCCTATCCTGGTGATGTGTTCTACCTACACTCC
CGGTTGCTGGAGAGAGCAGCCAAAATGAACGATGCTTTTGGTGGTGGCTCCTTGACTGCTTTGCCAGTCATAGAAACACAGGCTGGTGATG
TGTCTGCTTACATTCCAACAAATGTCATTTCCATCACTGACGGACAGATCTTCTTGGAAACAGAATTGTTCTACAAAGGTATCCGCCCTGC
AATTAACGTTGGTCTGTCTGTATCTCGTGTCGGATCCGCTGCCCAAACCAGGGCTATGAAGCAGGTAGCAGGTACCATGAAGCTGGAATTG
GCTCAGTATCGTGAGGTTGCTGCTTTTGCCCAGTTCGGTTCTGACCTCGATGCTGCCACTCAACAACTTTTGAGTCGTGGCGTGCGTCTAA
CTGAGTTGCTGAAGCAAGGACAGTATTCTCCCATGGCTATTGAAGAACAAGTGGCTGTTATCTATGCGGGTGTAAGGGGATAICTTGATAA
ACTGGAGCCCAGCAAGATTACAAAGTTTGAGAATGCTTTCTTGTCTCATGTCGTCAGCCAGCACCAAGCCTTGTTGGGCACTATCAGGGCT
GATGGAAAGATCTCAGAACAATCAGATGCAAAGCTGAAAGAGATTGTAACAAATTTCTTGGCTGGATTTGAAGCTTAAACTCCTGTGGATT
CACATCAAATACCAGTTCAGTTTTGTCATTGTTCTAGTAAATTAGTTCCATTTGTAAAAGGGTTACTCTCATACTCCTTATGIACAGAAAT
CACATGAAAAATAAAGGTTCCATAATGCATAGTT

Figura 2n

>gi|20428653|ref|NM_001743.3| Calmodulina 2 de *Homo sapiens* (fosforilasa quinasa, delta), (CALM2), ARNm

AGTCCGAGTGGAGAGAGCGAGCTGAGTGGTTGTGTGGTCGCGTCTCGGAAACCGGTAGCGCTTGCAGCATGGCTGACCAACTGACTGAAGA
GCAGATTGCAGAATTCAAAGAAGCTTTTTCACTATTTGACAAAGATGGTGATGGAACTATAACAACAAAGGAATTGGGAACTGTAATGAGA
TCTCTTGGGCAGAATCCCACAGAAGCAGAGTTACAGGACATGATTAATGAAGTAGATGCTGATGGTAATGGCACAATTGACTTCCCTGAAT
TTCTGACAATGATGGCAAGAAAAATGAAAGACACAGACAGTGAAGAAGAAATTAGAGAAGCATTCCGTGTGTTTGATAAGGATGGCAATGG
CTATATTAGTGCTGCAGAACTTCGCCATGTGATGACAAACCTTGGAGAGAAGTTAACAGATGAAGAAGTTGATGAAATGATCAGGGAAGCA
GATATTGATGGTGATGGTCAAGTAAACTATGAAGAGTTTGTACAAATGATGACAGCAAAGTGAAGACCTTGTACAGAATGTGTTAAATTTC
TTGTACAAAATTGTTTATTTGCCTTTTCTTTGTTTGTAACTTATCTGTAAAAGGTTTCTCCCTACTGTCAAAAAAATATGCATGTATAGTA
ATTAGGACTTCATTCCTCCATGTTTTCTTCCCTTATCTTACTGTCATTGTCCTAAAACCTTATTTTAGAAAATTGATCAAGTAACATGTTG
CATGTGGCTTACTCTGGATATATCTAAGCCCTTCTGCACATCTAAACTTAGATGGAGTTGGTCAAATGAGGGAACATCTGGGTTATGCCTT
TTTTAAAGTAGTTTTCTTTAGGAACTGTCAGCATGTTGTTGTTGAAGTGTGGAGTTGTAACTCTGCGTGGACTATGGACAGTCAACAATAT
GTACTTAAAAGTTGCACTATTGCAAAACGGGTGTATTATCCAGGTACTCGTACACTATTTTTTTGTACTGCTGGTCCTGTACCAGAAACAT
TTTCTTTTATTGTTACTTGCTTTTTAAACTTTGTTTAGCCACTTAAAATCTGCTTATGGCACAATTTGCCTCAAAATCCATTCCAAGTTGT
ATATTTGTTTTCCAATAAAAAAATTACAATTTACCC

Figura 2o

>gi|5031856|ref|NM_005566.1| Lactato deshidrogenasa A de *Homo sapiens* (LDHA), ARNm

TGCTGCAGCCGCTGCCGCCGATTCCGGATCTCATTGCCACGCGCCCCCGACGACCGCCCGACGTGCATCCCGATTCCTTTTGGTTCCAAG
TCCAATATGGCAACTCTAAAGGATCAGCTGATTTATAATCTTCTAAAGGAAGAACAGACCCCCCAGAATAAGATTACAGTTGTTGGGGTTG
GTGCTGTTGGCATGGCCTGTGCCATCAGTATCTTAATGAAGGACTTGGCAGATGAACTTGCTCTTGTTGATGTCATCGAAGACAAATTGAA
GGGAGAGATGATGGATCTCCAACATGGCAGCCTTTTCCTTAGAACACCAAAGATTGTCTCTGGCAAAGACTATAATGTAACTGCAAACTCC
AAGCTGGTCATTATCACGGCTGGGGCACGTCAGCAAGAGGGAGAAAGCCGTCTTAATTTGGTCCAGCGTAACGTGAACATATTTAAATTCA
TCATTCCTAATGTTGTAAAATACAGCCCGAACTGCAAGTTGCTTATTGTTTCAAATCCAGTGGATATCTTGACCTACGTGGCTTGGAAGAT
AAGTGGTTTTCCCAAAAACCGTGTTATTGGAAGTGGTTGCAATCTGGATTCAGCCCGATTCCGTTACCTGATGGGGGAAAGGCTGGGAGTT
CACCCATTAAGCTGTCATGGGTGGGTCCTTGGGGAACATGGAGATTCCAGTGTGCCTGTATGGAGTGGAATGAATGTTGCTGGTGTCTCTC
TGAAGACTCTGCACCCAGATTTAGGGACTGATAAAGATAAGGAACAGTGGAAAGAGGTTCACAAGCAGGTGGTTGAGAGTGCTTATGAGGT
GATCAAACTCAAAGGCTACACATCCTGGGCTATTGGACTCTCTGTAGCAGATTTGGCAGAGAGTATAATGAAGAATCTTAGGCGGGTGCAC
CCAGTTTCCACCATGATTAAGGGTCTTTACGGAATAAAGGATGATGTCTTCCTTAGTGTTCCTTGCATTTTGGGACAGAATGGAATCTCAG
ACCTTGTGAAGGTGACTCTGACTTCTGAGGAAGAGGCCCGTTTGAAGAAGAGTGCAGATACACTTTGGGGGATCCAAAAGGAGCTGCAATT
TTAAAGTCTTCTGATGTCATATCATTTCACTGTCTAGGCTACAACAGGATTCTAGGTGGAGGTTGTGCATGTTGTCCTTTTTATCTGATCT
GTGATTAAAGCAGTAATATTTTAAGATGGACTGGGAAAAACATCAACTCCTGAAGTTAGAAATAAGAATGGTTTGTAAAATCCACAGCTAT
ATCCTGATGCTGGATGGTATTAATCTTGTGTAGTCTTCAACTGGTTAGTGTGAAATAGTTCTGCCACCTCTGAGCACCACTGCCAATGCT
GTACGTACTGCATTTGCCCCTTGAGCCAGGTGGATGTTTACCGTGTGTTATATAACTTCCTGGCTCCTTCACTGAACATGCCTAGTCCAAC
ATTTTTTCCCAGTGAGTCACATCCTGGGATCCAGTGTATAAATCCAATATCATGTCTTGTGCATAATTCTTCCAAAGGATCTTATTTTGTG
AACTATATCAGTAGTGTACATTACCATATAATGTAAAAAGATCTACATACAAACAATGCAACCAACTATCCAAGTGTTATACCAACTAAAA
CCCCCAATAAACCTTGAACAGTG

Figura 2p

>gi|52851446|ref|NM_000365.4| Triosafosfato isomerasa 1 de *Homo sapiens* (TPI1), ARNm

```
CCTTCAGCGCCTCTGGTCCAGCGCCATGGCCCTCCAGGAAGTTCTTGTTGGCGGAAACTGGAAGATGAAAGTGGCGAAGCAGATCTG
GGGGAGCTCATCGGCACTCTGAACGCGGCCAAGGTGCCTGCCGACACCGAGGTGGTTTGTCTCCCCTACTGCCTATATCGACTTCGCCC
GGCAGAAGCTAGATCCCAAGATTGCTGTGGCTGCGCAGAACTGCTACAAAGTGACTAATGGGGCTTTTACTGGGGAGATCAGCCTGGCAT
GATCAAAGACTGCGGAGCCACGTGGGTGCTCTGAGAGAAGGCATGTCTTTGGGGAGTCAGATGAGCTGATTGGGCAGAAA
GTGGCCCATGCTCTGGCAGTAATCGCCTGCATTGGGGAGAAGCTAGATGAAAGGGAAGCTGGCATCACTGAGAAGGTTG
TTTTCGAGCAGACAAAGGTCATCGCAGATAACGTGAAGGACTGGAGCAAGGTCGTCCTGGCCTATGAGCCTGTGTGGGCCATTGGTACTGG
CAAGACTGCAACACCCCAACAGGCCCAGGAAGTACACGAGAAGCTCCGAGAAGTCCGAAGTCCAACGCTCTCTGATGCGTGAGCTCAGAGC
ACCCGTATCATTTATGGAGGCTCTGTGACTGGGGCAACCTGCAAGGAGCTGGCCAGCCAGCCTGATGTGGATGCTTCCTTGTGGTGGTG
CTTCCCTCAAGCCCGAATTCGTGGACATCATCAATTGCCAAACAATGAGCCCCATCCATCTTCCCTACCCTTCCTGCCAAGCCAGGGACTAA
GCAGCCCAGAAGCCCAGTACTGCCCTTTCCCTGCATATGCTTCTGATGGTCATCTGCTTCCTCTGTGGGCTCATCGAAACTGTATCT
TCCTTTACTGTTTATATCTTCACCCTGTAATGGTTGGGACCAGGCCAATCCCTTCCACTTACTATAATGGTTGGAACTAAACGTCACCA
AGGTGGCTTCTCCTTGGCTGAGAGATGGAAGGCTGGTGGGATTTGCTCCTGGGTTCCCTAGGTGAGGGCAGAAGAGAAACCATC
CTCTCCCTTCTTACACCGTGAGGCCAAGATCCCCTCAGAAGGCCAGGAGTGCTGCCCTCCATGGTGCCCGTGCCTCTTGTGCTGTATG
TGAACCACCCATGTGAGGGAATAAACCTGGCACTAGG
```

Figura 2q

TPT1 = Proteína tumoral, traduccionalmente controlada, 1, 830 pb, chr 13q12-q14.

VIM = Vimentina, 1847 pb, chr 10p13.

FTH1 = Ferritina, polipéptido pesado 1, 1228 pb, chr 11q13.

RPS4X = Proteína ribosómica S4, unida a X, 955 pb, chr X q13.1.

RPL7A = Proteína ribosómica L7a, 890 pb, chr 9q34.

ENO1 = Enolasa 1, 1812 pb, chr 1p36.3-p36.2.

HSPA8 = Proteína 8 de choque térmico de 70 kDa, 2260 pb, chr 11q24.1.

GAPDH = Gliraldehído-3-fosfato deshidrogenasa, 1310 pb, chr 12p13.

TMSB4X = Timosina, beta 4, unida a X, 627 pb, chr X q21.3-q22.

RPS6 = Proteína ribosómica S6, 829 pb, chr 9p21.

FTL = Ferritina, polipéptido ligero, 870 pb, chr 19q13.3-13.4.

ALB = Albúmina, 2215 pb, chr 4q11-q13.

ALDOA = Aldolasa A, fructos-bisfosfato, 2303 pb, chr 16 q22-124.

ATP5A1 = ATP sintasa, transportador de H+, complejo F1 mitocondrial, subunidad alfa 1, músculo cardíaco, 1945 pb, chr 18q12-q21.

CALM2 = Calmodulina 2 (fosforilasa quinasa, delta), 1128 pb, chr 2p21.

LDHA = Lactato deshidrogenasa A, 1661 pb, chr 11p15.4.

TPI1 = Triosafosfato isomerasa 1, 1220 pb, chr 12p13.

Figura 2r

>gi|47009767|emb|BX437546.2| BX437546 *Homo sapiens* TIMO *Homo sapiens* ADNc clon CS0CAP007YK05
5-PRIMA, secuencia de ARNm
Longitud = 868

Puntuación = 1531 bits (829), Esperado = 0,0
Identidades = 829/829 (100%)
Hebra = +/+

```
Problema: 2    ccccccgagcgccgctccggctgcaccgcgctcgctccgagtttcaggctcgtgctaagc 61
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 12   ccccccgagcgccgctccggctgcaccgcgctcgctccgagtttcaggctcgtgctaagc 71

Problema: 62   tagcgccgtcgtcgtctcccttcagtcgccatcatgattatctaccgggacctcatcagc 121
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 72   tagcgccgtcgtcgtctcccttcagtcgccatcatgattatctaccgggacctcatcagc 131

Problema: 122  cacgatgagatgttctccgacatctacaagatccgggagatcgcggacgggttgtgcctg 181
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 132  cacgatgagatgttctccgacatctacaagatccgggagatcgcggacgggttgtgcctg 191

Problema: 182  gaggtggaggggaagatggtcagtaggacagaaggtaacattgatgactcgctcattggt 241
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 192  gaggtggaggggaagatggtcagtaggacagaaggtaacattgatgactcgctcattggt 251

Problema: 242  ggaaatgcctccgctgaaggccccgagggcgaaggtaccgaaagcacagtaatcactggt 301
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 252  ggaaatgcctccgctgaaggccccgagggcgaaggtaccgaaagcacagtaatcactggt 311

Problema: 302  gtcgatattgtcatgaaccatcacctgcaggaaacaagtttcacaaaagaagcctacaag 361
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 312  gtcgatattgtcatgaaccatcacctgcaggaaacaagtttcacaaaagaagcctacaag 371
```

Figura 3a

Pregunta: 362 aagtacatcaaagattacatgaaatcaatcaaagggaaacttgaagaacagagaccagaa 421
Sujeto: 372 aagtacatcaaagattcatgaaatcaatcaaagggaaacttgaagaacagagaccagaa 431
Pregunta: 422 agagtaaaacctttatgacagggctgcagaacaaatcaagcacatccttgctaatttc 481
Sujeto: 432 agagtaaaacctttatgacagggctgcagaacaaatcaagcacatccctgcttaatttc 491
Pregunta: 482 aaaaactaccagttctttattggtgaaaacatgaatccagatgggcatggttgctctattg 541
Sujeto: 492 aaaaactaccagttctttattggtgaaaacatgaatccagatgggcatggttgctctattg 551
Pregunta: 542 gactaccgtgaggatgggtgtgaccccatatatgatttctttaaggatgggtttagaaatg 601
Sujeto: 552 gactaccgtgaggatgggtgtgaccccatatatgatttctttaaggatgggtttagaaatg 611
Pregunta: 602 gaaaaatgttaacaaatgtggcaatttatttttggatctatccacctgtcatccataactggct 661
Sujeto: 612 gaaaaatgttaacaaatgtggcaatttatttttggatctatccacctgtcatccataactggct 671
Pregunta: 662 tctgcttgtcatccacacaaccaggacttaagacaaatgggactgatgtccatcttgag 721
Sujeto: 672 tctgcttgtcatccacacaaccaggacttaagacaaatgggactgatgtccatcttgag 731
Pregunta: 722 ctcttcatttatttgactgtgatttatttggagtgggaggcattgttttttaagaaaaaca 781
Sujeto: 732 ctcttcatttatttgactgtgatttatttggagtgggaggcattgttttttaagaaaaaca 791
Pregunta: 782 tgtcatgtaggtgtctaaaaataaaatgcatttaaactcatttgagag 830
Sujeto: 792 tgtcatgtaggtgtctaaaaataaaatgcatttaaactcatttgagag 840

Figura 3b

>gi|45716222|emb|AL540618.3| AL540618 *Homo sapiens* PLACENTA *Homo sapiens* ADNc clon CS0DE002YF2 (
5-PRIMA, secuencia de ARNm
Longitud = 952

Puntuación = 1531 bits (829), Esperado = 0,0
Identidades = 829/829 (100%)
Hebra = +/+

```
Pregunta: 2   ccccccgagcgccgctccggctgcaccgcgctcgctccgagtttcaggctcgtgctaagc 61
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 18  ccccccgagcgccgctccggctgcaccgcgctcgctccgagtttcaggctcgtgctaagc 77

Pregunta: 62  tagcgccgtcgtcgtctcccttcagtcgccatcatgattatctaccgggacctcatcagc 121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 78  tagcgccgtcgtcgtctcccttcagtcgccatcatgattatctaccgggacctcatcagc 137

Pregunta: 122 cacgatgagatgttctccgacatctacaagatccgggagatcgcggacgggttgtgcctg 181
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 138 cacgatgagatgttctccgacatctacaagatccgggagatcgcggacgggttgtgcctg 197

Pregunta: 182 gaggtggaggggaagatggtcagtaggacagaaggtaacattgatgactcgctcattggt 241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 198 gaggtggaggggaagatggtcagtaggacagaaggtaacattgatgactcgctcattggt 257

Pregunta: 242 ggaaatgcctccgctgaaggccccgagggcgaaggtaccgaaagcacagtaatcactggt 301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 258 ggaaatgcctccgctgaaggccccgagggcgaaggtaccgaaagcacagtaatcactggt 317

Pregunta: 302 gtcgatattgtcatgaaccatcacctgcaggaaacaagtttcacaaaagaagcctacaag 361
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 318 gtcgatattgtcatgaaccatcacctgcaggaaacaagtttcacaaaagaagcctacaag 377

Pregunta: 362 aagtacatcaaagattacatgaaatcaatcaaagggaaacttgaagaacagagaccagaa 421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 378 aagtacatcaaagattacatgaaatcaatcaaagggaaacttgaagaacagagaccagaa 437
```

Figura 3c

```
Pregunta: 482 aaaaactaccagttctttattggtgaaaacatgaatccagatggcatggttgctctattg 541
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 498 aaaaactaccagttctttattggtgaaaacatgaatccagatggcatggttgctctattg 557

Pregunta: 542 gactaccgtgaggatggtgtgacccatatatgattttctttaaggatggtttagaaaatg 601
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 558 gactaccgtgaggatggtgtgacccatatatgattttctttaaggatggtttagaaaatg 617

Pregunta: 602 gaaaaatgttaacaaatgtggcaattattttggatctatcacctgtcatcataactggct 661
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 618 gaaaaatgttaacaaatgtggcaattattttggatctatcacctgtcatcataactggct 677

Pregunta: 662 tctgcttgtcatccacacaacaccaggacttaagacaaatgggactgatgtcatcttgag 721
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 678 tctgcttgtcatccacacaacaccaggacttaagacaaatgggactgatgtcatcttgag 737

Pregunta: 722 ctcttcatttattttgactgtgatttatttggagtggaggcattgttttaagaaaaaca 781
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 738 ctcttcatttattttgactgtgatttatttggagtggaggcattgttttaagaaaaaca 797

Pregunta: 782 tgtcatgtaggttgtctaaaaataaaatgcatttaaactcatttgagag 830
          ||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 798 tgtcatgtaggttgtctaaaaataaaatgcatttaaactcatttgagag 846
```

Figura 3d

>gi|45696213|emb|AL520698.3| AL520698 *Homo sapiens* NEUROBLASTOMA COT 10-normalizado *Homo sapiens* ADNc clon CS0D8002YF24 5-PRIMA, secuencia de ARNm
Longitud = 909

Puntuación = 1530 bits (828), Esperado = 0,0
Identidades = 828/829 (99%)
Hebra = +/+

```
Pregunta: 2    ccccccgagcgccgctccggctgcaccgcgctcgctccgagtttcaggctcgtgctaagc 61
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sujeto: 18     ccccccgagcgccgctccggctgcaccgcgctcgctccgagtttcaggctcgtgctaagc 77

Pregunta: 62   tagcgccgtcgtcgtctcccttcagtcgccatcatgattatctaccgggacctcatcagc 121
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sujeto: 78     tagcgccgtcgtcgtctcccttcagtcgccatcatgattatctaccgggacctcatcagc 137

Pregunta: 122  cacgatgagatgttctccgacatctacaagatccgggagatcgcggacgggttgtgcctg 181
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sujeto: 138    cacgatgagatgttctccgacatctacaagatccgggagatcgcggacgggttgtgcctg 197

Pregunta: 182  gaggtggaggggaagatggtcagtaggacagaaggtaacattgatgactcgctcattggt 241
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sujeto: 198    gaggtggaggggaagatggtcagtaggacagaaggtaacattgatgactcgctcattggt 257

Pregunta: 242  ggaaatgcctccgctgaaggcccgagggcgaaggtaccgaaagcacagtaatcactggt 301
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sujeto: 258    ggaaatgcctccgctgaaggcccgagggcgaaggtaccgaaagcacagtaatcactggt 317

Pregunta: 302  gtcgatattgtcatgaaccatcacctgcaggaaacaagtttcacaaaagaagcctacaag 361
                |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Sujeto: 318    gtcgatattgtcatgaaccatcacctgcaggaaacaagtttcacaaaaraagcctacaag 377

Pregunta: 362  aagtacatcaaagattacatgaaatcaatcaaagggaaacttgaagaacagagaccagaa 421
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sujeto: 378    aagtacatcaaagattacatgaaatcaatcaaagggaaacttgaagaacagagaccagaa 437
```

Figura 3e

```
Pregunta: 422 agagtaaaaccttttatgacaggggctgcagaacaaatcaagcacatccttgctaatttc 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 438 agagtaaaaccttttatgacaggggctgcagaacaaatcaagcacatccttgctaatttc 497

Pregunta: 482 aaaaactaccagttctttattggtgaaaacatgaatccagatggcatggttgctctattg 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 498 aaaaactaccagttctttattggtgaaaacatgaatccagatggcatggttgctctattg 557

Pregunta: 542 gactaccgtgaggatggtgtgaccccatatatgatttctttaaggatggtttagaaatg 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 558 gactaccgtgaggatggtgtgaccccatatatgatttctttaaggatggtttagaaatg 617

Pregunta: 602 gaaaaatgttaacaaatgtggcaattatttggatctatcacctgtcatcataactggct 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 618 gaaaaatgttaacaaatgtggcaattatttggatctatcacctgtcatcataactggct 677

Pregunta: 662 tctgcttgtcatccacacaacaccaggacttaagacaaatgggactgatgtcatcttgag 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 678 tctgcttgtcatccacacaacaccaggacttaagacaaatgggactgatgtcatcttgag 737

Pregunta: 722 ctcttcatttatttgactgtgatttatttggagtggaggcattgttttaagaaaaaca 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 738 ctcttcatttatttgactgtgatttatttggagtggaggcattgttttaagaaaaaca 797

Pregunta: 782 tgtcatgtaggttgtctaaaaataaaatgcatttaaactcatttgagag 830
              |||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 798 tgtcatgtaggttgtctaaaaataaaatgcatttaaactcatttgagag 846
```

Figura 3f

>gi|10105790|gb|BE717525.1| RC4-HT0778-190700-014-e04 HT0778 *Homo sapiens* ADNc, secuencia de ARNm
Longitud = 341

Puntuación = 564 bits (305), Esperado = e-158
Identidades = 314/318 (98%), Huecos = 1/318 (0%)
Hebra = +/+

```
Pregunta: 347 aaagaagcctacaagaagtacatcaaagattacatgaaatcaatcaaagggaaacttgaa 406
          ||||||||  ||||||||||||||||||||  ||||||||||||||||||||||||||||||
  Sujeto: 12  aaagaagcatacaagaagtacatcaaaga-tacatgaaatcaatcaaagggaaacttgaa 70

Pregunta: 407 gaacagagaccagaaagagtaaaacctttatgacaggggctgcagaacaaatcaagcac 466
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 71  gaacagagaccagaaagagtaaaacctttatgacaggggctgcagaacaaatcaagcac 130

Pregunta: 467 atccttgctaatttcaaaaactaccagttctttattggtgaaaacatgaatccagatggc 526
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 131 atccttgctaatttcaaaaactaccagttctttattggtgaaaacatgaatccagatggc 190

Pregunta: 527 atggttgctctattggactaccgtgaggatggtgtgaccccatatatgattttctttaaag 586
          ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 191 atggttgctctattgggctaccgtgaggatggtgtgaccccatatatgattttctttaag 250

Pregunta: 587 gatggtttagaaatggaaaaatgttaacaaatgtggcaattattttggatctatcacctg 646
          ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||
  Sujeto: 251 gatggtttagaaatggaaaaatgttaacaaatgtggcaattatttgggatctatcacctg 310

Pregunta: 647 tcatcataactggcttct 664
          ||||||||||||||||||
  Sujeto: 311 tcatcataactggcttct 328
```

Figura 3g

>gi|10224966|gb|BE771308.1| QV1-FT0082-210600-249-a12 FT0082 *Homo sapiens* ADNc, secuencia de ARNm
Longitud = 338

Puntuación = 562 bits (304), Esperado = e-157
Identidades = 316/321 (98%), Huecos = 3/321 (0%)
Hebra = +/+

```
Pregunta: 346 aaaagaagcctacaagaa-gtacatca-aagattacatgaaatcaatcaaagggaaacct 403
            ||||||||||||||||| ||||||||  |||||||| ||||||||||||||||||||
  Sujeto: 6   aaaagaagcctacaagaacgtacatcagttgattacatg-aatcaatcaaagggaaacct 64

Pregunta: 404 gaagaacagagaccagaaagagtaaaacctttatgacaggggctgcagaacaaatcaag 463
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 65  gaagaacagagaccagaaagagtaaaacctttatgacaggggctgcagaacaaatcaag 124

Pregunta: 464 cacatccttgctaatttcaaaaactaccagttctttattggtgaaaacatgaatccagat 523
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 125 cacatccttgctaatttcaaaaactaccagttctttattggtgaaaacatgaatccagat 184

Pregunta: 524 ggcatggttgctctattggactaccgtgaggatggtgtgaccccatatatgattttcttt 583
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 185 ggcatggttgctctattggactaccgtgaggatggtgtgaccccatatatgattttcttt 244

Pregunta: 584 aaggatggtttagaaatggaaaaatgttaacaaatgtggcaattatttggatctatcac 643
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 245 aaggatggtttagaaatggaaaaatgttaacaaatgtggcaattatttggatctatcac 304

Pregunta: 644 ctgtcatcataactggcttct 664
            ||||||||||||||||||||
  Sujeto: 305 ctgtcatcataactggcttct 325
```

Figura 3h

>gi|10246833|gb|BE814599.1| MR0-BN0070-070800-033-b04_1 BN0070 *Homo sapiens* ADNc, secuencia de ARNm
Longitud = 140

Puntuación = 259 bits (140), Esperado = 2e-066
Identidades = 140/140 (100%)
Hebra = +/+

```
Pregunta: 451 agaacaaatcaagcacatccttgctaatttcaaaaactaccagttctttattggtgaaaa 510
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 1   agaacaaatcaagcacatccttgctaatttcaaaaactaccagttctttattggtgaaaa 60

Pregunta: 511 catgaatccagatggcatggttgctctattggactaccgtgaggatggtgtgaccccata 570
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Sujeto: 61  catgaatccagatggcatggttgctctattggactaccgtgaggatggtgtgaccccata 120

Pregunta: 571 tatgattttctttaaggatg 590
          ||||||||||||||||||||
  Sujeto: 121 tatgattttctttaaggatg 140
```

Figura 3i

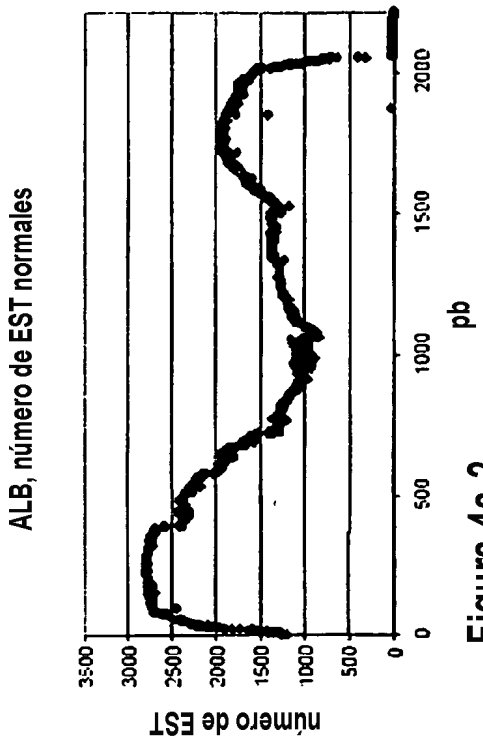


Figura 4a-1

Figura 4a-2

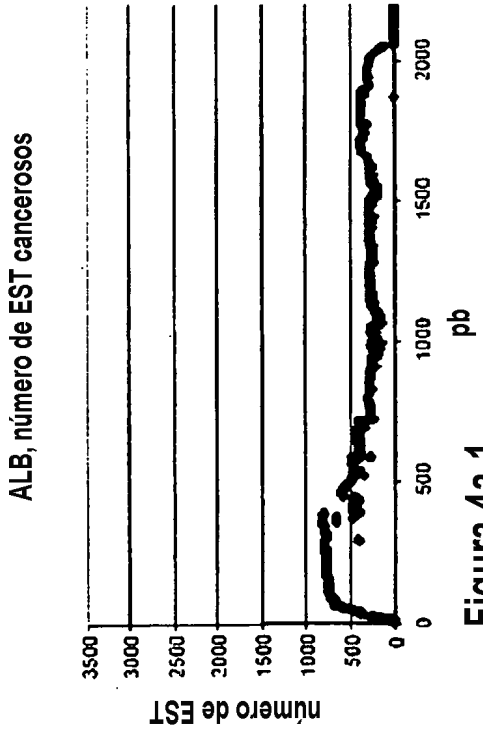


Figura 4a-3

Figura 4a-4

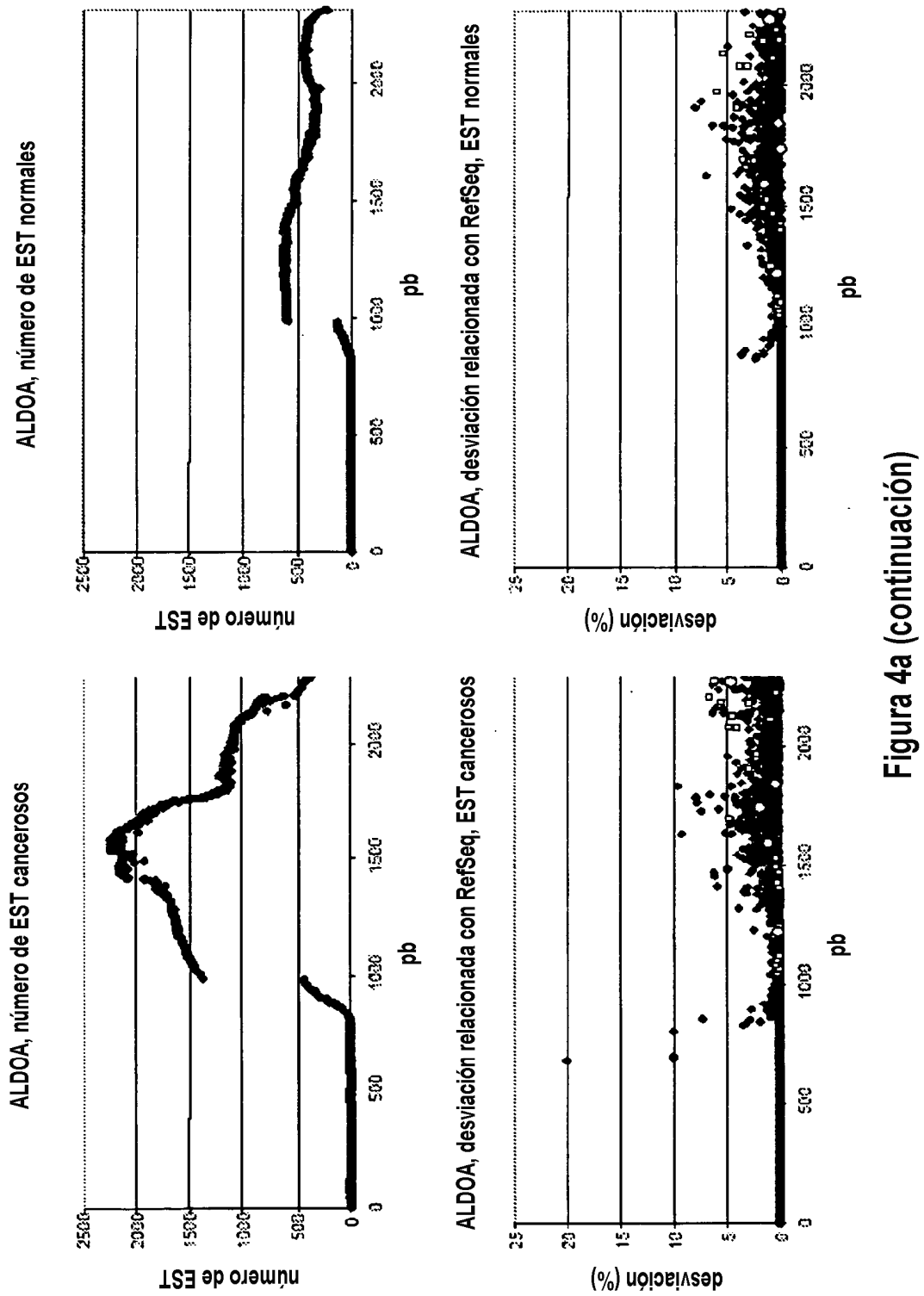


Figura 4a (continuación)

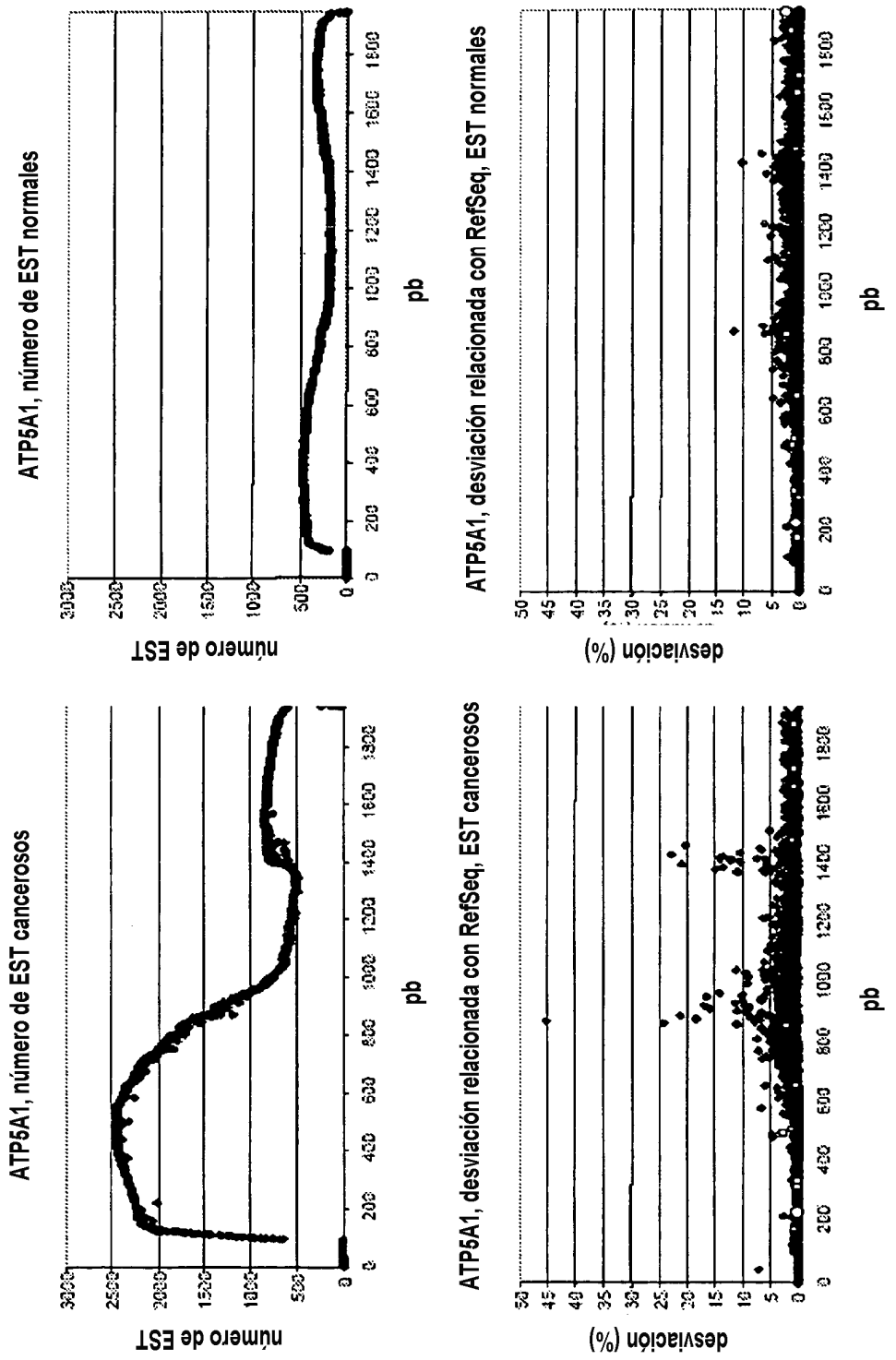


Figura 4a (continuación)

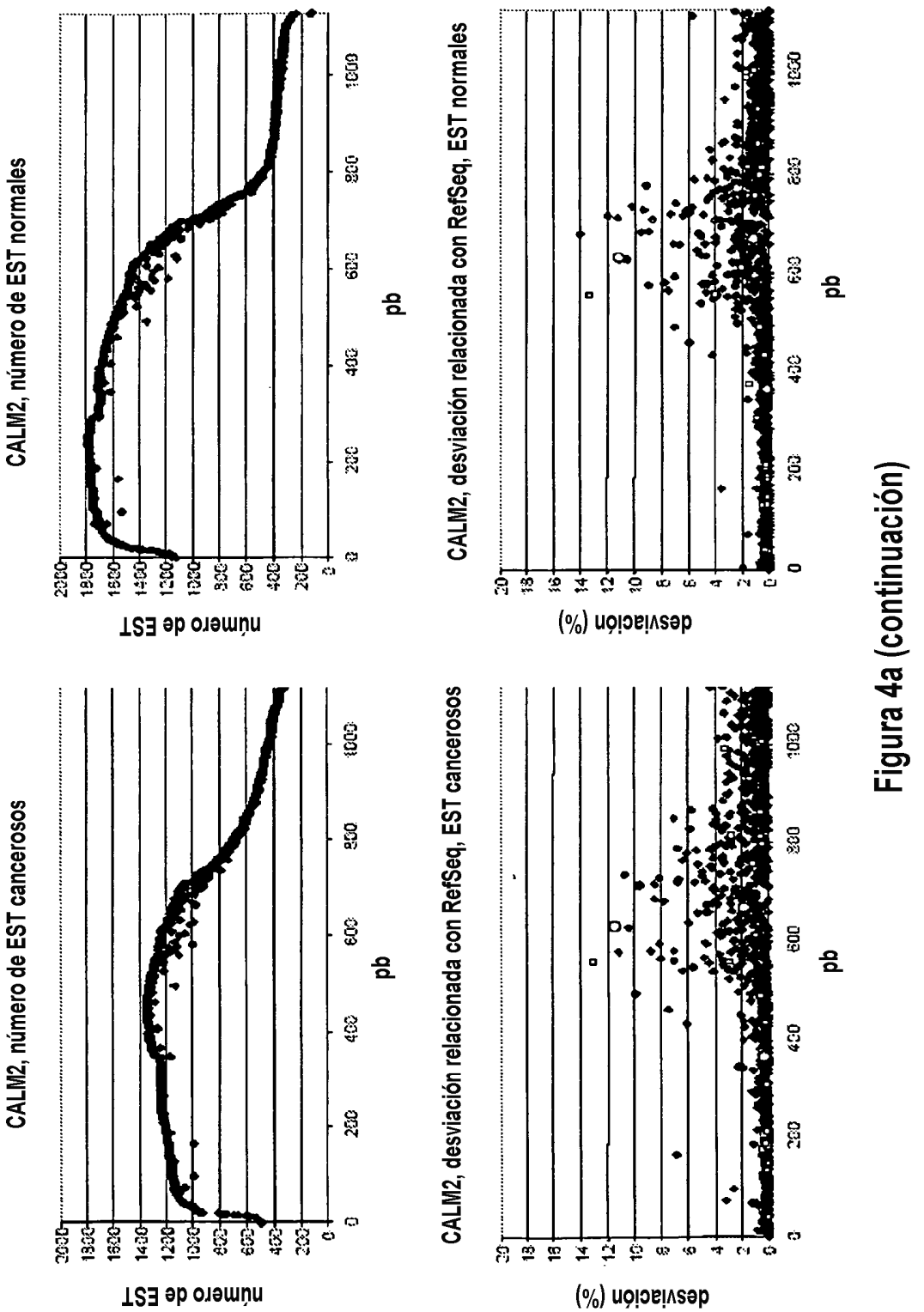


Figura 4a (continuación)

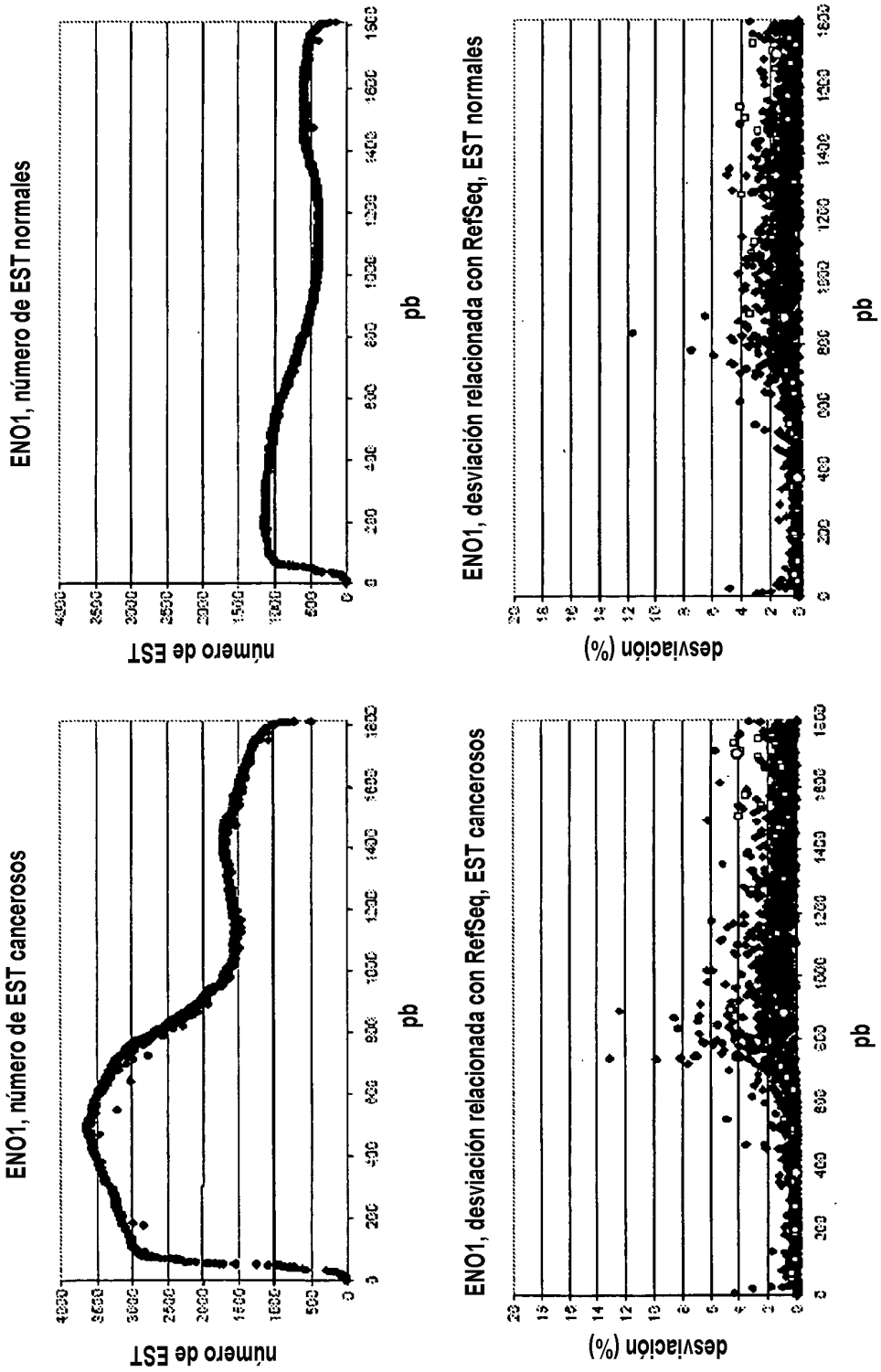

ENO1, número de EST normales
número de EST
pb
ENO1, desviación relacionada con RefSeq, EST normales
desviación (%)
pb
ENO1, número de EST cancerosos
número de EST
pb
ENO1, desviación relacionada con RefSeq, EST cancerosos
desviación (%)
pb

Figura 4b

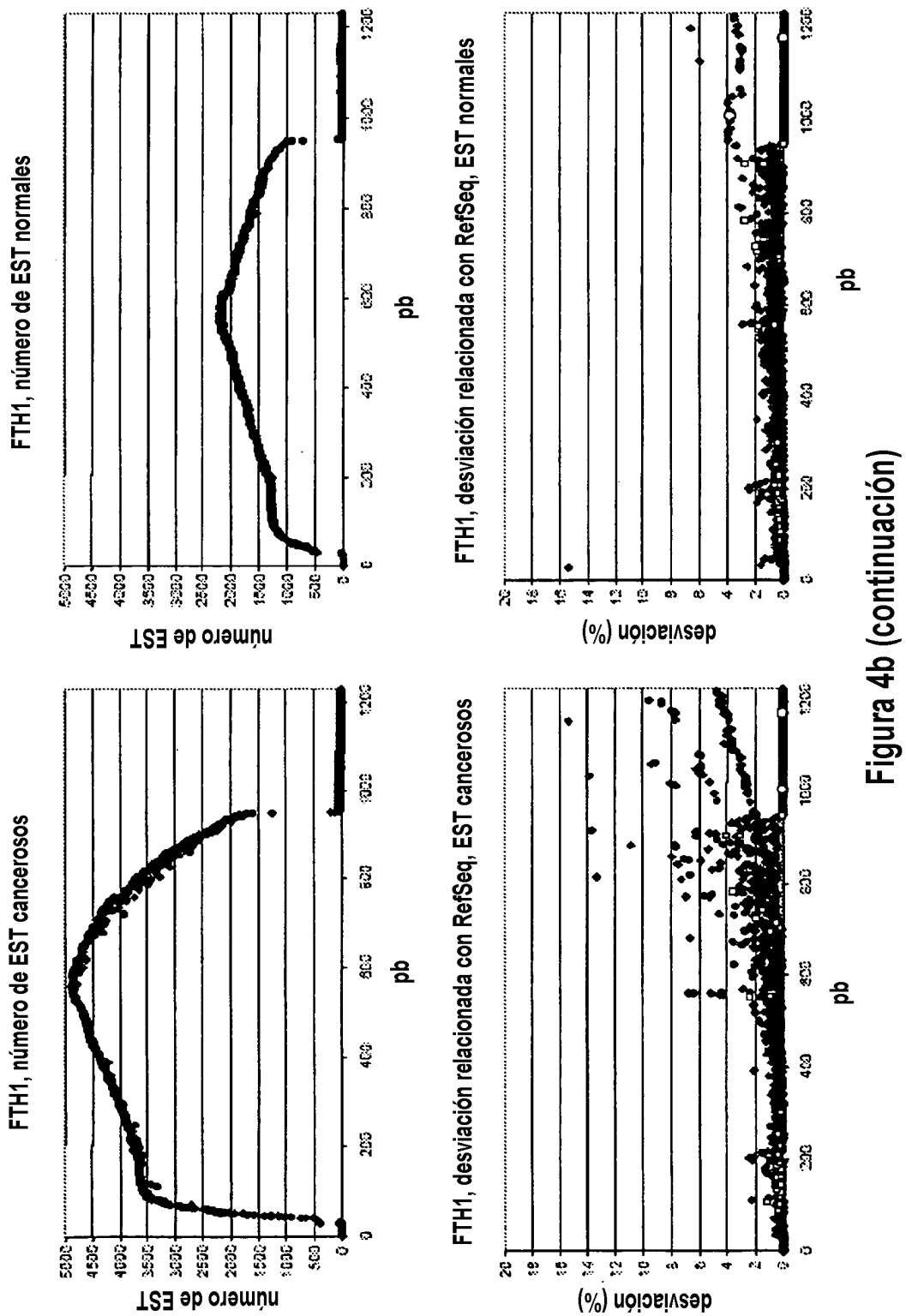


Figura 4b (continuación)

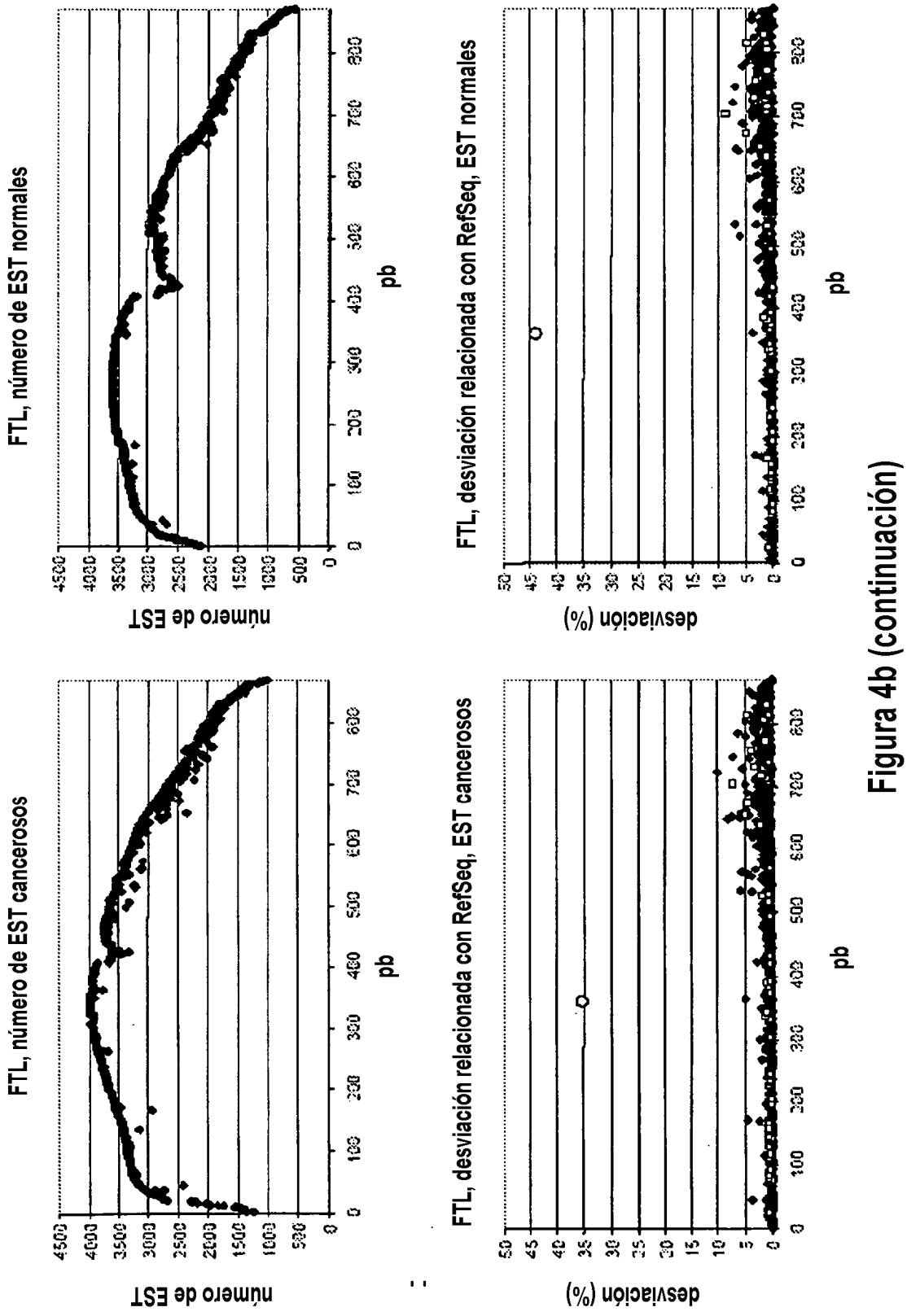


Figura 4b (continuación)

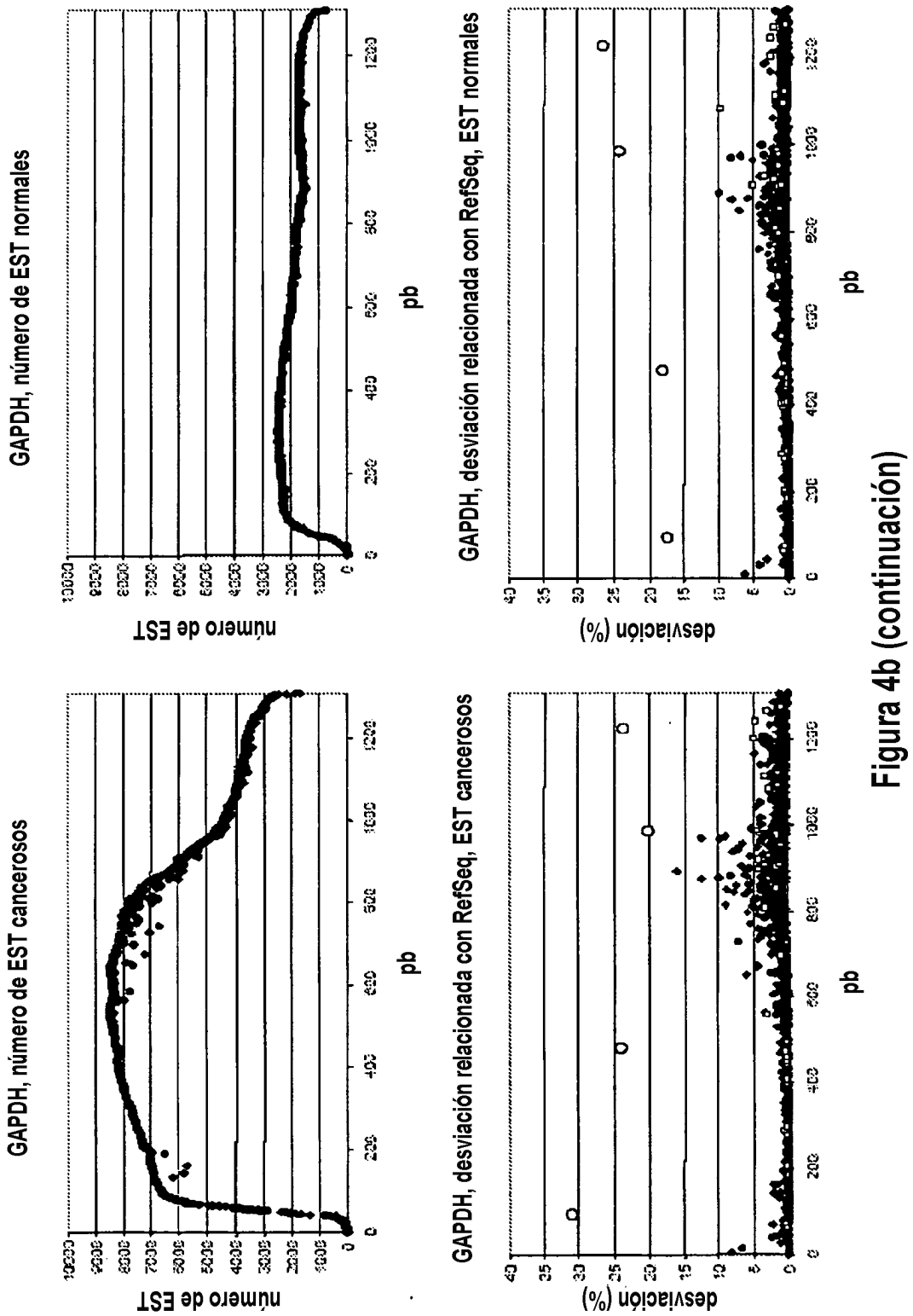


Figura 4b (continuación)

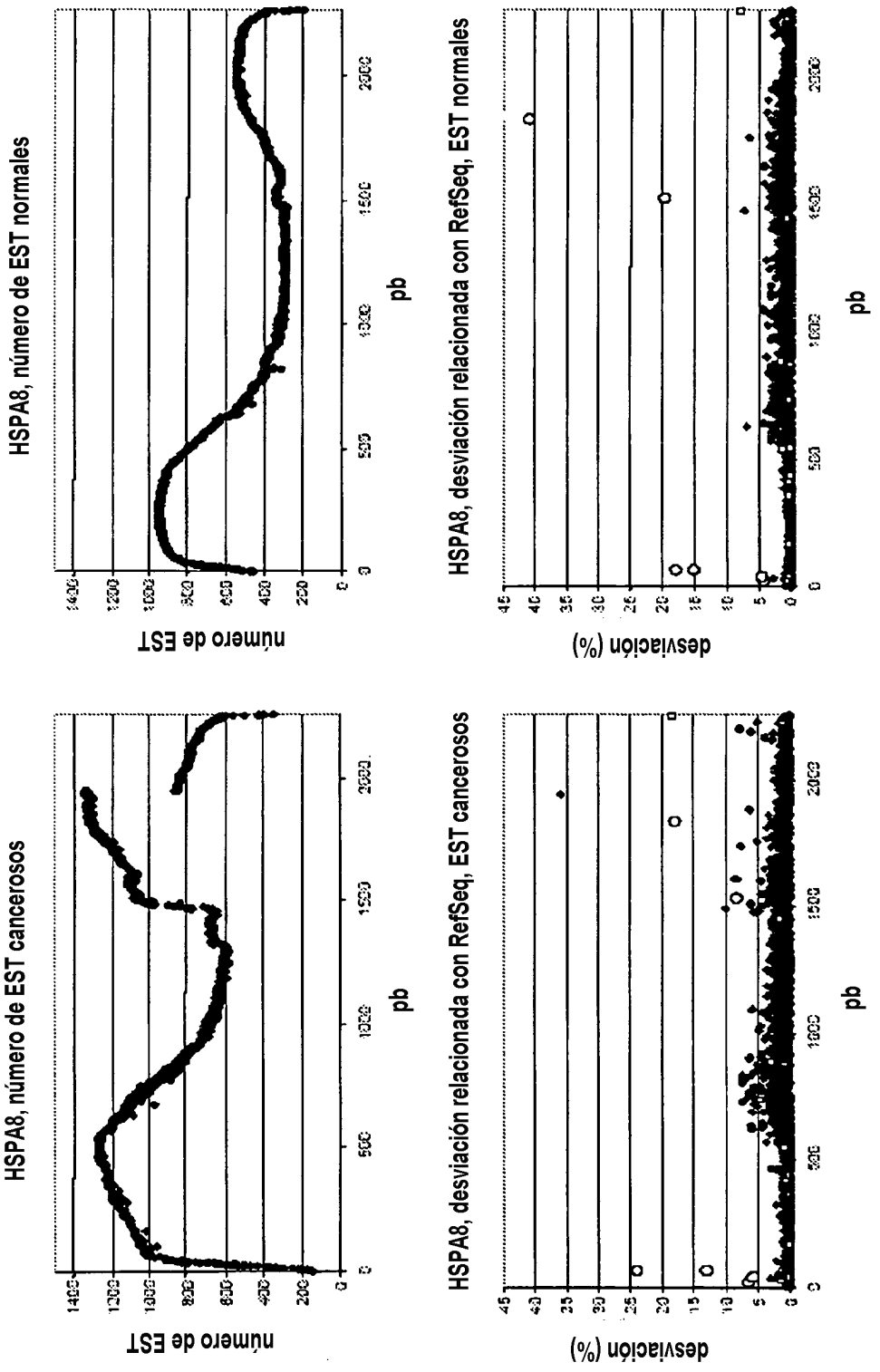

HSPA8, número de EST normales
número de EST
pb
HSPA8, desviación relacionada con RefSeq, EST normales
desviación (%)
pb
HSPA8, número de EST cancerosos
número de EST
pb
HSPA8, desviación relacionada con RefSeq, EST cancerosos
desviación (%)
pb

Figura 4c

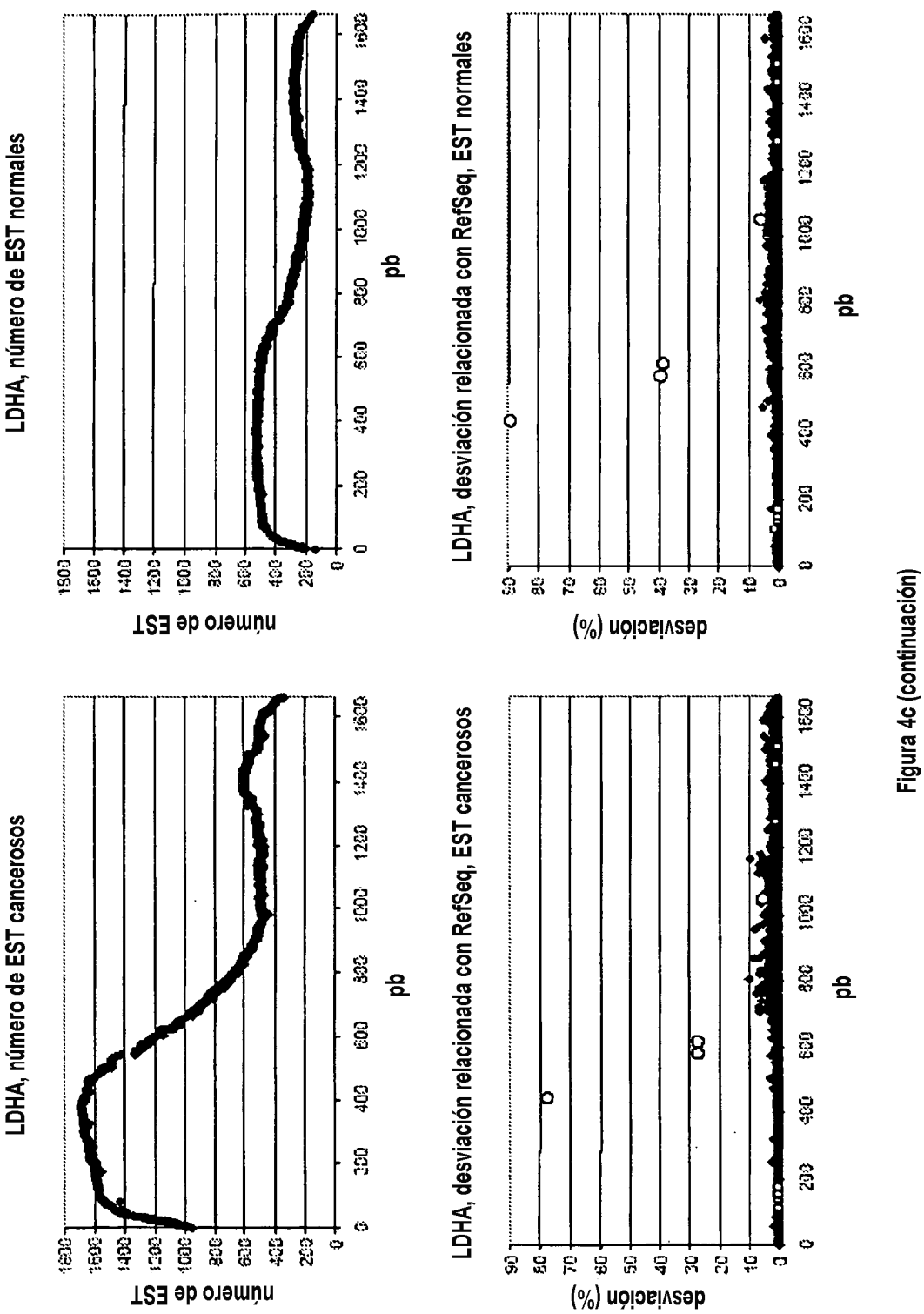


Figura 4c (continuación)

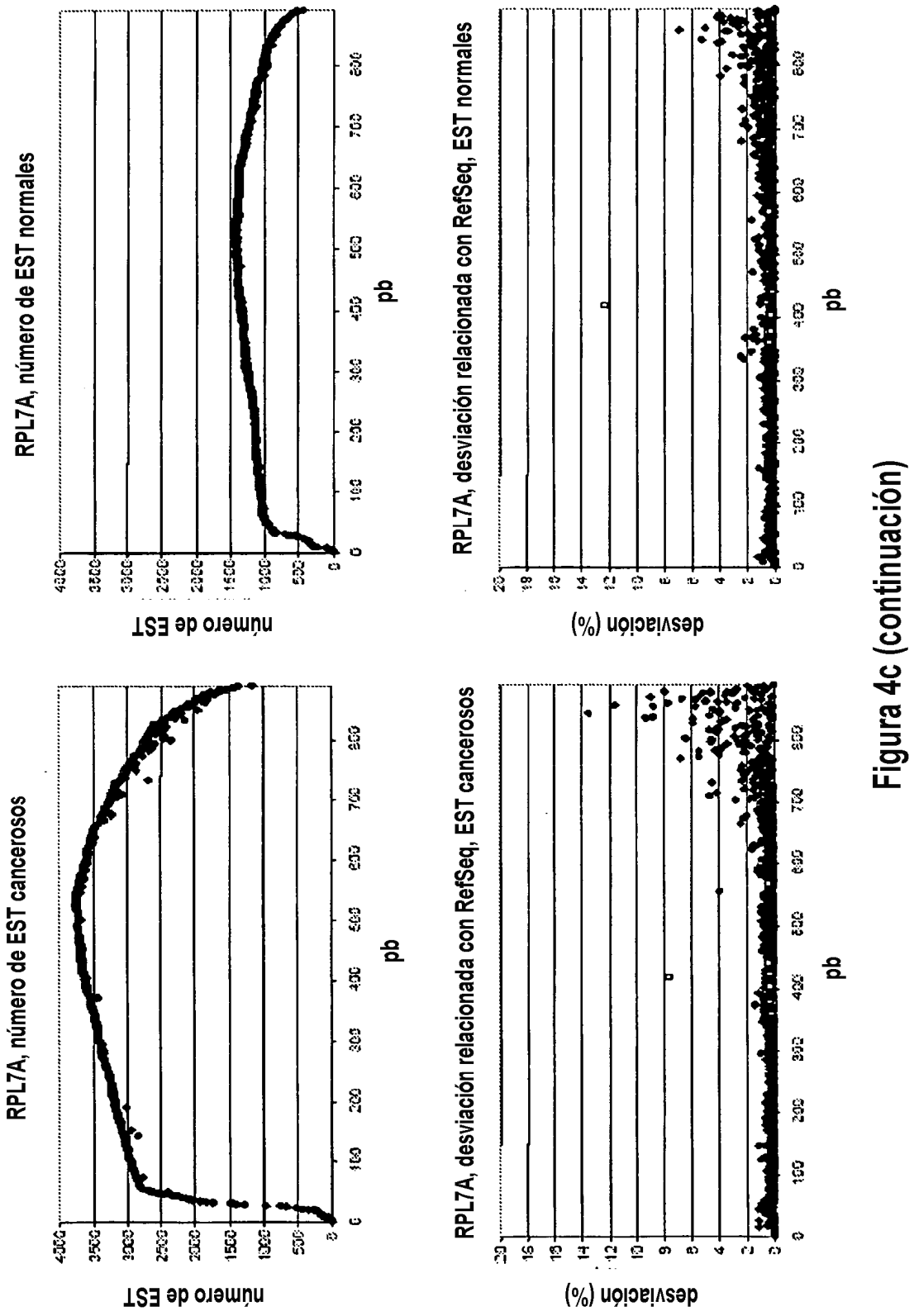


Figura 4c (continuación)

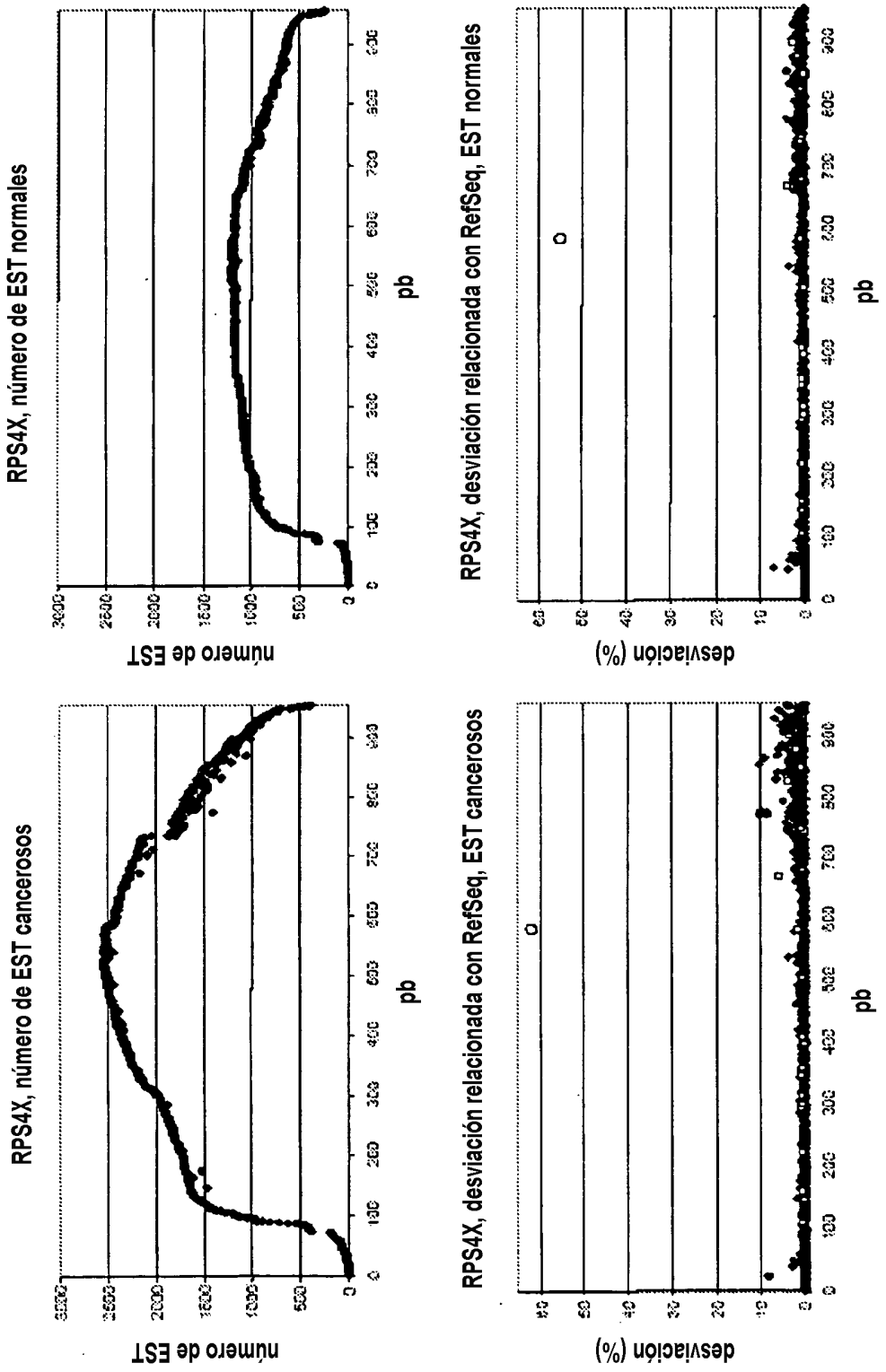


Figura 4c (continuación)

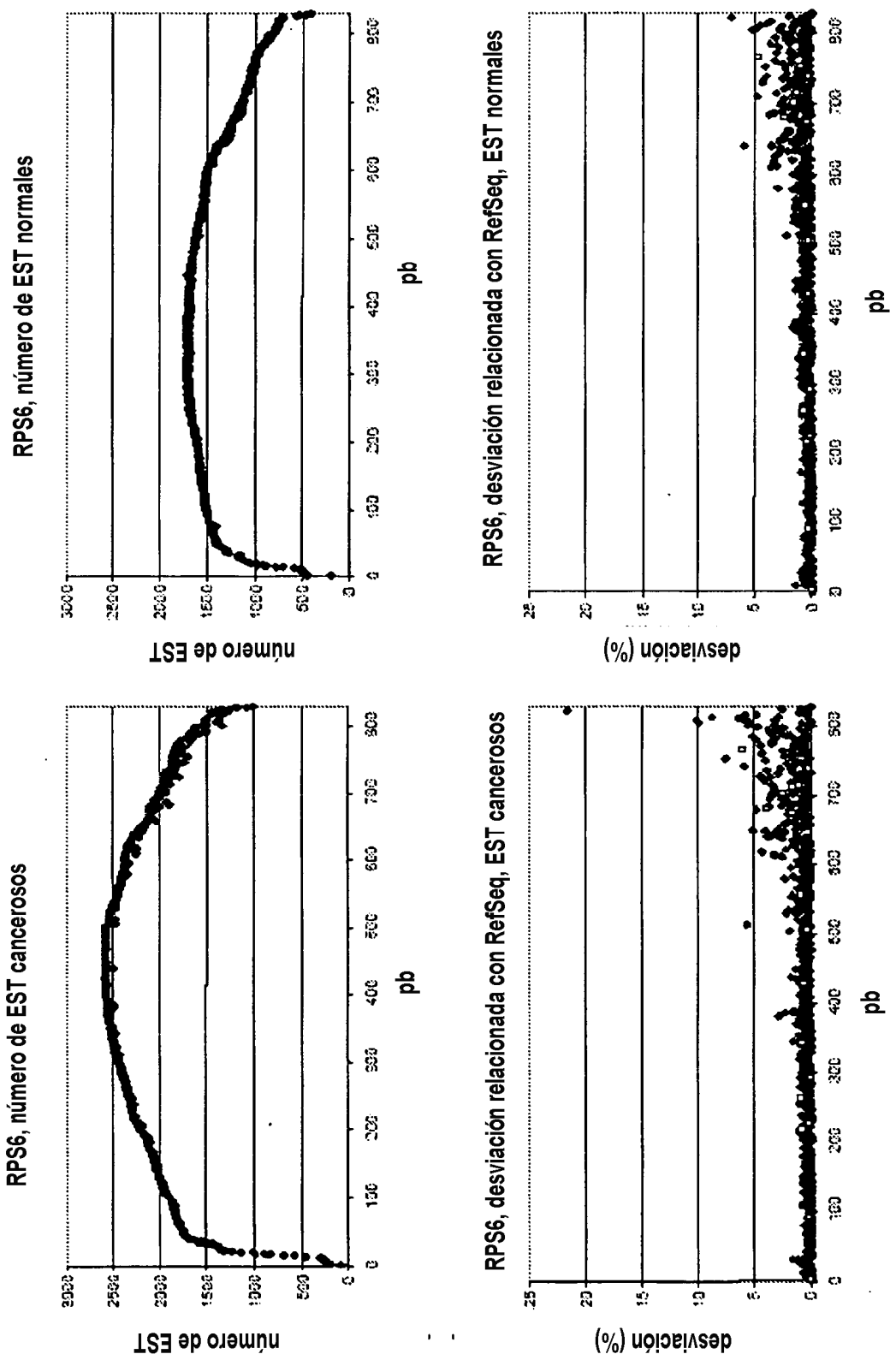

RPS6, número de EST normales
número de EST
pb
RPS6, desviación relacionada con RefSeq, EST normales
desviación (%)
pb
RPS6, número de EST cancerosos
número de EST
pb
RPS6, desviación relacionada con RefSeq, EST cancerosos
desviación (%)
pb

Figura 4d

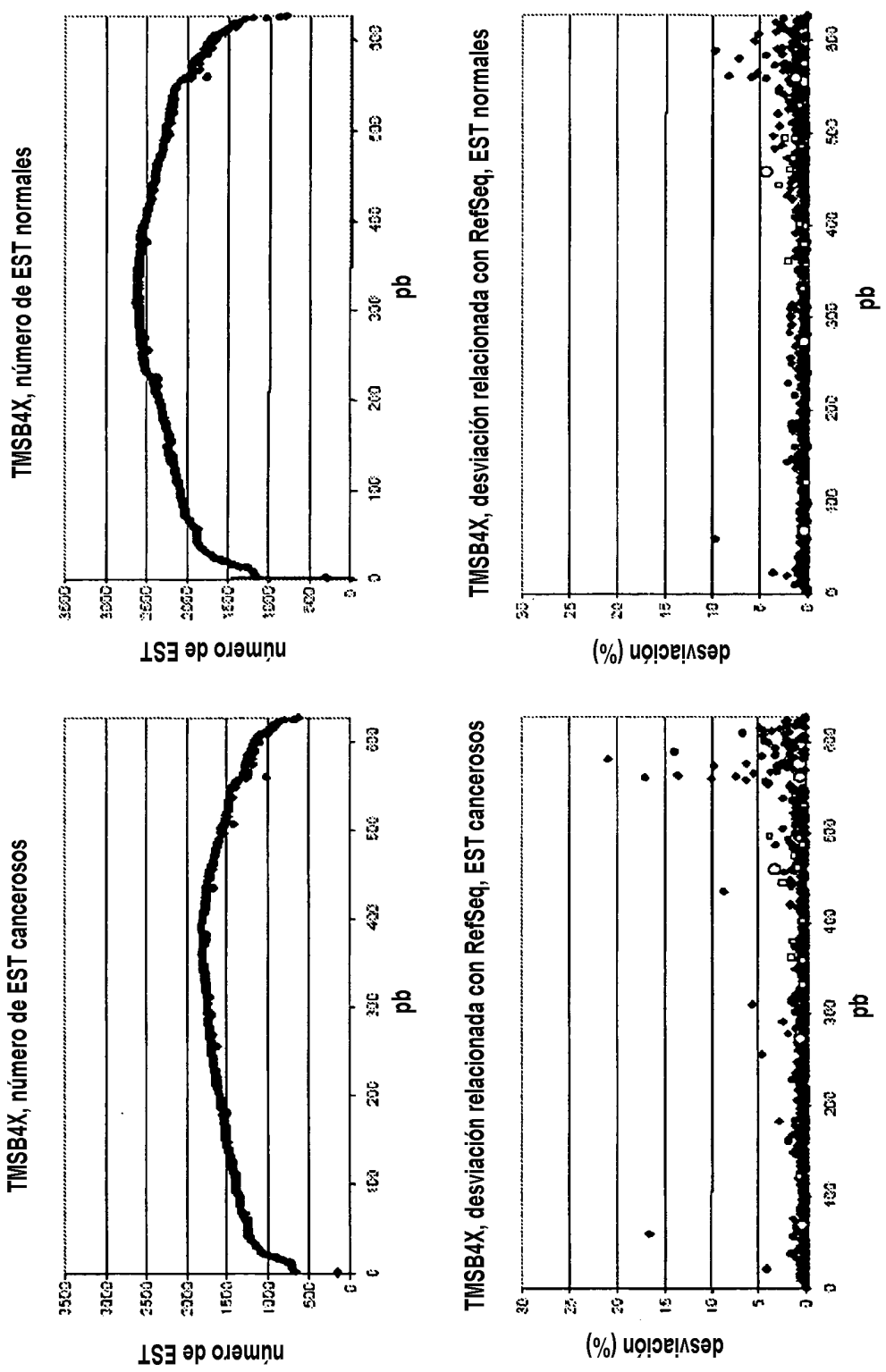


Figura 4d (continuación)

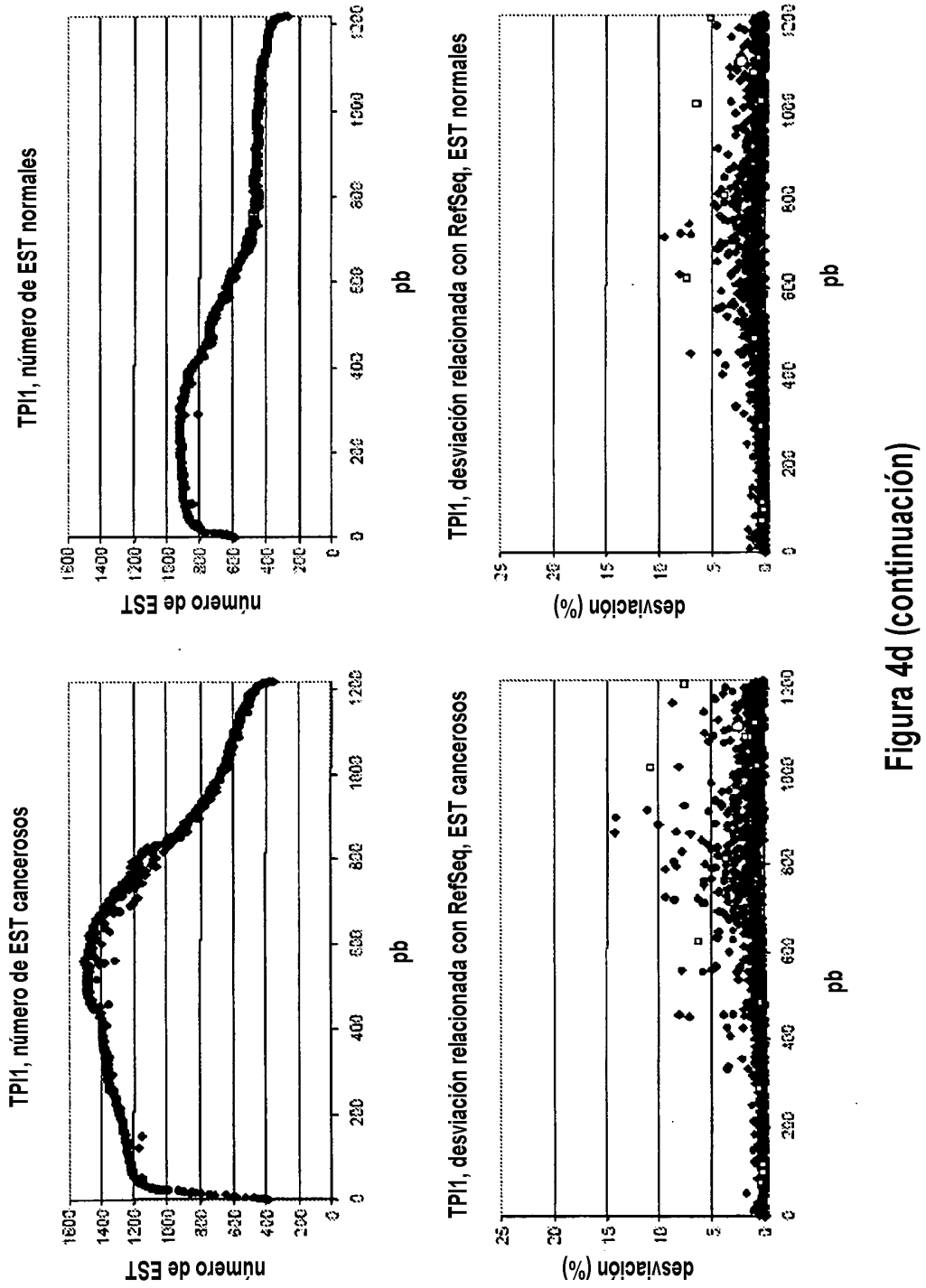


Figura 4d (continuación)

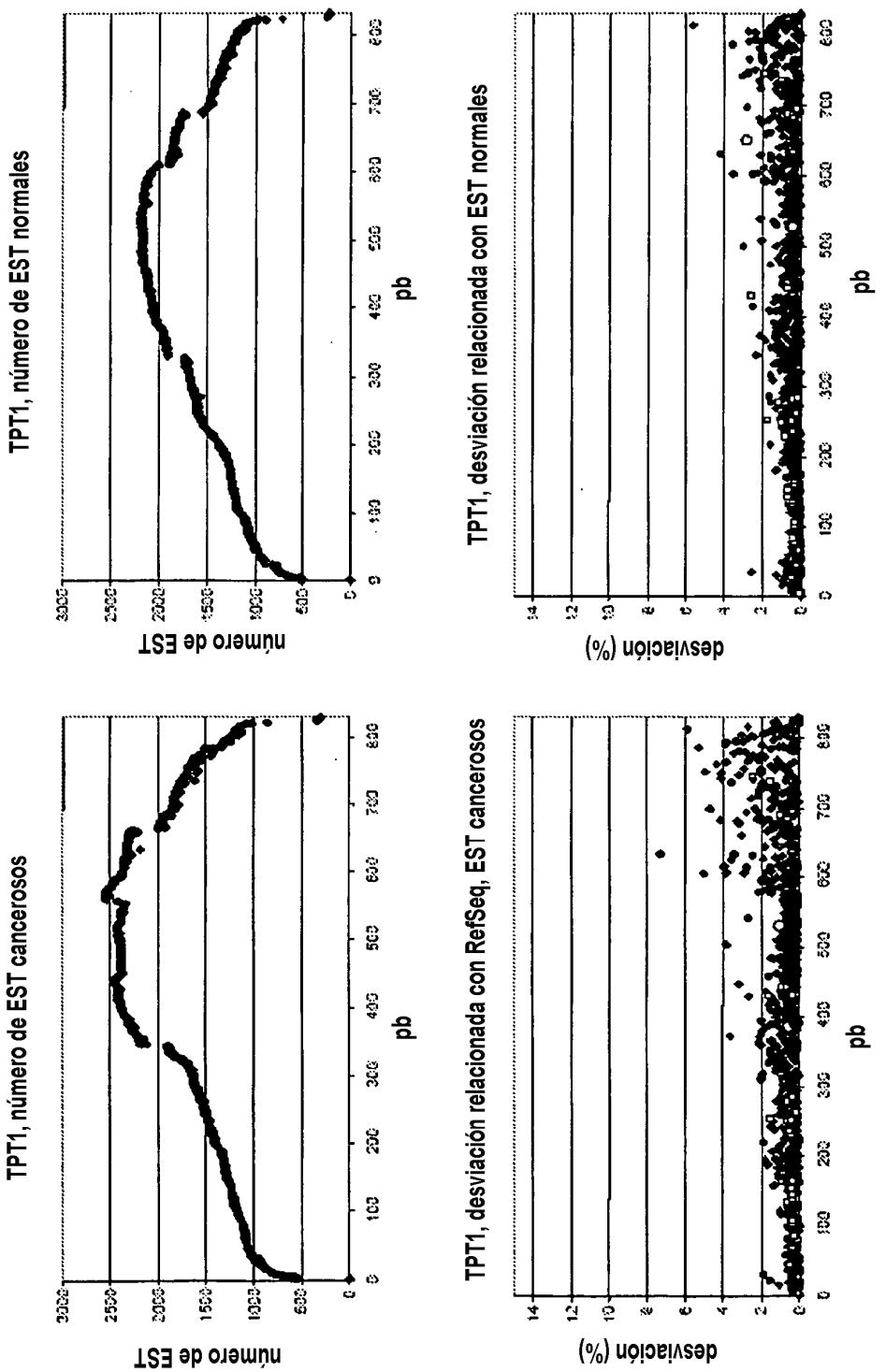


Figura 4d (continuación)

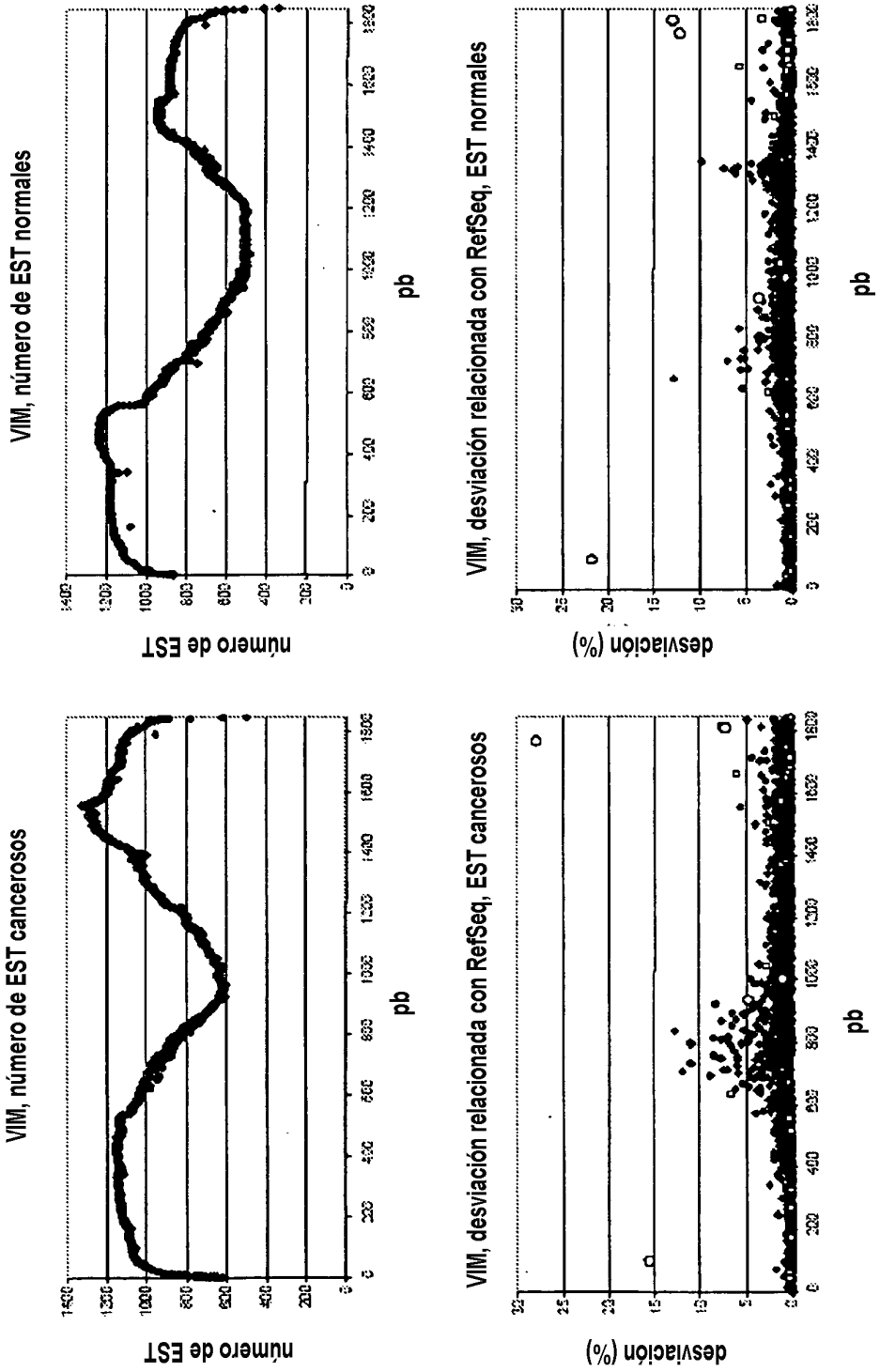


Figura 4d (continuación)

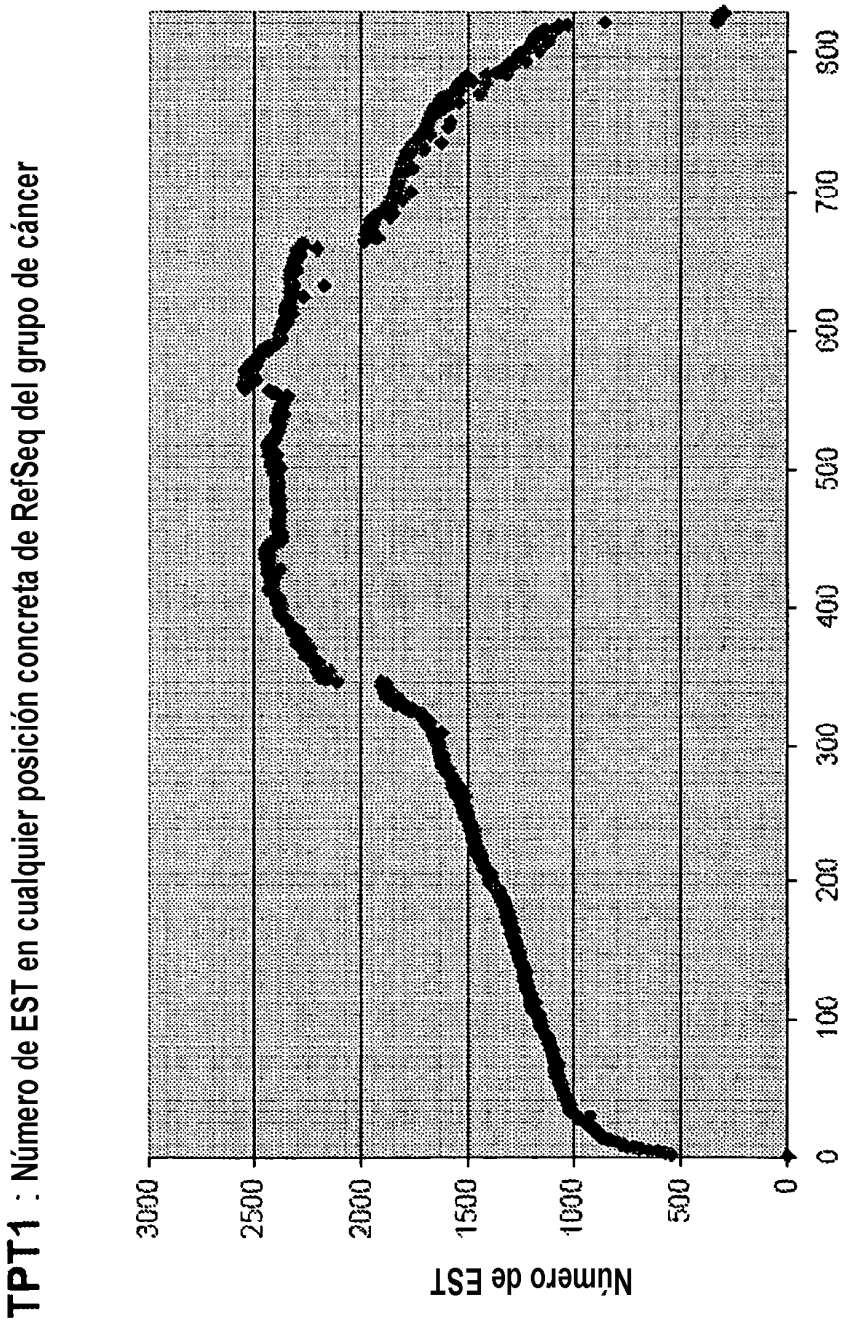

TPT1 : Número de EST en cualquier posición concreta de RefSeq del grupo de cáncer
Número de EST
3000
2500
2000
1500
1000
500
0
0
100
200
300
400
500
600
700
800
pb

Figura 5a

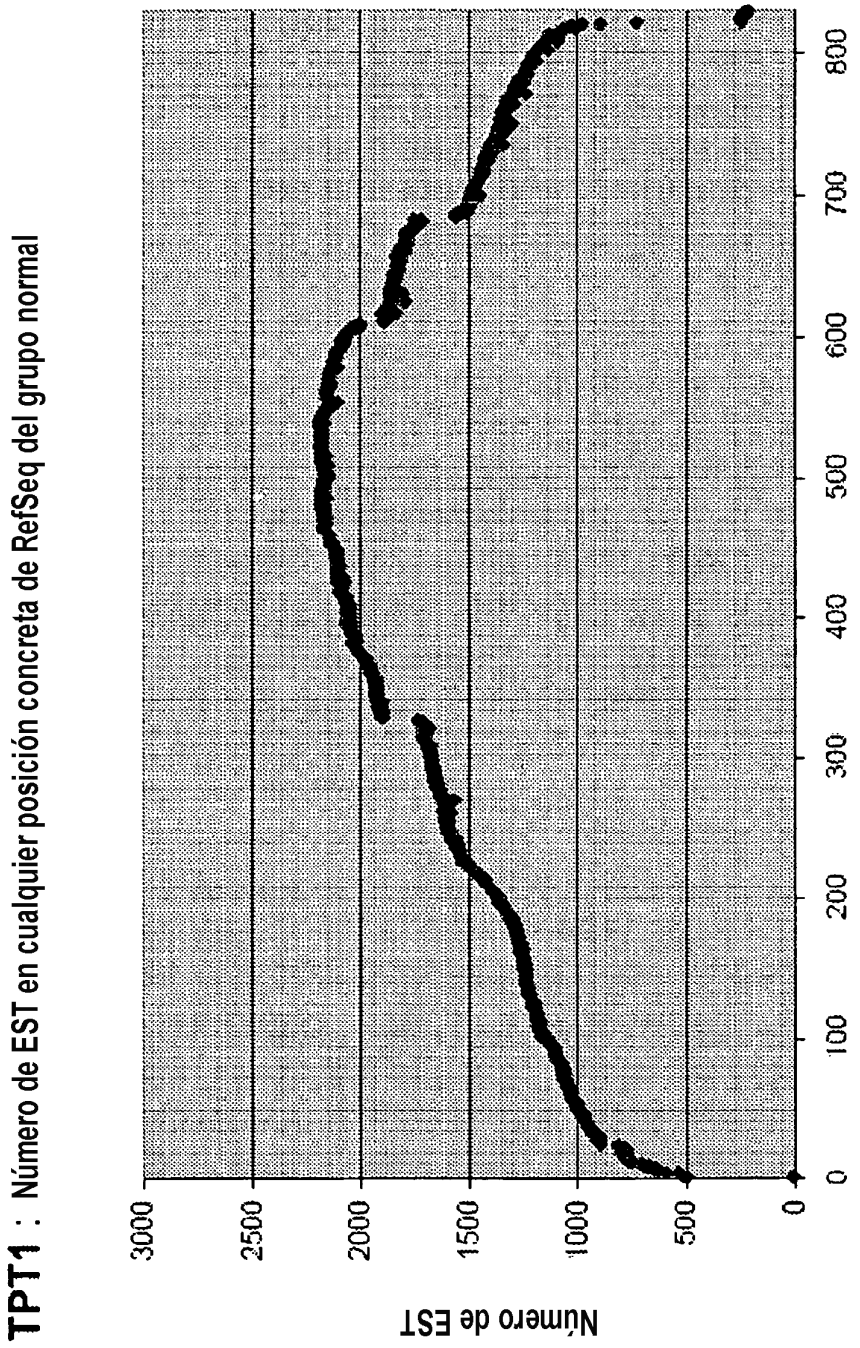

TPT1 : Número de EST en cualquier posición concreta de RefSeq del grupo normal
Número de EST
3000
2500
2000
1500
1000
500
0
0
100
200
300
400
500
600
700
800
pb

Figura 5b

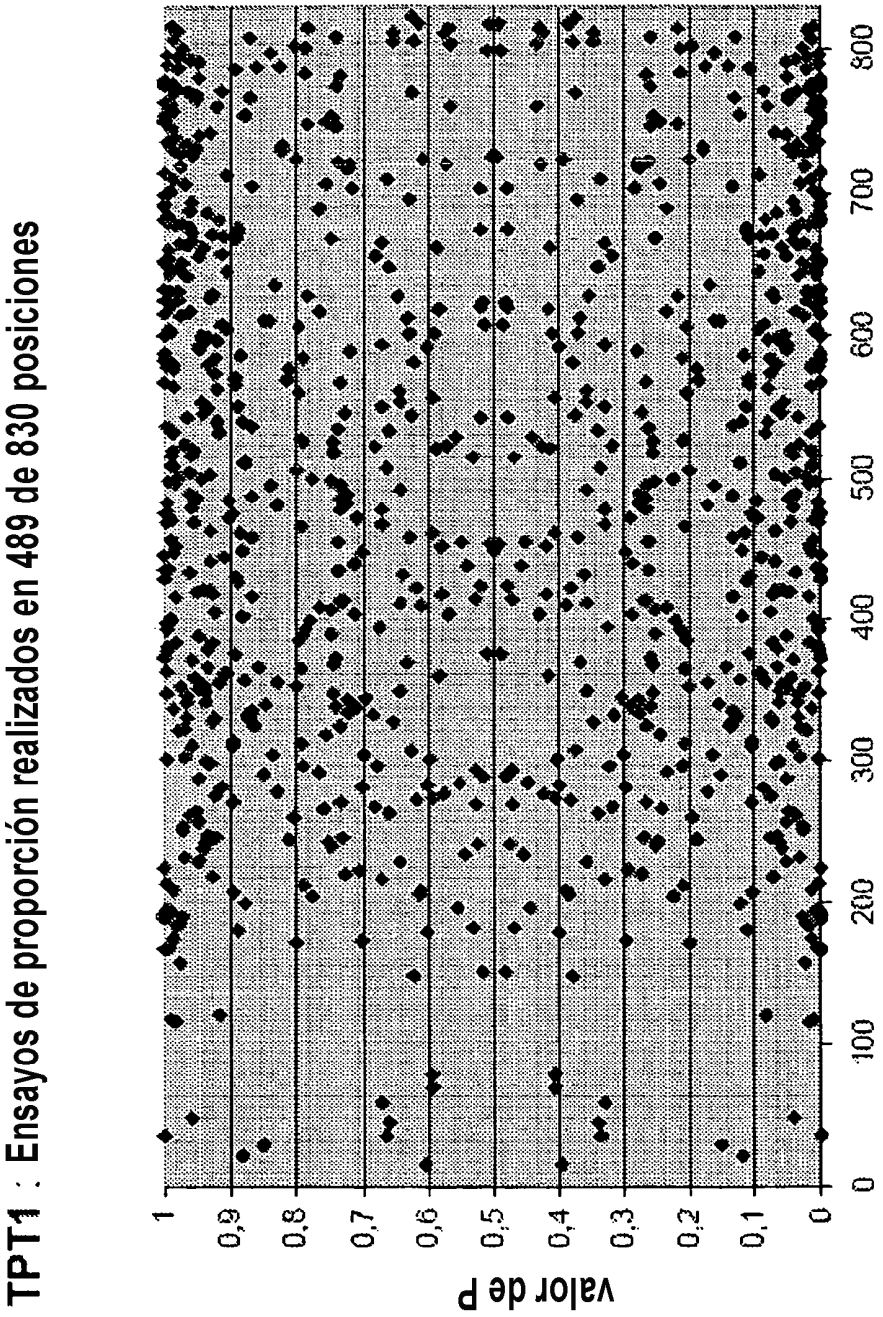

TPT1 : Ensayos de proporción realizados en 489 de 830 posiciones
valor de P
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
0
100
200
300
400
500
600
700
800
pb

Figura 5c

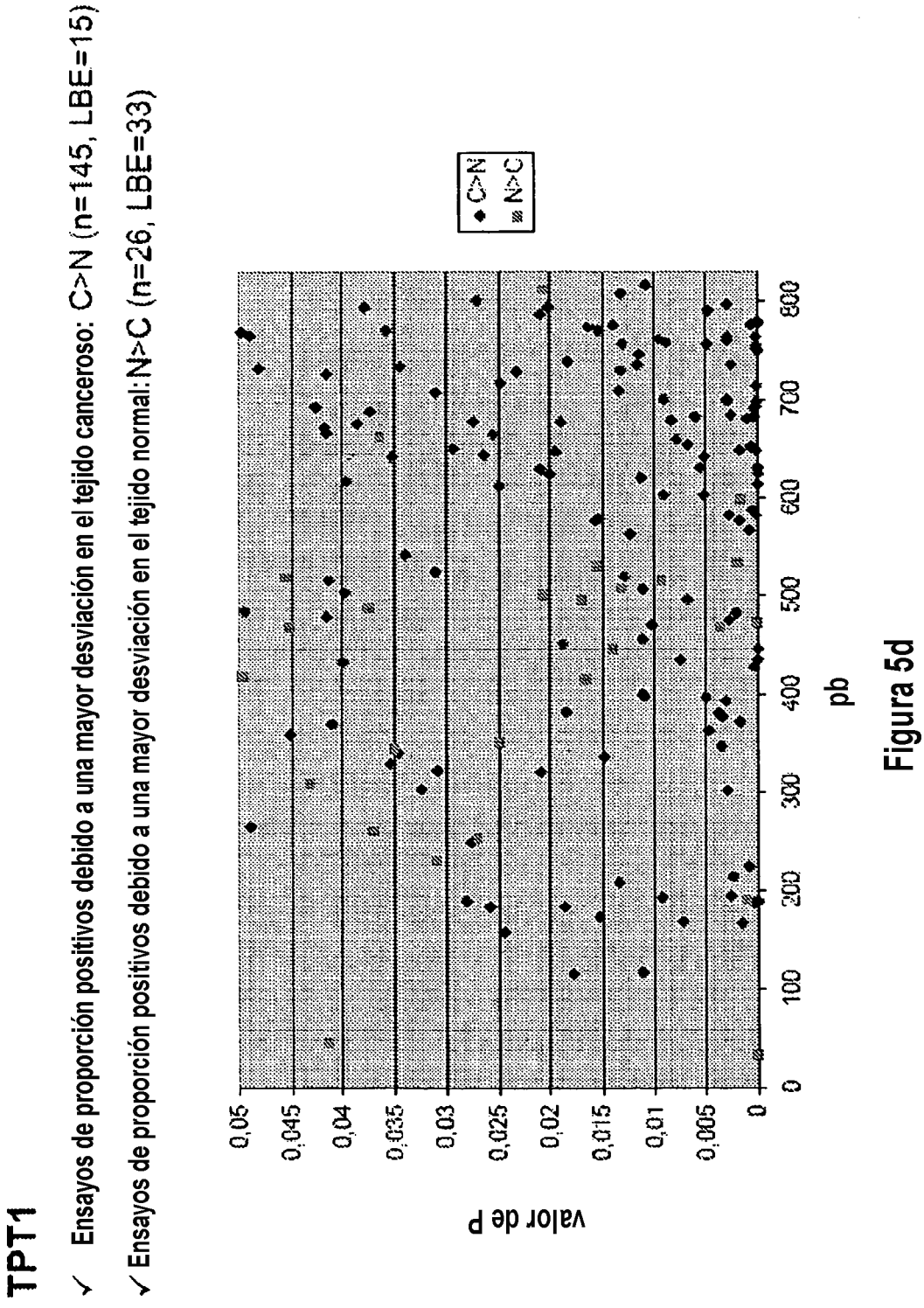

TPT1
✓ Ensayos de proporción positivos debido a una mayor desviación en el tejido canceroso: C>N (n=145, LBE=15)
✓ Ensayos de proporción positivos debido a una mayor desviación en el tejido normal: N>C (n=26, LBE=33)
C>N
N>C
valor de P
0,05
0,045
0,04
0,035
0,03
0,025
0,02
0,015
0,01
0,005
0
0
100
200
300
400
500
600
700
800
pb

Figura 5d

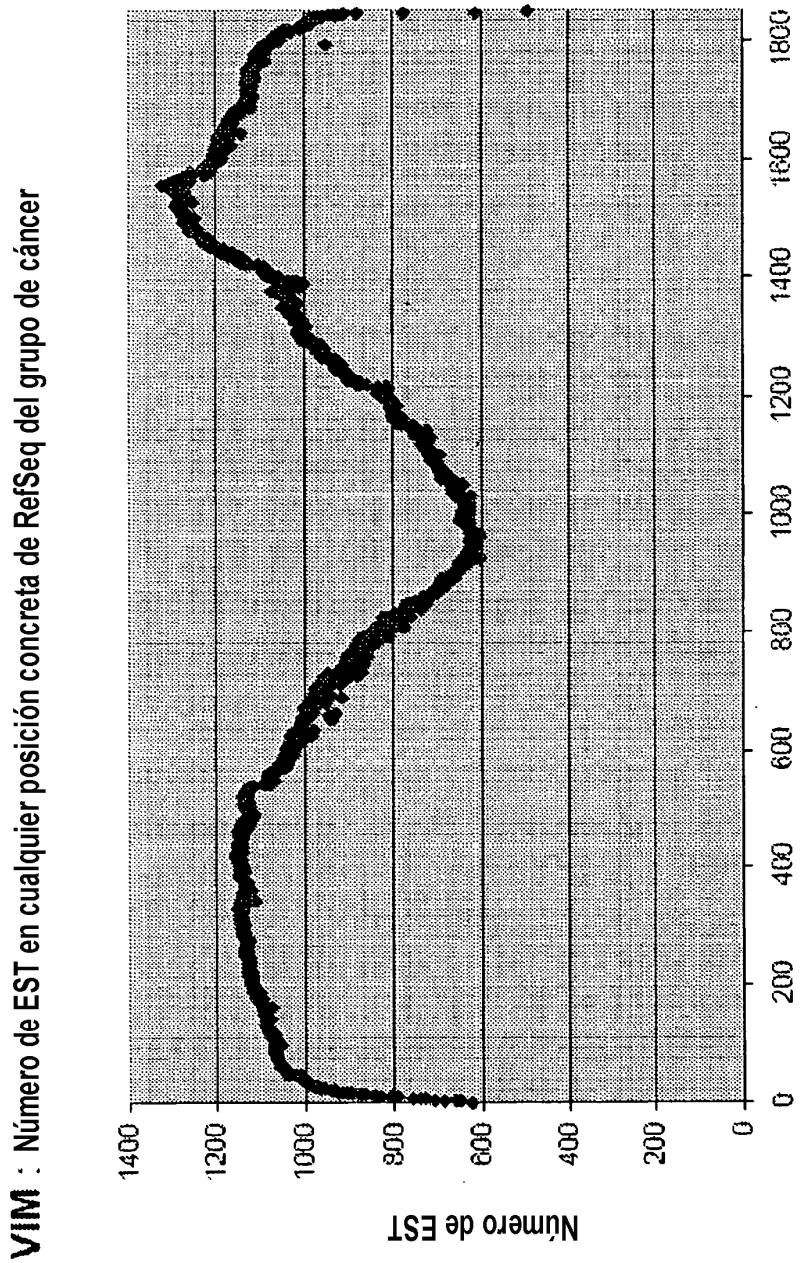

VIM : Número de EST en cualquier posición concreta de RefSeq del grupo de cáncer
Número de EST
0
200
400
600
800
1000
1200
1400
pb
0
200
400
600
800
1000
1200
1400
1600
1800

Figura 5e

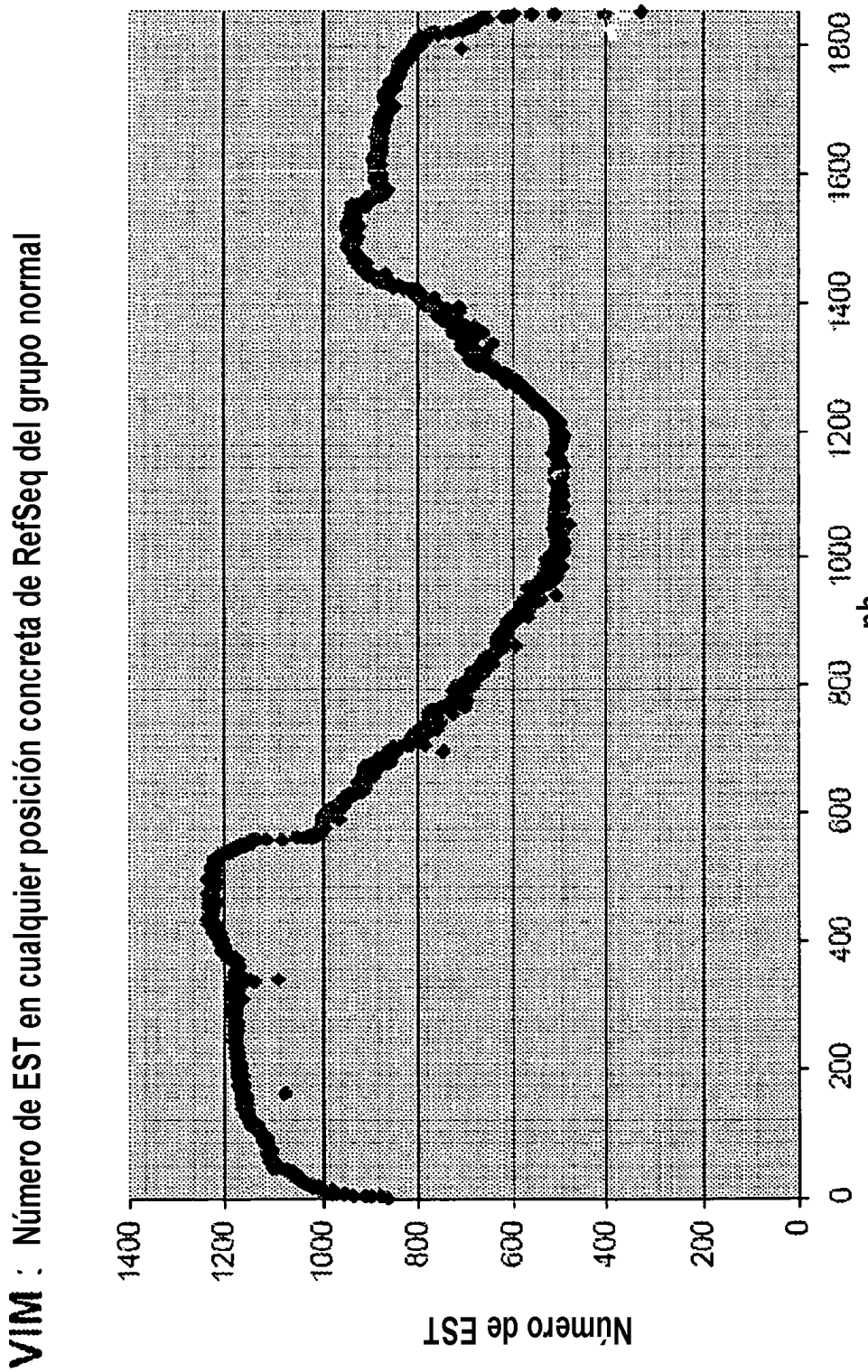

VIM : Número de EST en cualquier posición concreta de RefSeq del grupo normal
Número de EST
0
200
400
600
800
1000
1200
1400
0
200
400
600
800
1000
1200
1400
1600
1800
pb

Figura 5f

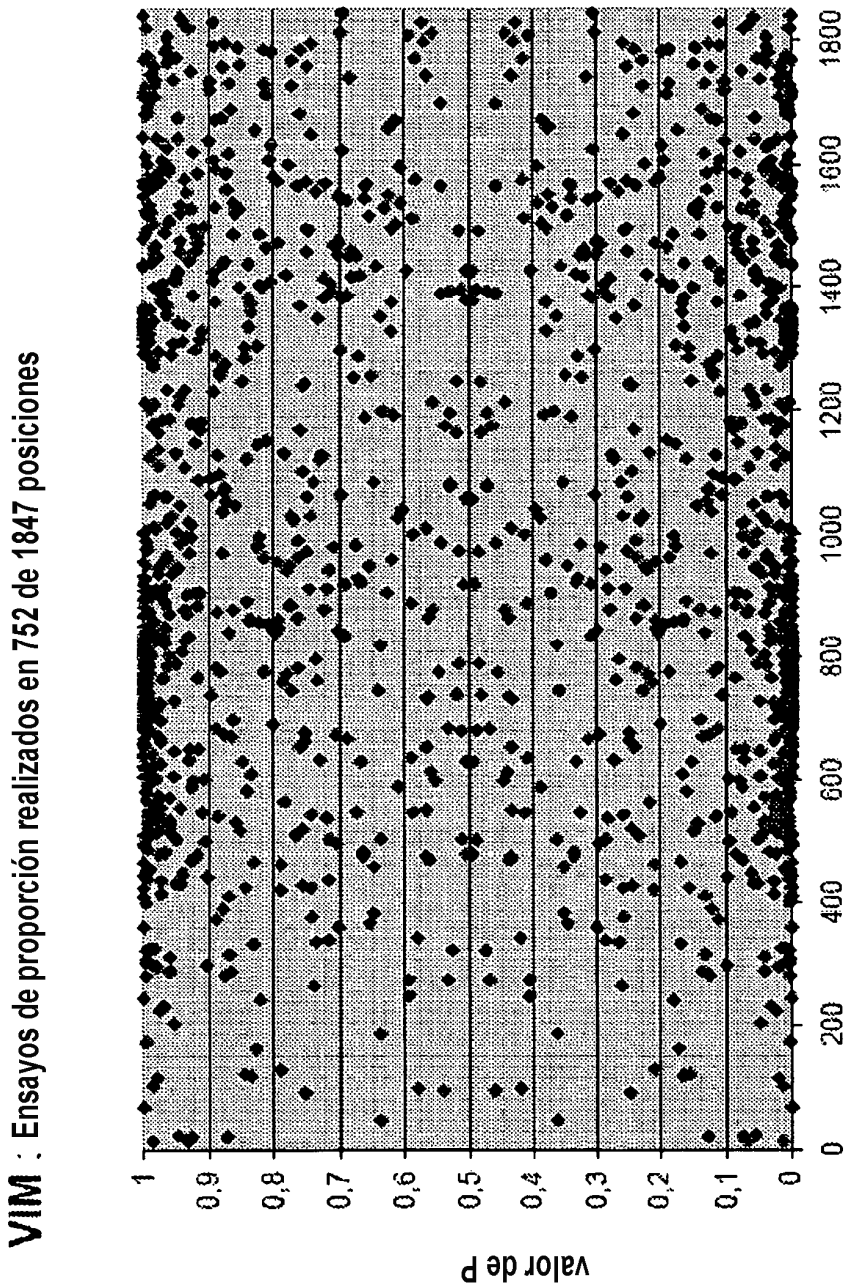

VIM : Ensayos de proporción realizados en 752 de 1847 posiciones
valor de P
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
0
200
400
600
800
1000
1200
1400
1600
1800
pb

Figura 5g

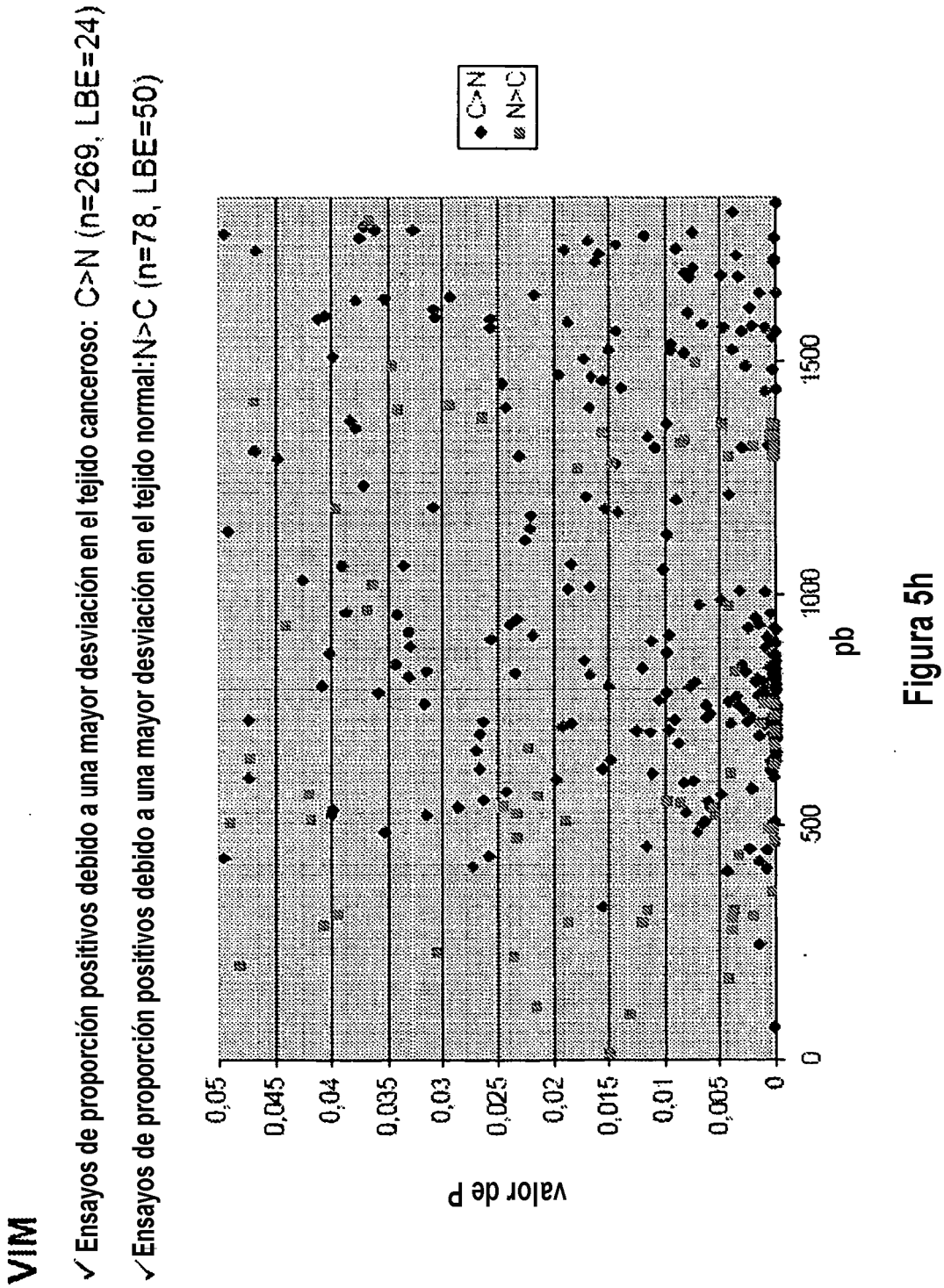

VIM
✓ Ensayos de proporción positivos debido a una mayor desviación en el tejido canceroso: C>N (n=269, LBE=24)
✓ Ensayos de proporción positivos debido a una mayor desviación en el tejido normal:N>C (n=78, LBE=50)
C>N
N>C
valor de P
0,05
0,045
0,04
0,035
0,03
0,025
0,02
0,015
0,01
0,005
0
0
500
1000
1500
pb

Figura 5h

Análisis de los ensayos de proporción: resultados

Gen	nº de ensayos de proporción	% de cobertura	C>N	LBE	N>C	LBE
ALB	194	9	**38**	10	**56**	8
ALDOA transcripción 1	336	30	**81**	11	**35**	21
ATP5A1 transcripción 1	238	38	**123**	3	**6**	20
CALM2	374	23	**69**	16	**40**	21
ENO1	614	27	**186**	18	**31**	42
FTH1	592	71	**226**	20	**57**	38
FTL	678	56	**118**	31	**86**	36
GAPDH	812	91	**311**	23	**61**	57
HSPA8 transcripción 1	513	59	**163**	13	**18**	37
LDHA	181	9	**70**	4	**10**	13
RPL7A	337	35	**103**	11	**33**	22
RPS4X	371	45	**124**	11	**27**	25
RPS6	432	19	**86**	17	**40**	25
TMSB4X	352	19	**70**	18	**74**	17
TPI1	379	31	**99**	14	**47**	23
TPT1	489	26	**145**	15	**26**	33
VIM	752	57	**269**	24	**78**	50

Figura 5i

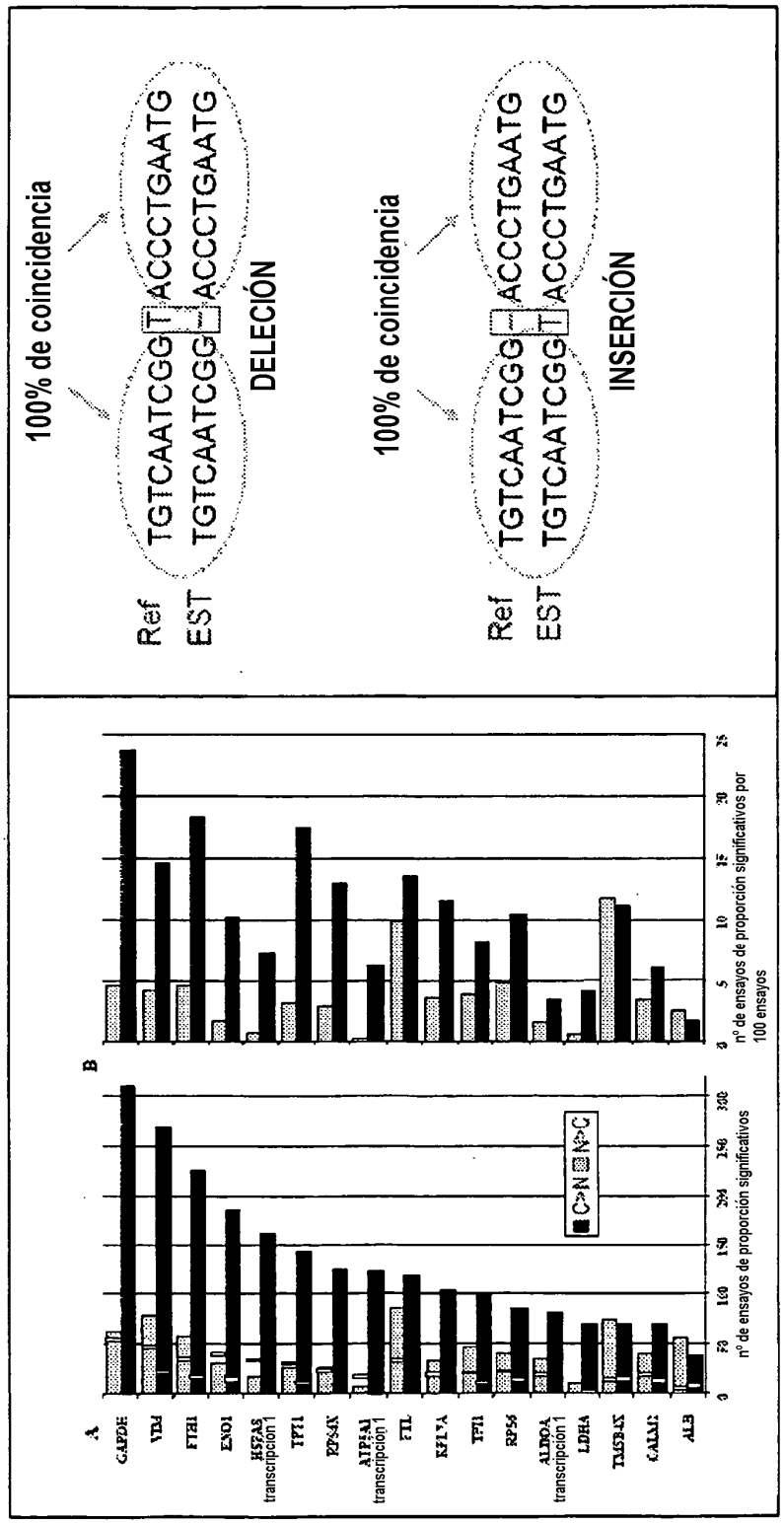

100% de coincidencia
Ref TGTCAATCGG T ACCCTGAATG
EST TGTCAATCGG – ACCCTGAATG
DELECIÓN
100% de coincidencia
Ref TGTCAATCGG – ACCCTGAATG
EST TGTCAATCGG T ACCCTGAATG
INSERCIÓN
A
B
C>N
N>C
nº de ensayos de proporción significativos
nº de ensayos de proporción significativos por 100 ensayos

Figura 5j

Análisis de los ensayos de proporción: resultados *(inserciones y deleciones)*

DELECIONES

Genes	nº de ensayos de proporción	C>N	LBE	N>C	LBE	C>N / N>C
ALB	17	1	1	13	0	0,1
ALDOA	34	8	1	7	1	1,1
ATP5A1	24	16	0	0	2	
CALM2	50	14	2	4	2	3,5
ENO1	41	29	0	5	3	5,8
FTH1	67	46	0	4	5	11,5
FTL	100	42	2	9	7	4,7
GAPDH	99	59	2	17	7	3,5
HSPA8	30	27	0	0	2	
LDHA	14	6	0	3	0	2,0
RPL7A	45	26	0	1	3	26,0
RPS4X	35	29	0	1	3	29,0
RPS6	45	24	0	1	3	24,0
TMSB4X	28	16	0	4	1	4,0
TPI1	30	17	0	1	2	17,0
TPT1	43	27	0	4	3	6,8
VIM	26	11	1	9	1	1,2
Total	728	398	9	83	45	4,8

INSERCIONES

Genes	nº de ensayos de proporción	C>N	LBE	N>C	LBE	C>N / N>C
ALB	9	0	0	9	0	0,0
ALDOA	5	1	0	1	0	1,0
ATP5A1	9	6	0	0	0	
CALM2	72	10	3	10	3	1,0
ENO1	40	14	0	2	3	7,0
FTH1	114	54	1	3	9	18,0
FTL	188	26	9	33	9	0,8
GAPDH	137	46	4	18	9	2,6
HSPA8	14	7	0	2	1	3,5
LDHA	2	0	0	0	0	
RPL7A	56	12	1	1	3	12,0
RPS4X	29	9	1	2	1	4,5
RPS6	44	12	1	1	3	12,0
TMSB4X	22	5	1	6	0	0,8
TPI1	26	6	1	5	1	1,2
TPT1	41	7	1	4	2	1,8
VIM	27	10	1	3	1	3,3
Total	835	225	24	100	45	2,3

Figura 5k

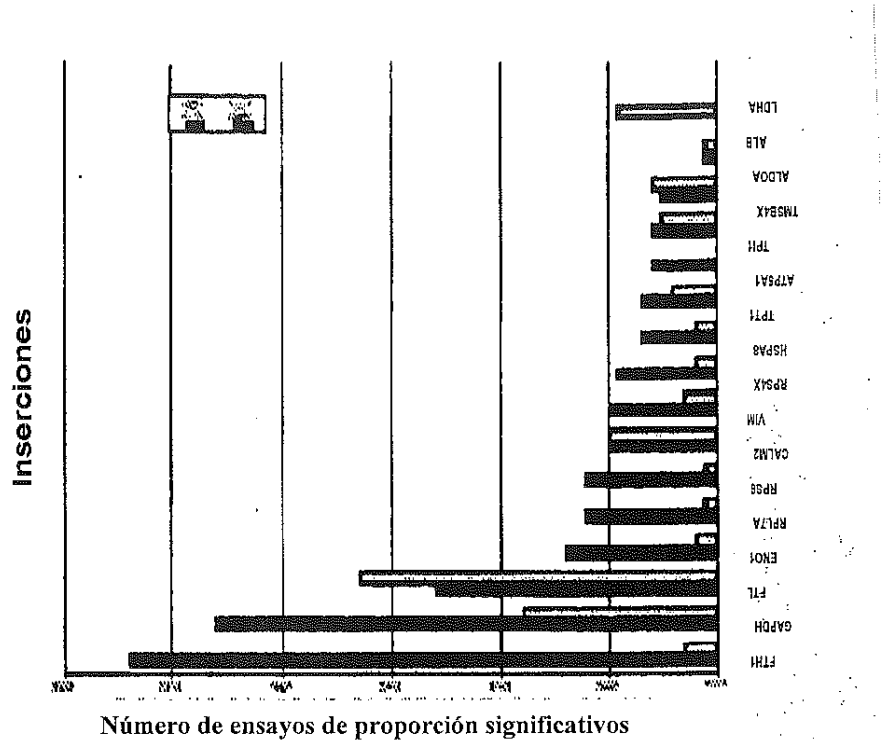
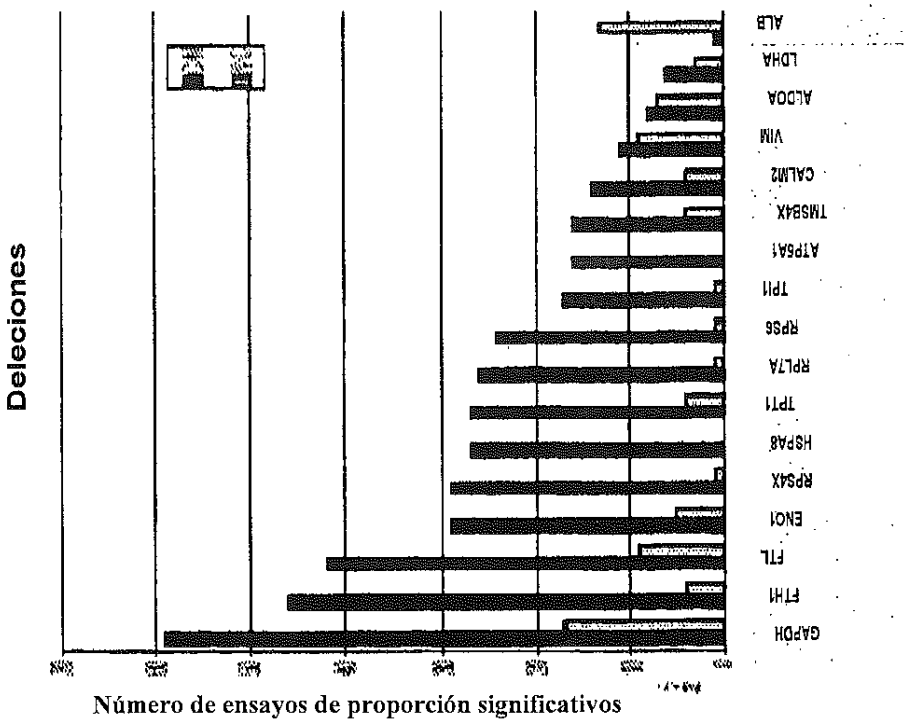
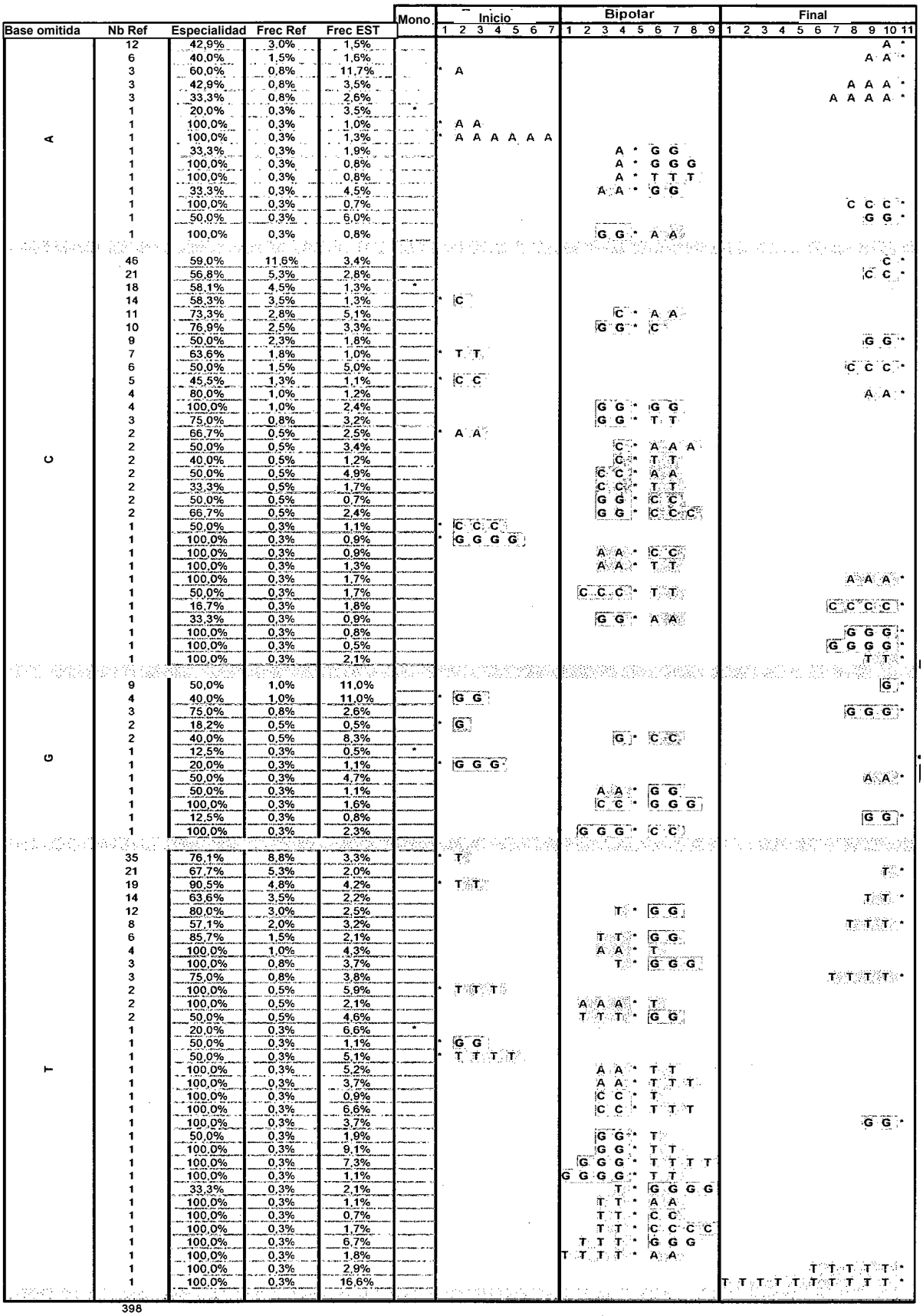

Inserciones
FTH1
GAPDH
FTL
ENO1
RPLZA
RPS6
CALM2
YIM
RPS4X
HSPA8
TPT1
ATP5A1
TPI1
TMSB4X
ALDOA
ALB
LDHA
Número de ensayos de proporción significativos

Figura 5I


Deleciones
GAPDH
FTH1
FTL
ENO1
RPS4X
HSPA8
TPT1
RPLZA
RPS6
TPI1
ATP5A1
TMSB4X
CALM2
YIM
ALDOA
LDHA
ALB
Número de ensayos de proporción significativos

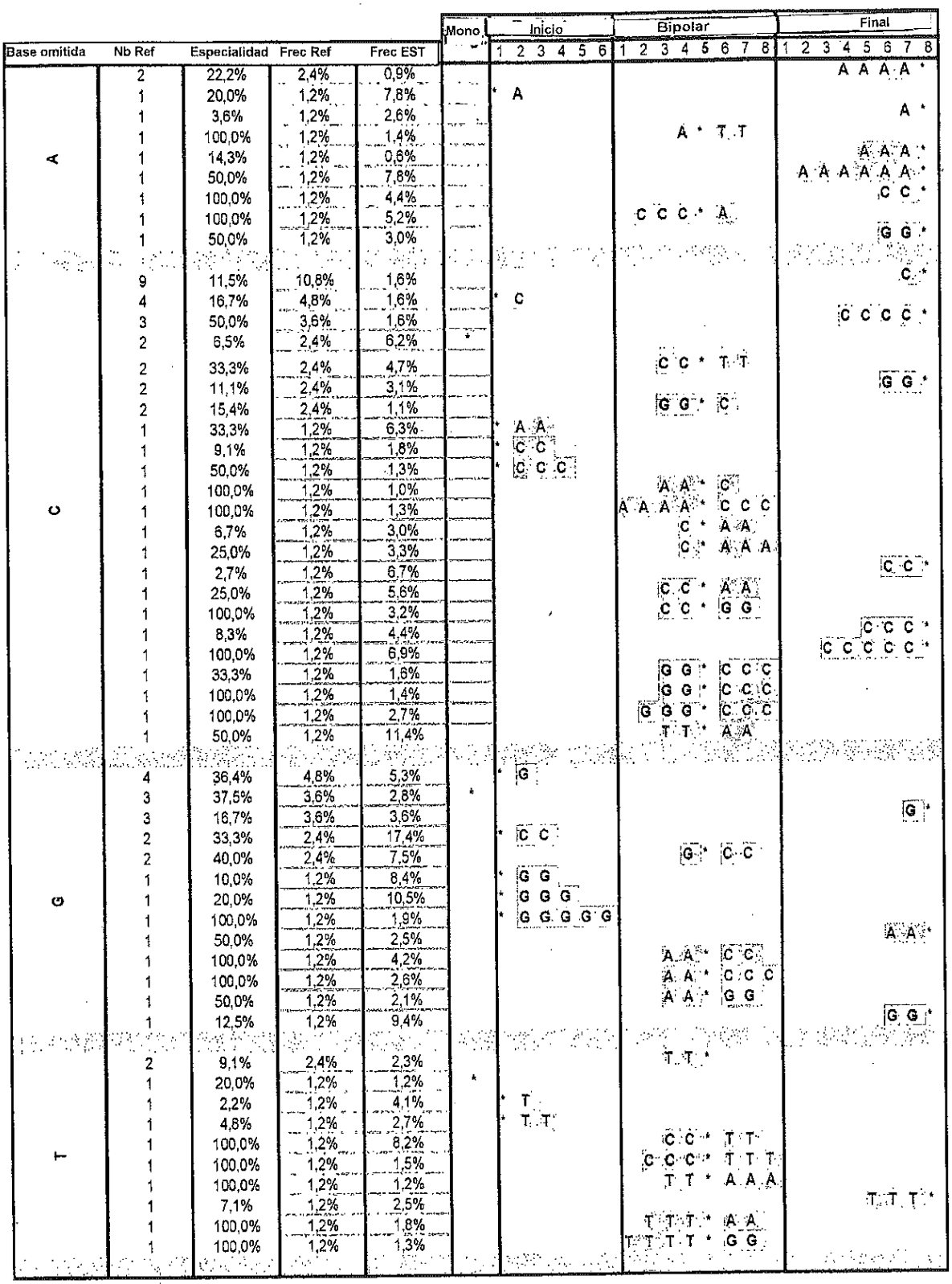
Posiciones C>N (398)



Base omitida	Nb Ref	Especialidad	Frec Ref	Frec EST
A	12	42,9%	3,0%	1,5%
	6	40,0%	1,5%	1,6%
	3	60,0%	0,8%	11,7%
	3	42,9%	0,8%	3,5%
	3	33,3%	0,8%	2,6%
	1	20,0%	0,3%	3,5%
	1	100,0%	0,3%	1,0%
	1	100,0%	0,3%	1,3%
	1	33,3%	0,3%	1,9%
	1	100,0%	0,3%	0,8%
	1	100,0%	0,3%	0,8%
	1	33,3%	0,3%	4,5%
	1	100,0%	0,3%	0,7%
	1	50,0%	0,3%	6,0%
	1	100,0%	0,3%	0,8%
C	46	59,0%	11,6%	3,4%
	21	56,8%	5,3%	2,8%
	18	58,1%	4,5%	1,3%
	14	58,3%	3,5%	1,3%
	11	73,3%	2,8%	5,1%
	10	76,9%	2,5%	3,3%
	9	50,0%	2,3%	1,8%
	7	63,6%	1,8%	1,0%
	6	50,0%	1,5%	5,0%
	5	45,5%	1,3%	1,1%
	4	80,0%	1,0%	1,2%
	4	100,0%	1,0%	2,4%
	3	75,0%	0,8%	3,2%
	2	66,7%	0,5%	2,5%
	2	50,0%	0,5%	3,4%
	2	40,0%	0,5%	1,2%
	2	50,0%	0,5%	4,9%
	2	33,3%	0,5%	1,7%
	2	50,0%	0,5%	0,7%
	2	66,7%	0,5%	2,4%
	1	50,0%	0,3%	1,1%
	1	100,0%	0,3%	0,9%
	1	100,0%	0,3%	0,9%
	1	100,0%	0,3%	1,3%
	1	100,0%	0,3%	1,7%
	1	50,0%	0,3%	1,7%
	1	16,7%	0,3%	1,8%
	1	33,3%	0,3%	0,9%
	1	100,0%	0,3%	0,8%
	1	100,0%	0,3%	0,5%
	1	100,0%	0,3%	2,1%
G	9	50,0%	1,0%	11,0%
	4	40,0%	1,0%	11,0%
	3	75,0%	0,8%	2,6%
	2	18,2%	0,5%	0,5%
	2	40,0%	0,5%	8,3%
	1	12,5%	0,3%	0,5%
	1	20,0%	0,3%	1,1%
	1	50,0%	0,3%	4,7%
	1	50,0%	0,3%	1,1%
	1	100,0%	0,3%	1,6%
	1	12,5%	0,3%	0,8%
	1	100,0%	0,3%	2,3%
T	35	76,1%	8,8%	3,3%
	21	67,7%	5,3%	2,0%
	19	90,5%	4,8%	4,2%
	14	63,6%	3,5%	2,2%
	12	80,0%	3,0%	2,5%
	8	57,1%	2,0%	3,2%
	6	85,7%	1,5%	2,1%
	4	100,0%	1,0%	4,3%
	3	100,0%	0,8%	3,7%
	3	75,0%	0,8%	3,8%
	2	100,0%	0,5%	5,9%
	2	100,0%	0,5%	2,1%
	2	50,0%	0,5%	4,6%
	1	20,0%	0,3%	6,6%
	1	50,0%	0,3%	1,1%
	1	50,0%	0,3%	5,1%
	1	100,0%	0,3%	5,2%
	1	100,0%	0,3%	3,7%
	1	100,0%	0,3%	0,9%
	1	100,0%	0,3%	6,6%
	1	100,0%	0,3%	3,7%
	1	50,0%	0,3%	1,9%
	1	100,0%	0,3%	9,1%
	1	100,0%	0,3%	7,3%
	1	100,0%	0,3%	1,1%
	1	33,3%	0,3%	2,1%
	1	100,0%	0,3%	1,1%
	1	100,0%	0,3%	0,7%
	1	100,0%	0,3%	1,7%
	1	100,0%	0,3%	6,7%
	1	100,0%	0,3%	1,8%
	1	100,0%	0,3%	2,9%
	1	100,0%	0,3%	16,6%
	398			

Figura 5m

Posiciones C>N (83)



Base omitida	Nb Ref	Especialidad	Frec Ref	Frec EST
A	2	22,2%	2,4%	0,9%
	1	20,0%	1,2%	7,8%
	1	3,6%	1,2%	2,6%
	1	100,0%	1,2%	1,4%
	1	14,3%	1,2%	0,6%
	1	50,0%	1,2%	7,8%
	1	100,0%	1,2%	4,4%
	1	100,0%	1,2%	5,2%
	1	50,0%	1,2%	3,0%
C	9	11,5%	10,8%	1,6%
	4	16,7%	4,8%	1,6%
	3	50,0%	3,6%	1,6%
	2	6,5%	2,4%	6,2%
	2	33,3%	2,4%	4,7%
	2	11,1%	2,4%	3,1%
	2	15,4%	2,4%	1,1%
	1	33,3%	1,2%	6,3%
	1	9,1%	1,2%	1,8%
	1	50,0%	1,2%	1,3%
	1	100,0%	1,2%	1,0%
	1	100,0%	1,2%	1,3%
	1	6,7%	1,2%	3,0%
	1	25,0%	1,2%	3,3%
	1	2,7%	1,2%	6,7%
	1	25,0%	1,2%	5,6%
	1	100,0%	1,2%	3,2%
	1	8,3%	1,2%	4,4%
	1	100,0%	1,2%	6,9%
	1	33,3%	1,2%	1,6%
	1	100,0%	1,2%	1,4%
	1	100,0%	1,2%	2,7%
	1	50,0%	1,2%	11,4%
G	4	36,4%	4,8%	5,3%
	3	37,5%	3,6%	2,8%
	3	16,7%	3,6%	3,6%
	2	33,3%	2,4%	17,4%
	2	40,0%	2,4%	7,5%
	1	10,0%	1,2%	8,4%
	1	20,0%	1,2%	10,5%
	1	100,0%	1,2%	1,9%
	1	50,0%	1,2%	2,5%
	1	100,0%	1,2%	4,2%
	1	100,0%	1,2%	2,6%
	1	50,0%	1,2%	2,1%
	1	12,5%	1,2%	9,4%
T	2	9,1%	2,4%	2,3%
	1	20,0%	1,2%	1,2%
	1	2,2%	1,2%	4,1%
	1	4,8%	1,2%	2,7%
	1	100,0%	1,2%	8,2%
	1	100,0%	1,2%	1,5%
	1	100,0%	1,2%	1,2%
	1	7,1%	1,2%	2,5%
	1	100,0%	1,2%	1,8%
	1	100,0%	1,2%	1,3%

Figura 5n

Posiciones C>N (226)

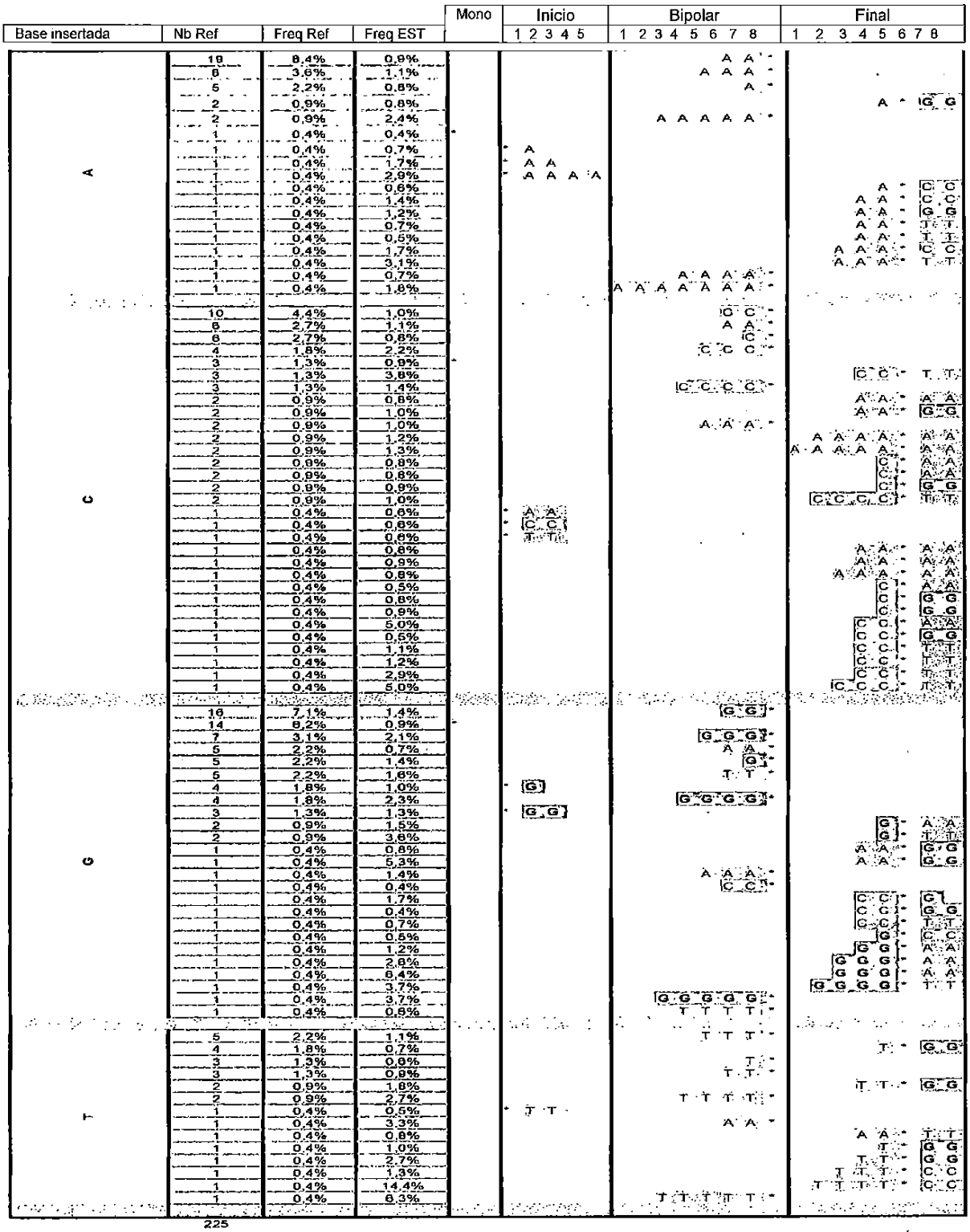


Figura 5o

Posiciones N>C (100)

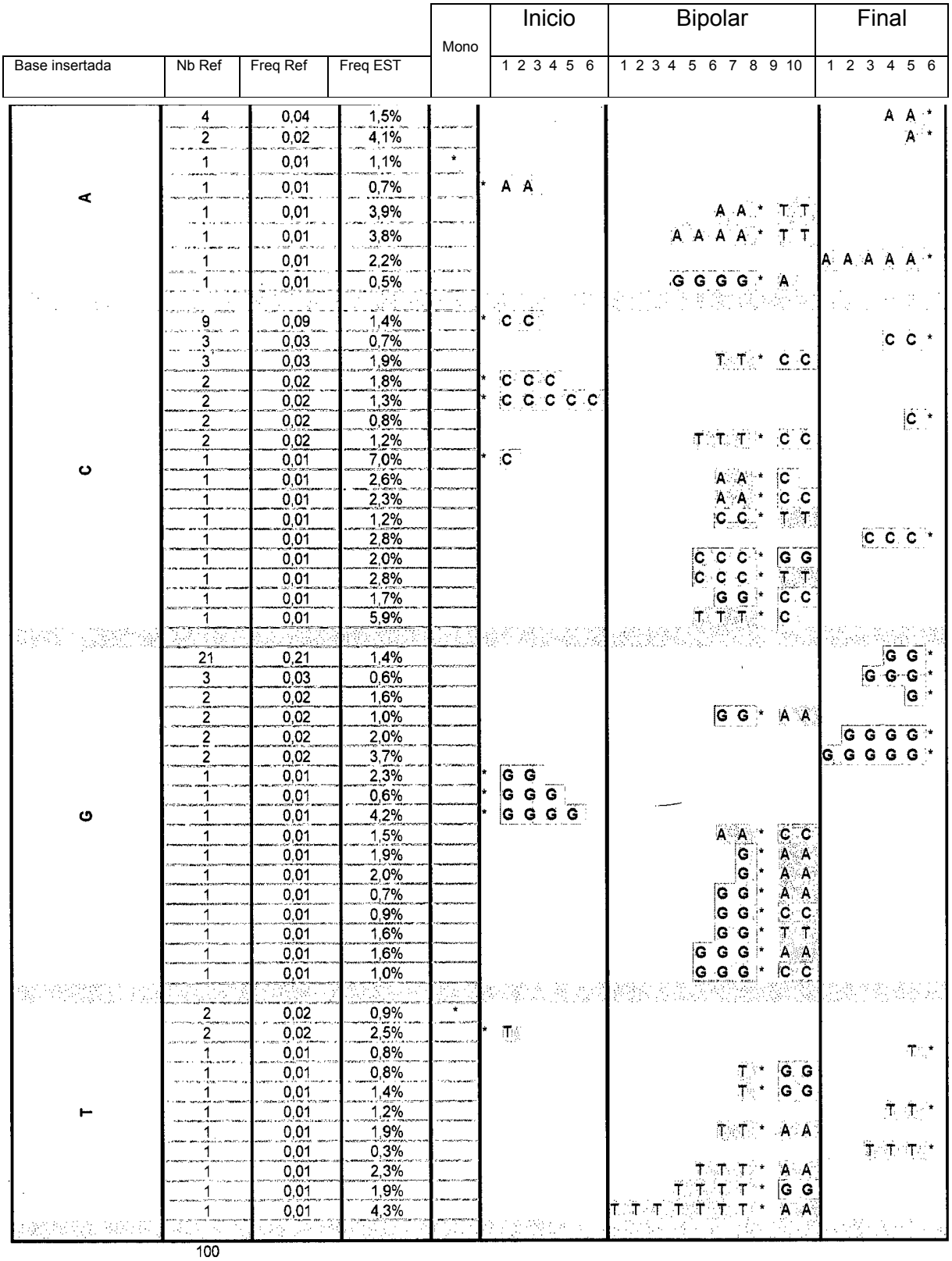
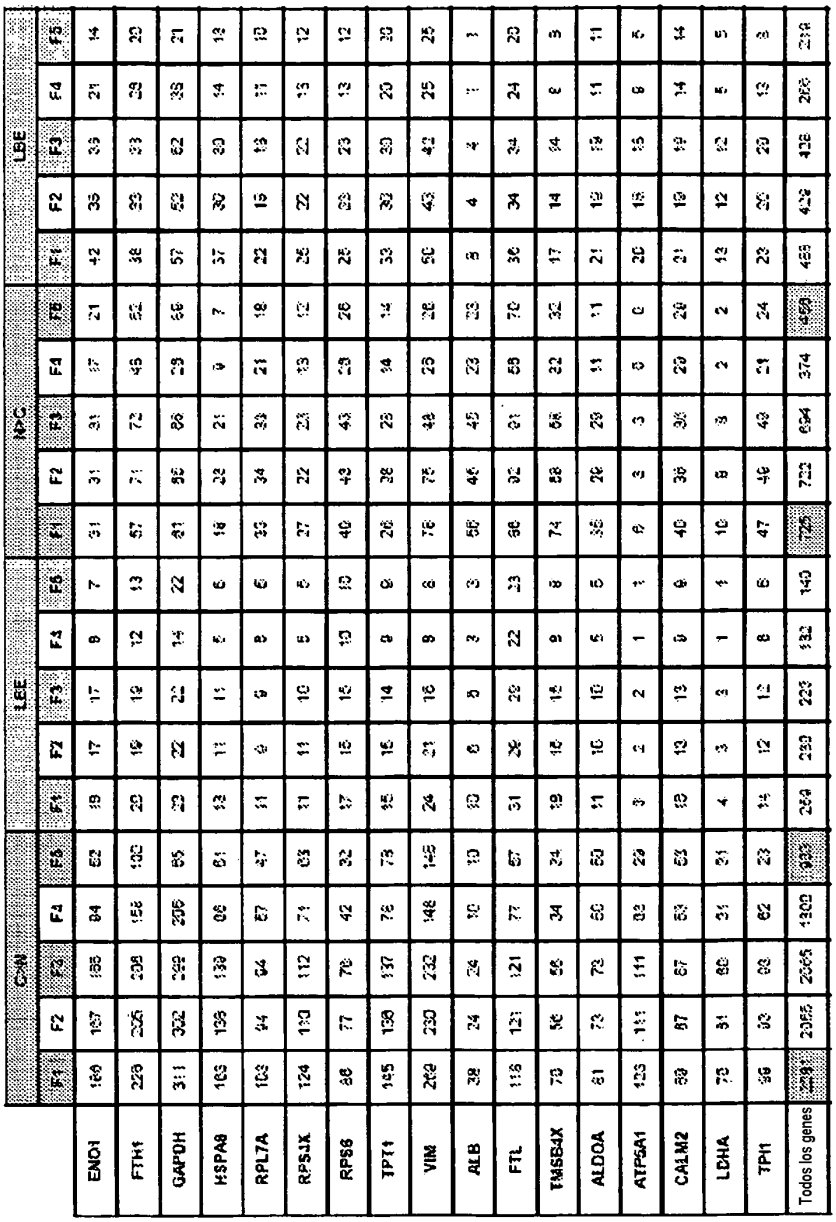


Figura 5p

	nº de ensayos					CAN					LDE				
	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5
ENO1	614	506	557	351	220	217	185	159	111	72	40	37	37	19	14
FTH1	592	536	632	320	335	282	279	276	208	152	39	25	25	15	18
GAPDH	812	752	748	324	440	372	367	365	232	164	42	37	37	25	28
HSPA8	513	423	415	197	190	181	181	159	75	56	32	28	28	11	10
RPL7A	337	294	281	175	164	139	129	127	78	65	19	14	14	8	9
RPS4X	371	340	338	193	170	181	132	135	94	76	21	19	19	10	10
RPS6	432	397	397	230	225	129	129	119	88	68	31	25	24	17	18
TPT1	436	452	455	300	300	171	164	165	92	82	31	28	23	23	20
VIM	752	844	593	337	337	347	365	250	174	124	34	31	30	15	15
ALB	194	117	117	49	49	94	69	66	32	32	9	4	4	1	1
FTL	678	844	644	464	444	204	213	212	133	107	46	45	44	35	31
TMSB4X	352	303	302	163	163	144	124	124	85	85	20	16	16	8	9
ALDOA	338	297	297	174	174	118	162	162	91	81	23	16	16	11	11
ATP6A1	238	211	211	109	88	129	114	114	82	29	9	9	6	4	3
CALM2	374	335	335	245	245	160	103	103	82	81	26	23	23	16	13
LDHA	161	186	186	75	73	53	66	66	33	33	6	7	7	4	4
TPI1	376	331	335	233	150	149	142	142	83	47	20	16	16	12	10
Todos los genes	7644	6782	6701	4140	3159	3022	2787	2750	1974	1369	448	381	378	238	223

F1 EST alineados >25 bp
F2 >=70% EST alineados
F3 Eliminación de parálogos y pseudogenes
F4 EST creados en cada extremidad del alineamiento
F5 Longitud normalizada

Figura 6a



	C>N					LSE					N>C					LSE				
	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5
ENO1	169	197	185	84	62	59	17	17	8	7	31	31	31	87	21	42	35	33	21	14
FTH1	229	256	208	156	105	29	19	19	12	13	57	71	72	46	52	38	33	33	29	29
GAPDH	311	302	292	206	85	29	22	23	14	22	41	95	86	29	56	57	52	62	38	21
HSPA8	103	135	139	66	61	13	11	11	5	6	16	32	21	9	7	37	32	30	14	13
RPL7A	103	94	64	67	47	11	9	9	8	6	33	34	33	21	18	22	15	13	11	10
RPS3X	124	110	112	71	63	11	11	10	5	5	27	22	23	13	12	25	22	22	13	12
RPS8	86	77	70	42	32	17	15	15	10	10	40	43	43	23	26	28	23	23	13	12
TPT1	145	130	137	76	73	15	16	14	8	8	26	38	28	14	14	33	32	30	20	19
VIM	240	230	232	148	145	24	21	16	8	8	76	75	48	25	28	50	43	42	25	25
ALB	38	24	24	10	10	10	6	6	3	3	56	46	45	23	23	8	4	4	1	1
FTL	118	121	121	77	67	31	26	25	22	23	96	92	81	69	70	36	34	34	24	23
TMSB4X	78	90	66	34	24	19	16	16	9	8	74	68	66	22	32	17	14	14	8	8
ALDOA	81	73	73	60	50	11	10	10	5	5	35	26	29	11	11	21	19	19	11	11
ATP5A1	123	111	111	82	29	9	2	2	1	1	6	3	3	3	0	20	15	15	6	5
CALM2	69	67	67	53	53	13	13	13	6	6	40	36	36	20	20	21	19	19	14	14
LDHA	73	91	80	31	31	4	3	3	1	1	10	8	3	2	2	13	12	12	5	5
TPI1	96	92	88	62	23	14	12	12	8	6	47	46	46	21	24	28	26	20	13	3
Todos los genes	2791	2955	2565	1800	939	269	269	223	132	140	725	722	694	374	459	455	432	432	268	259

(C>N)/(N>C)

F1	F2	F3	F4	F5
3,15	2,86	2,98	3,48	2,95

Figura 6b

	C>N-LBE					N>C-LBE				
	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5
ENO1	109	150	145	86	45	0	0	0	0	7
FTH1	208	128	157	146	87	19	38	33	22	32
GAPDH	288	250	277	192	43	4	13	14	6	68
HSPA8	150	127	128	81	66	0	0	6	0	0
RPLP0	92	85	86	51	42	11	10	15	10	6
RPS4X	113	99	102	66	63	2	0	1	0	0
RPS6	89	82	81	32	22	15	20	20	13	14
TPT1	130	121	123	69	60	0	0	0	0	0
VIM	245	200	210	140	140	23	32	6	1	1
ALB	28	16	18	7	7	48	41	41	22	22
FTL	87	92	92	55	34	50	69	57	32	50
TMSB4X	52	41	41	26	26	67	64	64	24	24
ALDOA	70	63	62	45	45	14	10	10	0	0
ATP5A1	120	109	109	82	23	0	0	0	0	0
CALM2	53	54	54	44	44	10	17	17	15	15
LDHA	68	58	67	30	30	0	0	0	0	0
TPI1	95	81	81	66	17	24	29	26	8	18
Todos los genes	2022	1835	1842	1168	793	291	329	303	147	257

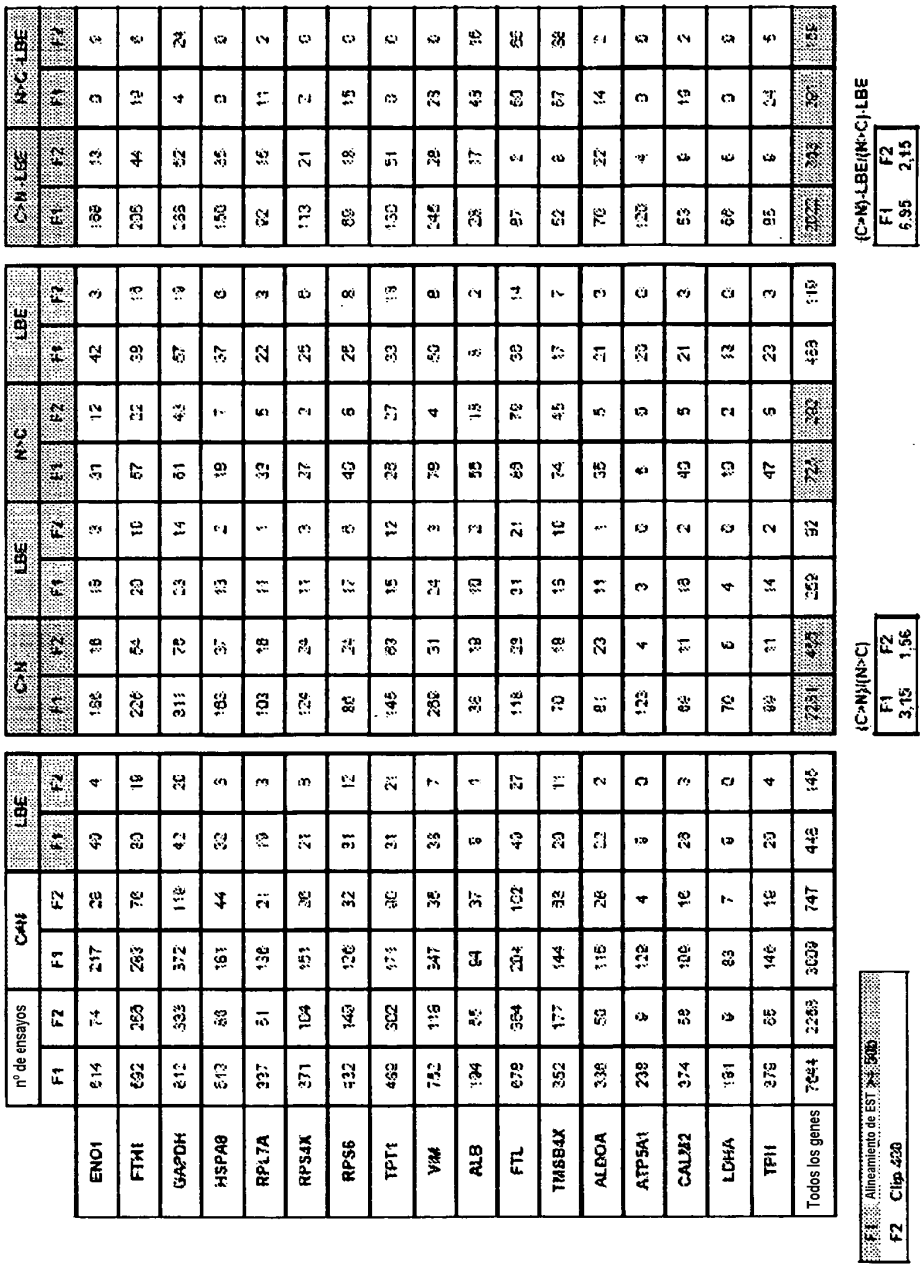
(C>N)-LBE/(N>C)-LBE

F1	F2	F3	F4	F5
6,95	5,59	6,08	7,95	3,09

Figura 6c-1

	C>N:	LBE	N>C:	LBE	(C>N)/(N>C)	[(C>N)-LBE]/[(N>C)-LBE]
F1	2281	259	725	488	3,15	6,95
F2	2066	230	722	429	2,86	5,59
F3	2066	223	694	428	2,98	6,08
F4	1300	132	374	266	3,48	7,95
F5	933	140	456	219	2,05	3,09

Figura 6c-2



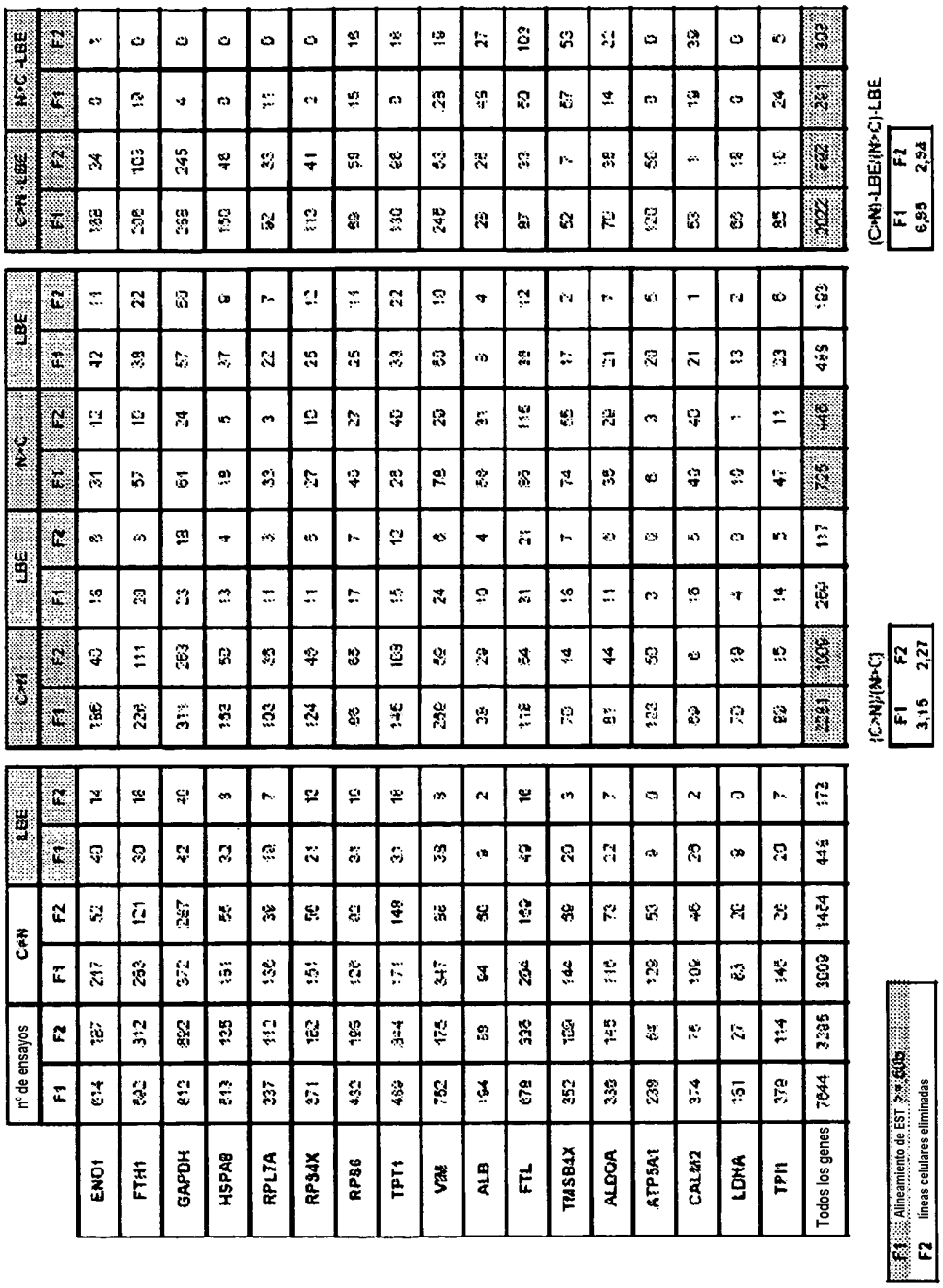
	nº de ensayos		C>N		LBE		N>C		LBE		C>N-LBE		N>C-LBE	
	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
ENO1	814	74	217	28	40	4	31	12	42	3	136	13	0	0
FTH1	692	269	263	76	80	16	67	22	39	13	235	44	12	6
GAPDH	812	333	372	118	42	26	61	43	67	19	233	52	4	24
HSPA8	513	88	161	44	32	5	13	7	37	6	150	35	0	0
RPL7A	337	51	136	21	19	3	32	5	22	3	82	16	11	2
RPS4X	371	104	151	36	21	5	27	2	25	6	113	21	2	0
RPS6	432	140	126	32	31	12	40	6	25	8	69	18	15	0
TPT1	492	302	173	80	21	21	28	27	33	13	130	51	0	0
VIM	782	116	347	35	35	7	79	4	50	8	245	28	23	0
ALB	194	55	64	37	6	1	55	15	8	2	28	17	45	16
FTL	678	364	204	102	40	27	69	78	38	14	87	2	60	65
TMSB4X	352	177	144	32	20	11	74	45	17	7	52	8	67	38
ALDOA	338	58	115	26	22	2	35	5	21	3	76	22	14	2
ATP5A1	238	6	122	4	6	0	6	5	20	0	120	4	6	0
CALR2	374	58	109	16	23	3	40	5	21	3	53	6	16	2
LDHA	191	6	83	7	6	0	10	2	13	0	68	6	6	0
TPI1	378	55	146	16	20	4	47	5	23	3	95	6	24	5
Todos los genes	7644	2269	3609	747	448	145	722	293	469	119	3577	748	307	156

F1 Alineamiento de EST ≥ 500
F2 Clip 420

{C>N}/{N>C}
F1 3,15
F2 1,56

{C>N}-LBE/{N>C}-LBE
F1 6,95
F2 2,15

Figura 6d



nº de ensayos	C+N		LBE		C+N		LBE		N+C		LBE		C+N-LBE		N+C-LBE	
	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
ENO1	614	787	247	52	40	14	186	40	16	8	31	12	42	11	8	1
FTH1	842	312	263	121	30	16	226	111	29	9	57	19	39	22	18	0
GAPDH	812	892	372	287	42	40	311	263	33	18	61	24	57	60	4	0
HSPA8	813	435	151	54	33	9	192	50	13	4	19	5	37	6	3	0
RPL7A	337	112	136	36	18	7	103	35	11	3	33	3	22	7	11	0
RPS4X	371	162	151	50	21	12	124	46	11	5	27	10	25	12	2	0
RPS6	432	195	128	62	31	10	96	65	17	7	43	27	25	11	15	16
TPT1	469	364	171	149	30	16	146	103	15	12	28	46	33	22	3	16
VIM	752	475	347	56	33	3	236	59	24	6	78	29	63	10	23	16
ALB	194	59	64	50	6	2	39	29	10	4	58	31	9	4	29	27
FTL	678	335	294	169	49	16	118	54	21	21	85	115	38	12	29	102
TMSB4X	352	159	144	39	20	3	79	14	16	7	74	55	17	2	7	53
ALDOA	338	145	116	73	22	7	81	44	11	3	38	29	21	7	14	22
ATP5A1	239	64	129	50	9	0	123	50	3	0	6	3	28	5	3	0
CALM2	374	75	109	46	29	2	89	6	16	5	43	40	21	1	19	39
LDHA	151	27	83	20	9	0	79	19	4	0	19	1	13	2	9	0
TPI1	379	114	145	36	29	7	99	15	14	5	47	11	33	6	24	5
Todos los genes	7644	3285	3009	1464	446	172	2281	1006	264	117	725	446	465	183	281	303

F1 Alineamiento de EST > 95,60%
F2 líneas celulares eliminadas

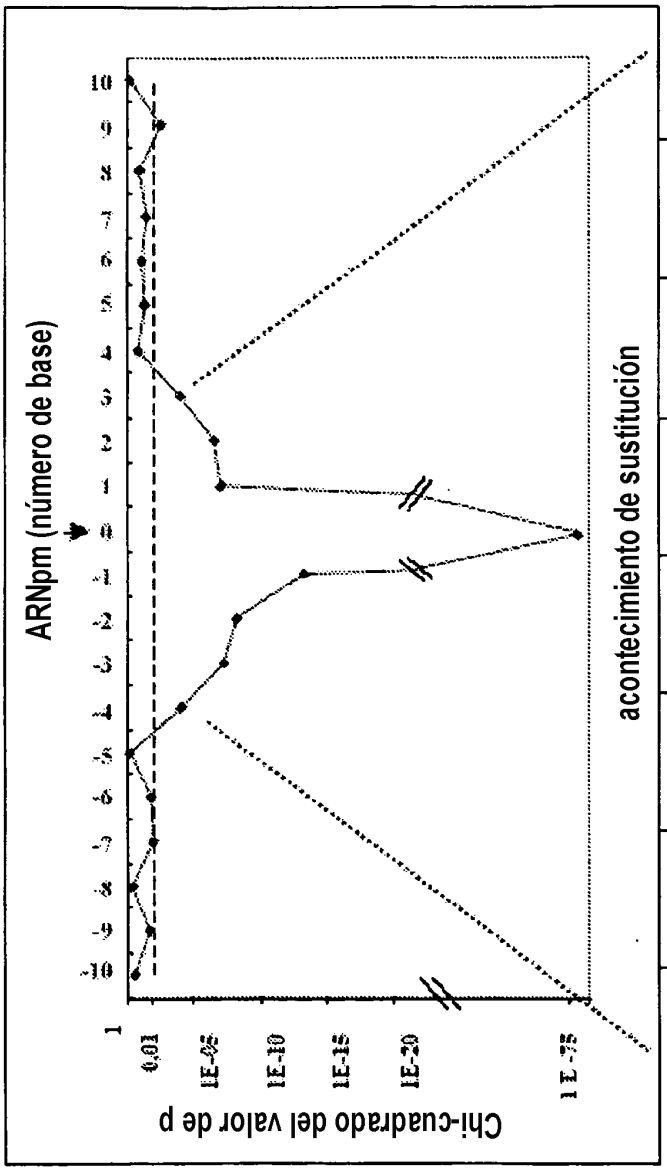
(C+N)/(N+C)

F1	F2
3,15	2,27

(C+N)-LBE/(N+C)-LBE

F1	F2
6,89	2,94

Figura 6e



	b-4		b-3		b-2		b-1		b0		b+1		b+2		b+3	
	H	NH	H	NH	H	NH	H	NH	H	NH	H	NH	H	NH	H	NH
A	28	24	24	27	25	27	30	22	33	23	29	25	27	26	28	24
T	23	25	25	23	26	23	28	27	32	22	22	24	24	26	24	24
C	24	25	25	24	21	26	23	26	21	26	21	25	20	24	22	26
G	25	26	23	26	28	24	25	25	14	29	28	26	29	24	26	26

Figura 7a

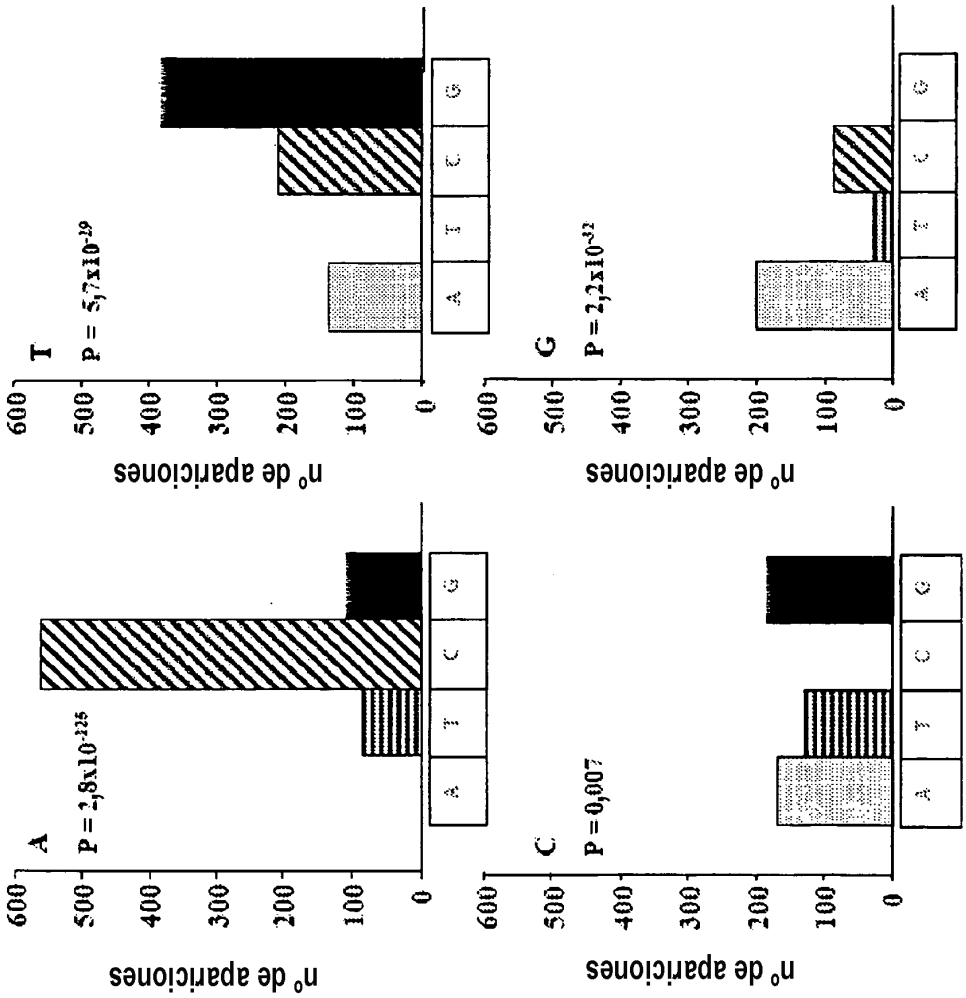

T
P = 5,7x10-19
G
P = 2,2x10-12
A
P = 2,8x10-225
C
P = 0,007
nº de apariciones
600
500
400
300
200
100
0
A
T
C
G

Figura 7b

A. Todas las desviaciones (n = 2281)

afectadas \ sustituciones	A	T	C	G	%
A		85	560	107	33
T	139		210	384	32
C	169	130		185	21
G	199	26	85		14
%	22	11	37	30	

B. Desviaciones hacia b-1 (n = 799)

afectadas \ sustituciones	A	T	C	G	%
A		41	201	42	36
T	28		56	83	21
C	77	41		126	30
G	75	9	20		13
%	23	11	35	31	

C. Desviaciones hacia b+1 (n = 530)

afectadas \ sustituciones	A	T	C	G	%
A		24	52	37	21
T	27		45	177	47
C	38	48		15	19
G	40	11	16		13
%	20	16	21	43	

D. Desviaciones ni hacia b-1 ni hacia b+1 (n = 347)

afectadas \ sustituciones	A	T	C	G	%
A		7	148	8	47
T	42		50	9	29
C	17	10		23	14
G	7	2	24		10
%	19	5	64	12	

E. Desviaciones hacia b-1 = b+1 (n = 339)

afectadas \ sustituciones	A	T	C	G	%
A		5	52	12	20
T	11		12	95	35
C	25	30		17	21
G	71	1	8		24
%	31	11	21	37	

F. Desviaciones dentro del triplete (n = 266)

afectadas \ sustituciones	A	T	C	G	%
A		8	107	8	46
T	31		47	20	37
C	12	1		4	6
G	6	5	17		11
%	19	5	64	12	

Figura 7c

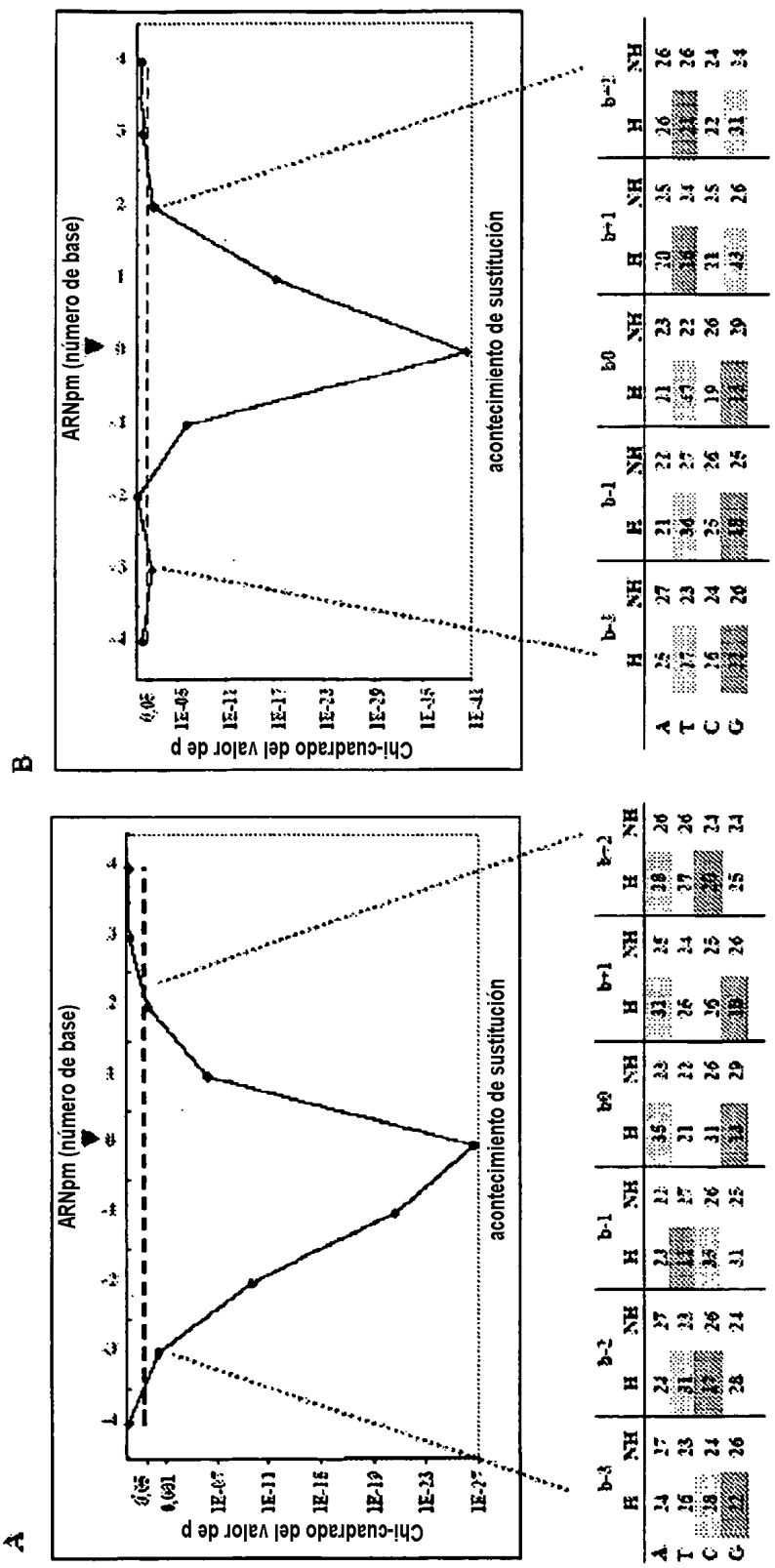

A
B
ARNpm (número de base)
Chi-cuadrado del valor de p
acontecimiento de sustitución

Figura 7d

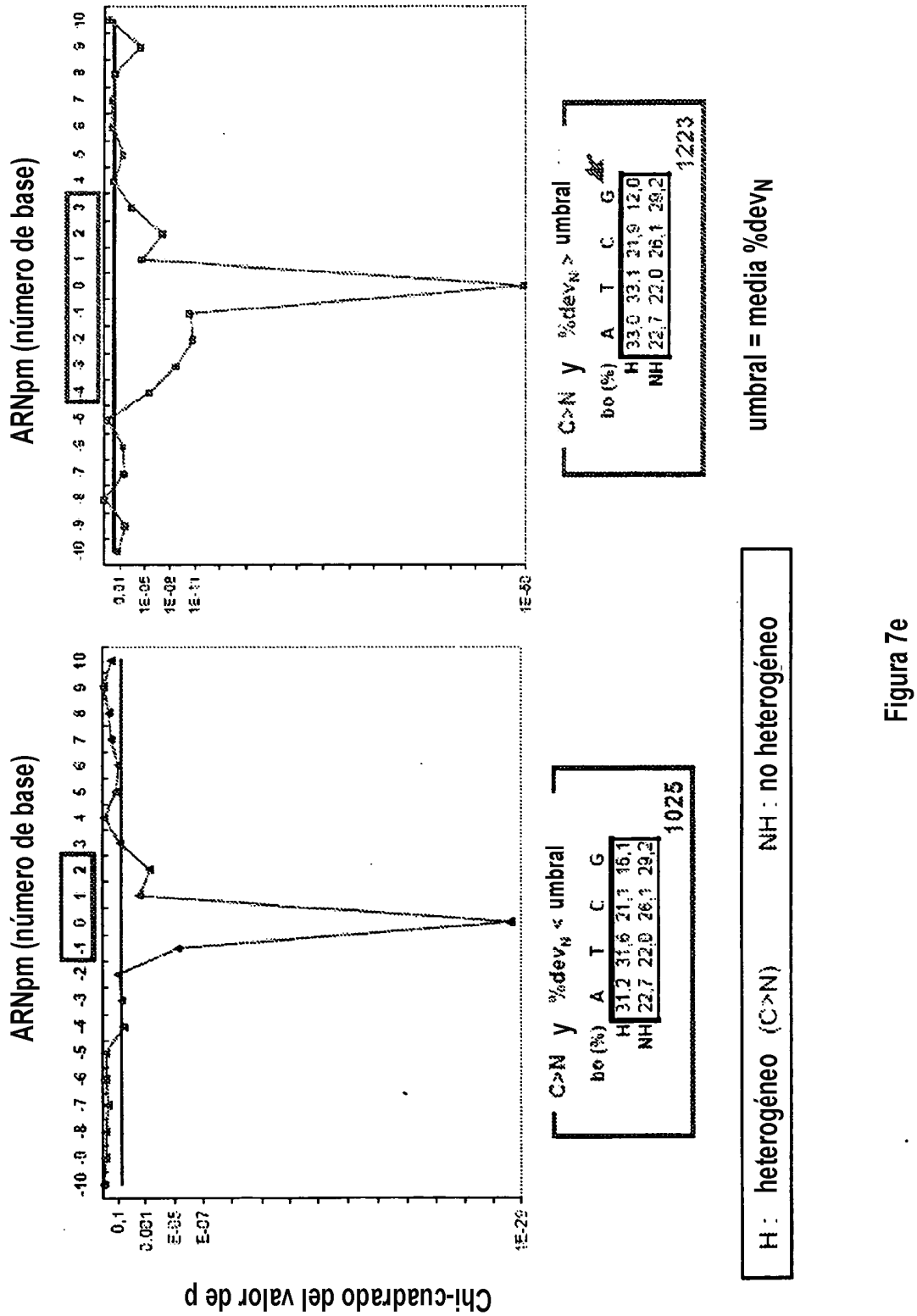

ARNpm (número de base)
Chi-cuadrado del valor de p
C>N y %dev$_N$ < umbral
bo (%) A T C G
H 31,2 31,6 21,3 16,1
NH 22,7 22,0 26,1 29,2
1025
C>N y %dev$_N$ > umbral
bo (%) A T C G
H 33,0 33,1 21,9 12,0
NH 22,7 22,0 26,1 29,2
1223
umbral = media %dev$_N$
H : heterogéneo (C>N)
NH : no heterogéneo

Figura 7e

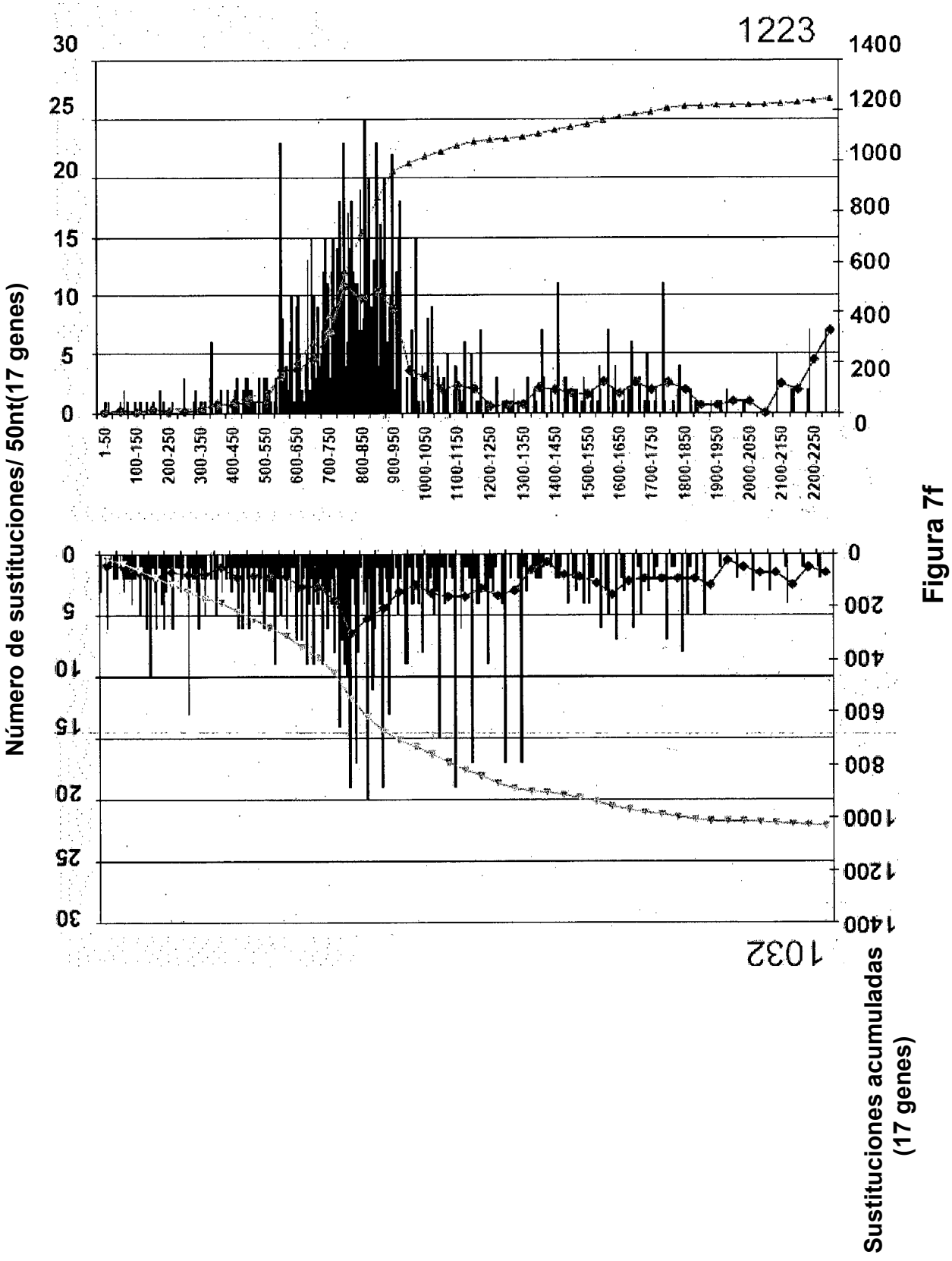

1223
1032
Número de sustituciones/ 50nt(17 genes)
Sustituciones acumuladas (17 genes)
1-50
100-150
200-250
300-350
400-450
500-550
600-650
700-750
800-850
900-950
1000-1050
1100-1150
1200-1250
1300-1350
1400-1450
1500-1550
1600-1650
1700-1750
1800-1850
1900-1950
2000-2050
2100-2150
2200-2250

Figura 7f

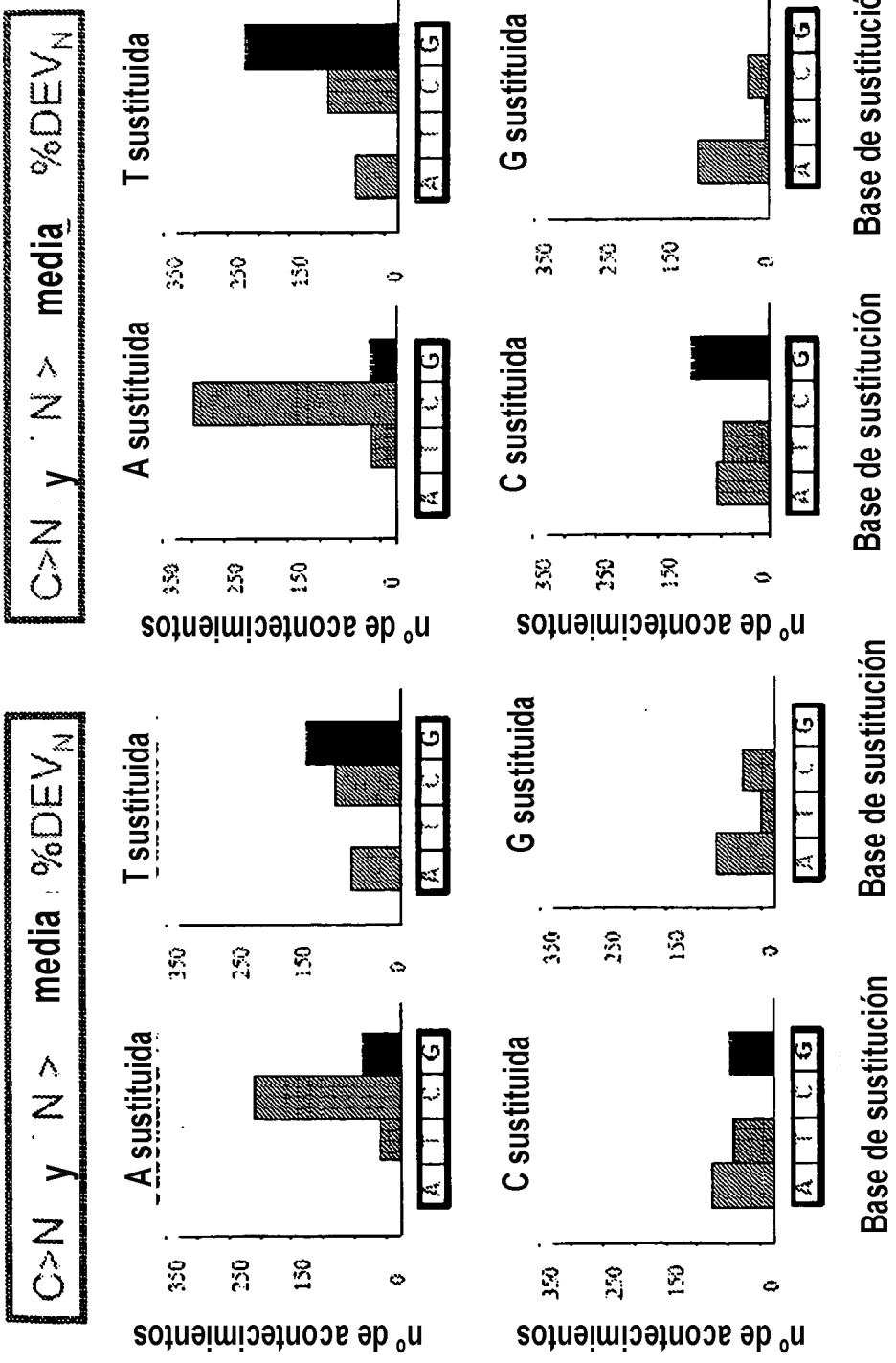

C>N y N> media %DEV_N
A sustituida
T sustituida
C sustituida
G sustituida
nº de acontecimientos
Base de sustitución
A T C G

Figura 7g

>vim nat, 1847 bases, 3DE7CB32

GTCCCCGCGCCAGAGACGCAGCCGCGGCTCCCACCACCCACACGCACCGCGGCCTCGTTCGCCTCTTCTCCGGGAGCCAGTCCGCGCCACCGCCGCCGCCCAGGCCATCGC
CACCCTCCGCAGCCATGTCCACCAGGTCCGTGTCCTCGTCCTCCTACCGCAGGATGTTCGGCGGCCCGGGCACCGCGAGCCGGCCGAGCTCCAGCCGGAGCTACGTGACT
ACGTCCACCCGCACCTACAGCCTGTGCAGCGCGCTGCGCCCCAGCACCAGCCGCAGCCTCTACGCCTCGTCCCCGGGCGGCGTGTATGCCACGCGCTCCTCTGCCGGGCG
CCTGCGGAGCAGCGTGCCCGGGGTGCGGCTCCTGCAGGACTCGGTGGACTTCTCGCTGGCCGACGCCATCAACACCGAGTTCAAGAACACCCCGACCAACGAGAAGGTGG
ACCTGCAGGAGCTGAATGACCGCTTCGCCAACTACATCGCCAAGGTGCGCTTCCTGGAGCAGCAGAATAAGATCCTGCTGGCCGAGCTCGAGCAGCTCAAGGGCCAAGGC
AAGTCGCGCCTGGGGGACCTCTATGAGGAGGAGATGCGGGAGCTGCGCCGCCAGGTGGACCAGCTAACCAACGACAAAGCCCGCGTCGAGGTGGAGCGCGACAACCTGGC
CGAGGACATCATGCGCCTCCGGGAGAAATTGCAGGAGGAGATGCTTCAGAGAGAGGAAGCCGAAAACACCCTGCAATCTTTCAGACAGGATGTTGACAATGCGTCTCTGG
CAGTCTTGACCTTGAACGCAAAGTGGAATCTTTGCAAGAAGAGATTGCCTTTTTGAAGAAACTCCACGAAGAGGAAATCCAGGAGCTGCAGGCTCAGATTCAGGAACAG
CATGTCCAAATCGATGTGGATGTTTCCAAGCCTGACCTCACGGCTGCCTGCTGTGACGTACGTCAGCAATATGAAAGTGTGGCTGCCAAGAACCTGCAGGAGGCAGAAGA
ATGGTACAAATCCAAGTTTGCTGAGCTCTCTGAGGCTGCCAACGGAACAATGACGCCCTGCGCCAGGCAAAGCAGGAGTCCACTGAGTACCGGAGACAGGTGCAGTCGC
TCACCTGTGAAGTGGATGCCCTTAAAGGAACCAATGAGTCCCTGGAACGCCAGATGCGTGAAATGAAGAGAACTTTGCCGTTGAAGCTGCTAACTACCAAGACACTATT
GGTCGCCTGCAGGATGAGATTCAGAATATGAAGGAGGAAATGGCTCGTCACCTTCGTGAATACCAAGACCTGCTCAATGTTAAGATGGCCCTTGATATTGAGATTGCCAC
CTACAGAAAGCTGCTGGAAGGCGAGGAGAGCAGGATTTCTCTGCCTCTTTCCAAACTTTTCCTCCCTGAACCTGAAGGAAACTAATCTGGATTCACTCCCTCTGGTTGATA
CCCACTCAAAAAGGACACTTCTGATTAAGACGGTTGAAACTAGAGATGGACAGGTTATCAACGAAACTTCTCAGCATCACGATGACCTTGAATAAAAATTGCACACACTC
AGTGCAGCAATATATTACCAGCAAGAATAAAAAAGAAATCCATATCTTAAAGAAACAGCTTTCAAGTGCCTTTCTGCAGTTTTTCAGGAGCGCAAGATAGATTTGGAATA
GGAATAAGCTCTAGTTCTTAACAACCGCAGCTCCTCCAAGATTTTGAAAAAAGTTTCCCAGATTATCTAGTTTCCCGAAAAATTTGGTGGTAGAAAACTTTTTACAAGGT
ATTTTGAATACCATAAAACTGCTTTTTTTTTCCAGCAAGTATCCAACCAACTTGGTTCTGCTTCAATAAATCTTTGGAAAAACTC

>vim variante, **1847 bases, 3DE7CB32**

GTCCCCGCGCCAGAGACGCAGCCGCGGCTCCCACCACCCACACGCACCGCGGCCTCGTTCGCCTCTTCTCCGGGAGCCAGTCCGCGCCACCGCCGCCGCCCAGGCCATCGC
CACCCTCCGCAGCCATGTCCACCAGGTCCGTGTCCTCGTCCTCCTACCGCAGGATGTTCGGCGGCCCGGGCACCGCGAGCCGGCCGAGCTCCAGCCGGAGCTACGTGACT
ACGTCCACCCGCACCTACAGCCTGTGCAGCGCGCTGCGCCCCAGCACCAGCCGCAGCCTCTACGCCTCGTCCCCGGGCGGCGTGTATGCCACGCGCTCCTCTGCCGGGCG
CCTGCGGAGCAGCGTGCCCGGGGTGCGGCTCCTGCAGGACTCGGTGGACTTCTCGCTGGCCGACGCCATCAACACCGAGTTCAAGAACACCCCGACCAACGAGAAGGTGG
AGCTGCAGGAGCTGAATGACCGCTTCGCCAACTACATCGCCAAGGTGCGCTTCCTGGAGCAGCAGAATAAGATCCTGCTGGCCGAGCTCGAGCAGCTCAAGGGCCAAGGC
AAGTCGCGCCTGGGGGACCTCTATGAGGAGGAGATGCGGGAGCTGCGCCGCCAGGTGGACCAGCTAACCAACGACAAAGCCCGCGTCGAGGTGGAGCGCGACAACCTGGC
CGAAGACATCATGCGCCTCCGGGAGAAATTGCAGGAGGAGATGCTTCAGAGAGAGGAAGCCGAAAACACCCTGCAATCTTTCAGACAGGATGTTGACAATGCGTCTCTGG
GCAGTCTTGACCTTGAACGCAAAGTGGAATCTTTGCAAGAAGAGATTGCCTTTTTGAAGAAACTCCACGAAGAGGAAATCCAGGAGCTGCAGGCTCAGATTCAGGAACAG
CATGTCCAAATCGATGTGGATGTTTCCAAGCCTGACCTCACGGCTGCCTGCTGTGACGTACGTCAGCAATATGAAAGTGTGGCTGCCAAGAACCTGCAGGAGGCAGAAGA
ATGGTACAAATCCAAGTTTGCTGAGCTCTCTGAGGCTGCCAACGGAACAATGACGCCCTGCGCCAGGCAAAGCAGGAGTCCACTGAGTACCGGAGACAGGTGCAGTCGC
TCACCTGTGAAGTGGATGCCCTTAAAGGAACCAATGAGTCCCTGGAACGCCAGATGCGTGAAATGAAGAGAACTTTGCCGTTGAAGCTGCTAACTACCAAGACACTATT
GGTCGCCTGCAGGATGAGATTCAGAATATGAAGGAGGAAATGGCTCGTCACCTTCGTGAATACCAAGACCTGCTCAATGTTAAGATGGCCCTTGATATTGAGATTGCCAC
CTACAGAAAGCTGCTGGAAGGCGAGGAGAGCAGGATTTCTCTGCCTCTTTCCAAACTTTTCCTCCCTGAACCTGAAGGAAACTAATCTGGATTCACTCCCTCTGGTTGATA
CCCACTCAAAAAGGACACTTCTGATTAAGACGGTTGAAACTAGAGATGGACAGGTTATCAACGAAACTTCTCAGCATCACGATGACCTTGAATAAAAATTGCACACACTC
AGTGCAGCAATATATTACCAGCAAGAATAAAAAAGAAATCCATATCTTAAAGAAACAGCTTTCAAGTGCCTTTCTGCAGTTTTTCAGGAGCGCAAGATAGATTTGGAATA
GGAATAAGCTCTAGTTCTTAACAACCGCAGCTCCTCCAAGATTTTGAAAAAAGTTTCCCAGATTATCTAGTTTCCCGAAAAATTTGGTGGTAGAAAACTTTTTACAAGGT
ATCCGGAAAACCCTTACACCTGCTCTTTTTTTTCCAGCAAGTATCCAACCAACTGGGTTCTGCTTCAATAAATCTTTGGAAAAACTC

Figura 8a

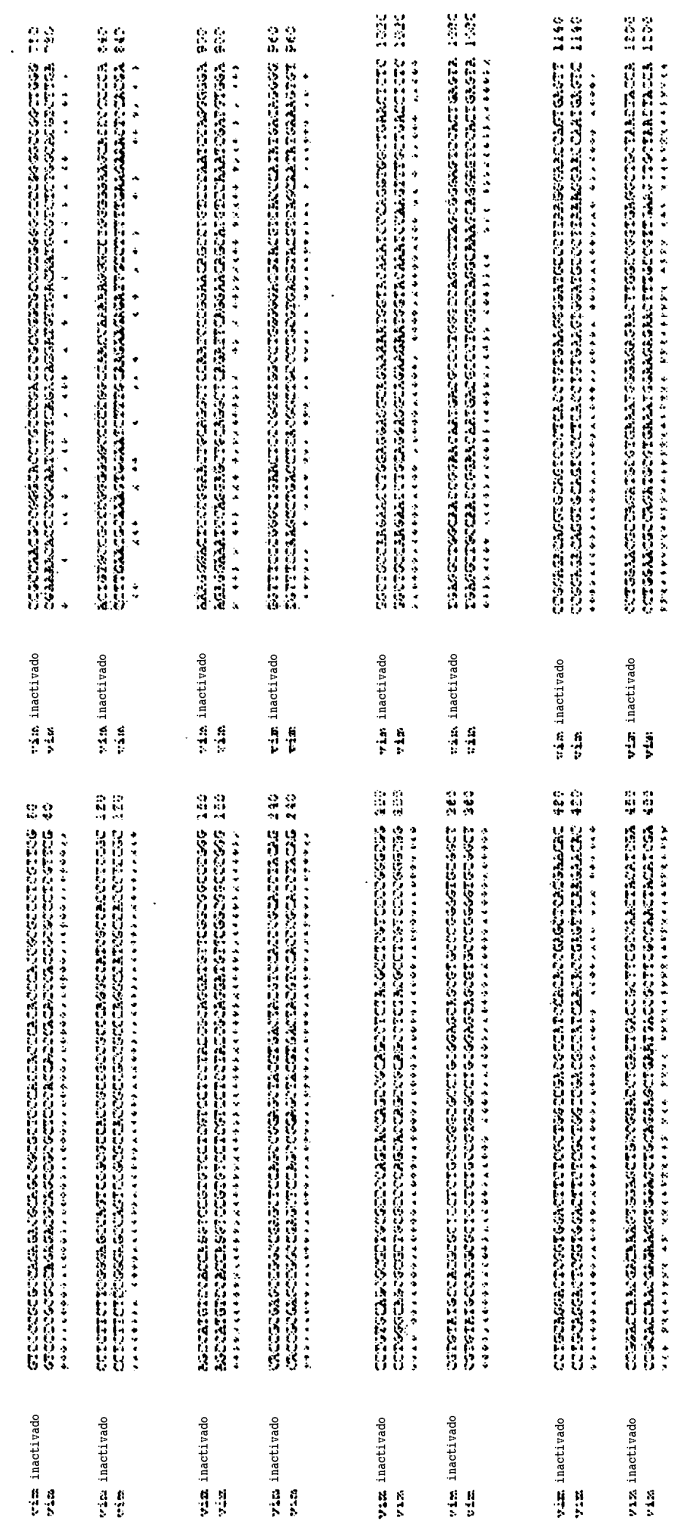


Figura 8b

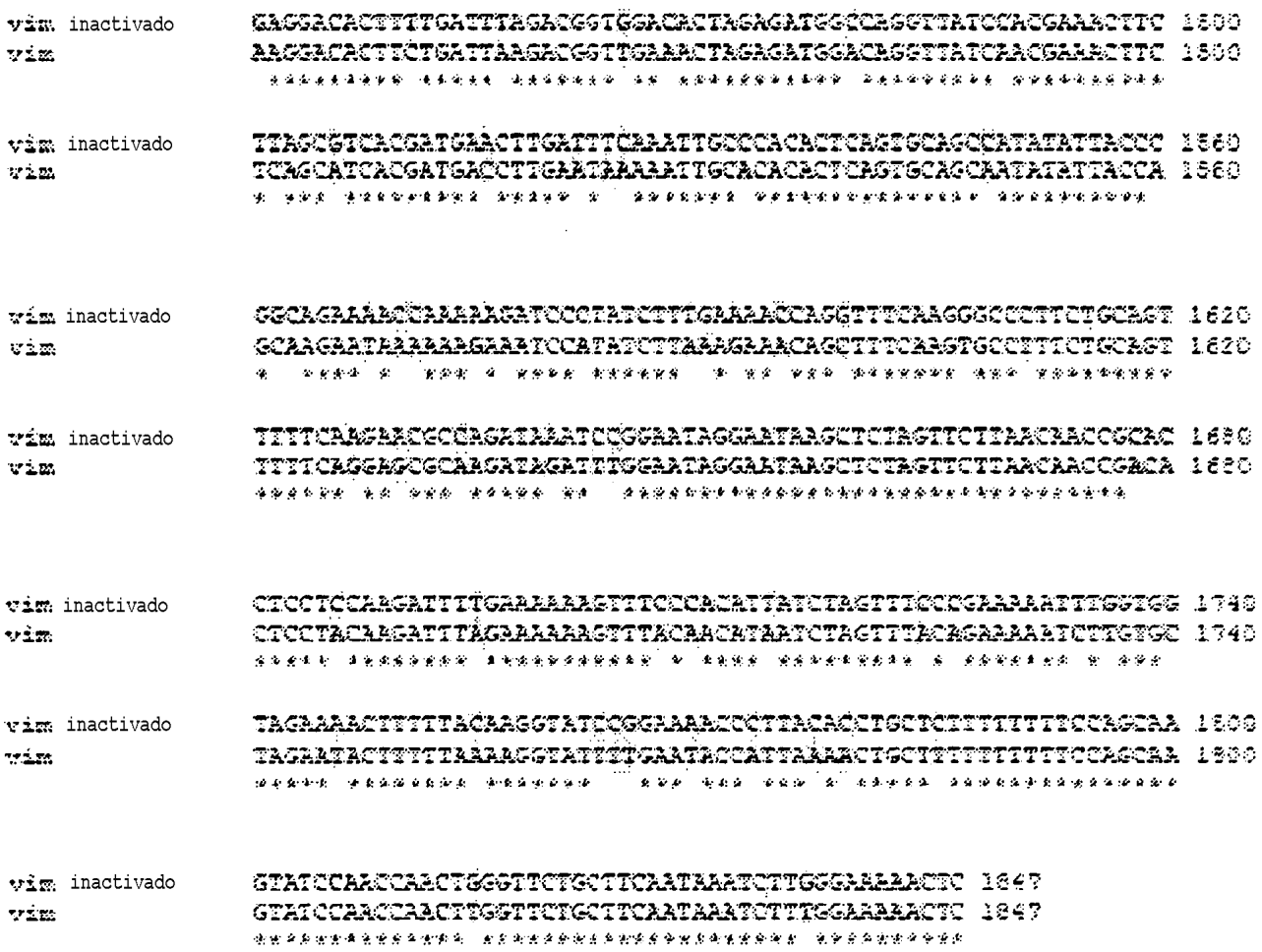

vim inactivado
vim
vim inactivado
vim
vim inactivado
vim
vim inactivado
vim
vim inactivado
vim
vim inactivado
vim

Figura 8c

>VIM 1 a 1401 Fase 1 (sentido directo) **467 aa**

MSTRSVSSSSYRRMFGGPGTASRPSSSRSYVTTSTRTYSLGSALRPSTSRSLYASSPGGVYATRSSAVRLRSSVPGVRLLQDSVDFSLADAINTEFKNTRTNEKVELQEL
NDRFANYIDKVRFLEQQNKILLAELEQLKGQGKSRLGDLYEEEMRELRRQVDQLTNDKARVEVERDNLAEDIMRLREKLQEEMLQREEAENTLQSFRQDVDNASLARLDL
ERKVESLQEEIAFLKKLHEEEIQELQAQIQEQHVQIDVDVSKPDLTAALRDVRQQYESVAAKNLQEAEEWYKSKFADLSEAANRNNDALRQAKQESTEYRRQVQSLTCEV
DALKGTNESLERQMREMEENFAVEAANYQDTIGRLQDEIQNMKEEMARHLREYQDLLNVKMALDIEIATYRKLLEGEESRISLPLPNFSSLNLRETNLDSLPLVDTHSKR
TLLIKTVETRDGQVINETSQHHDDLE*

>VIM variante **1 a 1723** Fase 1 (sentido directo) **574 aa**

MSTRSVSSSSYRRMFGGPGTASRPSSSRSYVTTSTRTYSLGSALRPSTSRSLYASSPGGVYATRSSAGRLRSSVPGVRLLQDSVDFSLADAIHIELINTKTNEKVELPDL
IDRFANYIDKGRFLEQQTPILLAKLQHLQGQGKSRLGDLYQEEMRELRRPGDQLNNEQARVEGKREKLGEDLMRLGEPREPKPPFKKGSSRQRRAPARPAGAPDPRGRIGL
CRPGGFRPTKKGLGEAFPQKGLPELQAPIPEQPVFIQGEVSRAELPGGLGEVRHPYDRGAAANLEEREKWYKSQVAELSERGNRNNDALGQA*PESTEYRRQVQSLTCEG
DALKGTSEFLERQMREMGENLAGEAANYQETMGRLQDEIQYMKEEMARHLREYQELLNVQLALDIEIATYRKLLEGEESRISLALPNFSSLNLRETNNLDSLPLVDTQSKR
TLLI*TVDTRDGQVINETS*RHDELDEKLPTLSAAIYYPAENQKDFYL*KPGEKGPSAVFQERQINPE*E*ALVLNNRTSSKILKKVSHII*EPEKPGGRKLFTRYPENP
YTCSFFSSKYPTNWVLLQ*ILGKI

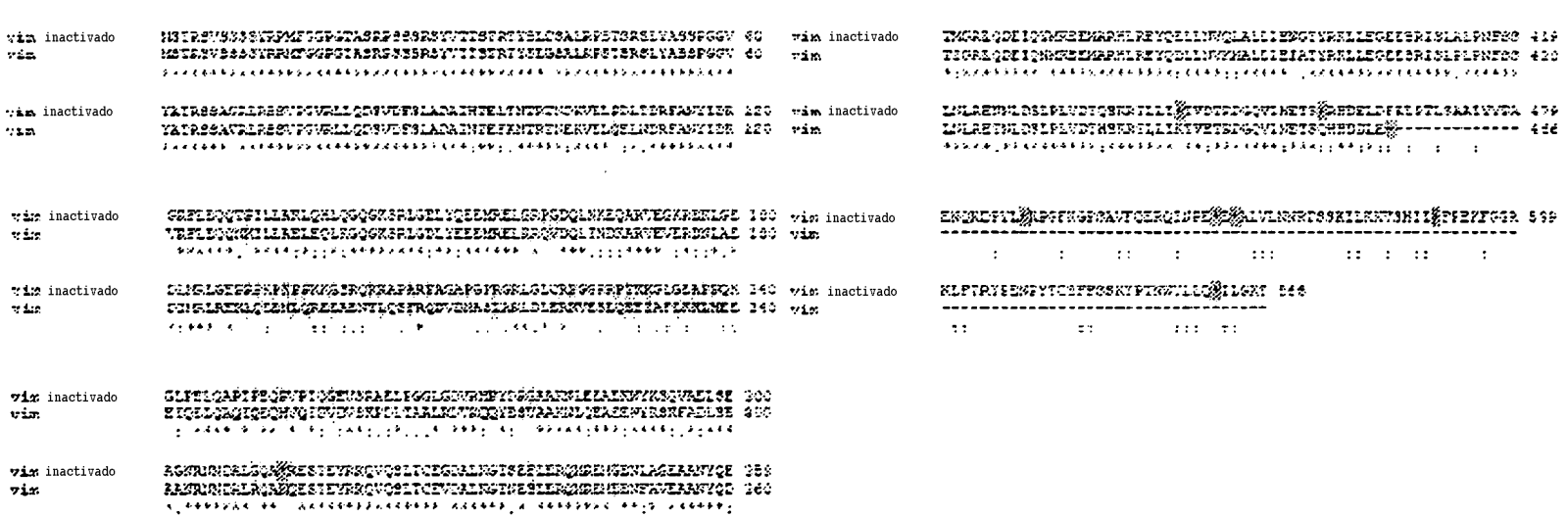
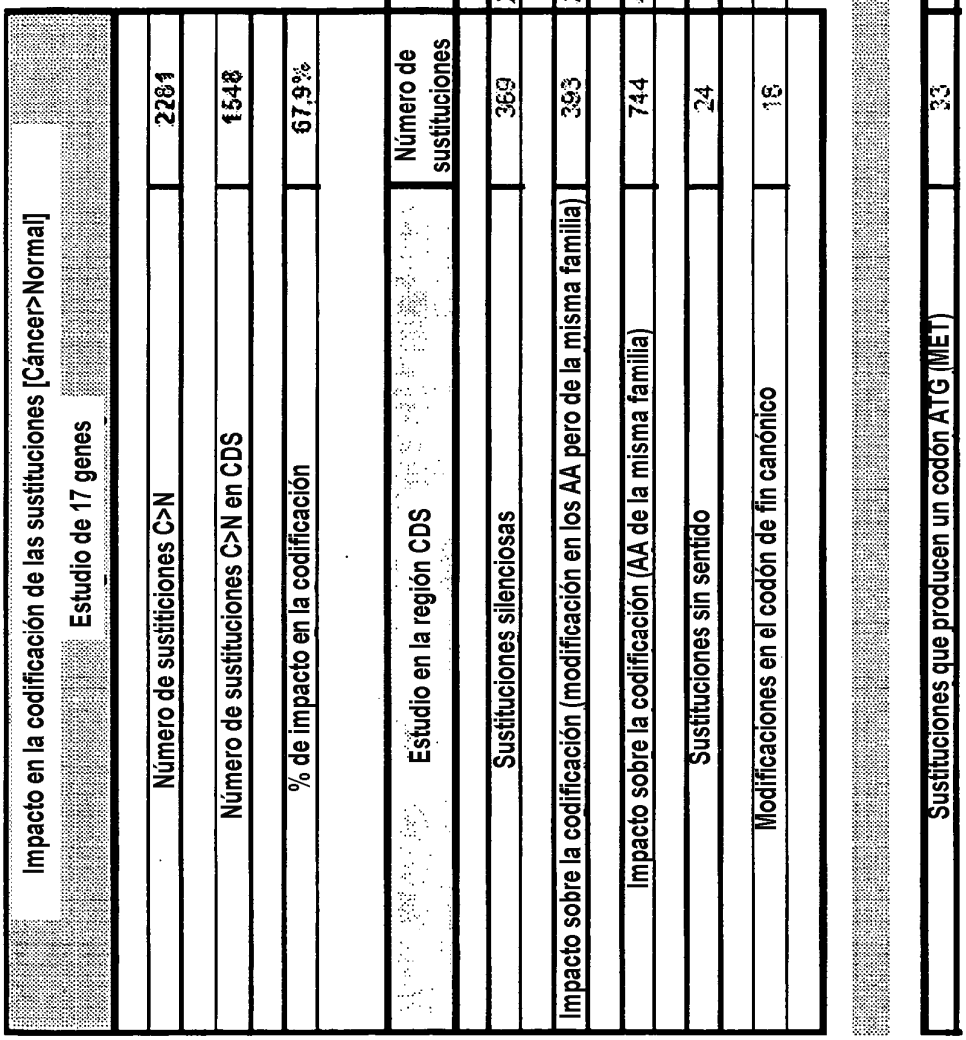


Figura 8d



Impacto en la codificación de las sustituciones [Cáncer>Normal]	
Estudio de 17 genes	
Número de sustituciones C>N	2281
Número de sustituciones C>N en CDS	1548
% de impacto en la codificación	67,9%

Estudio en la región CDS	Número de sustituciones	%
Sustituciones silenciosas	369	23,8%
Impacto sobre la codificación (modificación en los AA pero de la misma familia)	393	25,4%
Impacto sobre la codificación (AA de la misma familia)	744	48,1%
Sustituciones sin sentido	24	1,6%
Modificaciones en el codón de fin canónico	18	1,2%
Sustituciones que producen un codón ATG (MET)	33	2,1%

Figura 9

>gi|4507669|ref|NP_003286.1| proteína tumoral, traduccionalmente controlada 1 [Homo sapiens]

>gi|62414289|ref|NP_003371.2| vimentina [Homo sapiens]

>gi|17158044|ref|NP_001001.2| proteína ribosómica S6 [Homo sapiens]

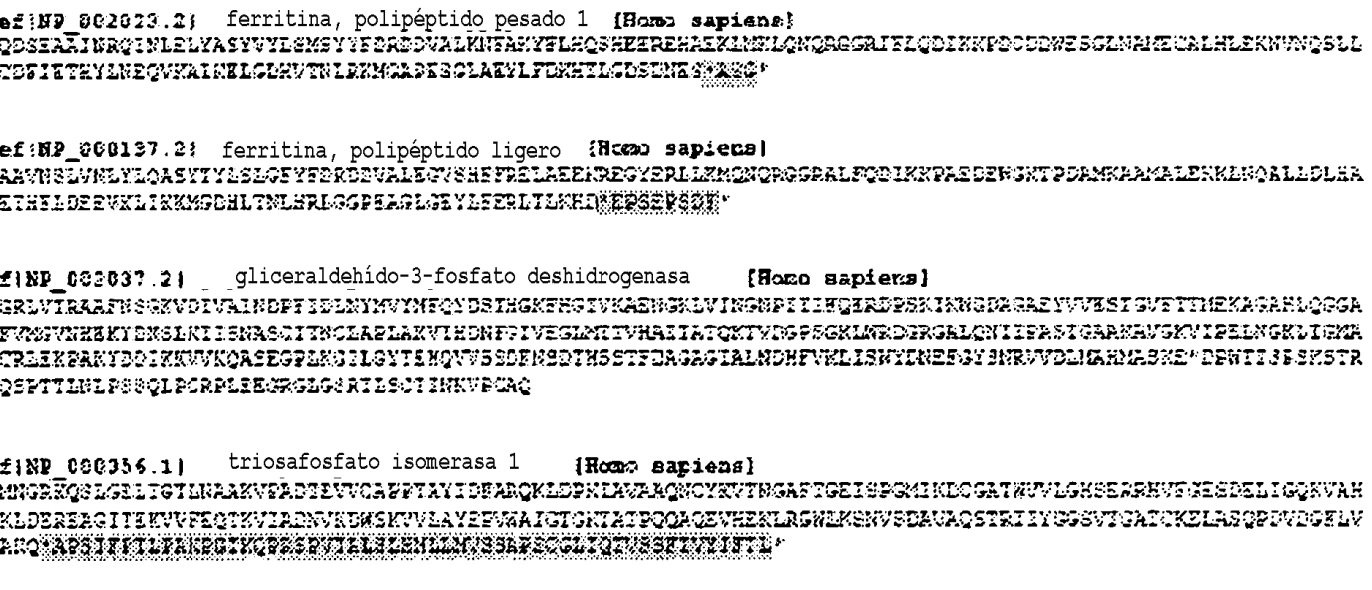
>gi|4506661|ref|NP_000963.1| proteína ribosómica L7a [Homo sapiens]

>gi|4506725|ref|NP_000998.1| proteína ribosómica S4, isoforma X unida a X [Homo sapiens]

proteínas normales

otros péptidos de proteínas cuyo codón de fin se ha visto afectado

Figura 10a



péptidos normales

otros péptidos de proteínas cuyo codón de fin se ha visto afectado

Figura 10b

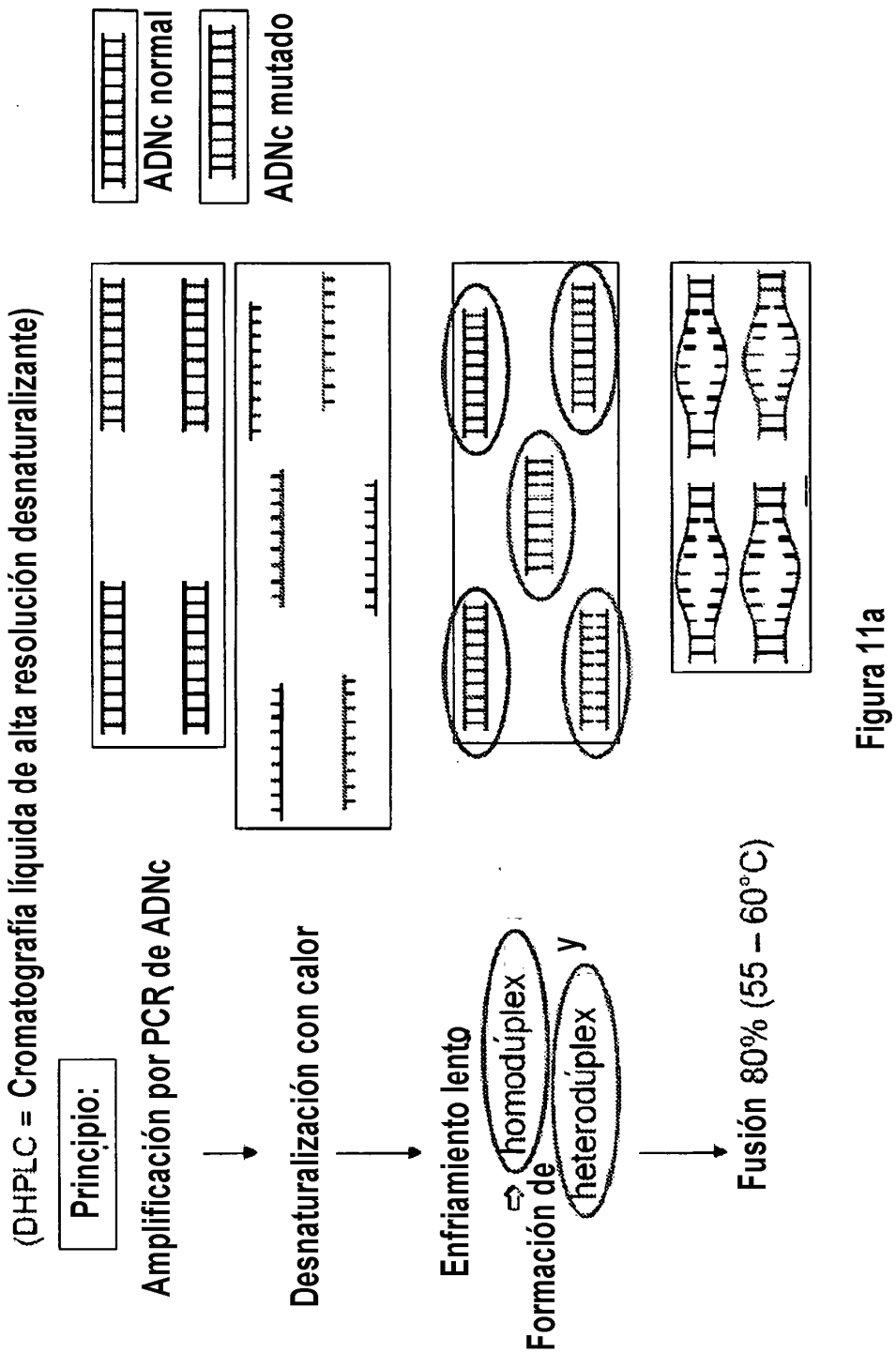

(DHPLC = Cromatografía líquida de alta resolución desnaturalizante)
Principio:
Amplificación por PCR de ADNc
Desnaturalización con calor
Enfriamiento lento
⇨ Formación de homodúplex y heterodúplex
Fusión 80% (55 – 60°C)
ADNc normal
ADNc mutado

Figura 11a

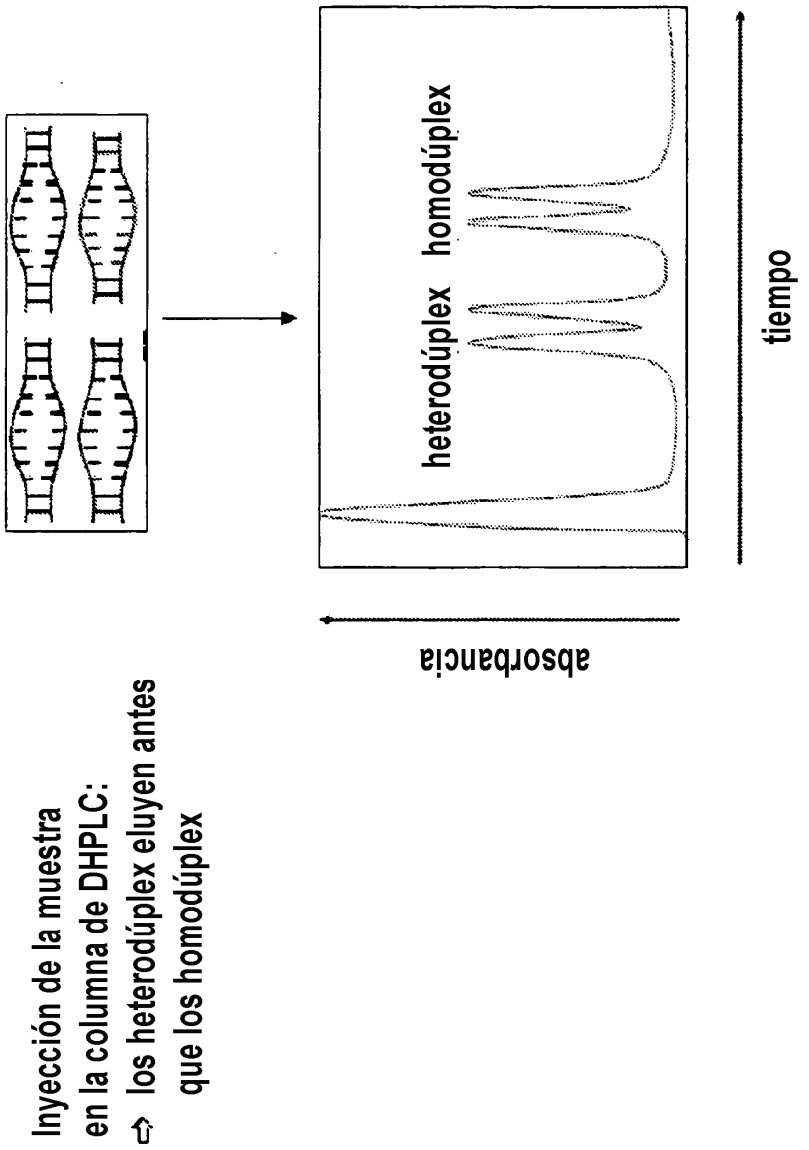

heterodúplex
homodúplex
tiempo
absorbancia
Inyección de la muestra
en la columna de DHPLC:
⇧ los heterodúplex eluyen antes
que los homodúplex

Figura 11b

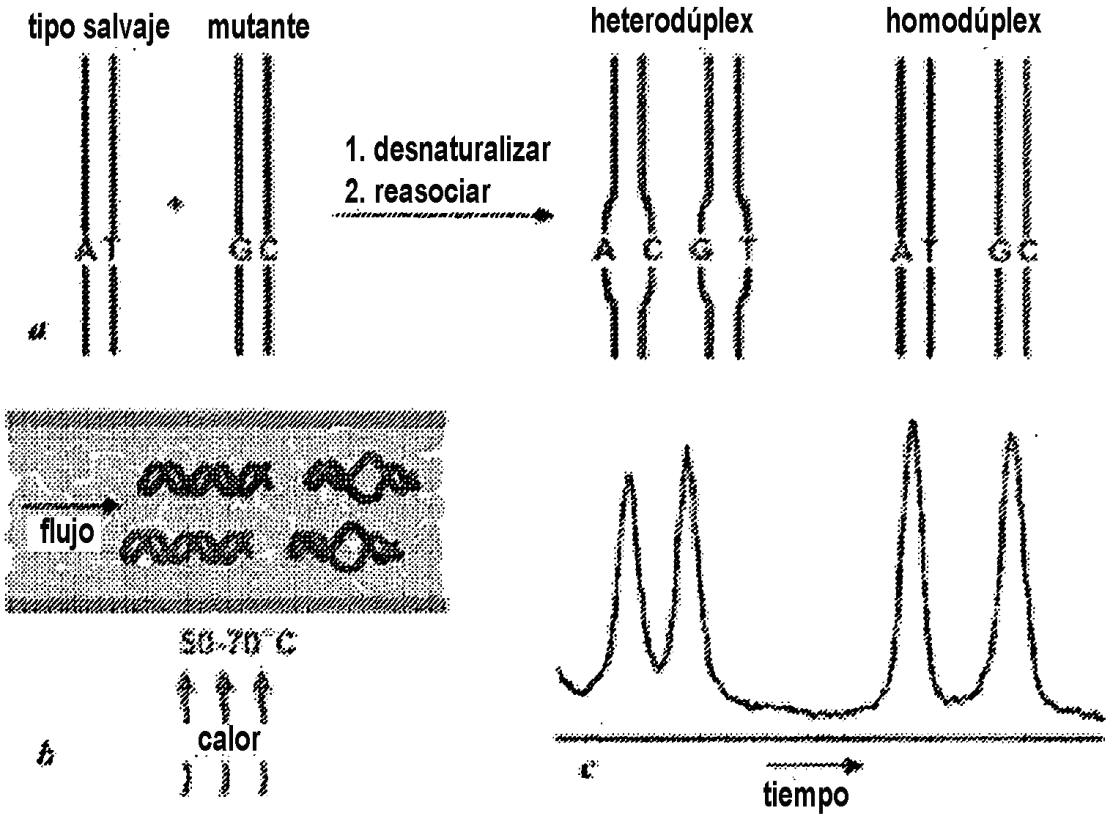


Figura 2. Principio de mutación detectada mediante DHPLC. a. Una DHPLC compara dos cromosomas generalmente como una mezcla de productos de la PCR desnaturalizados a 95 ºC durante 3 min y reasociados a lo largo de 30 min mediante un enfriamiento gradual desde 95 ºC a 65 ºC antes del análisis. En presencia de un desapareamiento, no sólo se forman los homodúplex originales sino también las hebras sentido y antisentido de cualquiera de los homodúplex forman heterodúplex. B. Los heterodúplex se desnaturalizan en mayor grado a las temperaturas del análisis (que varían de 50 ºC a 70ºC), y eluyen antes que los homodúplex en la columna de separación de ADN. c. Los correspondientes patrones cromatográficos muestran cuatro picos diferentes que pertenecen a cuatro especies diferentes de ADN.

Fuente: Theru A. Sivakumaran et al., Current Science, 84:291-296, 2003

Figura 11c

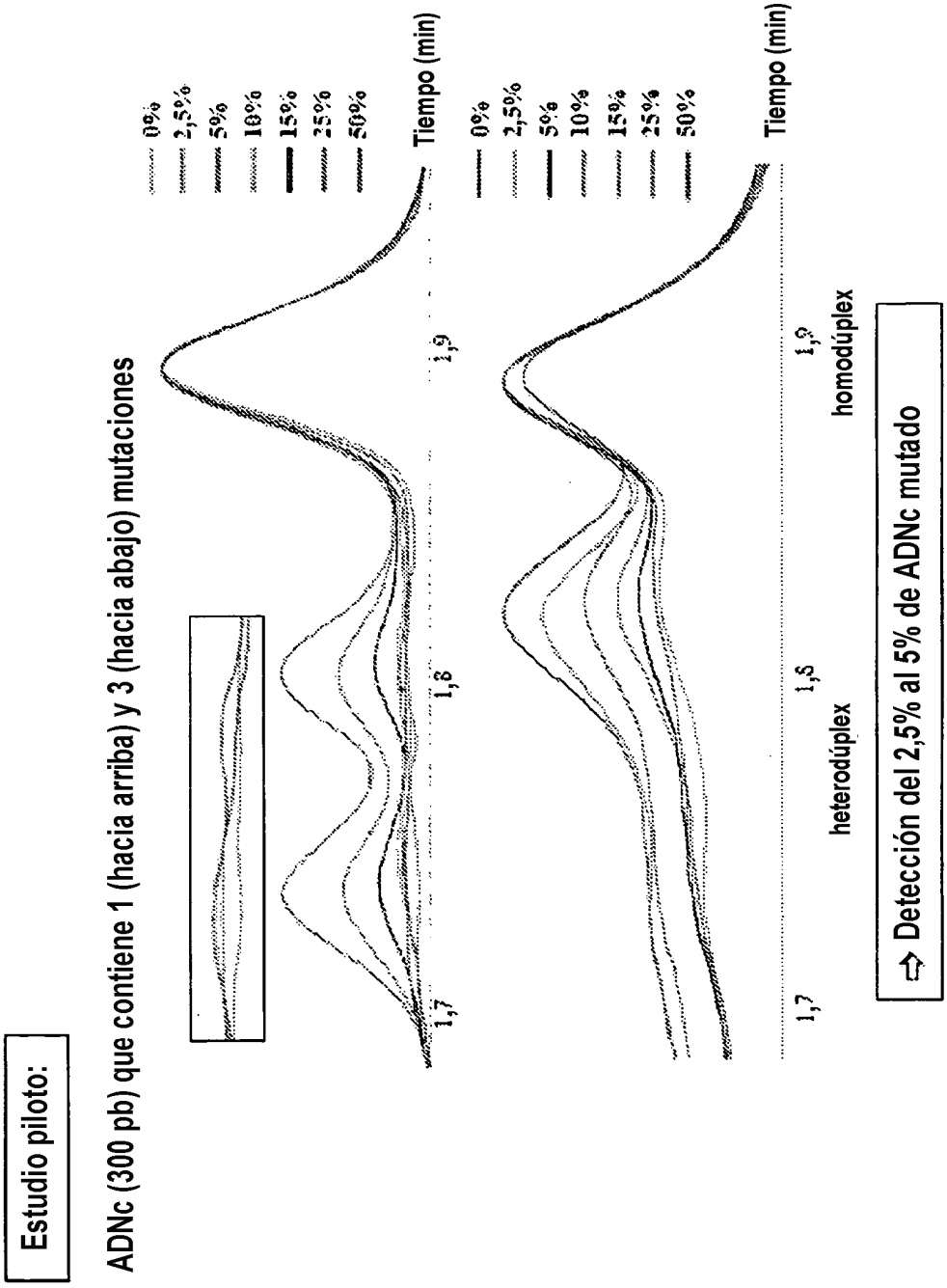

Estudio piloto:
ADNc (300 pb) que contiene 1 (hacia arriba) y 3 (hacia abajo) mutaciones
0%
2,5%
5%
10%
15%
25%
50%
Tiempo (min)
1,7
1,8
1,9
heterodúplex
homodúplex
⇨ Detección del 2,5% al 5% de ADNc mutado

Figura 11d

```
   5'CTGCCAAGGCTGTGGGCAAG 3'     713-732
   5'GAAATGAGCTTGACAAAGTGGTCGT 3'     986-1010
Longitud del producto de la amplificación : 298     713-1010
   GC% : 58,0%
         Temperatura de hibridación sugerida (TA) : 58,9
   Hebra del cebador + :
   Hebra del cebador - :
   Cebador  :  :     5' CTGCCAAGGCTGTGGGCAAG 3'
                     ||||||||||||||||||||
    Diana   : 713 : 5' CTGCCAAGGCTGTGGGCAAG 3'
puntuación : 174
   TM : 74,4 / 52,8 / 60,9
   Cebador  :  :     3' ACGACCACTTTGTCAAGCTCATTTC 5' (complemento)
                     ||||||||||||||||||||||||
    Diana   : 986 : 5' ACGACCACTTTGTCAAGCTCATTTC 3'
puntuación : 181
   TM : 72,5 / 50,9 / 58,314
```

cebadores

productos de la amplificación

>gi|8364890|ref|NM_001046.3| Gliceraldehído-3-fosfato deshidrogenasa de *Homo sapiens* (GAPDH), ARNm

```
1
CTCTCTGCTCCTCCTGTTCGACAGTCAGCCGCATCTTCTTTTGCGTCGCCAGCCGAGCCACATCGCTCAGACACCATGGGGAAGGTGAAGGTCGGAGTCAACGGATTTGGTCGTATTGG
GGCGCCTGGTCACCAGGGCTGCTTTTAACTCTGGTAAAGTGGATATTGTTGCCATCAATGACCCCTTCATTGACCTCAACTACATGGTTTACATGTTCCAATATGATTCCACCCATGGG
CAAATTCCATGGCACCGTCAAGGCTGAGAACGGGAAGCTTGTCATCAATGGAAATCCCATCACCATCTTCCAGGAGCGAGATCCCTCCAAAATCAAGTGGGGCGATGCTGGCGCTGAG
TACGTCGTGGAGTCCACTGGCGTCTTCACCACCATGGAGAAGGCTGGGGCTCATTTGCAGGGGGGAGCCAAAAGGGTCATCATCTCTGCCCCCTCTGCTGATGCCCCCATGTTCGTCA
TGGGTGTGAACCATGAGAAGTATGACAACAGCCTCAAGATCATCAGCAATGCCTCCTGCACCACCAACTGCTTAGCACCCCTGGCCAAGGTCATCCATGACAACTTTGGTATCGTGGAA
AGGACTCATGACCACAGTCCATGCCATCACTGCCACCCAGAAGACTGTGGATGGCCCCTCCGGGAAACTGTGGCGTGATGGCCGCGGGGCTCTCCAGAACATCATCCCTGCCTCTACT
GGCGCTGCCAAGGCTGTGGGCAAGGTCATCCCTGAGCTGAACGGGAAGCTCACTGGCATGGCCTTCCGTGTCCCCACTGCCAACGTGTCAGTGGTGGACCTGACCTGCCGTCTAGAAA
AACCTGCCAAATATGATGACATCAAGAAGGTGGTGAAGCAGGCGTCCGAGGGCCCCCTCAAGGGCATCCTGGGCTACACTGAGCACCAGGTGGTCTCCTCTGACTTCAACAGCGACAC
CCACTCCTCCACCTTTGACGCTGGGGCTGGCATTGCCCTCAACGACCACTTTGTCAAGCTCATTTCCTGGTATGACAACGAATTTGGCTACAGCAACAGGGTGGTGGACCTCATGGCCC
CACATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCTGGGGAGTCCCTGCCACACTCAGTCCCCCACCACACTGA
ATCTCCCCTCCTCACAGTTGCCATGTAGACCCCTTGAAGAGGGGAGGGGCCTAGGGAGCCGCACCTTGTCATGTACCATCAATAAAGTACCCTGTGCTCAACCT
```

Figura 12a

```
   5' TCCCCTCGTGTCAGCTCAGG 3' 1505-1524
   5' CTCATGGGTCACTGAGGCTTTTTAT 3' 1788-1812
Longitud del producto de la amplificación : 308    1505-1812
   GC% : 56,0%
        Temperatura de hibridación sugerida (TA) : 58,1
  Hebra del cebador + :
  Hebra del cebador - :
  Cebador : :  :     5' TCCCCTCGTGTCAGCTCAGG 3'
                        ||||||||||||||||||||
    Diana : : 1505 : 5' TCCCCTCGTGTCAGCTCAGG 3'
puntuación : 174             TM : 74,4 / 52,8 / 60,
    Cebador :  :     3' ATAAAAAGCCTCAGTGACCCATGAG 5' (complemento)
                        |||||||||||||||||||||||||
    Diana   : 1788 : 5' ATAAAAAGCCTCAGTGACCCATGAG 3'
puntuación : 181             TM : 72,5 / 56,9 / 57,1
```

>gi|16507965|ref|NM_001428.2| Enolasa 1 de *Homo sapiens*, (alfa) (ENO1), ARNm

1

TAGCTAGGCAGGAAGTCGGCGCGGGCGGCGCGGACAGTATCTGTGGGTACCCGGAGCACGGACATCTCGCCGGCTTTACGTTCACCTCGGTGTCTGCAGCACCCTCCGCTTCCTCTCCTA
GGCGACGCAGACCCAGTGGCTAGAAGTTCACCATGTCTATTCTCAAGATCCATGCCAGGCAGATCTTTGACTCTCGCGGGAATCCCACTGTTGAGGTTGATCTCTTCACGTCAAAGGTCT
CTTCAGAGCTGCTGTGCCCAATGGTGCTTCAACTGGTATCTATGAGGCCCTAGAGCTCCGGGACAATGATAAGACTCGCTATATGGCGAAGGGTGTCTCAAAGGCTGTTGAGCACATCAA
TAAAACTATTGCGCCTGCCCTCGTTAGCAAGAAACTGAACGTCACAGAACAAGAGAAGATTGACAAACTCATGATCGAGATGGATGGAACAGAAAATAAATCTAAGTTTGGTGCGAACGC
CATTCTGCGCGGTGTCCGTTGCCGTCTGCAAAGCTGGTGCCGTTGAGAAGGGGTCCCCCTGTACCGCCACATCGCTCACTTGCCTGCCAACTCTCAAGTCATCCTGCCAGTCCCGCGTT
CAATGTCATCAATGGCGGTTCTCATGCTGGCAACAAGCTGGCCATGCAGGAGTTCATGATCCTCCCAGTCGGTGCAGCAAACTTCAGGGAAGCCATGCGCATTGGAGCAGAGGTTTACCA
CAACCTGAAGAATGTCATCAAGGAGAAATATGGGAAGATGCCACCAATGTGGGGGATGAAGGCGGGTTTGCTCCCAACATCCTGGAGAATAAAGAAGGCCTGGAGCTGCTGAAGACTGC
TATTGGGAAAGCTGGCTACACTGATAAGGTGGTCATCGGCATGGACGTAGCGGCCTCCGAGTTCTTCAGGTCTGGGAAGTATGACCTGGACTTCAAGTCTCCCGATGACCCCACCAGGTA
CATCTCGCCTGACCAGCTGGCTGACCTGTACAAGTCCTTCATCAAGGACTACCCAGTGGTGTCTATCGAAGATCCCTTTGACCAGGATGACTGGGGAGCTTGGCAGAAGTTCACAGCCAG
TGCAGGAATCCAGGTAGTGGGGGATGATCTCACAGTGACCAACCCCAAAAGAGGATCGGCCAAGGCGGTGAACGAGAAGTCCTGCAACTGCCTCCTGCTCAAAGTCAACCAGATTGGCTCCGT
GACCGAGTCTCTTCAGGCGTGCAAGCTGGCCCAGGCCAATGGTTGGGGCGTCATGGTGTCTCATCGTTCGGGGGAGACTCAAGATACCTTCATCGCTGACCTGGTTGTGGGGCTGTGCAC
TGGGCAGATCAAGACTGGTGCCCCTTGCCGATCTGAGCGCTTGGCCAAGTACAACCAGCTCCTCAGAATTGAAGAGGAGCTGGGCAGCAAGGCTAAGTTTGCCGGCAGGAACTTCAGAAA
CCCCTTGCCCAAGTAAGCTGTGGGCAGGCAAGCCCTTCGGTCACCTGTTGCCTACACAGACCCCTCCCCTCGTGTCAGCTCAGGCACCTCGACGCCCCTGACCAAGTCTTGCACGGCTCC
CTGCTAGTTAGTCGCGGACCGGGGTGGAGTTCGTACCGCTTGCTTAGAGTTCTACAGAAGGGAAGCTCCCTGGAGCCTGTTGCAGGTCTAGCTTTGCAGTCGTGTAATTGGGGCAAG
TCATTGTTTTTCTTGGCTCAGTTTCCAGCAAACTGTCTACAGTCATCTTAGCCTCGTGTCATCTCAGGGGTGGCCAGAGGCTAGATCCCGCGTGGTTTTGTGCTCAATATAAAAAGCTGA
GTGACCCATGAGT
```

**Figura 12b**
```

```
    3' GGATCAAGTTTAAATGACTGTGCTG 3'     311-335
    5' TTTTATTTGCATATTAAAAAAAATTGTGC 3'     590-614
Longitud del producto de la amplificación : 304
    GC% : 42,0%
           Temperatura de hibridación sugerida (IA) : 53,3
    Hebra del cebador + :
    Hebra del cebador - :
     Cebador :  :     5' GGATCAAGTTTAAATGACTGTGCTG 3'
                         ||||||||||||||||||||||||||
      Diana  : 311 : 5' GGATCAAGTTTAAATGACTGTGCTG 3'
  puntuación : 181
    TM : 70,9 / 48,3 / 56,
     Cebador :  :     3' GCACAATTTTTTTAATATGCAAATAAAA 5' (complemento)
                         ||||||||||||||||||||||||||||
      Diana  : 587 : 5' GCACAATTTTTTTAATATGCAAATAAAA 3'
  puntuación : 184
    TM : 64,4 / 42,8 / 56,5
```

>gi|34328943|ref|NM_021109.2| Timosina de *Homo sapiens*, beta 4, unida a X (TMSB4X), ARNm

¶

ACAACTCGGTGGTGGCCACTGCGCAGACCAGACTTCGCTCGTACTCGTGCGCCTCGCTTCGCTTTTCCTCCGCAACCATGTCTGACAAACCCGATATGGCTGAGATCGAGAAATTCGATA
AGTCGAAACTGAAGAAGACAGAGACGCAAGAGAAAAATCCACTGCCTTCCAAAGAAACGATTGAACAGGAGAAGCAAGCAGGCGAATCGTAATGAGGCGTGCGCCGCCAATATGCACTGT
ACATTCCACAAGCATTGCCTTCTTATTTTACTTCTTTTAGCTGTTTAACTTTGTAAGATGCAAAGAGCTTCGATCAACTTTAAATGACTCTGCTGCCGCTTTGAGATCAAAGAACTACTG
GCAACGAAGCGCGCGCTCTCCTTTCCTTCTGTCTATCTATCTGGCTGCCAGGGAAGCAAAGAACTTCAATGTTGCTCAAGCAAGAAGTGCGGTGGCAATGAAGTGCGGTGGCAACGACAGTG
AAATCTAGAGTAAAACCAAGCTTGCCCATGCTTCCTGCCTGTAATGCAGTTAATTGAGTGCCATTTTTTTTTTTGTTCAAACTCATTTTAATTAATTGCTATGCACAATTTTTTTTA
ATATCCAAATAAAAAGTTTAAAAACTT¶

Figura 12c

		Expasy	base de datos de proteosoma plasmático	
		proteína		ARNm.
Símbolo	Nombre	SP nº de registro	PPD nº de registro	PPD nº de registro
A2M	alfa-2-macroglobulina	P01023	NP_000005	NM_000014
AHSG	alfa-2-HS-glicoproteína	P02765	NP_001613	NM_001622
ALB	albúmina	P02768	AAH39235	BC039235
APCS	componente P amiloide	P02743	NP_001630	NM_001639
APOA1	apolipoproteína A-I	P02647	NP_000030	NM_000039
APOA2	apolipoproteína A-II	P02652	NP_001634	NM_001643
APOC1	apolipoproteína C-I	P02654	NP_001636	NM_001645
APOC2	apolipoproteína C-II	P02655	NP_000483	NM_000474
APOC3	apolipoproteína C-III	P02656	NP_000031	NM_000040
APOD	apolipoproteína D	P05090	NP_001638	NM_001647
APOE	apolipoproteína E	P02649	NP_000032	NM_000041
AZGP1	Zn-2-alfa-glicoproteína	Q5XKQ4	CAA42438	X59766
B2M	beta-2- microglobulina	P61769	NP_004039	NM_004048
C1S	componente del complemento f, subcomponente s	P09871	NP_958850	NM_201442
C3	componente del complemento 3	Q6LDJ0	NP_000055	NM_000064
C7	componente del complemento 7	P10643	NP_000578	NM_000587
CDH1	E- cadherina	P12830	BAA88957	AB025106
CFB	factor del complemento B	P00751	NP_001701	NM_001710
CHI3L1	1 tipo quinitasa 3	P36222	NP_001267	NM_001276

Figura 13a

Símbolo	Nombre	Expasy	base de datos de proteosoma plasmático	
		proteínas		ARNm
		SP nº de registro	PPD nº de registro	PPD nº de registro
CLEC3B	familia del dominio de la lectina de tipo C 3, miembro B	P05452	NP_003269	NM_003278
CLU	clusterina	P10909	NP_001822	NM_001831
CP	ceruloplasmina	P00450	NP_000087	NM_000096
CRP	proteína reactiva C	P02741	AAH20766	BC020766
EDN1	endotelina I	P05305	NP_001946	NM_001955
EGFR	receptor del factor del crecimiento epidérmico	P00533	NP_958441	NM_201284
F13A1	factor de coagulación XIII, polipéptido A1	P00488	NP_000120	NM_000129
F13B	factor de coagulación XIII, polipéptido B	P05160	NP_001985	NM_001994
FGA	cadena alfa del fibrinógeno	P02671	NP_000499	NM_000508
FGB	cadena beta del fibrinógeno	Q32P65	NP_005132	NM_005141
FGG	cadena gamma del fibrinógeno	P02679	NP_066556	NM_021870
GPI	glucosa fosfato isomerasa	P06744	NP_000166	NM_000175
GSTP1	glutatión S-transferasa pi	P09211	NP_000843	NM_000852
HDLBP	ADNc FLJ45936 fis, muy similar a la vigilina	Q00341	BAC87153	AK127833
HP	haptoglobina	P00738	NP_005134	NM_005143
HPX	hemopexina	P02790	NP_000604	NM_000613
HRG	glicoproteína rica en histidina	P04196	NP_000403	NM_000412
IGFBP3	proteína de unión del factor del crecimiento de tipo insulínico 3	P17936	NP_000589	NM_000598
IGJ	polipéptido J de inmunoglobulina	P01591	NP_653247	NM_144646

Figura 13b

		Expasy	base de datos de proteosoma plasmático	
		proteínas		ARNm
Símbolo	Nombre	SP nº de registro	PPD nº de registro	PPD nº de registro
INHA	inhibina, alfa	P05111	NP_002182	NM_002191
KLK11	peptidasa 11 relacionada con calicreína	Q9UBX7	NP_659196	NM_144947
LDHA	lactato deshidrogenasa A	P00338	NP_005557	NM_005566
LGALS3BP	proteína de unión 3 de lectina, de unión a galactósido, soluble	Q08380	NP_005558	NM_005567
LRG1	alfa-2-glicoproteína 1 rica en leucina	P02750	NP_443204	NM_052972
ORM1	orosomucoide 1 de *Homo sapiens*	P02763	NP_000598	NM_000607
PKM2	piruvato quinasa, músculo	P14618	NP_872270	NM_182470
PLG	plasminógeno	P00747	NP_000292	NM_000301
PON1	paraoxonasa 1	P27169	NP_000437	NM_000446
PROC	proteína C	P04070	NP_000303	NM_000312
PZP	proteína de la zona de embarazo	P20742	NP_002855	NM_002864
RBP4	proteína 4 de unión a retinol	P02753	P02753	BC020633
SAA1	amiloide sérico A1	P02735	NP_000322	NM_000331
SERPINA1	inhibidor de serpín peptidasa, clado A	P01009	NP_000286	NM_000295
SERPINA3	inhibidor de serpín peptidasa, clado A, miembro 3	P01011	NP_001076	NM_001085
SERPINA5	inhibidor de serpín peptidasa, clado B, miembro 4	P05154	NP_033966	NM_002974
TF	transferrina	P02787	NP_001054	NM_001063
TFRC	receptor de transferrina	P02786	NP_003225	NM_003234
TGFB1	factor del crecimiento transformante, beta 1	P01137	NP_000651	NM_000660

Figura 13c

		Expasy	base de datos de proteosoma plasmático	
		proteínas		ARNm
Símbolo	Nombre	SP nº de registro	PPD nº de registro	PPD nº de registro
TIMP1	inhibidor de metalopeptidasa 1 TIMP	P01033	NP_003245	NM_003254
TTR	trantiretina	P02766	NP_000362	NM_000371
VTN	vitronectina	P04004	NP_000629	NM_000638

Figura 13d

Selección de 3 proteínas plasmáticas:

APOA1

RLREGQPPRPFRMKAAVLTLAVLFLTGSQARHFWQQDEPPQSPWDRVKDLATVYVDVLKDSGRDYVSQFEGSALGKQLNLKLLD
NWDSVTSTFSKLREQLGPVTQEFWDNLEKETEGLRQEMSKDLEEVKAKVQPYLDDFQKKWQEEMELYRQKVEPLRAELQEGAR
QKLHELQEKLSPLGEEMRDRARAHVDALRTHLAPYSDELRQRLAARLEALKENGGARLAEYHAKATEHLSTLSEKAKPALEDLRQG
LLPVLESFKVSFLSALEEYTKKLNTQ*GARRRPPSRCSE*TFPKW

Péptido de post-fin

APOA2

GTDTKDRDAQ*AALPTVTNMKLLAATVLLLTICSLEGALVRRQAKEPCVESLVSQYFQTVTDYGKDLMEKVKSPELQAEAKSYFEKS
KEQLTPLIKKAGTELVNFLSYFVELGTQPATQ*SVQTIVFQPQLASRTPTGQS*SSCPYPLFATINAE*!

Péptido de post-fin

APOC2

LWLWSQSQSQPL*ILSVPVRSW*GQPARVWSLDTMGTRLLPALFLVLLVLGFEVQGTQQPQQDEMPSPTFLTQVKESLSSYWESA
KTAAQNLYEKTYLPAVDEKLRDLYSKSTAAMSTYTGIFTDQVLSVLKGEE*QPDPPSVDKGRVPYSPDPPGSD*APPSQ*LLHPPPN
SSLNSFQ*KIQFKLLLMDGTAFLRTQGPRWRG*LSPANI**APTLCMEL*PMGPWE*SSE**D*IIVTKKKKKKKGKK

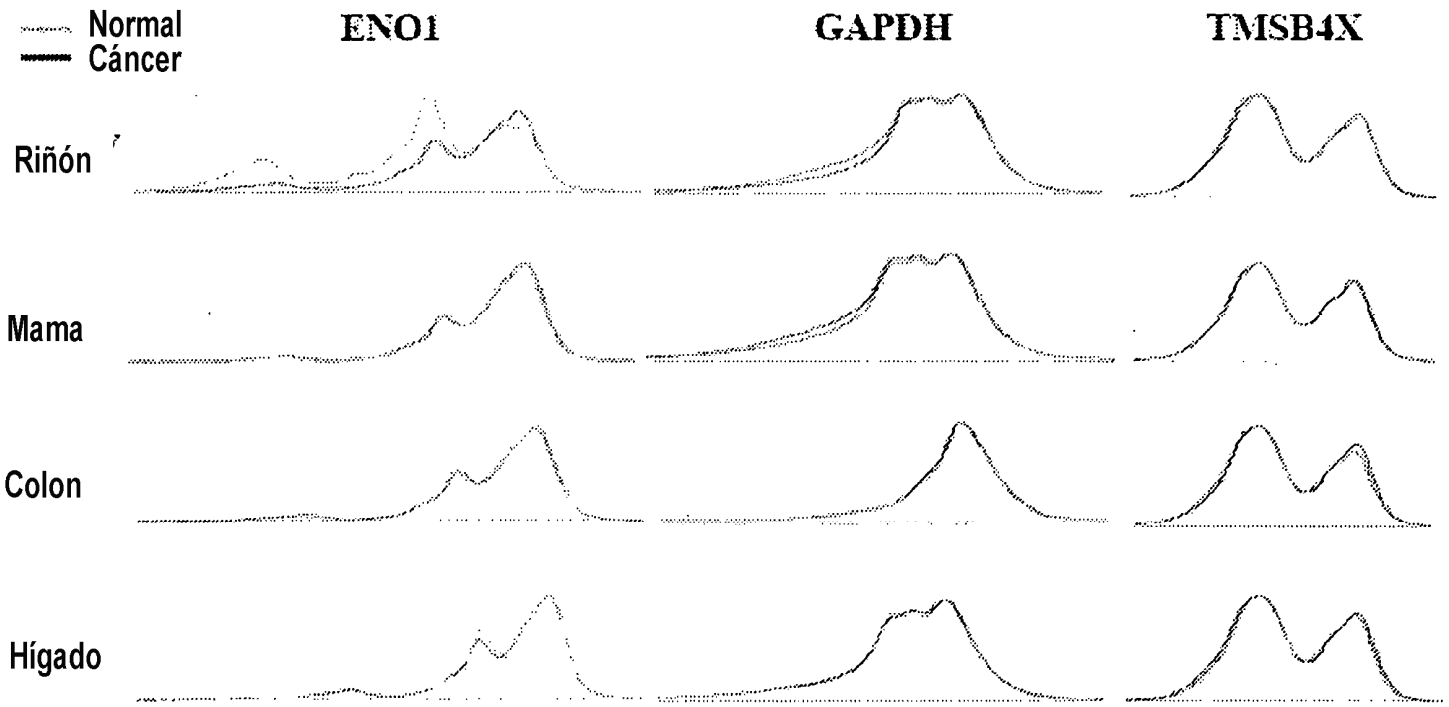
Péptido de post-fin

Producción de anticuerpos antipéptido contra:

APOA1 GARRRPPSRCSE
APOA2 IVFQPQLASRTPTGQS
APOC2 DPPSVDKGRVPYSPD
(perfil antigénico positivo)

NH_2 — Variante de infidelidad de la transcripción — COOH

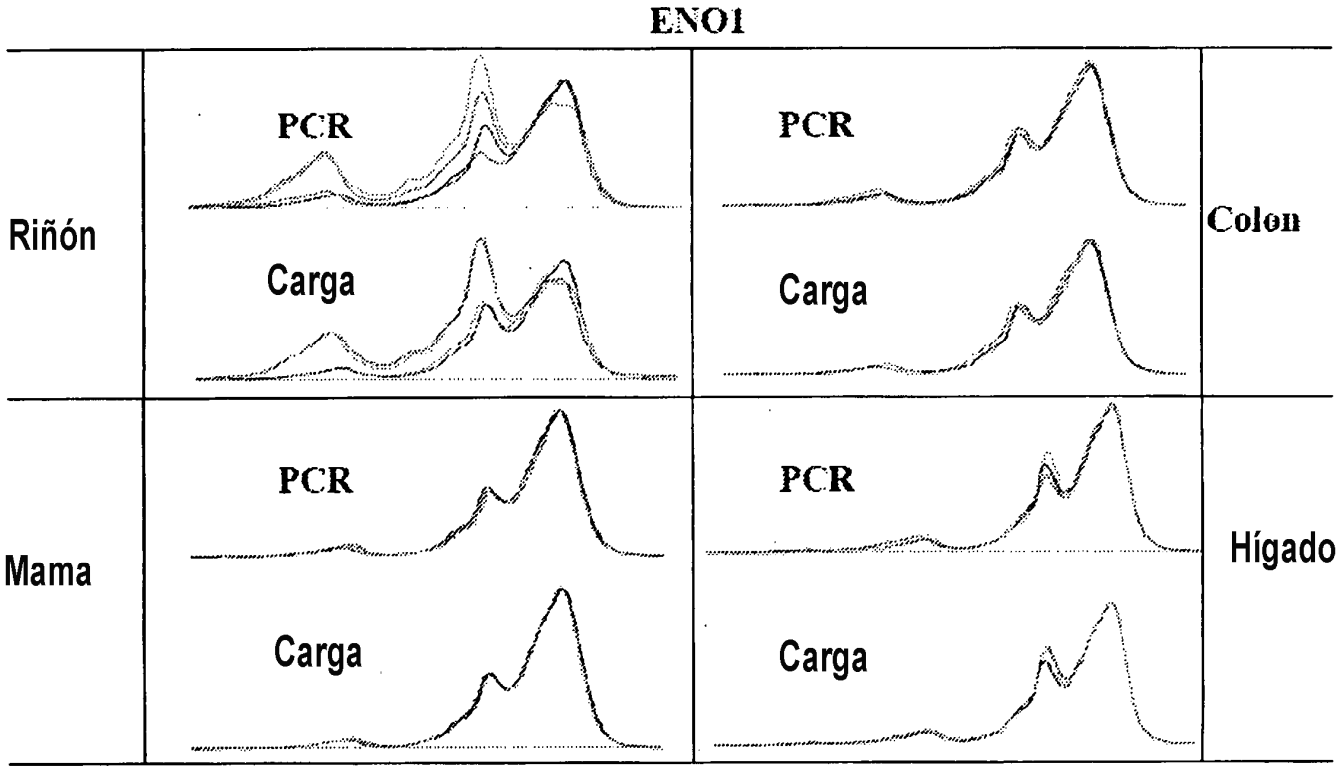
Figura 14



Perfiles de DHPLC de tejido tumoral frente a tejido adyacente normal para los genes ENO1, GAPDH y TMSB4X.

El ADN de pacientes cancerosos (Biochian Inc.) se amplificó mediante PCR utilizando oligonucleótidos complementarios con los genes ENO1, GAPDH y TMSB4X y la polimerasa *pfx*. El ADNc se obtuvo de tejidos de riñón, mama, colon e hígado. El perfil de DHPLC normal se representa en verde y el perfil canceroso se representa en negro. Las muestras de ADN bicatenario se inyectaron en un sistema DHPLC (Transgenomic). La temperatura de la estufa es de 62,5 ºC para el gen GAPDH, de 61,5 ºC para el gen ENO1, y de 55 ºC para el gen TMSB4X. Las curvas se visualizaron y se normalizaron con el programa informático Navigator (Transgenomic).

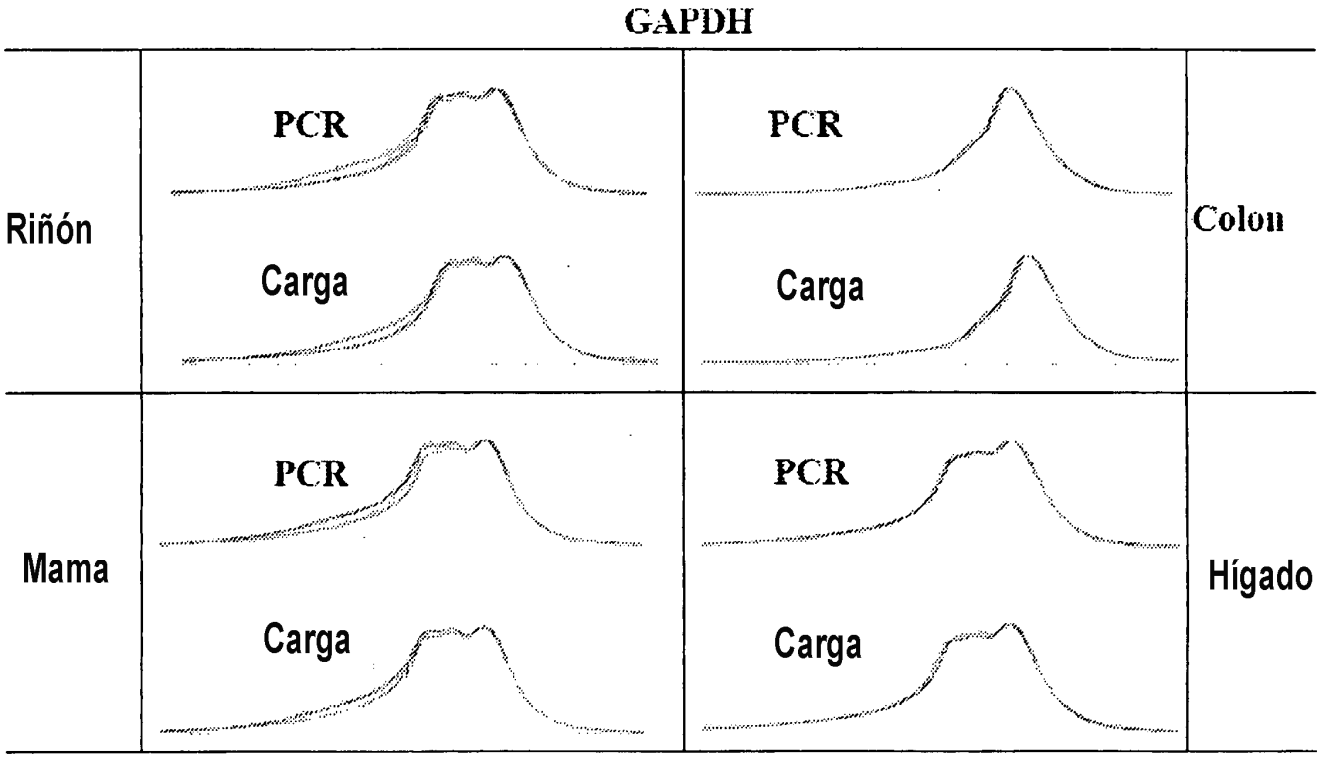
Figura 15a



Diferentes perfiles de PCR y de carga de DHPLC de tejido tumoral frente a tejido adyacente normal para el gen ENO1.

Se prepararon dos productos de PCR diferentes para los 4 tejidos y cada producto de PCR se inyectó dos veces en el sistema DHPLC. El perfil de DHPLC normal se representa en verde y el perfil canceroso se representa en negro. El perfil de DHPLC obtenido para el duplicado de PCR se representa en rojo para normal y azul para canceroso. Para el duplicado de carga, los perfiles se representan en morado para normal y azul claro para canceroso. Las curvas se visualizaron y se normalizaron con el programa informático Navigator (Transgenomic).

Figura 15b



Diferentes perfiles de PCR y de carga de DHPLC de tejido tumoral frente a tejido adyacente normal para el gen GAPDH.

Se prepararon dos productos de PCR diferentes para los 4 tejidos y cada producto de PCR se inyectó dos veces en el sistema DHPLC. El perfil de DHPLC normal se representa en verde y el perfil canceroso se representa en negro. El perfil de DHPLC obtenido para el duplicado de PCR se representa en rojo para normal y azul para canceroso. Para el duplicado de carga, los perfiles se representan en morado para normal y azul claro para canceroso. Las curvas se visualizaron y se normalizaron con el programa informático Navigator (Transgenomic).

Figura 15c

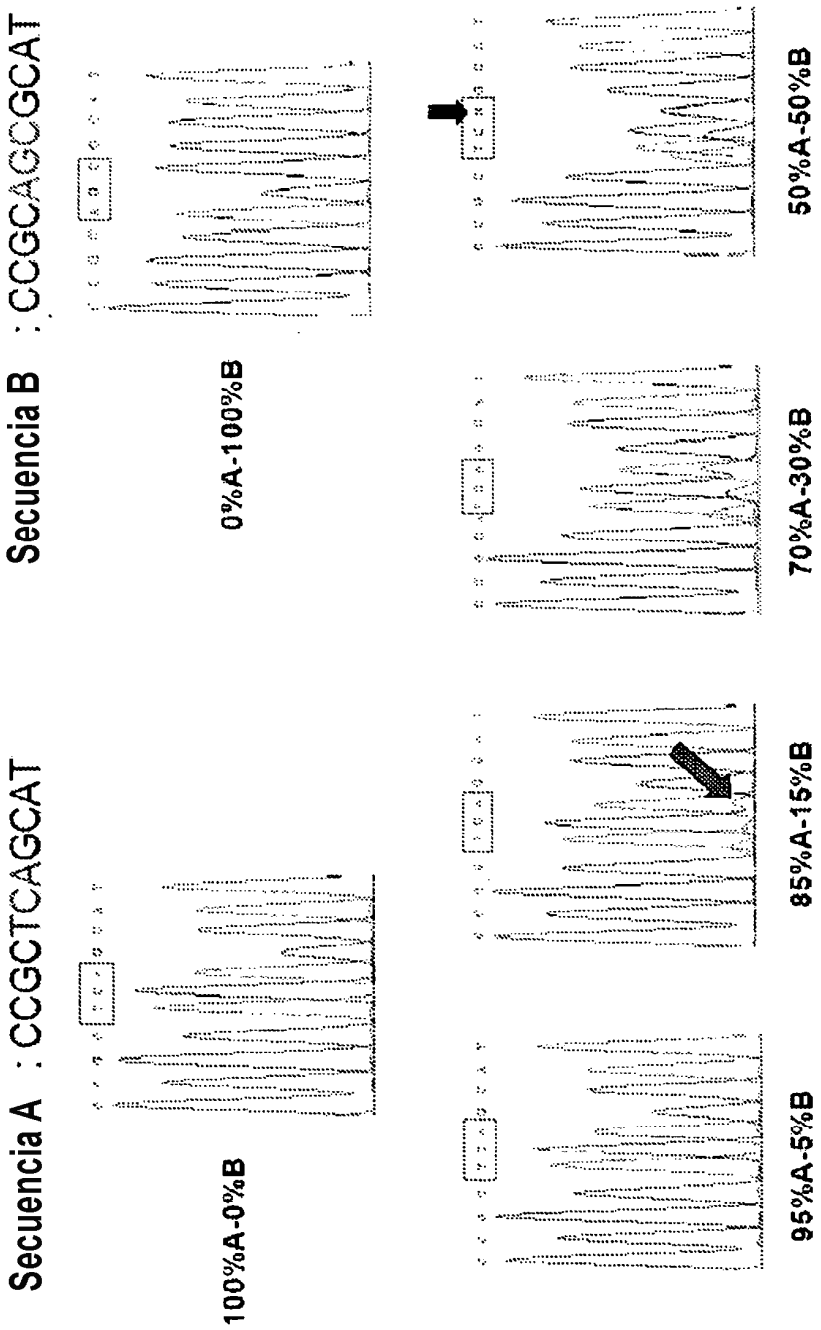

Secuencia A : CCGCTCAGCAT
Secuencia B : CCGCAGCGCAT
100%A-0%B
0%A-100%B
95%A-5%B
85%A-15%B
70%A-30%B
50%A-50%B

Figura 16

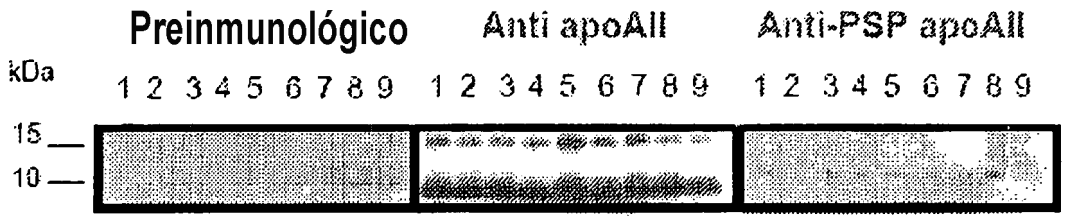


Figura 17a

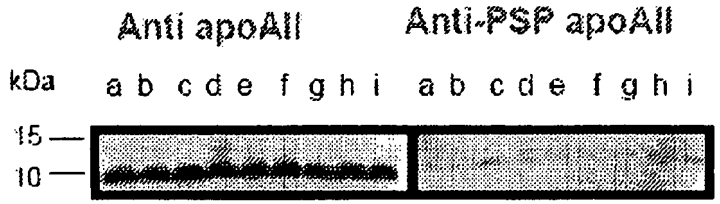


Figura 17b

Se aplicó plasma procedente de pacientes con cáncer de próstata (a) o con cáncer en la fase metastásica (b) a geles de SDS al 4-12%-poliacrilamida bajo condiciones reductoras. Después de una transferencia sobre membranas de PVDF y de un bloque en TBST que contenía leche en polvo desnatada al 5% (1 h, temperatura ambiente), se realizarón análisis de la transferencia Western utilizando suero preinmunológico de conejo (dilución 1/100, panel izquierdo), anticuerpo anti-ApoAII humana de cabra disponible en el mercado (1/250, panel central), y anticuerpo anti-PSP de ApoAII policlonal de conejo (1/100, panel derecho). Después de 1 h de incubación a temperatura ambiente, las membranas se lavaron y después se incubaron con el segundo anticuerpo conjugado apropiado. Entonces se revelaron las bandas mediantes quimioluminiscencia y exposición a películas de rayos X.

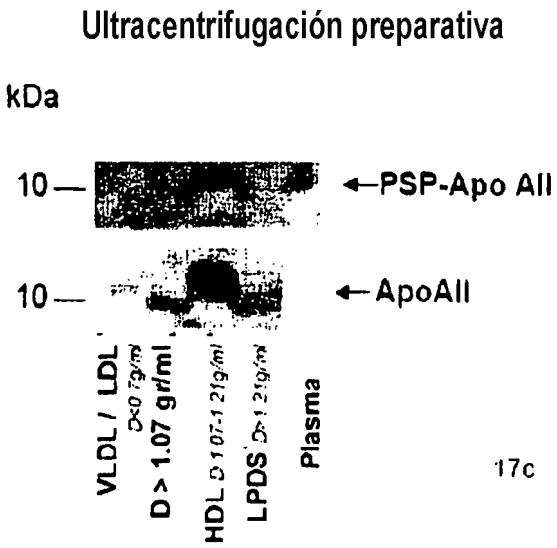


Figura 17c

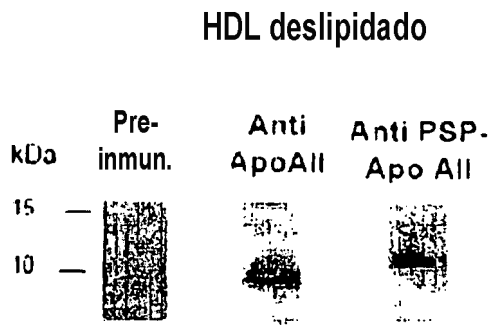


Figura 17d

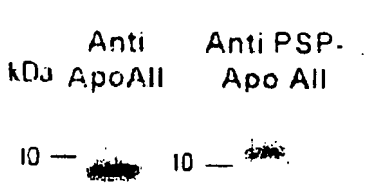


Figura 17e

c. Se preparó HDL (d 1,07-1,21 g/ml) mediante ultracentrifugación secuencial del plasma reunido de 10 pacientes con cáncer en la fase metastásica. Las diferentes fracciones recuperada después de cada etapa se dializaron contra NaCl 0,15 M, pH 7,4, que contenía EDTA al 0,01%. Se realizaron análisis de la transferencia Western en las fracciones indicadas con anticuerpo anti-ApoAII comercial (panel inferior) o anticuerpo anti-PSP de ApoAII (panel superior) utilizando las mismas diluciones que las indicadas en las figuras 17a a 17b. Los geles se sobrecargaron, lo cual explica la distorsión y las bandas tan grandes. También se realizó un análisis de la transferencia Western en el plasma reunido original o un volumen correspondiente de suero deficiente en lipoproteínas (d > 1,21 g/ml). d. El HDL purificado se dializó contra EDTA al 0,01%, pH 7,4, se congeló a -80 ºC y después se liofilizó. Entonces se realizó la deslipidación del HDL liofilizado utilizando cloroformo-metanol enfriado en hielo (2:1, v/v, 5 ml por 20 mg de proteína HDL). Después se realizó una segunda deslipidación utilizando éter dietílico, seguido de una centrifugación (1000 rpm, 5 min, 4 ºC) para sedimentar la proteína deslipidada. Esto se repitió y el sedimento final se secó en una atmósfera de N2 (gaseoso). El sedimento se redisolvió en Tris 10 mM que contenía urea 6 M, pH 8. Se realizó un análisis de la transferencia Western en 10 ug de HDL deslipidado utilizando suero preinmunológico, anti-ApoAII comercial o anti-PSP de ApoAII, tal como se describió anteriormente. e. Se realizaron experimentos de cromatografía de afinidad para aislar la forma de PSP de ApoAII utilizando el anticuerpo anti-PSP inmovilizado sobre esferas de matriz. La fracción eluida se analizó mediante transferencia Western utilizando los anticuerpos anti-ApoAII comercial y anti-PSP de ApoAII.

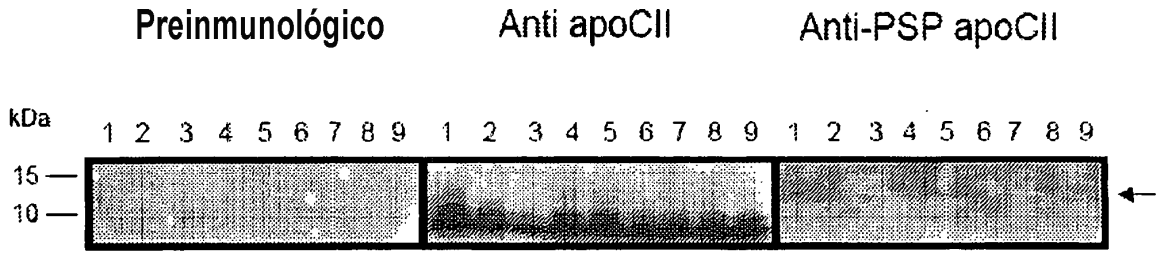


Figura 17f

Se realizó un experimento similar al de la figura a., con la excepción de que se utilizó un anticuerpo anti-PSP de ApoCII policlonal de conejo.

Nombre	id ARNm	PSP	Decisión
AHSG	NM_001622.1	ARHQRDEEVYWHRHSHHFVQAWAVV33LVCWFRKCHMRSTLSSLESKLPVIPHRTEAEWVVVMPDRAH	SÍ
ALB	BC033235	RSKAYESWFEFWCKANTLSKNKKEL	SÍ
APOB	NM_001638.2	GLDSTRALENEMTV	SÍ
APOA1	NM_000039.1	GARRRPPSRCSE	SÍ
APOA2	NM_001643.1	SVQTIVFQPQLASSTPTSGS	SÍ
APOC2	NM_000483.3	QPDPFSVDKGRVPYSPDFPSD	SÍ
APOC3	NM_000040.1	DLNTPEPPAYPSCELLSSCNLQGCPCELLVRDSILSALPHLMPGPFPGVLASQ	SÍ
APOD	NM_001647.2	PGSTGRLHPLHVTSASLSPTPPPPHQMPIHHDHGS	SÍ
APOE	NM_000041.2	TPKPAAMRPHATPCLSPRSLCRETLSPSQPSSWGGP	SÍ
AZGP1	X56786.1	EARVGGNVGSGTQ	SÍ
C4BL1	NM_001278.3	PSVLHTARGFRMPPPPLAPAGREPDHLPC	SÍ
CLU	NM_001831.2	DVDVAFAPTGASESSGPGQELQPPRESSARHQVTRPQFPGPGRPASPSGGSLTLTLDSAAHGKNSIAPACN	SÍ
HRG	NM_000412.2	NVPLARKMNMTLN	SÍ
IGFBP3	NM_000598.4	TPAARLMWSSMPYFAGKTAKDMTSSWLQPRSIFLFVVI	SÍ
IKHA	NM_002161.2	GWGVFLLNPWAGSHAPTIGNEEROSWEIDGSHSSLGLLCLWATLPTPLLSC	SÍ
KUK11	NM_144947.3	TGPTHKGPSVSTWCLVPVHSWNKP	SÍ
PKHG	NM_152479.3	WTPEPLLQFLSHPLFPAHPLSGCRL	SÍ
PLG	NM_000301.3	LDGRQSDALTHEAGTWVGI	SÍ
SERPINA3	NM_001085.4	SLPSSSGALSKELGMQAGCLGLWAGPGPCAPGGHGMQGPWCLSLEGPDSLCSSHMPRGPWTLGSGGSWAS	SÍ
TF	NM_001063.2	NLRGRAATKVKMGTGMHFEALVSLAGVVCANRVVCLHGSVLPCVLNKK	SÍ
TGFB1	NM_000660.3	SPAPPRPAPAGPAPPFPAPAALPMGAVPKGDTAAFSPPGAPLKMERGLSNGVSILSACLGSPSLTFPHSHSLSLPLCLLL PVCTIPLPGMAQGTSGEHYCS	SÍ
TTR	NM_000371.3	GTSPVDLKDEGWDFM	SÍ

Figura 18a

>P02765|AHSG Alfa-2-HS-glicoproteína
>P02768| Albumina
>P02743|APCS Componente P amiloide sérico
>P02647|APOA1 Apolipoproteína A-I
>P02652|APOA2 Apolipoproteína A-II
>P02655|APOC2 Apolipoproteína C-II
>P02656|APOC3 Apolipoproteína C-III
>P05090|APOD Apolipoproteína D
proteínas normales
otros péptidos de proteínas cuyo codón de fin se ha visto afectado

Figura 18b

>P02649|APOE Apolipoproteina E
>Q5XKQ4|AZGP1 Alfa-2-glicoproteina 1, de unión a cinc
>P36222|CHI3L1 Proteina 1 de tipo quitinasa 3
>P10909|CLU Clusterina
>P04196|HRG Glicoproteina rica en histidina

Figura 18c

>P17936|IGFBP3 Proteína 3 de unión al factor de crecimiento de tipo insulínico

MQRARPTLWAAALTLLVLLRGPPVARAGASSAGLGPVVRCEPCDARALAQCAPPPAVCAELVREPGCGCCLTCALSEGQPCGIYTERCGSGLR
CQPSPDEARPLQALLDGRGLCVNASAVSRLRAYLLPAPPAPGNASESEEDRSAGSVESPSVSSTHRVSDPKFHPLHSKIIIIKKGHAKDSQRYKVD
YESQSTDTQNFSSESKRETEYGPCRREMEDTLNHLKFLNVLSPRGVHIPNCDKKGFYKKKQCRPSKGRKRGFCWCVDKYGQPLPGYTTKGKED
VHCYSMQSK**>P05111|INHA** Cadena alfa de inhibina

MVLHLLLFLLLTPQGGHSCQGLELARELVLAKVRALFLDALGPPAVTREGGDPGVRRLPRRHALGGFTHRGSEPEEEEDVSQAILFPATDASCED
KSAARGLAQEAEEGLFRYMFRPSQHTRSRQVTSAQLWFHTGLDRQGTAASNSSEPLLGLLALSPGGPVAVPMSLGHAPPHWAVLHLATSALSLL
THPVLVLLLRCPLCTCSARPEATPFLVAHTRTRPPSGGERARRSTPLMSWPWSPSALRLLQRPPEEPAAHANCHRVALNISFQELGWERWIVYP
PSFIFHYCHGGCGLHIPPNLSLPVPGAPPTPAQPYSLLPGAQPCCAALPGTMRPLHVRTTSDGGYSFKYETVPNLLTQHCACIWGYGVFLLMPMA
GGHAPTISWEERQSWEDQSHSSLLSLEQLWATEPTPLESQ**>Q9UBX7|KLK11** Calicreína-11

MQRLRWLRDWKSSGRGLTAAKEPGARSSPLQAMRILQLILLALATGLVGGETRIIKGFECKPHSQPWQAALFEKTRLLCGATLIAPRWLLTAAHCL
KPRYIVHLGQHNLQKEEGCEQTRTATESFPHPGFNNSLPNKDHRNDIMLVKMASPVSITWAVRPLTLSSRCVTAGTSCLISGWGSTSSPQLRLPH
TLRCANITIIEHQKCENAYPGNITDTMVCASVQEGGKDSCQGDSGGPLVCNQSLQGIISWGQDPCAITRKPGVYTKVCKYVDWIQETMKNN**>P14618|PKM2** Isoenzimas de piruvato quinasa **M1/M2**

MSKPHSEAGTAFIQTQQLHAAMADTFLEHMCRLDIDSPPITARNTGIICTIGPASRSVETLKEMIKSGMNVARLNFSHGTHEYHAETIKNVRTATES
FASDPILYRPVAVALDTKGPEIRTGLIKGSGTAEVELKKGATLKITLDNAYMEKCDENILWLDYKNICKVVEVGSKIYVDDGLISLQVKQKGADFLVTE
VENGGSLGSKKGVNLPGAAVDLPAVSEKDIQDLKFGVEQDVDMVFASFIRKASDVHEVRKVLGEKGKNIKIISKIENHEGVRRFDEILEASDGIMVA
RGDLGIEIPAEKVFLAQKMMIGRCNRAGKPVICATQMLESMIKKPRPTRAEGSDVANAVLDGADCIMLSGETAKGDYPLEAVRMQHLIAREAEAA
MFHRKLFEELVRASSHSTDLMEAMAMGSVEASYKCLAAALIVLTESGRSAHQVARYRPRAPIIAVTRNPQTARQAHLYRGIFPVLCKDPVQEAWA
EDVDLRVNFAMNVGKARGFFKKGDVVIVLTGWRPGSGFTNTMRVVPVP**>P00747|PLG** Plasminógeno

MEHKEVVLLLLLFLKSGQGEPLDDYVNTQGASLFSVTKKQLGAGSIEECAAKCEEDEEFTCRAFQYHSKEQQCVIMAENRKSSIIIRMRDVVLFEK
KVYLSECKTGNGKNYRGTMSKTKNGITCQKWSSTSPHRPRFSPATHPSEGLEENYCRNPDNDPQGPWCYTTDPEKRYDYCDILECEEECMHC
SGENYDGKISKTMSGLECQAWDSQSPHAHGYIPSKFPNKNLKKNYCRNPDRELRPWCFTTDPNKRWELCDIPRCTTPPPSSGPTYQCLKGTGE
NYRGNVAVTVSGHTCQHWSAQTPHTHNRTPENFPCKNLDENYCRNPDGKRAPWCHTTNSQVRWEYCKIPSCDSSPVSTEQLAPTAPPELTPV
VQDCYHGDGQSYRGTSSTTTTGKKCQSWSSMTPHRHQKTPENYPNAGLTMNYCRNPDADKGPWCFTTDPSVRWEYCNLKKCSGTEASVVAP
PPVVLLPDVETPSEEDCMFGNGKGYRGKRATTVTGTPCQDWAAQEPHRHSIFTPETNPRAGLEKNYCRNPDGDVGGPWCYTTNPRKLYDYCD
VPQCAAPSFDCGKPQVEPKKCPGRVVGGCVAHPHSWPWQVSLRTRFGMHFCGGTLISPEWVLTAAHCLEKSPRPSSYKVILGAHQEVNLEPHV
QEIEVSRLFLEPTRKDIALLKLSSPAVITDKVIPACLPSPNYVVADRTECFITGWGETQGTFGAGLLKEAQLPVIENKVCNRYEFLNGRVQSTELCAG
HLAGGTDSCQGDSGGPLVCFEKDKYILQGVTSWGLGCARPNKPGVYVRVSRFVTWIEGVMRNN**Figura 18d**

>P01011|SERPINA3 Alfa-1-antiquimotripsina

MERMLPLLALGLLAAGFCPAVLCHPNSPLDEENLTQENQDRGTHVDLGLASANVDFAFSLYKQLVLKAPDKNVIFSPLSISTALAFLSLGAHNTTLT
EILKGLKFNLTETSEAEIHQSFQHLLRTLNQSSDELQLSMGNAMFVKEQLSLLDRFTEDAKRLYGSEAFATDFQDSAAAKKLINDYVKNGTRGKITD
LIKDLDSQTMMVLVNYIFFKAKWEMPFDPQDTHQSRFYLSKKKWVMVPMMSLHHLTIPYFRDEELSCTVVELKYTGNASALFILPDQDKMEEVEA
MLLPETLKRWRDSLEFREIGELYLPKFSISRDYNLNDILLQLGIEEAFTSKADLSGITGARNLAVSQVVHKAVLDVFEEGTEASAATAVKITLLSALVE
TRTIVRFNRPFLMIIVPTDTQNIFFMSKVTNPKQA*SLPSSSGALSKELGMQAGCLGLWAQPGPCAPSGHGMCGPVCLSLEGDSDSLCSSHMHR
GPWTLQSGGGSVAS*

>P02787|TF Serotransferrina

MRLAVGALLVCAVLGLCLAVPDKTVRWCAVSEHEATKCQSFRDHMKSVIPSDGPSVACVKKASYLDCIRAIAANEADAVTLDAGLVYDAYLAPNN
LKPVVAEFYGSKEDPQTFYYAVAVVKKDSGFQMNQLRGKKSCHTGLGRSAGWNIPIGLLYCDLPEPRKPLEKAVANFFSGSCAPCADGTDFPQL
CQLCPGCGCSTLNQYFGYSGAFKCLKDGAGDVAFVKHSTIFENLANKADRDQYELLCLDNTRKPVDEYKDCHLAQVPSHTVVARSMGGKEDLIW
ELLNQAQEHFGKDKSKEFQLFSSPHGKDLLFKDSAHGFLKVPPRMDAKMYLGYEYVTAIRNLREGTCPEAPTDECKPVKWCALSHHERLKCDE
WSVNSVGKIECVSAETTEDCIAKIMNGEADAMSLDGGFVYIAGKCGLVPVLAENYNKSDNCEDTPEAGYFAVAVVKKSASDLTWDNLKGKKSCH
TAVGRTAGWNIPMGLLYNKINHCRFDEFFSEGCAPGSKKDSSLCKLCMGSGLNLCEPNNKEGYYGYTGAFRCLVEKGDVAFVKHQTVPQNTGG
KNPDPWAKNLNEKDYELLCLDGTRKPVEEYANCHLARAPNHAVVTRKDKEACVHKILRQQQHLFGSNVTDCSGNFCLFRSETKDLLFRDDTVCL
AKLHDRNTYEKYLGEEYVKAVGNLRKCSTSSLLEACTFRRP*NLRGRAATKVKMGTQMHREFALVSLAQVVCANPVCLHSSVLPDVLNKK*

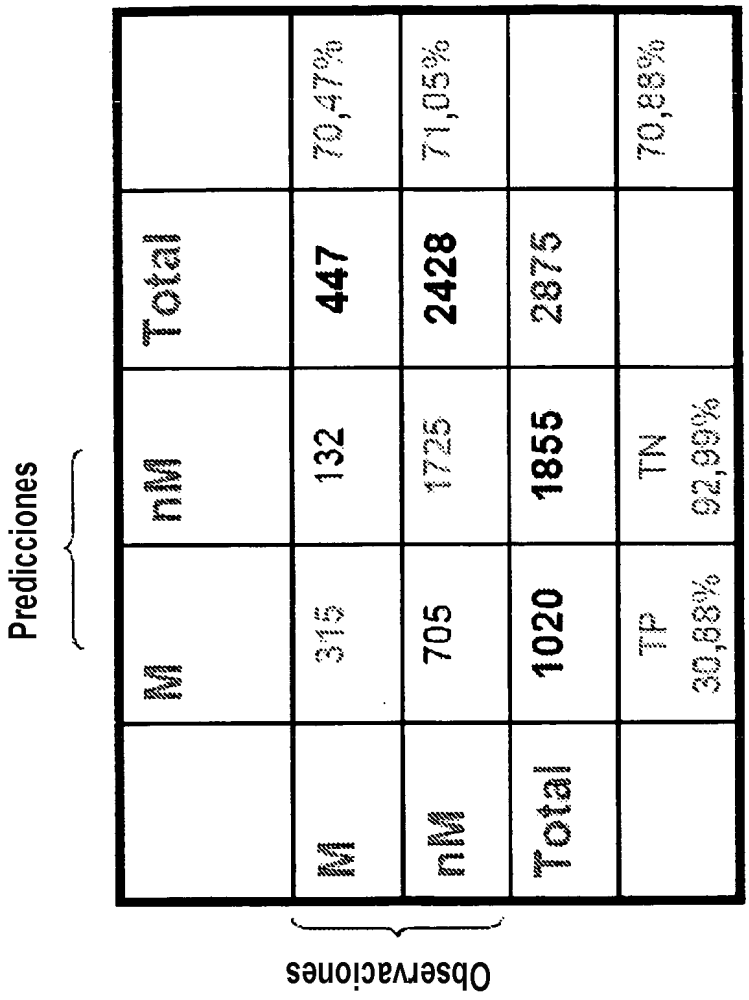
>P01137|TGFB1 Factor del crecimiento transformante beta-1

MPPSGLRLLPLLLPLLWLLVLTPGRPAAGLSTCKTIDMELVKRKRIEAIRGQILSKLRLASPPSQGEVPPGPLPEAVLALYNSTRDRVAGESAEPEP
EPEADYYAKEVTRVLMVETHNEIYDKFKQSTHSIYMFFNTSELREAVPEPVLLSRAELRLLRLKLKVEQHVELYQKYSNNSWRYLSNRLLAPSDSP
EWLSFDVTGVVRQWLSRGGEIEGFRLSAHCSCDSRDNTLQVDINGFTTGRRGDLATIHGMNRPFLLLMATPLERAQHLQSSRHRRALDTNYCFS
STEKNCCVRQLYIDFRKDLGWKWIHEPKGYHANFCLGPCPYIWSLDTQYSKVLALYNQHNPGASAAPCCVPQALEPLPIVYYVGRKPKVEQLSN
MIVRSCKCS*GPAPPRPADAGPAPPRPAPAALPMGAVPKDTRAPSPPGAPLKMERGLRISVSLGAQLGSPSLTFPHSHSLSLPLCLLLPVCTIPLP
GIKAQGTSGEHYCS*

>P02766|TTR Transtiretina

MASHRLLLLCLAGLVFVSEAGPTGTGESKCPLMVKVLDAVRGSPAINVAVHVFRKAADDTWEPFASGKTSESGELHGLTTEEEFVEGIYKVEIDTK
SYWKALGISPFHEHAEVVFTANDSGPRRYTIAALLSPYSYSTTAVVTNPKE*GTSPPVDLKDEGWDFM*

Figura 18e



	M	nM	Total	
M	315	132	**447**	70,47%
nM	705	1725	**2428**	71,05%
Total	**1020**	**1855**	2875	
	TP 30,88%	TN 92,99%		70,88%

TP **:Positivos verdaderos;** ***TN*** **: Negativos verdaderos**

Figura 19