

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 644 057**

51 Int. Cl.:

**C12Q 1/68** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **08.12.2010 PCT/IB2010/003397**

87 Fecha y número de publicación internacional: **16.06.2011 WO11070441**

96 Fecha de presentación y número de la solicitud europea: **08.12.2010 E 10816306 (4)**

97 Fecha y número de publicación de la concesión europea: **26.07.2017 EP 2510124**

54 Título: **Categorización de muestras de ADN**

30 Prioridad:

**19.02.2010 US 306201 P**  
**20.04.2010 US 325977 P**  
**11.12.2009 US 285758 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**27.11.2017**

73 Titular/es:

**NUCLEIX (100.0%)**  
**3 Pekeris Street**  
**7670203 Rehovot , IL**

72 Inventor/es:

**FRUMKIN, DAN y**  
**WASSERSTROM, ADAM**

74 Agente/Representante:

**CAMPELLO ESTEBARANZ, Reyes**

**ES 2 644 057 T3**

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

## DESCRIPCIÓN

Categorización de muestras de ADN

## 5 CAMPO DE LA INVENCION

La presente tecnología inventiva se refiere a métodos para la categorización precisa y rentable y el análisis de muestras de ADN para identificar la fuente de origen y la identidad de la muestra de ADN o muestras de ADN mixto.

## 10 RESUMEN DE LA INVENCION

La presente invención se refiere a métodos para la categorización precisa y rentable de muestras de ADN en diferentes tipos de ADN generado *in vitro* o diferentes tipos de ADN natural tales como de diferentes tejidos y/o estados fisiológicos/patológicos. La invención logra la categorización comparando "relaciones de señal" que se correlacionan con relaciones de niveles de metilación en loci genómicos específicos, pero no se basa en el cálculo de los niveles reales de metilación en ningún locus genómico. Por lo tanto, la configuración de la invención elimina el requisito de especies y/o controles externos de ADN, lo que simplifica y aumenta de este modo la precisión del ensayo. La tecnología de la invención también permite realizar la categorización de ADN junto con el perfil de ADN en la misma reacción, permitiendo así la categorización concomitante y la determinación de la identidad de las muestras.

Un aspecto de la presente invención es un método para categorizar una muestra de ADN de acuerdo con la reivindicación 1 El método comprende:

- 25 (A) digerir una muestra de ADN con una endonucleasa de restricción sensible a la metilación y/o dependiente de la metilación;  
 (B) realizar un análisis de PCR en el ADN digerido amplificando al menos dos loci genómicos de los cuales al menos uno es un locus de restricción;  
 (C) determinar la intensidad de la señal de cada producto de amplificación;  
 30 (D) calcular "relaciones de señal" entre las intensidades de las señales producidas por los loci; y  
 (E) comparar las relaciones de señal con los valores de referencia correspondientes a diferentes categorías de ADN, en el que se determina que la categoría cuyos valores de referencia corresponden mejor a las relaciones de señal es la categoría de la muestra de ADN. Es decir, por ejemplo, la aproximación del valor de una relación de señal al valor de un valor de referencia de una categoría de ADN particular indica la fuente categórica de la muestra de ADN. Entonces, si una relación de señal que tiene un valor cercano al valor de una categoría de referencia, entonces la probabilidad de que la muestra de ADN probada se origine a partir de ese tipo particular de categoría, en base a la similitud en los valores entre la relación de señal y el valor de referencia, es alta. Por lo tanto, la relación o relaciones de señal que mejor se corresponden, o mejor se aproximan, a un valor o valores de referencia particular para una categoría de ADN particular hace que sea probable que esa muestra de ADN se origine a partir de esa fuente categórica de ADN particular. Para ejemplos sobre cómo comparar las relaciones de señal y los valores de referencia de categoría, consultar la divulgación en otra parte de esta solicitud, tal como en la sección titulada Algoritmo y Software, que explica cómo calcular las distribuciones de probabilidad y las probabilidades de verosimilitud de que una muestra de ADN se categorice en una o más categorías de referencia.

45 En una realización, la digestión del ADN y el análisis de PCR se realizan en una única reacción bioquímica en la que la plantilla de ADN, las enzimas de digestión y amplificación, tampones, cebadores e ingredientes accesorios se colocan juntos en un tubo de ensayo y después el tubo de ensayo se cierra y se coloca en un termociclador, donde tiene lugar la reacción.

50 En una realización, la endonucleasa de restricción sensible a metilación es incapaz de cortar o digerir un ADN si su secuencia de reconocimiento está metilada. En una realización, una endonucleasa de restricción sensible a la metilación se selecciona del grupo que consiste en AatII, Acc65I, Accl, Acil, ACII, Afel, Agel, Apal, ApaLI, Ascl, AsiSI, Aval, Avall, Bael, BanI, Bbel, BceAI, Bcgl, BfuCI, BgII, BmgBI, BsaAI, BsaBI, BsaHI, Bsal, BseYI, BsiEI, BsiWI, BsII, BsmAI, BsmBI, BsmFI, BspDI, BsrBI, BsrFI, BssHII, BssKI, BstAPI, BstBI, BstUI, BstZ17I, Cac8I, Clal, Dpnl, Drdl, Eael, Eagl, Eagl-HF, EciI, EcoRI, EcoRI-HF, Faul, Fnu4HI, Fsel, Fspl, HaelI, Hgal, Hhal, HincII, HincIII, HinfI, HinPII, Hpal, Hpall, Hpy166ii, Hpy188iii, Hpy99I, HpyCH4IV, KasI, MluI, Mmel, MspAII, MwoI, Nael, NarI, NgoNIV, Nhe-HFI, NheI, NlaIV, NotI, NotI-HF, Nrul, Nt.BbvCI, Nt.BsmAI, Nt.CviPII, PaeR7I, P1el, Pmel, PmII, PshAI, PspOMI, Pvul, Rsal, RsrII, SacII, Sall, Sall-HF, Sau3AI, Sau96I, ScrFI, Sfil, SfoI, SgrAI, Smal, SnaBI, Tfil, TscI, Tsel,

TspMI, y Zral.

En otra realización, puede usarse una enzima de restricción dependiente de metilación para digerir una muestra de ADN. Una endonucleasa de restricción dependiente de la metilación digiere solamente el ADN metilado. Los 5 ejemplos de dichas enzimas de restricción incluyen, pero sin limitación, McrBC, McrA y MrrA.

En una realización, se sabe que al menos un locus de restricción está metilado en ADN natural.

En otra realización, se sabe que al menos un locus de restricción está sin metilar en ADN natural.

10

En otra realización, se sabe que al menos dos loci de restricción están metilados de forma diferente en ADN natural.

En otra realización, se sabe que la relación de señal entre al menos dos loci es diferente para al menos dos categorías potenciales.

15

En otra realización, los pares de cebadores usados para amplificar loci se seleccionan para amplificar los loci de repetición corta en tándem (STR) del núcleo identificados en el Sistema de Índice Combinado de ADN de la Oficina Federal de Investigación (CODIS).

20 En una realización, los loci del CODIS son loci genómicos humanos seleccionados del grupo que consiste en D16S539 (SEQ ID NO. 1), D7S820 (SEQ ID NO. 2), D13S317 (SEQ ID NO. 3), D5S818 (SEQ ID NO. 4), CSF1PO (SEQ ID NO. 5), TPOX (SEQ ID NO. 6), TH01 (SEQ ID NO. 7), vWA (SEQ ID NO. 8), FGA (SEQ ID NO. 9), D21S11 (SEQ ID NO. 10), D8S1179 (SEQ ID NO. 11), D18S51 (SEQ ID NO. 12), y D3S1358 (SEQ ID NO. 13).

25 En una realización particular, los cebadores directo e inverso de un locus de restricción y/o perfil están diseñados para hibridar con la SEQ ID NO. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, o 25.

En una realización, los cebadores directo e inverso de un locus están disponibles en un kit comercial de Promega Corporation seleccionado del grupo que consiste en PowerPlex® 16 HS (Cat. N.º DC2100, DC2101), PowerPlex® 16 30 (Cat. N.º DC6530, DC6531), PowerPlex® 2.1 (Cat. N.º DC6470, DC6471), PowerPlex® 16 BIO (Cat. N.º DC6540, DC6541), y PowerPlex® ES Systems (Cat. N.º DC6730, DC6731).

En una realización, los cebadores directo e inverso de un locus están disponibles en un kit comercial de Applied Biosystems/Life seleccionado del grupo que consiste en SGM, SGM+, AmpFISTR Identifier, AmpFISTR Profiler, 35 AmpFISTR ProfilerPlus, AmpFISTR ProfilerPlusID, AmpFISTR SEfiler, AmpFISTR SEfiler Plus, AmpFISTR Cofiler, AmpFISTR Identifier Direct, AmpFISTR Identifier Plus, AmpFISTR NGM, AmpFISTR Y-filer, AmpFISTR Minifiler.

En una realización, los cebadores directo o inverso de un locus están disponibles en un kit comercial Qiagen 40 seleccionado del grupo que consiste en Investigator ESSPlex, Investigator ESSPlex SE, Investigator Nonaplex ESSPlex, Investigator Hexaplex ESSPlex, Investigator Triplex AFS QS, Investigator Triplex DSF, Investigator IDplex, Investigator Decaplex SE, Investigator HDplex, Investigator Argus X-12, Investigator Y-12 QS, Investigator DIPplex.

En otra realización, los loci de restricción contienen una secuencia de reconocimiento de HpaII y se seleccionan del grupo que consiste en ADD6, ADD10, ADD 17 (SEQ ID NOs. 26-28) e Hypo23, Hypo28, Hypo33 (SEQ ID NOs. 29- 45 31). En una realización, la muestra de ADN es un ADN de mamífero o un ADN vegetal. En una realización, el ADN de mamífero es ADN de un mamífero seleccionado del grupo que consiste en un ser humano, simio, mono, rata, ratón, conejo, vaca, cerdo, oveja y caballo. En una realización particular, el ADN de mamífero es ADN humano. En una realización, el ADN humano es de un macho. En otra realización, el ADN humano es de una hembra.

50 En una realización, la muestra de ADN humano se obtiene de un macho o una hembra que tiene un trastorno caracterizado por la metilación de ADN anormal. En una realización, el trastorno es un trastorno del neurodesarrollo de ICF (inmunodeficiencia, inestabilidad centromérica y anomalías faciales), síndrome de Rett y síndrome de X frágil.

En una realización, el ADN vegetal es ADN de una planta. En una realización, la planta es una planta 55 monocotiledónea seleccionada del grupo que consiste en trigo, hierba de césped, cereal, maíz, arroz, avena, cebada, sorgo, orquídea, iris, lirio, cebolla, plátano, caña de azúcar, sorgo y palma. En otra realización, la planta es una planta dicotiledónea seleccionada del grupo que consiste en aguacate, patata, tabaco, tomate, remolacha azucarera, brócoli, mandioca, boniato, pimienta, algodón, poinsetia, leguminosas, alfalfa, soja, zanahoria, fresa, lechuga, roble, arce, nuez, rosa, menta, calabaza, margarita y cactus.

En una realización del método, al menos dos loci se amplifican por PCR con cebadores marcados de forma fluorescente y en el que se realizan comparaciones por parejas de la intensidad de las señales fluorescentes correspondientes a los loci para calcular relaciones de señal para cada par de loci.

5 En una realización, se usan pares de cebadores para amplificar pares de loci seleccionados del grupo que de pares que consisten en D3S1358/D18S51, D3S1358/D7S820, D3S1358/Penta\_D, D3S1358/TPOX, D3S1358/FGA, TH01/Penta\_D, D21S11/D18S51, D21S11/D7S820, D21S11/Penta\_D, D21S11/AMEL, D21S11/TPOX, D21S11/FGA, D5S818/D18S51, vWA/D18S51, D5S818/Penta\_E, vWA/Penta\_E, D5S818/D7S820,  
 10 D5S818/Penta\_D, D5S818/TPOX, D5S818/FGA, D13S317/D7S820, D13S317/Penta\_D, D13S317/TPOX, D13S317/FGA, D16S539/D7S820, CSF1PO/D7S820, vWA/D7S820, D8S1179/D7S820, D16S539/TPOX, D16S539/FGA, CSF1PO/Penta\_D, CSF1PO/TPOX, vWA/Penta\_D, AMEL/TPOX, AMEL/FGA, vWA/D8S1179, vWA/TPOX, vWA/FGA, D8S1179/TPOX, D8S1179/FGA.

15 En otra realización, los amplicones de los loci de restricción son más pequeños que los amplicones más pequeños usados en el perfil de ADN, que es de aproximadamente 100 pb de tamaño.

En un aspecto de la presente tecnología inventiva, se realiza la cuantificación de señales de amplicón y no necesariamente la cuantificación de la cantidad de plantillas de ADN. Al calcular relaciones de señales de amplicón, no se necesita una curva estándar o un ADN de referencia, ya que no es necesario calcular las concentraciones de ADN reales. En esta realización, se calculan las relaciones de señales de amplicón. Las señales pueden ser señales detectadas de cualquier tipo de etiquetado usado para detectar productos de amplificación de PCR, tales como, pero sin limitación, tinción con bromuro de etidio, tinción sybr green, tinción de plata o por fluorescencia.

20

25 Con respecto a esto último, en una realización, la reacción de PCR se realiza con cebadores marcados de forma fluorescente, y la determinación de la intensidad de las señales de los productos de amplificación se logra mediante la separación de productos por electroforesis capilar y cuantificación de señales de fluorescencia.

En otra realización, la PCR es PCR en tiempo real, y la determinación de la intensidad de las señales de los productos de amplificación se logra mediante la cuantificación de las señales de fluorescencia.

30

En una realización particular, la endonucleasa de restricción sensible a la metilación es HhaI.

En una realización particular, la endonucleasa de restricción sensible a la metilación es HhaI y el primer locus de restricción es el locus TPOX.

35

En una realización particular, la endonucleasa de restricción sensible a la metilación es HhaI y el primer locus de restricción es el locus FGA.

40 En una realización particular en la que la categorización es según el semen o no según el semen, la endonucleasa de restricción sensible a la metilación es HhaI, los loci de restricción son SD1, SD2, SD3 y SD4 (SEQ ID NOS. 33-36), un locus de control digerido es SD5 (SEQ ID NO. 37), y un locus de control no digerido es SD6 (SEQ ID NO. 38).

45 En una realización particular en la que la categorización es según la sangre, la saliva, el semen o la epidermis de la piel, la endonucleasa de restricción sensible a la metilación es HhaI y los loci de restricción son las SEQ ID NOS. 39-53.

En una realización particular en la que la categorización se realiza junto con el perfil de ADN, y es según el semen o no según el semen, la endonucleasa de restricción sensible a la metilación es HhaI, y los loci de restricción son SEQ ID NO. 40 y SEQ ID NO. 54.

50

En una realización, se selecciona un par de loci de loci con niveles de metilación conocidos en al menos dos categorías de ADN, de tal forma que se sabe que uno de los loci en el par tiene un nivel de metilación más alto que el segundo locus en ese par en una categoría de ADN pero no en la otra.

55

En otra realización, se selecciona un par de loci sin conocimiento de los niveles de metilación como se indica a continuación: seleccionar un par de loci genómicos aleatorios; diseñar cebadores marcados de forma fluorescente para amplificar cada locus; amplificar ambos loci en muestras de ADN obtenidas de diferentes categorías;

amplificones separados con electroforesis capilar; calcular las relaciones de señal en las diferentes muestras de ADN; si las relaciones calculadas a partir de muestras correspondientes al menos a una categoría son significativamente diferentes (por ejemplo, según se determina por la prueba de Kolmogorov-Smirnov con un umbral de  $p = 0,05$ ) que las obtenidas de muestras correspondientes al menos a otra categoría, seleccionar el par para su uso en el ensayo.

- 5 En una realización, las relaciones calculadas se comparan con un conjunto de relaciones obtenidas de muestras naturales y con un conjunto de relaciones obtenidas a partir de muestras artificiales.

Con respecto a esto último, en una realización, se muestran a continuación las respectivas relaciones medias de señal de muestras de ADN naturales y artificiales digeridas con HhaI para pares específicos de loci determinados en el presente documento.

	ADN natural	ADN artificial
FGA(PowerPlex16)/Hypo12	10,3	0,89
FGA(PowerPlex16)/Hypo28	12,4	0,76
TPOX(PowerPlex16)/Hypo12	9,7	1,12

En otra realización, las relaciones calculadas se comparan con un conjunto de relaciones obtenidas a partir de muestras de ADN extraídas de diferentes categorías de tejidos. Véanse las Tablas 1-4 presentadas en las Figuras 7-10 para ejemplos adicionales de secuencias de loci y relaciones de señal ilustrativas y resultados usados y obtenidos de acuerdo con los métodos de categorización de la invención descritos en el presente documento.

En una realización, este método, o cualquier método descrito en el presente documento, comprende además analizar simultáneamente una muestra de ADN y determinar si la categoría del ADN. En una realización, la etapa de perfilar una muestra de ADN se realiza de acuerdo con las instrucciones proporcionadas en un kit de análisis de ADN disponible comercialmente. En otra realización, el kit de análisis de ADN disponible comercialmente es el kit de análisis PowerPlex®16 (Promega Corporation).

En una realización, con respecto a esto último, la categorización del ADN es según los tipos de tejido. Se puede usar cualquier tipo de tejido tal como, pero sin limitación, sangre, saliva, semen, epidermis, orina, plasma y cabello.

25 Por consiguiente, el método para categorizar una muestra de ADN, que comprende:

- (A) digerir una muestra de ADN con al menos una de una endonucleasa de restricción sensible a la metilación y una endonucleasa de restricción dependiente de la metilación;
- (B) amplificar al menos dos loci genómicos del ADN digerido, en el que al menos uno de los loci es un locus de restricción;
- (C) determinar la intensidad de señal de cada producto de amplificación;
- (D) calcular relaciones de señal entre las intensidades de señal de los productos de amplificación; y
- (E) comparar las relaciones de señal con valores de referencia de diferentes categorías de ADN,

35 en el que la aproximación del valor de una relación de señal a los valores de las relaciones de referencia de una categoría de ADN particular indica la fuente categórica de la muestra de ADN.

En una realización, los valores de referencia con los que se comparan las relaciones de señal son de al menos una categoría de ADN seleccionada del grupo que consiste en tipo de tejido, tipo de célula, condición fisiológica, condición patológica, edad, etnia, sexo, nivel de metilación, especie, líneas celulares, ADN natural y ADN sintetizado artificialmente, riesgo de desarrollar una condición patológica, condiciones prenatales, pronóstico, propensión a responder al tratamiento, efectos del subcultivo de la línea celular, respuesta a la medicación.

45 En otra realización, el tipo de tejido o tipo de célula se selecciona del grupo que consiste en sangre, saliva, semen, epidermis, orina, plasma, cabello, sangre menstrual, células y/o secreción vaginal, sudor, heces, cerebro, esófago, pulmón, estómago, corazón, duodeno, hígado, vesícula biliar, intestino, riñón, glándula suprarrenal, vejiga urinaria, uretra, colon, testículo, ovario, útero, vagina, músculo, tendón, ligamento, grasa, cartílago, hueso, células endoteliales, cuello uterino, linfoma, tiroides, glándula pituitaria, cerebelo y mama.

50 En una realización particular, el tipo de tejido o tipo de célula es sangre, saliva, semen o epidermis.

En otra realización, la afección patológica es cáncer, inflamación, trastorno autoinmune, trastorno metabólico, infección, enfermedad degenerativa, desequilibrios hormonales, un trastorno caracterizado por metilación anormal del ADN, trastorno del desarrollo neurológico de ICF (inmunodeficiencia, inestabilidad centromérica y anomalías faciales), síndrome de Rett y síndrome de X frágil.

5

En una realización, la afección prenatal es síndrome de Prader-Willi, síndrome de Angelman, síndrome de Beckwith-Wiedemann, síndrome de X frágil, síndrome de Russell-Silver, diabetes mellitus neonatal transitoria, osteodistrofia hereditaria de Albright, síndrome de McCune-Albright, paraganglioma familiar no cromafínico, síndromes UPD14 maternos y paternos.

10

En otra realización, los valores de referencia contra los cuales se comparan las relaciones de señal representan categorías que son mezclas de otras categorías.

En otra realización, los valores de referencia contra los que se comparan las relaciones de señal son de mezclas de semen y diferentes de semen en diferentes relaciones.

15

En una realización, la endonucleasa de restricción sensible a la metilación se selecciona del grupo que consiste en AatII, Acc65I, Accl, Acil, ACII, Afel, Agel, Apal, ApaLI, Ascl, AsiSI, Aval, Avall, Bael, BanI, Bbel, BceAI, Bcgl, BfuCI, BgII, BmgBI, BsaAI, BsaBI, BsaHI, Bsal, BseYI, BsiEI, BsiWI, BslI, BsmAI, BsmBI, BsmFI, BspDI, BsrBI, BsrFI, BssHII, BssKI, BstAPI, BstBI, BstUI, BstZ17I, Cac8I, Clal, Dpnl, Drdl, Eael, Eagl, Eagl-HF, Ecil, EcoRI, EcoRI-HF, Faul, Fnu4HI, Fsel, Fspl, Haell, Hgal, Hhal, HincII, HincII, HinfI, HinP1I, Hpal, HpalI, Hpy166ii, Hpy188iii, Hpy99I, HpyCH4IV, KasI, MluI, Mmel, MspA1I, MwoI, Nael, NarI, NgoNIV, Nhe-HFI, NheI, NlaIV, NotI, NotI-HF, Nrul, Nt.BbvCI, Nt.BsmAI, Nt.CviPII, PaeR7I, PleI, Pmel, PmlI, PshAI, PspOMI, PvuI, RsaI, RsrII, SacII, Sall, Sall-HF, Sau3AI, Sau96I, ScrFI, Sfil, SfoI, SgrAI, Smal, SnaBI, Tfil, TscI, Tsel, TspMI, y ZraI.

20

En una realización particular, la endonucleasa de restricción sensible a la metilación es Hhal.

En otra realización, la endonucleasa de restricción dependiente de la metilación se selecciona del grupo que consiste en McrBC, McrA y MrrA.

25

En una realización, la categoría de ADN de referencia es ADN natural o ADN sintetizado artificialmente, en el que al menos un locus genómico en el ADN natural o ADN sintetizado artificialmente comprende un locus de repetición corta en tándem (STR) de núcleo usado para el perfil de ADN.

En otra realización, el locus genómico comprende un locus humano seleccionado del grupo que consiste en D16S539 (SEQ ID NO. 1), D7S820 (SEQ ID NO. 2), D13S317 (SEQ ID NO. 3), D5S818 (SEQ ID NO. 4), CSF1PO (SEQ ID NO. 5), TPOX (SEQ ID NO. 6), TH01 (SEQ ID NO. 7), vWA (SEQ ID NO. 8), FGA (SEQ ID NO. 9), D21S11 (SEQ ID NO. 10), D8S1179 (SEQ ID NO. 11), D18S51 (SEQ ID NO. 12), y D3S1358 (SEQ ID NO. 13), Penta D (SEQ ID NO. 14), Penta E (SEQ ID NO. 15), y Amelogenina (SEQ ID NOs. 16 y 17), D2S1338 (SEQ ID No. 18), D19S433 (SEQ ID No. 19), ACTBP2SE33 (SEQ ID No. 20), D10S1248 (SEQ ID No. 21), D1S1656 (SEQ ID No. 22), D22S1045 (SEQ ID No. 23), D2S441 (SEQ ID No. 24), y D12S391 (SEQ ID No. 25).

En una realización, los cebadores para la amplificación de al menos uno de los loci genómicos están disponibles en:

(1) un kit comercial de Promega Corporation seleccionado del grupo que consiste en PowerPlex® 16 HS (Cat. N.º DC2100, DC2101), PowerPlex® 16 (Cat. N.º DC6530, DC6531), PowerPlex® 2.1 (Cat. N.º DC6470, DC6471), PowerPlex® 16 BIO (Cat. N.º DC6540, DC6541), y PowerPlex® ES Systems (Cat. N.º DC6730, DC6731);

(2) un kit comercial de Applied Biosystems seleccionado del grupo que consiste en SGM, SGM+, AmpFISTR Identifiler, AmpFISTR Profiler, AmpFISTR ProfilerPlus, AmpFISTR ProfilerPlusID, AmpFISTR SEfiler, AmpFISTR SEfiler Plus, AmpFISTR Cofiler, AmpFISTR Identifiler Direct, AmpFISTR Identifiler Plus, AmpFISTR NGM, AmpFISTR Y-filer, y AmpFISTR Minifiler; o

(3) Investigator ESSPlex, Investigator ESSPlex SE, Investigator Nonaplex ESSPlex, Investigator Hexaplex ESSPlex, Investigator Triplex AFS QS, Investigator Triplex DSF, Investigator IDplex, Investigator Decaplex SE, Investigator HDplex, Investigator Argus X-12, Investigator Y-12 QS, Investigator DIPlex.

En otra realización, se sabe que uno de los loci genómicos está sin metilar en todos los tejidos de ADN natural. En una realización, uno de los loci genómicos se selecciona del grupo que consiste en Hypo23, Hypo28, e Hypo33.

En otra realización, el método comprende además calcular al menos una de las siguientes relaciones:

- 5 (1) D3S1358/D18S51,
- (2) D3S1358/D7S820,
- (3) D3S1358/Penta\_D,
- (4) D3S1358/TPOX,
- (5) D3S1358/FGA,
- (6) TH01/Penta\_D,
- 10 (7) D21S11/D18S51,
- (8) D21S11/D7S820,
- (9) D21S11/Penta\_D,
- (10) D21S11/AMEL,
- (11) D21S11/TPOX,
- (12) D21S11/FGA,
- 15 (13) D5S818/D18S51,
- (14) vWA/D18S51,
- (15) D5S818/Penta\_E,
- (16) vWA/Penta\_E,
- (17) D5S818/D7S820,
- 20 (18) D5S818/Penta\_D,
- (19) D5S818/TPOX,
- (20) D5S818/FGA,
- (21) D13S317/D7S820,
- (22) D13S317/Penta\_D,
- 25 (23) D13S317/TPOX,
- (24) D13S317/FGA,
- (25) D16S539/D7S820,
- (26) CSF1PO/D7S820,
- (27) vWA/D7S820,
- 30 (28) D8S1179/D7S820,
- (29) D16S539/TPOX,
- (30) D16S539/FGA,
- (31) CSF1PO/Penta\_D,
- (32) CSF1PO/TPOX,
- 35 (33) vWA/Penta\_D,
- (34) AMEL/TPOX,
- (35) AMEL/FGA,
- (36) vWA/D8S1179,
- (37) vWA/TPOX,
- 40 (38) vWA/FGA,
- (39) D8S1179/TPOX, y
- (40) D8S1179/FGA.

En una realización particular, al menos uno de los loci genómicos es TPOX (SEQ ID NO. 6) o FGA (SEQ ID NO. 9).

45

En otra realización, los loci amplificado son de 90 pb o de menor tamaño.

En una realización, la intensidad de la señal es el nivel de fluorescencia del producto de amplificación medido durante la electroforesis capilar.

50

En una realización, las etapas (B) y (C) del método descrito anteriormente se realizan mediante PCR en tiempo real. Es decir, en una realización, todos los reactivos requeridos, cebadores y enzimas que se necesitan para la digestión del ADN y la amplificación por PCR están presentes juntos en el mismo tubo o recipiente.

55 En una realización, se usan cebadores marcados por fluorescencia para amplificar por PCR los loci elegidos del ADN digerido, en el que la intensidad de señal de cada producto de amplificación es el nivel de fluorescencia de cada producto de amplificación.

En otra realización, el método comprende además hibridar una sonda marcada por fluorescencia con un producto de

amplificación, en el que la intensidad de señal del producto de amplificación es el nivel de fluorescencia medido después de la hibridación de la sonda con el producto.

En otra realización, el método comprende además asignar niveles de confianza a categorías potenciales de una muestra de ADN, en el que el nivel de confianza refleja la probabilidad de que una categoría de ADN particular indique la fuente categórica de la muestra de ADN; en el que la probabilidad se calcula:

- 10 (A) asignando una puntuación de probabilidad a cada comparación de una relación de señal a los valores de referencia que corresponden a las diversas categorías de ADN, en el que la puntuación de probabilidad se obtiene (i) asignando una función de distribución de probabilidad gamma a los valores de referencia de cada categoría potencial; en el que (ii) la puntuación de probabilidad es igual al valor de la función de distribución de probabilidad gamma asignada en la etapa (A) a la relación de señal observada;
- 15 (B) para cada categoría, calcular la puntuación de probabilidad de la categoría, que es el producto de las puntuaciones de probabilidad obtenidas en la etapa (A); y
- (C) normalizar las puntuaciones de probabilidad dividiendo la puntuación de probabilidad de categoría de cada categoría potencial por la suma de todas las puntuaciones de probabilidad de categoría;

en el que el nivel de confianza de cada categoría es la puntuación de probabilidad normalizada de esa categoría.

20 En otra realización, el método comprende además realizar simultáneamente el perfil de ADN con la categorización de ADN.

En una realización, las etapas de la digestión del ADN y la reacción en cadena de la polimerasa se realizan juntas en un solo tubo.

25 En una realización, los loci genómicos que se amplifican se escogen para producir relaciones de señal distintas para diferentes categorías potenciales.

También se describe un kit para categorizar una muestra de ADN en al menos dos categorías predeterminadas y para obtener un nivel de confianza de categorización asociado, que comprende (1) cebadores para la amplificación de loci genómicos específicos; y al menos uno o más reactivos seleccionados del grupo que consiste en (2) un tampón de reacción, (3) ADN de control, (4) una endonucleasa de restricción sensible a la metilación y/o una endonucleasa de restricción dependiente de la metilación, (5) un protocolo escrito para realizar la categorización. En una realización de la presente invención, el protocolo escrito puede comprender instrucciones para realizar cualquiera de los métodos descritos en el presente documento, incluyendo, pero sin limitación, parámetros de digestión de ADN, parámetros de ciclado de PCR, análisis de relación de señal, y análisis de categorización, así como métodos de perfil de ADN.

40 En otra realización, el kit comprende además un software de análisis para realizar los análisis de categorización.

Otro aspecto de la presente invención es un kit para categorizar una muestra de ADN como semen o diferentes de semen y para obtener un nivel de confianza de categorización asociado, que comprende:

- 45 (a) mezcla de cebadores, que comprende los siguientes cebadores:  
 SD1f (AAGAGCCCATCAGGCAGGTC);  
 SD1r (GTTTCTTGTCGAGCAGCACGTGGATGATG);  
 SD2f (CTCCAGAACTGGAATTCCTG);  
 SD2r (GTTTCTTAACCTTGGAGACGACGGCATC);  
 SD3f (TGGAGGACAATGCCCTGGTG);  
 50 SD3r (GTTTCTTGGCTTCACCTGCGACCGTCTC);  
 SD4f (CCCTCCGAGTGGCCAGCAG);  
 SD4r (GTTTCTGACCACTGCCGTGGGAATG);  
 SD5f (CTTCTCAGCCAATGGGAAGAG);  
 SD5r (ACGTAGAAGGACCCGAGGAC);  
 55 SD6f (TACAGACAAATCACTCAGCAGC); y  
 SD6r (GTTTCTTGTCTGACACTCGGTTGTAGGTATT);  
 (b) tampón de reacción;  
 (c) endonucleasa de restricción HhaI;  
 (d) ADN polimerasa

(d) un protocolo escrito para realizar la categorización.

En otra realización, el kit comprende además muestras de ADN de control.

5 En otro aspecto de la presente invención se encuentra un kit para categorizar una muestra de ADN como semen o diferente de semen y para obtener un nivel de confianza de categorización asociado, que comprende al menos un par de par de combinaciones de cebadores directo (f) e inverso (r) seleccionadas del grupo que consiste en:

- 10 (1) SD1f (AAGAGCCCATCAGGCAGGTC) y SD1r (GTTTCTTGTCGAGCAGCACGTGGATGATG);  
 (2) SD2f (CTCCAGAACTGGAACCTCCTG) y SD2r (GTTTCTTAACTTGGAGACGACGGCATC);  
 (3) SD3f (TGGAGGACAATGCCCTGGTG) y SD3r (GTTTCTTGGCTTACCTGCGACCGTCTC);  
 (4) SD4f (CCCTCCGAGTGGCCAGCAG) y SD4r (GTTTCTGACCACTGCCGTGGGAATG);  
 (5) SD5f (CTTCTCAGCCAATGGGAAGAG) y SD5r (ACGTAGAAGGACCCGAGGAC);  
 15 (6) SD6f (TACAGACAAATCACTCAGCAGC) y SD6r (GTTTCTTGTCTGACTCGGTTGTAGGTATT).

15 En una realización, la concentración de los cebadores en la mezcla de cebadores son: SD1f 0,6 µM, SD1r 0,6 µM, SD2f 1,75 µM, SD2r 1,75 µM, SD3f 1,25 µM, SD3r 1,25 µM, SD4f 1,75 µM, SD4r 1,75 µM, SD5f 1,75 µM, SD5r 1,75 µM, SD6f 0,9 µM, y SD6r 0,9 µM.

20 En una realización, el tampón de reacción comprende TRIS-HCl 150 mM, MgCl<sub>2</sub> 15 mM, cada dntp 0,2 mM, y 0,1 µg/µl de BSA.

En una realización, el kit comprende además una escalera de ADN.

25 En una realización, el kit comprende además un software de análisis para realizar los análisis de categorización.

También se describe un kit para analizar, categorizar una muestra de ADN como semen o diferente de semen, y obtener un nivel de confianza de categorización asociado, que comprende:

- 30 (a) cebadores para amplificar al menos un locus específico de semen;  
 (b) cebadores para amplificar al menos un locus usado para el perfil de ADN;  
 (c) tampón de reacción;  
 (d) endonucleasa de restricción HhaI,  
 (e) ADN polimerasa  
 35 (f) un protocolo escrito para realizar la categorización.

En una realización, el kit comprende además muestras de ADN de control.

40 Otro aspecto de la presente invención es un kit para analizar, categorizar una muestra de ADN como semen o diferente de semen, y obtener un nivel de confianza de categorización asociado, que comprende:

- (a) cebadores para amplificar al menos un locus específico de semen; y  
 (b) cebadores para amplificar al menos un locus usado para el perfil de ADN;

45 al menos un locus específico de semen amplificado mediante un kit descrito en el presente documento es el locus L68346 y en el que los cebadores para amplificar un producto de amplificación específico para semen de 70 pb a partir de L68346 son un cebador directo que comprende la secuencia de CAGCAACAGCACCCAGCTTG (FAM) y un cebador inverso que comprende la secuencia de CACAGGCTCAGTCGCGGATC; o,  
 al menos un locus específico de semen amplificado mediante un kit descrito en el presente documento es el locus  
 50 L16264 y en el que los cebadores para amplificar un producto de amplificación específico para semen de 95 pb a partir de L16264 son un cebador directo que comprende la secuencia de GGACGAGTTAACTTCCTTAATTTTC (FAM) y un cebador inverso que comprende la secuencia de GTTTCTTTCGCGGAACCTGGTTAACTTC.

55 En otra realización, el tampón de reacción comprende TRIS-HCl 150 mM, MgCl<sub>2</sub> 15 mM, cada dntp 0,2 mM, y 0,1 µg/µl de BSA.

En otra realización, el kit comprende además al menos uno de (a) una escalera de ADN, (b) una ficha de datos de seguridad (MSDS), y (c) un software de análisis para realizar análisis de categorización.

Otro aspecto de la presente invención es un kit para categorizar una muestra de ADN como sangre, saliva, semen o epidermis de la piel, y para obtener un nivel de confianza de categorización asociado, que comprende:

- 5 (a) mezcla de cebadores que comprende cebadores directo e inverso para amplificar los loci representados como se indica a continuación:
1. L91762 (GCAGCAGGCCGCGGAGAAG directo (FAM); AGCAGCTGTGCCGGGCCAG inverso)  
 2. L68346 (CAGCAACAGCACCCAGCTTG directo (JOE); CACAGGCTCAGTCGCGGATC inverso)  
 10 3. L50468 (AGGAAACCTCAGTAGCAAATTG directo (JOE); GCGAGACTTTAGGTGTGCATC inverso)  
 4. L14432 (CGTAGGCTGCGGTGAGCTC directo (FAM); GATCCATGCCCGCTGGGATG inverso)  
 5. L4648 (CAGCCTAGACGTC AAGTTACAG directo (JOE); ACGACCTCCGGATCCAACCTG inverso)  
 6. L39664 (CCCAGCTGGTTGGACATGTTG directo (FAM); CACTTCCTTCGTGGACGCC inverso)  
 15 7. L30139 (GAGAAGCGGGAGGATGAGAC directo (FAM); CCGCATCTCCTCCGTCCTG inverso)  
 8. L55429 (GCCTTCAGCAGGAAGTCCAC directo (JOE); CCTGTGCCTCACACAGACTC inverso)  
 9. L62086 (GTGCATGGTGTCTGGTACTTC directo (FAM); GAAGCTCTCGTGGACTACTTG inverso)  
 10. L76138 (CAGCCTGCTCTTCACTGCAG directo (JOE); AGAGGCCGATGAAGCCGTAG inverso)  
 20 11. L15952 (CTCCCTGATTTACGACAAGTTC directo (FAM); GACAGTATGCTGATGCTTCTTG inverso)  
 12. L36599 (AAGGGCAGAGTTCCGCTGTC directo (FAM); CGGATGCAGGAGGATCCTAG inverso)  
 25 13. L26688 (CGGACCAGATTGCTGGTCAC directo (JOE); CGACCTTGCCAGATGTTTGAC inverso)  
 14. L81528 (AGCCTCATCCACACTGACCAG directo (JOE); TCAGAGCTCTCCTATCTGGAC inverso)  
 15. L36556 (GCCAGGCCGTTGATGATGAC directo (JOE); GAATATGGAGCCCTGGGCAG inverso)  
 30
- (b) tampón de reacción;  
 (c) endonucleasa de restricción HhaI;  
 (d) ADN polimerasa  
 35 (e) un protocolo escrito para realizar la categorización.

En otra realización, el kit comprende además muestras de ADN de control.

40 En una realización, el tampón de reacción comprende TRIS-HCl 150 mM, MgCl<sub>2</sub> 15 mM, cada dntp 0,2 mM, y 0,1 µg/µl de BSA.

En otra realización, el kit comprende además al menos uno de (a) una escalera de ADN, (b) una ficha de datos de seguridad (MSDS), y (c) un software de análisis para realizar análisis de categorización.

45 En una realización en cualquiera de los métodos descritos en el presente documento, la muestra de ADN comprende una mezcla de muestras de ADN.

También se describe un método para calcular una medida de distancia de metilación diferencial entre muestras de ADN, que comprende:

- 50 (A) digerir cada muestra de ADN con una endonucleasa de restricción sensible a la metilación y/o dependiente de la metilación;  
 (B) realizar un análisis de PCR en cada ADN digerido amplificando al menos dos loci genómicos de los cuales al menos uno es un locus de restricción;  
 55 (C) para cada muestra de ADN, determinar la intensidad de la señal de cada producto de amplificación;  
 (D) para cada muestra de ADN, calcular "relaciones de señal" entre las intensidades de las señales producidas por los loci;  
 (E) calcular la medida de metilación diferencial entre muestras de ADN realizando una comparación cuantitativa de sus relaciones de señal correspondientes.

En una realización de este método, la medida de distancia es la suma de las diferencias absolutas entre las relaciones de señal de las muestras de ADN.

5 En otra realización de este método, la medida de distancia es la raíz cuadrada de la suma de las diferencias cuadráticas entre las relaciones de señal de las muestras de ADN.

En una realización de este método, la distancia medida entre una muestra de sujeto y una muestra sana de referencia indica la cantidad de medicamento requerida para tratar una patología.

10

En otra realización de este método, la distancia medida entre una muestra de ADN obtenida de células cultivadas a una muestra de referencia indica el número de procedimientos de subcultivo que las células experimentaron.

También se describe un cebador seleccionado del grupo que consiste en:

15

1. GCAGCAGGCCGCGGAGAAG;
2. AGCAGCTGTGCCGGGCCAG;
3. CAGCAACAGCACCCAGCTTG;
4. CACAGGCTCAGTCGCGGATC;
- 20 5. AGGAAACCTCAGTAGCAAAATTG;
6. GCGAGACTTTAGGTGTGCATC;
7. CGTAGGCTGCGGTGAGCTC;
8. GATCCATGCCCGCTGGGATG;
9. CAGCCTAGACGTCAAGTTACAG;
- 25 10. ACGACCTCCGGATCCAAGT;
11. CCCAGCTGGTTGGACATGTTG;
12. CACTTCCTTCGTGGACGCC;
13. GAGAAGCGGGAGGATGAGAC;
14. CCGCATCTCCTCCGTCCTG;
- 30 15. GCCTTCAGCAGGAAGTCCAC;
16. CCTGTGCCTCACACAGACTC;
17. GTGCATGGTGTCTGGTACTTC;
18. GAAGCTCTCGTGGACTACTTG;
19. CAGCCTGCTCTTCACTGCAG;
- 35 20. AGAGGCCGATGAAGCCGTAG;
21. CTCCCTGATTTACGACAAGTTC;
22. GACAGTATGCTGATGCTTCTTG;
23. AAGGGCAGAGTTCCGCTGTC;
24. CGGATGCAGGAGGATCCTAG;
- 40 25. CGGACCAGATTGCTGGTCAC;
26. CGACCTTGCCAGATGTTTGAC;
27. AGCCTCATCCACACTGACCAG;
28. TCAGAGCTCTCCTATCTGGAC;
29. GCCAGGCCGTTGATGATGAC; y
- 45 30. GAATATGGAGCCCTGGGCAG
31. AAGAGCCCATCAGGCAGGTC
32. GTTTCTTGTGCGAGCAGCACGTGGATGATG
33. CTCCAGAAGTGGAACTTCCTG
34. GTTTCTTAACTTGGAGACGACGGCATC
- 50 35. TGGAGGACAATGCCCTGGTG
36. GTTTCTTGGCTTACCTGCGACCGTCTC
37. CCCTCCGAGTGGCCAGCAG
38. GTTTCTGACCACTGCCGTGGGAATG
39. CTTCTCAGCCAATGGGAAGAG
- 55 40. ACGTAGAAGGACCCGAGGAC
41. TACAGACAAATCACTCAGCAGC
42. GTTTCTTGTCTGACACTCGGTTGTAGGTATT
43. GGACGAGTTAACTTCCTTAATTTT
44. GTTTCTTCGCGGAACCTGGTTTAACTTC

También se describe un par de cebadores para amplificar un locus genómico humano específico seleccionado del grupo que consiste en:

- 5 1. L91762 (GCAGCAGGCCGCGGAGAAG directo (FAM); AGCAGCTGTGCCGGGCCAG inverso);
2. L68346 (CAGCAACAGCACCCAGCTTG directo (JOE); CACAGGCTCAGTCGCGGATC inverso);
3. L50468 (AGGAAACCTCAGTAGCAAAATTG directo (JOE); GCGAGACTTTAGGTGTGCATC inverso);
4. L14432 (CGTAGGCTGCGGTGAGCTC directo (FAM); GATCCATGCCCGCTGGGATG inverso);
- 10 5. L4648 (CAGCCTAGACGTCAAAGTTACAG directo (JOE); ACGACCTCCGGATCCAAGT inverso);
6. L39664 (CCCAGCTGGTTGGACATGTTG directo (FAM); CACTTCCTTCGTGGACGCC inverso);
7. L30139 (GAGAAGCGGGAGGATGAGAC directo (FAM); CCGCATCTCCTCCGTCCTG inverso);
8. L55429 (GCCTCAGCAGGAAGTCCAC directo (JOE); CCTGTGCCTCACACAGACTC inverso);
9. L62086 (GTGCATGGTGTCTGGTACTTC directo (FAM); GAAGCTCTCGTGGACTACTTG inverso);
10. L76138 (CAGCCTGCTCTTCACTGCAG directo (JOE); AGAGGCCGATGAAGCCGTAG inverso);
- 15 11. L15952 (CTCCCTGATTTACGACAAGTTC directo (FAM); GACAGTATGCTGATGCTTCTTG inverso);
12. L36599 (AAGGGCAGAGTTCCGCTGTC directo (FAM); CGGATGCAGGAGGATCCTAG inverso);
13. L26688 (CGGACCAGATTGCTGGTCAC directo (JOE); CGACCTTGCCAGATGTTTGAC inverso);
14. L81528 (AGCCTCATCCCACTGACCAG directo (JOE); TCAGAGCTCTCCTATCTGGAC inverso);
15. L36556 (GCCAGGCCGTTGATGATGAC directo (JOE); GAATATGGAGCCCTGGGCAG inverso).
- 20 16. SD1 (AAGAGCCCATCAGGCAGGTC directo (FAM); GTTTCTTGTCTGAGCAGCACGTGGATGATG inverso);
17. SD2 (CTCCAGAACTGGAAGTTCCTG directo (FAM); GTTTCTTAACTTGGAGACGACGGCATC inverso);
18. SD3 (TGGAGGACAATGCCCTGGTG directo (FAM); GTTTCTTGGCTTCACCTGCGACCGTCTC inverso);
- 25 19. SD4 (CCCTCCGAGTGCCAGCAG directo (FAM); GTTCTGACCACTGCCGTGGGAATG inverso);
20. SD5 (CTTCTCAGCCAATGGGAAGAG directo (FAM); ACGTAGAAGGACCCGAGGAC inverso);
21. SD6 (TACAGACAAATCACTCAGCAGC directo (FAM); GTTCTTGTCTGACTCGGTTGTAGGTATT inverso); y
- 30 22. L16264 (GGACGAGTTAACTTCTTAATTTTC directo (FAM); GTTCTTCCGCGAACCTGGTTAACTTC inverso).

#### BREVE DESCRIPCIÓN DE LOS DIBUJOS

- 35 **Figuras 1A-D.** Ensayo de identificación tisular. (A) Descripción esquemática del ensayo. (B) Procedimiento bioquímico - Los loci metilados permanecen intactos durante la digestión y posteriormente se amplifican de manera eficiente en la PCR, produciendo una fuerte señal (locus A) mientras que los loci no metilados se digieren y posteriormente se amplifican ineficientemente en la PCR, produciendo una señal débil (Locus B). La relación de señal (SR; rfu del locus A/rfu del locus B) refleja el nivel de metilación diferencial entre los loci A y B. (C) Las relaciones de señal entre el locus 1 y el locus 2 son diferentes en sangre, saliva, piel y semen, reflejando los patrones de metilación diferencial en estos tejidos. (D) Las diferencias observadas en las relaciones de señal entre la sangre y el semen (es decir, la señal) son más que una magnitud mayor que las diferencias observadas en las relaciones de señal obtenidas de diferentes PCR y de diferentes cantidades de ADN de entrada (es decir, el ruido).
- 40 **Figura 2.** Identificación de tejidos de diferentes tipos de tejidos. Se representan electroferogramas de 8 muestras (2 muestras de cada uno de 4 tejidos diferentes) que demuestran el patrón de amplificación distinto observado para cada tipo de tejido.
- 45 **Figuras 3A-D.** Detección combinada de semen y perfil de ADN. (A, B) Electroferogramas del ensayo combinado realizados en una muestra de semen (A) y orina (B) del mismo individuo. Los perfiles ProfilerPlus de las muestras son idénticos, y las diferencias en la fuente tisular se indican mediante la amplificación diferencial de dos loci de detección de semen (flechas). (C) Datos del canal FAM del ensayo combinado realizado en muestras de diferentes tipos de tejidos. La relación de señal entre los loci de detección de semen (flechas) es 25,04 en la muestra de semen, y menos de 0,1 en todos los demás tipos de tejido. El panel superior incluye ADN no digerido como referencia. (D) Datos del canal FAM del ensayo combinado realizado en mezclas de semen y saliva (de diferentes individuos) con diversas relaciones. La relación de señal se correlaciona con el porcentaje de semen en la mezcla.
- 50 **Figura 4.** Evaluación de la digestión de HhaI. Se sometieron muestras de ADN de diferentes fuentes a la misma reacción de digestión-amplificación individual utilizada en los experimentos de identificación de tejidos. En este experimento, el panel de loci consistió en dos loci: un locus de control no digerido (L98328)
- 55

que carece de un sitio Hhal y un locus de control digerido (SW14) que no está metilado en estos tejidos. Todas las muestras mostraron la presencia del locus de control no digerido y la ausencia completa del locus de control digerido, lo que indica la digestión completa por Hhal en contraste con la muestra no digerida que

5 **Figura 5.** Muestras maduras. El ensayo de identificación de tejido independiente se ensayó en muestras de ADN extraídas de manchas de sangre y semen de 20 meses de antigüedad. Ambas muestras se amplificaron bien, de forma similar a sus homólogos recientes, y su fuente se identificó correctamente por el software de análisis.

10 **Figura 6.** Simulación de muestras degradadas. Las muestras de sangre (superior) y semen (inferior) se digirieron parcialmente usando DNaseI y se analizaron con el ensayo combinado de detección y análisis del semen. Aunque se obtuvieron solamente perfiles parciales para ambas muestras, los loci de identificación de tejido más pequeños (los dos loci en el canal superior a la izquierda de D3S1358) se amplificaron con éxito con el patrón esperado, y el algoritmo identificó correctamente la presencia/ausencia de semen en estas muestras.

15 **Figura 7.** Tabla 1.

**Figura 8.** Tabla 2.

**Figura 9.** Tabla 3.

**Figura 10.** Tabla 4.

**Figura 11.** Tabla 5.

20 **Figura 12.** Ensayo de detección de semen independiente. Se representan gráficos de muestras de 6 muestras de ADN de diferentes tipos de tejidos. Se logró una clasificación correcta (semen frente a no semen) para todas las muestras. Para cada muestra, se muestra el nivel de confianza de la categoría menos probable (por ejemplo, para una categorización de "semen", se muestra el valor p diferente de semen)

25

## DESCRIPCIÓN DETALLADA

### 1. Introducción

30 La presente invención se refiere a métodos y ensayos que permiten distinguir entre fuentes de ADN que usan cebadores específicos de loci y enzimas disponibles comercialmente. Un aspecto subyacente de un ensayo de la invención es la comparación de señales de al menos dos loci amplificados a partir de una fuente particular de ADN, lo que finalmente proporciona una relación numérica que indica si esa fuente de la muestra de ADN es, en una realización, natural o artificial. Los ensayos de la invención también se pueden usar para distinguir, por ejemplo,

35 diferentes fuentes fisiológicas y patológicas de ADN y entre diferentes tejidos.

Las relaciones de señal empleadas por el ensayo se correlacionan con los niveles de metilación en loci genómicos específicos, pero no indican niveles reales de metilación en ningún locus genómico. Por lo tanto, la configuración de la invención elimina el requisito de especies y/o controles externos de ADN, lo que simplifica y aumenta de este

40 modo la precisión del ensayo.

Los ensayos de categorización de ADN de la invención que se describen en el presente documento son, por lo tanto, técnicas potentes, múltiples, precisas y económicas aplicables en cualquier configuración que requiera la categorización de una muestra de ADN. Por lo tanto, los ensayos de la invención pueden usarse, por ejemplo, por la

45 policía en una investigación forense; la industria de la atención sanitaria con fines diagnósticos y terapéuticos; en la industria de los seguros para verificar las reclamaciones de conformidad con las leyes genéticas antidiscriminatorias, como la Ley de No Discriminación de Información Genética (H.R. 493); por parte de fiscales y abogados defensores con fines probatorios en juicios penales y procesos civiles y apelaciones; y la industria agroalimentaria para verificar la integridad de las carnes, los cultivos y las plantas, tales como las vides y las fuentes de café. La presente

50 invención no se limita a esta lista de aplicaciones no exclusiva, pero representativa.

Un principio subyacente de un método descrito en el presente documento es, por lo tanto, la medición de las intensidades de señal entre los loci genómicos amplificados y su posterior comparación por pares entre sí para producir relaciones de intensidades de señal entre los loci amplificados. Por ejemplo, si los loci A, B, C y D se

55 amplifican todos de acuerdo con las técnicas descritas en el presente documento, entonces se mide y se registra la intensidad de la señal para cada uno de A, B, C y D. Entonces, las intensidades de señal de A y B, A y C, y A y D se comparan respectivamente y se calculan las relaciones de intensidad de señal para cada una de las combinaciones A/B, A/C y A/D. Luego, se calculan las relaciones de intensidad de señal de B/C y B/D, y así sucesivamente.

El valor de cada una de esas relaciones de intensidad de señal puede compararse entonces con los valores de las relaciones de intensidad de señal de referencia conocidas de ADN cuya fuente de origen es conocida. Por lo tanto, por ejemplo, en el caso del ADN cuya fuente de origen se conoce como semen humano, se conocen las relaciones de intensidad de señal de los loci, tales como las relaciones por pares de los loci ilustrativos A, B, C y D, o pueden establecerse fácilmente de nuevo al analizar una muestra de ADN de ensayo. Por lo tanto, la categoría de ADN es "semen" y sus respectivas relaciones de intensidad de señal para esos loci son valores de referencia contra los que se puede comparar la muestra de ADN de ensayo.

En consecuencia, después de establecer las diversas relaciones de intensidad de señal de los loci A, B, C y D, a partir de la muestra de ADN, cada relación puede compararse con uno o más valores de referencia de categorías de ADN conocidas, por ejemplo, contra los valores de la relación de ADN del semen. La relación de cualquier combinación de loci de la muestra de ADN que más se aproxima al valor de referencia de la relación de intensidad de señal de una categoría de ADN particular significa que es altamente probable que la muestra de ADN ensayada se originase a partir de esa categoría particular de ADN a la que más se aproxima al compararse. Por lo tanto, una relación de intensidad de señal de la muestra de ADN ensayada de los loci A/B que es igual o cercana al valor de la relación de intensidad de señal de los loci A/B del semen indica que la muestra de ADN ensayada se originó probablemente a partir del semen. La realización del mismo análisis comparativo en muchos loci diferentes, y contra muchas categorías de referencia diferentes, aumenta la probabilidad de que la muestra de ADN ensayada se originase a partir de una fuente biológica particular, o que no se originase a partir de una fuente biológica particular.

De manera similar, en una situación en la que una muestra puede comprender mezclas de ADN de diferentes fuentes, tales como saliva y sangre, o saliva, sangre, epidermis y semen, será posible, mediante la realización de los métodos descritos en el presente documento, identificar la presencia de diferentes categorías de ADN dentro de la muestra. Por lo tanto, la identificación de fuentes múltiples, así como únicas de ADN presentes en una muestra se puede lograr usando los métodos y reactivos descritos en el presente documento.

Por consiguiente, los presentes métodos de "categorización" también pueden considerarse como un ensayo de identificación de tejido. En un ejemplo, como se explica en el siguiente paso, un ensayo de identificación de tejido utiliza un panel de loci que se metilan de forma diferente entre los tejidos para determinar el tejido de origen más probable de una muestra de ADN. Véase la Tabla 1 (Figura 7). Un esquema del ensayo se presenta en la figura 1A. El ADN de una muestra forense se digiere con, por ejemplo, la enzima de restricción sensible a la metilación Hha1, que escinde ADN en su secuencia de reconocimiento GCGC solamente si no está metilado (dejando intactas las dianas metiladas).

A continuación, se amplifica un panel de loci de identificación de tejido mediante PCR a partir del ADN digerido usando cebadores marcados por fluorescencia, y una alícuota de productos amplificados se separa mediante electroforesis capilar. En la situación con el ADN digerido con Hha1, los loci con niveles de metilación más altos se amplifican con mayor eficiencia porque se protegen más moléculas de ADN de la digestión, produciendo una señal relativamente fuerte en el electroferograma (figura 1B, locus A). Por el contrario, los loci con un nivel de metilación inferior se amplifican con menor eficiencia, produciendo una señal relativamente débil en el electroferograma (figura 1B, locus B).

El software de análisis de señal automatizado descrito en el presente documento permite el análisis de un electroferograma de salida, y asigna alturas (en rfu) a los amplicones correspondientes a los loci de identificación de tejido. Por lo tanto, la altura de un único locus se correlaciona con su nivel de metilación. Las relaciones de los niveles de metilación entre los loci co-amplificados pueden calcularse como se ha descrito genéricamente anteriormente. Para cada par de loci, una relación de señal, calculada como la relación entre las alturas del primer y segundo loci, y esta relación refleja la relación entre los niveles de metilación de los loci correspondientes.

Todas las relaciones de señal calculadas se pueden combinar en un único valor numérico. Por ejemplo, se pueden calcular 105 relaciones por parejas a partir de un panel de 15 loci (por ejemplo, como entre el locus 1 y 2, el locus 1 y 3, etc.). A continuación, se pueden comparar con una base de datos de valores de referencia obtenidos a partir de un conjunto de datos de muestras de origen tisular conocido. El algoritmo de identificación tisular calcula para cada fuente de tejido potencial una puntuación de probabilidad, lo que refleja la probabilidad de que la muestra de ADN se originase a partir de ese tejido, y el resultado del algoritmo es el tejido más probable. A continuación, se describen en otra parte en el presente documento algoritmos y cálculos específicos para obtener la posibilidad y puntuaciones de probabilidad a partir de estas razones y comparaciones.

Como se muestra en el Ejemplo 1, por ejemplo, tal ensayo de identificación tisular se realizó en 50 muestras: 14 de

sangre, 14 de saliva, 11 de semen y 11 epidermis usando un panel de 15 loci de identificación tisular, que varían en tamaño de amplicón de 66 a 150 pb. La figura 2 representa electroferogramas de ocho de estas muestras, dos de cada tipo de tejido. Cada tipo de tejido tenía un perfil de metilación distinto. Por ejemplo, la relación de los loci L91762/L68346 en muestras de semen varió de 0,04-0,53 y fue mayor en todas las demás muestras de tejido (2,15-18,28; véase la Tabla 2 en la figura 8). Por lo tanto, se encontró que una relación *baja* de L91762/L68346 era distintiva de las muestras de semen. Sin embargo, la relación de los loci L76138/L26688 fue *baja* en sangre y saliva (0,08-1,54) y *mayor* en semen y epidermis (2,04-19,56; Tabla 2 en la figura 8). Por lo tanto, una relación alta de L91762/L68346 concomitante con una alta relación de L76138/L26688 fue distintiva de las muestras de epidermis de la piel.

10

El algoritmo de identificación tisular descrito en el presente documento identificó correctamente la verdadera fuente de tejido de las 50 muestras. El análisis adicional de los datos utilizando subconjuntos de loci de identificación de tejido reveló que el conjunto completo de 15 loci fue redundante y se logró una identificación del 100 % utilizando solamente 7 de los 15 loci (L91762, L68346, L50468, L14432, L30139, L15952, y L26688).

15

Por consiguiente, este ejemplo ilustrativo corrobora el aspecto inventivo genérico de la presente invención, concretamente, que los análisis de relación de intensidad de señal comparativos proporcionan una identificación y categorización de muestras de ADN exactas y correctas.

Además, otro aspecto significativo de la presente invención es que puede complementar y expandir fácilmente la utilidad de kits de análisis de ADN comerciales existentes para hacer más de un perfil del ADN de un sujeto particular. La combinación de los ensayos de la invención descritos en el presente documento, tal como el ensayo de relación de señal descrito en detalle a continuación, con el kit PowerPlex® 16 de Promega Corporation, por ejemplo, permite al usuario no solo analizar una composición de ADN de un individuo, sino también determinar si el perfil de ADN es de una fuente natural o se ha sintetizado artificialmente. En otro ejemplo, combinar el perfil de ADN con la categorización de ADN permite al usuario tanto analizar de ADN como determinar el tejido de origen de la muestra de ADN.

Por ejemplo, un ensayo descrito en el presente documento puede emplear *tanto* perfiles convencionales de una muestra de ADN, por ejemplo con el kit PowerPlex16 (Promega), *como* la categorización de ADN en categorías de natural frente a artificial. Véase, por ejemplo, el Ejemplo 6 a continuación.

Con fines ilustrativos solamente de cómo la categorización y el análisis combinados pueden funcionar en conjunto como se describe en el presente documento, se analizan dos muestras de ADN: (1) una muestra de ADN natural extraída de la sangre del sujeto A, y (2) una muestra de ADN artificial sintetizada mediante amplificación de desplazamiento múltiple *in vitro* usando un kit comercial a partir de una pequeña cantidad de ADN real del sujeto A como plantilla.

Además, esta ilustración demuestra cómo el método de la presente invención es muy superior a cualquier método existente porque es capaz de distinguir el ADN artificial de las muestras de ADN natural, *aunque* los kits de análisis convencionales identificarán ambas muestras como idénticas entre sí, ya que tanto el ADN natural como el artificial proceden del sujeto A. Por consiguiente, a pesar del hecho de que el kit de análisis identificará las muestras como iguales, los análisis de la relación de intensidad de señal de la presente invención pueden distinguirlos fácilmente e identificar un perfil como generado a partir de ADN artificial, mientras que el otro perfil se generó a partir de ADN natural.

Para una explicación adicional, las dos muestras se analizan en primer lugar con PowerPlex16 y se analizan usando un software de análisis convencional, tal como GeneMapperID-X. El esquema de este procedimiento se puede describir brevemente como (1) realizar PCR múltiple utilizando la mezcla de cebadores PowerPlex16 en una muestra de ADN; (2) separar los productos de amplificación en una máquina de electroforesis capilar; y después (3) analizar los datos de salida con, por ejemplo, el software GeneMapperID-X. Para cada una de las muestras, el producto final de este ensayo es un perfil. Los perfiles de las dos muestras son idénticos. Véanse las figuras 9A-B.

Además, el software GeneMapperID-X determinó que los perfiles de ambas muestras eran perfiles de un único contribuyente (es decir, de una única persona), sin anomalías. Esto muestra que un perfil obtenido a partir de ADN artificial puede ser idéntico a un perfil obtenido a partir de ADN natural. Véase D. Frumkin, et al., Authentication of forensic DNA samples, Forensic Sci. Int. Genet. (2009), doi:10.1016/j.fsigen.2009.06.009, que se incorpora en el presente documento por referencia.

Posteriormente, ambas muestras de ADN también se analizaron mediante un ensayo de análisis y categorización combinado basado en el siguiente esquema: digerir una muestra de ADN con HhaI y realizar una PCR múltiple usando la mezcla de cebadores PowerPlex16 con la adición de cebadores para amplificar un locus adicional, tal como Hypo23; separar los productos de amplificación en una máquina de electroforesis capilar; y analizar los datos de salida con el software del analizador capilar, tal como el algoritmo de software descrito en el presente documento.

Para cada muestra, el producto final de este ensayo es un perfil y una asignación de categoría. Como se puede observar en la figura 9C y D, el ADN natural se asignó como "natural", en contraste con la muestra artificial que se asignó como "artificial".

10

En este ejemplo particular, la asignación de categoría se realizó mediante el análisis de tres parámetros: (1) SR1 (relación de señal de TPOX/D8S1179), (2) SR2 (relación de señal de D3S1358/Hypo23), y (3) RR (relación de representación de OCA2/D3S1358).

15 Los valores de umbral para estos parámetros se proporcionan en la Tabla 6:

**Tabla 6**

	SR1	SR2	RR
Umbral para natural	$\geq 0,9$	$\geq 10$	$\geq 2$
Umbral para artificial	$< 0,9$	$< 10$	$< 2$

Los valores observados para estos parámetros, y la categoría asignada se proporcionan en la Tabla 7 (los valores que están por encima del umbral para el ADN natural están sombreados).

20

**Tabla 7**

	SR1	SR2	RR	Categoría
Muestra de ADN natural	1,23	17,54	3,31	Natural
Muestra de ADN artificial	0,06	2,06	2,47	Artificial

La asignación de la categoría se realizó de acuerdo con la siguiente regla:

25

Si todos los valores de los parámetros observados están por encima de sus umbrales respectivos para el ADN natural, entonces se asigna una categoría "natural", de lo contrario se asigna una categoría "artificial". Por lo tanto, los análisis de relación de intensidad de señal de la presente invención pueden identificar fácilmente fuentes de ADN, tales como, por ejemplo, fuentes naturales frente a artificiales, y junto con técnicas de perfil de ADN proporciona un método potente y sofisticado para identificar materiales genómicos.

30

Las composiciones, métodos o realizaciones específicos que se analizan pretenden ser solamente ilustrativos de la invención que se describe por esta memoria descriptiva. Las variaciones en estas composiciones, métodos o realizaciones son fácilmente evidentes para un experto en la técnica en base a las enseñanzas de esta memoria específica y, por lo tanto, pretenden incluirse como parte de las invenciones descritas en el presente documento.

35

En la práctica de la presente invención, se usan muchas técnicas convencionales en biología molecular y ADN recombinante. Estas técnicas se explican en, por ejemplo, Current Protocols in Molecular Biology, Vols. I-III, Ausubel, Ed. (1997); Sambrook et al., Molecular Cloning: A Laboratory Manual, Second Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989); DNA Cloning: A Practical Approach, Vols. I y II, Glover, Ed. (1985); Oligonucleotide Synthesis, Gait, Ed. (1984); Nucleic Acid Hybridization, Hames & Higgins, Eds. (1985); Transcription and Translation, Hames & Higgins, Eds. (1984); Perbal, A Practical Guide to Molecular Cloning; the series, Meth. Enzymol., (Academic Press, Inc., 1984); Gene Transfer Vectors for Mammalian Cells, Miller & Calos, Eds. (Cold Spring Harbor Laboratory, NY, 1987); y Meth. Enzymol., Vols. 154 and 155, Wu & Grossman, y Wu, Eds., respectivamente.

45

## 2. Definiciones

La presente tecnología se describe en el presente documento usando varias definiciones, como se expone a lo largo de la memoria descriptiva. A menos que se defina de otro modo, todos los términos técnicos y científicos utilizados en el presente documento generalmente tienen el mismo significado que comúnmente entiende un experto en la técnica a la que pertenece esta invención. Como se usa en el presente documento, a menos que se indique otra cosa, las formas singulares "un", "una" y "el/la" incluyen referencias plurales. Por lo tanto, por ejemplo, una referencia a "un ácido nucleico" es una referencia a uno o más ácidos nucleicos.

Como se usa en el presente documento, el término "alelo" pretende ser una variación genética asociada con un segmento de ADN, es decir, una de dos o más formas alternativas de una secuencia de ADN que ocupa el mismo lugar.

La expresión "ADN artificial" o "ADN sintetizado artificialmente" o "ácido nucleico artificial" como se usa en el presente documento, se refiere a un ácido nucleico que se sintetiza mediante *diversos métodos in vitro*. Se describen en el presente documento las caracterizaciones de diferentes métodos para sintetizar ADN *in vitro*. Dichos ácidos nucleicos generados *in vitro* incluyen, pero sin limitación:

1. Oligonucleótidos sintetizados químicamente.
2. Productos de amplificación por PCR de las secuencias diana.
3. Productos de amplificación en círculo rodante (RCA) de secuencias diana circulares.
4. Productos de clonación molecular (por ejemplo, plásmidos clonados en *E. coli*).
5. Fragmentos de ADN ensamblados a partir de otros fragmentos de ADN que se generaron por cualquiera de los métodos 1-4, o una combinación de ellos. Dicho ensamblaje se logra mediante cualquiera de los siguientes métodos (o una combinación de ellos): hibridación, ligadura, polimerización. El proceso de ensamblaje también puede incluir etapas de ruptura de moléculas de ADN (por ejemplo, mediante endonucleasas de restricción, cizallamiento mecánico, etc.).
6. Productos de amplificación genómica completa (WGA) basada en PCR, y/o métodos de WGA basados en PCR mediada por ligadura (LMP), que incluyen la PCR de preamplificación de extensión de cebador (PEP), PCR cebada por oligonucleótidos degenerados (DOP), amplificación lineal basada en T7 de ADN (TLAD), PCR de ligadura de adaptadores. El kit comercial Genomeplex (Sigma) utiliza PCR de ligadura de adaptadores.
7. Productos de WGA mediante amplificación de desplazamiento múltiple (MDA) y amplificación en círculo rodante asistida por restricción y circularización (RCA-RCA). Los kits comerciales Repli-G (Qiagen) y GenomiPhi (GE Healthcare) utilizan este método.
8. Una mezcla de productos de cualquiera de 1-7.
9. Productos de cualquiera de 1-8 en los que todos o algunos productos se metilaron *in vitro* después de su síntesis (por ejemplo, por Sss1 Methylase).
10. Productos de cualquiera de 1-8 mezclados con ADN natural.
11. Productos de 9 mezclados con ADN natural.

La expresión "muestra biológica" o "muestra de ensayo" como se usa en el presente documento, se refiere, pero sin limitación, a cualquier muestra biológica derivada, o obtenida a partir de, un sujeto. La muestra puede comprender ácidos nucleicos, tales como ADN o ARN. En algunas realizaciones, las muestras no se recuperan directamente del sujeto, sino que se recogen del entorno, por ejemplo, una escena del crimen o una víctima de violación. Los ejemplos de dichas muestras incluyen, pero sin limitación, fluidos, tejidos, muestras celulares, órganos, biopsias, etc. Las muestras adecuadas incluyen, pero sin limitación, sangre, plasma, saliva, orina, esperma, cabello, etc. La muestra biológica también puede ser gotas de sangre, manchas de sangre seca, manchas de saliva seca, manchas secas de ropa interior (por ejemplo, manchas en la ropa interior, compresas, tampones, pañales), ropa, seda dental, cera de oído, recortes de maquinilla de afeitar, encía, cabello, sobres lamidos, uñas, tejido embebido en parafina, tejido post mortem, maquinillas de afeitar, dientes, cepillo de dientes, mondadientes, cordón umbilical deshidratado. El ADN genómico se puede extraer de dichas muestras de acuerdo con los métodos conocidos en la técnica. (por ejemplo, utilizando un protocolo de Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Segunda Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989);

La expresión "histograma de electroforesis capilar", como se usa en el presente documento, se refiere a un histograma obtenido a partir de electroforesis capilar de productos de PCR en los que los productos se amplificaron a partir de loci genómicos con cebadores fluorescentes.

El término "metilado" como se usa en el presente documento, se refiere a metilado en un nivel de al menos el 80 % (es decir, al menos el 80 % de las moléculas de ADN metiladas) en el ADN de células de tejidos que incluyen

sangre, saliva, semen, epidermis, flujo nasal, células bucales, cabello, recortes de uñas, excreción menstrual, células vaginales, orina y heces.

La expresión "parcialmente metilado" como se usa en el presente documento, se refiere a metilado en un nivel entre el 20-80 % (es decir, entre el 20-80 % de las moléculas de ADN metiladas) en el ADN de células de tejidos que incluyen sangre, saliva, semen, epidermis, flujo nasal, células bucales, cabello, recortes de uñas, excreción menstrual, células vaginales, orina y heces.

La expresión "no metilado" como se usa en el presente documento, se refiere a metilado en un nivel menor del 20 % (es decir, menor del 20 % de las moléculas de ADN metiladas) en el ADN de células de tejidos que incluyen sangre, saliva, semen, epidermis, flujo nasal, células bucales, cabello, recortes de uñas, excreción menstrual, células vaginales, orina, hueso, y heces. Se ha demostrado que los métodos proporcionados en el presente documento distinguen formas metiladas y no metiladas de loci de ácidos nucleicos en diversos tejidos y tipos de células que incluyen sangre, saliva, semen, epidermis, secreción nasal, células bucales, cabello, recortes de uñas, excreción menstrual, células vaginales, orina, hueso y heces.

Los términos "determinar", "medir", "evaluar", "ensayar" y "evaluar" se usan de forma intercambiable para referirse a cualquier forma de medición cuantitativa o cualitativa, e incluyen determinar si una característica, rasgo o aspecto está presente o no. La evaluación puede ser relativa o absoluta. La "evaluación de la presencia de" incluye determinar la cantidad de algo presente, así como determinar si está presente o ausente.

El término "forense" o "ciencia forense", como se usa en el presente documento, se refiere a la aplicación de un amplio espectro de métodos destinados a responder preguntas de identidad que sean de interés para el sistema legal. Por ejemplo, la identificación de posibles sospechosos cuyo ADN pueda coincidir con la evidencia dejada en las escenas del crimen, la exoneración de personas incorrectamente acusadas de delitos, la identificación de víctimas de crímenes y catástrofes, o el establecimiento de paternidad y otras relaciones familiares.

El término "locus" (plural - loci) se refiere a una posición en un cromosoma de un gen u otro elemento genético. El locus también puede referirse al ADN en esa posición. Una variante de la secuencia de ADN en un locus dado se denomina un alelo. Los alelos de un locus se encuentran en sitios idénticos en cromosomas homólogos. Un locus de restricción es un locus que contiene la secuencia de reconocimiento de la enzima de restricción utilizada en el ensayo. La expresión locus de control digerido, como se usa en el presente documento, se refiere a un locus genómico que se sabe que no está metilado en ninguna de las categorías potenciales de ADN. La expresión locus de control no digerido, como se usa en el presente documento, se refiere a un locus genómico que se sabe que está metilado en todas las categorías potenciales de ADN, o es un locus genómico que carece de la secuencia de reconocimiento de la endonucleasa usada en el ensayo.

La expresión "ADN natural" o "ácido nucleico natural" como se usa en el presente documento, se refiere, pero sin limitación, a ácido nucleico que se origina directamente de las células de un sujeto sin modificación o amplificación.

El término "ácido nucleico" como se utiliza en el presente documento se refiere, pero sin limitación, a ADN genómico, ADNc, hnRNA, ARNm, ARNr, ARNt, ácido nucleico fragmentado, y ácido nucleico obtenido de orgánulos subcelulares tales como mitocondrias. Además, los ácidos nucleicos incluyen, pero sin limitación, ácidos nucleicos sintéticos o productos de transcripción *in vitro*.

La expresión "procedimientos de análisis basados en ácidos nucleicos" como se usa en el presente documento, se refiere a cualquier procedimiento de identificación que se basa en el análisis de ácidos nucleicos, por ejemplo, análisis de ADN.

La expresión "tartamudeo de reacción en cadena de la polimerasa (PCR)", como se usa en el presente documento, se refiere a los subproductos de PCR, obtenidos junto con el producto principal de la PCR. Estos subproductos de "tartamudeo" son habitualmente más cortos por múltiplos de la unidad repetida producida en el transcurso de la amplificación por PCR de las secuencias STR. Se entiende el mecanismo por el cual se forman estos artefactos, pero representa una limitación intrínseca de la tecnología de PCR y, por lo tanto, no se ha encontrado un remedio eficaz para eliminar estos productos espurios (Olejniczak M, Krzyzosiak WJ., Electrophoresis. 2006 Oct; 27(19): 3724-34). El término "tartamudeo -1" como se usa en el presente documento se refiere a un subproducto de tartamudeo que es una unidad de repetición más pequeña que su alelo asociado. De forma similar, "tartamudeo +1" se refiere a un subproducto de tartamudeo que es una unidad de repetición más grande que su alelo asociado. La expresión "fracción de tartamudeo -1" se refiere a la altura (o área) del pico de tartamudeo -1 dividido por la altura (o

área) del pico del alelo verdadero. De forma similar, la expresión "fracción de tartamudeo +1" se refiere a la altura (o área) del pico de tartamudeo +1 dividido por la altura (o área) del pico del alelo verdadero.

5 La expresión "amplificación en círculo rodante asistido por restricción y circularización (RCA-RCA)" se refiere a un procedimiento de amplificación de genoma completo que retiene las diferencias alélicas entre genomas amplificados degradados mientras se logra una cobertura del genoma casi completa. RCA-RCA utiliza la digestión de restricción y la circularización del genoma completo para generar secuencias genómicas susceptibles a la amplificación en círculo rodante.

10 La expresión "cebadores STR" como se usa en el presente documento, se refiere a cualquier cebador de nucleótidos disponible comercialmente o hecho en laboratorio que se puede usar para amplificar una secuencia de ácido nucleico diana a partir de una muestra biológica mediante PCR. Hay aproximadamente 1,5 millones de loci STR no CODIS. Los ejemplos no limitantes de lo anterior se presentan en el siguiente sitio web [www.cstl.nist.gov/biotech/strbase/str\\_ref.htm](http://www.cstl.nist.gov/biotech/strbase/str_ref.htm) que actualmente contiene 3156 referencias para los STR empleados en ciencias, medicina forense, etc. Además de las secuencias de cebadores publicadas, los cebadores STR pueden obtenerse a partir de kits comerciales para la amplificación de cientos de loci STR (por ejemplo, ABI Prism Linkage Mapping Set-MD10-Applied Biosystems), y para la amplificación de miles de loci SNP (por ejemplo, panel de mapeo de ligadura Illumina BeadArray). La expresión "cebadores STR del CODIS", como se usa en el presente documento, se refiere a cebadores STR que están diseñados para amplificar cualquiera de los trece loci STR de núcleo designados por el "sistema de índice combinado de ADN" del FBI, específicamente, las secuencias repetidas de TH01, TPOX, CSF1PO, VWA, FGA, D3S1358, D5S818, D7S820, D13S317, D16S539, D8S1179, D18S51, y D21S11, y el locus de Amelogenina.

20 La expresión "relación de señal", como se usa en el presente documento, se refiere a la relación entre las intensidades de las señales obtenidas a partir de la amplificación de dos loci.

### 3. Métodos para distinguir entre muestras de ADN naturales y artificiales.

También se describe un método para distinguir entre muestras de ADN naturales y artificiales. Un esquema general es como se indica a continuación: el método acepta como entrada una muestra de ADN. El ADN se somete a un procedimiento que incluye una o más etapas bioquímicas seguidas de la detección de señales. En la última etapa del procedimiento, la señal se analiza para determinar si el ADN es natural o artificial. En otro aspecto, la presente descripción proporciona un método para verificar que un perfil de ADN es de ADN natural. Un esquema general es como se indica a continuación: el método acepta como entrada una muestra de ADN que se sometió a análisis (por ejemplo, con Identificador). La muestra de ADN se somete a un procedimiento de verificación que incluye una o más etapas bioquímicas seguidas de detección de señal. En la última etapa de todo el procedimiento, se analizan los datos tanto del perfil como de la verificación de la muestra de ADN. El análisis de la señal determina si el perfil obtenido de la muestra de ADN representa el ADN natural (*in vivo*) o artificial (*in vitro*).

40 En diversos aspectos, los métodos de la presente descripción se refieren a la verificación de que los perfiles de ADN representan ADN natural. Los métodos se emplean en una muestra de ADN en cuestión, por ejemplo, el ADN de una muestra de sangre encontrada en la escena del crimen. El aislamiento de ácidos nucleicos (por ejemplo, ADN) de una muestra biológica se puede lograr mediante diversos métodos conocidos en la técnica (por ejemplo, véase Sambrook et al, (1989) *Molecular Cloning: A Laboratory Manual*, 2ª ed. Cold Spring Harbor, Nueva York). La distinción entre ADN natural y artificial, o la determinación de si un perfil de ADN representa el ADN natural, se puede lograr usando diversas estrategias, incluyendo las descritas en las siguientes secciones.

### 4. Metilación

50 La metilación en el genoma humano se produce en forma de 5-metilcitosina y se limita a residuos de citosina que son parte de la secuencia CG (los residuos de citosina que son parte de otras secuencias no están metilados).

Algunos dinucleótidos CG en el genoma humano están metilados, y otros no. Además, la metilación es específica de células y tejidos, de manera que un dinucleótido específico de CG puede metilarse en una determinada célula y al mismo tiempo no estar metilado en una célula diferente, o metilado en un determinado tejido y al mismo tiempo no metilado en diferentes tejidos. Dado que la metilación en un locus específico puede variar de célula a célula, al analizar el estado de metilación del ADN extraído de una pluralidad de células (por ejemplo, de una muestra forense), la señal puede mezclarse, mostrando las señales tanto metiladas como no metiladas en relaciones variables. Los inventores descubrieron que en algunas regiones genómicas todos los loci CG están metilados

- constitutivamente. Estas regiones se proporcionan en la Tabla 1 y en la sección en el presente documento titulada Secuencias. Los inventores también descubrieron que en algunas regiones genómicas todos los loci CG están parcialmente metilados. Estas regiones se proporcionan en la Tabla 1 y en la sección en el presente documento titulada Secuencias. Los inventores también descubrieron que en algunas regiones genómicas todos los loci CG están constitutivamente sin metilar. Estas regiones se proporcionan en la Tabla 1 y en la sección en el presente documento titulada Secuencias. Los inventores también descubrieron regiones genómicas contiguas que contenían loci CG constitutivamente metilados, parcialmente metilados y constitutivamente no metilados. Estas regiones se proporcionan en la Tabla 1 y en la sección en el presente documento titulada Secuencias. Existen varios métodos diferentes para determinar el nivel de metilación de los loci genómicos. Los ejemplos de métodos que se usan comúnmente son la secuenciación de bisulfito, la PCR específica de metilación, y la digestión con endonucleasa sensible a la metilación. Además, existen diversas fuentes de datos disponibles para recuperar o almacenar datos de metilación del ADN y hacer que estos datos estén fácilmente disponibles para el público, por ejemplo, MetDB (<http://www.methdb.net>).
- 15 Los métodos ejemplares para determinar el nivel de metilación de los ácidos nucleicos incluyen, pero sin limitación, los siguientes métodos:

- Secuenciación de bisulfito.* La secuenciación de bisulfito es la secuenciación de ADN tratado con bisulfito para determinar su patrón de metilación. El método se basa en el hecho de que el tratamiento del ADN con bisulfito sódico da como resultado la conversión de residuos de citosina no metilados en uracilo, sin dejar los residuos de citosina metilados sin afectar. Después de la conversión por bisulfito de sodio, las regiones específicas del ADN se amplifican por PCR, y los productos de PCR se secuencian. Dado que en la reacción en cadena de la polimerasa los residuos de uracilo se amplifican como si fueran residuos de timina, los residuos de citosina no metilados en el ADN original aparecen como residuos de timina en el producto de PCR secuenciado, mientras que los residuos de citosina metilados en el ADN original aparecen como residuos de citosina en la PCR secuenciada producto.

#### 5. PCR específica de metilación

- 30 La PCR específica de metilación es un método de análisis de metilación que, al igual que la secuenciación de bisulfito, también se realiza en ADN tratado con bisulfito, pero evita la necesidad de secuenciar la región genómica de interés. En cambio, la región seleccionada en el ADN tratado con bisulfito se amplifica mediante PCR usando dos conjuntos de cebadores que están diseñados para hibridar con las mismas dianas genómicas. Los pares de cebadores están diseñados para ser "específicos metilados" mediante la inclusión de secuencias que complementan solo las 5-metilcitosinas no convertidas o, por el contrario, "específicos no metilados", que complementan las timinas convertidas a partir de citosinas no metiladas. La metilación está determinada por la eficiencia relativa de los diferentes pares de cebadores para lograr la amplificación.

- Debe entenderse en el contexto de la presente invención que la PCR específica de metilación determina el nivel de metilación de dinucleótidos CG en las secuencias de cebadores solamente, y no en la región genómica completa que se amplifica mediante PCR. Por lo tanto, los dinucleótidos CG que se encuentran en la secuencia amplificada pero que no están en las secuencias del cebador no son parte del locus CG.

#### 6. Digestión de endonucleasa sensible a la metilación

- 45 La digestión de ADN con endonucleasas sensibles a la metilación representa un método para el análisis de metilación que se puede aplicar directamente al ADN genómico sin la necesidad de realizar la conversión de bisulfito. El método se basa en el hecho de que las endonucleasas sensibles a la metilación digieren solo el ADN no metilado, dejando intacto el ADN metilado. Después de la digestión, el ADN se puede analizar para determinar el nivel de metilación mediante una diversidad de métodos, incluyendo electroforesis en gel, y amplificación por PCR de loci específicos.

- En el procedimiento basado en la digestión con endonucleasa sensible a la metilación, cada locus CG está compuesto por uno o más dinucleótidos CG que son parte de la secuencia o secuencias de reconocimiento de la endonucleasa o endonucleasas de restricción sensible a la metilación que se usan en el procedimiento. Los dinucleótidos CG que se encuentran en la región genómica amplificada, pero que no están en la secuencia o secuencias de reconocimiento de la endonucleasa o endonucleasas no son parte del locus CG.

En una realización, el uno o más loci de CG que se detectan están parcialmente metilados en ADN natural, pero no

estarán metilados en ADN artificial. Se espera que la metilación parcial dé como resultado una mezcla de T y C en la posición que se está interrogando. La hibridación se observará tanto en las sondas/cebadores específicos de T como en las sondas/cebadores específicos de C, de forma similar a la detección de un SNP heterocigoto. Las cantidades relativas de hibridación se pueden usar para determinar la cantidad relativa de metilación. Como alternativa, tanto C como T se observarán tras la secuenciación de bisulfito. Como alternativa, pueden detectarse señales fluorescentes correspondientes a productos de amplificación de loci CG metilados o parcialmente metilados.

### 7. Descripción general del presente ensayo de la relación de señal

10 Como se ha mencionado anteriormente, un ensayo particular de la presente invención implica la comparación cuantitativa de la intensidad de las señales de un par de productos de amplificación específicos de locus producidos mediante la realización de una reacción en cadena de la polimerasa en el ADN digerido por restricción. Véase la figura 1A,B. La relación numérica de intensidades corresponde a la categoría de la plantilla de ADN, por ejemplo, si la muestra de ADN se extrajo del semen o de un tejido que no sea semen. Por ejemplo, en una realización, el locus 15 1 y el locus 2 pueden amplificarse usando cebadores marcados por fluorescencia, separados por electroforesis, y la intensidad de las señales son las unidades de fluorescencia relativa (rfu) de los picos correspondientes a los loci. Véase la figura 1B,C. La intensidad de las señales corresponderá al éxito de la amplificación del locus 1 y el locus 2 de la plantilla de ADN de origen. Al comparar rfu entre los dos productos de amplificación, se puede calcular una relación de señal, una relación que refleja la relación de las cantidades de los productos de amplificación.

20 Además, sin embargo, un aspecto de este ensayo incluye la predeterminación de las relaciones de señal esperadas de diversos tipos de categorías de ADN. Por lo tanto, el ADN molde que está sujeto a análisis se digiere en primer lugar con una endonucleasa de restricción sensible a la metilación antes de que se cicle a través del protocolo de amplificación por PCR de estilo cuantitativo. Si se conoce el nivel de metilación en el sitio/s de la secuencia/s de reconocimiento de restricción en el par de los dos loci en las diversas categorías de ADN, el experto puede predecir las relaciones de señal esperadas de los productos de amplificación del par de loci en las respectivas categorías. Como alternativa, si los niveles de metilación son desconocidos, el experto puede, sin embargo, determinar empíricamente las relaciones de señal entre un par de loci genómicos en las diversas categorías de ADN realizando el ensayo en muestras de categorías conocidas.

30 Con esta premisa, los ensayos de la presente invención comprenden digerir una muestra de ADN con una enzima sensible a la metilación y/o dependiente la metilación, realizar una reacción de amplificación por PCR sobre el ADN digerido, y determinar la intensidad de las señales procedentes de productos de amplificación específicos del locus. Como se ha mencionado, la intensidad de las señales se puede cuantificar o medir mediante el uso de PCR 35 fluorescente. La determinación de la categoría de la muestra de ADN se realiza comparando las relaciones de señal observadas con respecto a las relaciones de referencia obtenidas de muestras de ADN de categorías conocidas, y determinando qué categoría corresponde mejor a las relaciones observadas.

Este ensayo de relación de señal particular no depende de la identificación u obtención de mediciones de la fracción de metilación absoluta o el nivel de los loci seleccionados. Además, este ensayo de relación de señal particular no depende de las eficiencias de los pares de cebadores usados, no requiere que ambos pares de cebadores tengan eficiencias similares, no depende de la cantidad de ADN molde de entrada, no depende de la máquina de termociclador específica y las condiciones de reacción. En su lugar, el ensayo determina la relación entre dos 45 señales que corresponden a la relación de niveles de metilación en los diferentes loci. De esta manera, la cantidad o concentración de material de ADN de partida en la muestra es irrelevante para el análisis y no distorsiona los resultados de salida. Es decir, la relación de los niveles de señal entre un primer locus y un segundo locus se mantendrá constante, independientemente de la cantidad de ADN que se use como plantilla para la PCR e independientemente del número de ciclos de amplificación que se ejecuten en el termociclador. Por ejemplo, una relación de señal de 10 entre los loci 1 y 2 seguirá siendo la misma si el ADN de entrada representa niveles de 50 metilación del 0,9 y el 0,09 (metilación del 90 % en el locus 1 y el 9 % en 2), o el 0,5 y el 0,05 (metilación del 50 % en el locus 1 y el 5 % en 2), etc. En otro ejemplo, una relación de señal de 10 entre los loci 1 y 2 seguirá siendo la misma independientemente de la cantidad de plantilla de ADN utilizada en la reacción.

El hecho de que el ensayo de la invención no dependa de la determinación de los niveles de metilación en ningún locus genómico, y no requiera ningún control externo, permite que el ensayo sea más preciso, rentable y aplicable a 55 muestras con cantidades bajas de ADN.

En el contexto de la identificación de tejido forense, el ensayo de la invención representa un procedimiento estandarizado que es idéntico independientemente de la tarea específica de identificación de tejido que se está

realizando (excepto para el uso de cebadores específicos de panel), que requiere solamente procedimientos forenses estándar y equipos, lo que permite de este modo que el ensayo se automatice y se integre fácilmente en los laboratorios forenses.

5 Además, la configuración del ensayo de la invención hace posible combinar la identificación del tejido y el perfil de ADN en un único ensayo.

El hecho de que el ensayo de la invención realice la digestión y la amplificación en la misma reacción minimiza la complejidad (el número de tubos requeridos y las etapas de pipeteo del ensayo), minimizando así el riesgo de  
10 contaminación y/o error.

Por lo tanto, este ensayo de la invención proporciona un marcador bioquímico útil en forma de, en un ejemplo, una relación numérica, que se puede usar para diferenciar entre ADN natural y ADN artificial que se sintetizó *in vitro* tal como mediante PCR, síntesis química, o amplificación de desplazamiento múltiple (MDA). A continuación se  
15 encuentran más detalles particulares de este ensayo ejemplar.

### 8. Características de los cebadores para la amplificación específica de locus

Por consiguiente, un aspecto de la presente invención se refiere a la obtención de una "relación de señal" en la que  
20 las intensidades de fluorescencia de múltiples productos de amplificación de loci de ADN producidos a partir de PCR fluorescentes se comparan entre sí para calcular relaciones entre pares de loci sucesivos, por ejemplo, Loci N.º 1 frente a Loci N.º 2; Loci N.º 1 frente a Loci N.º 3; Loci N.º 1 frente a Loci N.º 4; Loci N.º 2 frente a Loci N.º 3, Loci N.º 2 frente a Loci N.º 4, etc. Cuando esta técnica se usa para ensayar la naturaleza natural frente a artificial de una muestra de ADN, los cebadores que se usan en las reacciones de amplificación de la relación de señal se escogen  
25 para amplificar un par de loci que están metilados de forma diferente en el ADN natural (por ejemplo, el primero está metilado al 90 % y el segundo está metilado al 30 % en ADN natural).

Una consideración para seleccionar qué dos pares de cebadores (un primer par y un segundo par) usar para amplificar dos loci (1) y (2) es el grado en el que los dos loci están metilados de forma diferente en el ADN *in vivo*.  
30 Por lo tanto, un par de loci en el que el locus (1) está más metilado en comparación con el locus (2), son ejemplos de loci que se pueden usar para diseñar cebadores para el ensayo de amplificación de la relación señal.

### 9. Selección de loci genómicos

35 1. La selección de los loci putativos se puede realizar en cualquiera de los tres modos:

(i) Seleccionar un locus más cercano al sitio de inicio de la transcripción de un gen que se ha mostrado en la bibliografía para expresarse de forma diferente (datos de ARNm) en las categorías  
40 relevantes (por ejemplo, diferentes tejidos).

(ii) Seleccionar un locus que se ha demostrado en la bibliografía que está metilado de forma diferente entre diferentes categorías (por ejemplo, tejidos).

(iii) Seleccionar un locus genómico aleatorio.

2. Diseño de cebadores para loci putativos.

45 3. Ensayar empíricamente la "capacidad informativa" de los loci putativos. Para cada locus putativo se calcula la relación de señal entre el locus putativo y otro locus putativo, o como alternativa, entre el locus putativo y un locus de control no digerido.

4. Escoger loci "informativos" (es decir, loci cuyas relaciones de señal son significativamente diferentes en las diferentes categorías de ADN) para usar en el ensayo.

50

### 10. Endonucleasas de restricción sensibles a la metilación

Una segunda consideración del presente método de la invención es la selección de loci que se cortan o se digieren mediante una endonucleasa de restricción sensible a la metilación y/o dependiente de la metilación, dependiendo de  
55 su estado de metilación. La endonucleasa se selecciona si, por ejemplo, no puede cortar la cadena de ADN si su secuencia de reconocimiento en ese lugar está metilada. Por lo tanto, en el contexto del locus (1), que está metilado, y el locus (2), que no está metilado, una endonucleasa como HhaI o HpaII no digerirá el locus (1) sino que digerirá el locus (2). Por consiguiente, la selección de loci para la amplificación en el ensayo de relación de señal también puede tener en cuenta la presencia de secuencias de reconocimiento de la endonucleasa de restricción sensibles a

la metilación dentro de cada locus.

Los cebadores directo e inverso pueden diseñarse entonces para hibridar con una región de ADN que flanquea la secuencia de reconocimiento de los loci.

5

Por consiguiente, en el caso de una enzima sensible a la metilación, si un locus está metilado, no se digierá (A) sino que (B) se amplificará. Por el contrario, si un locus no está metilado, se digierá (A) pero (B) no se amplificará. En el caso de una enzima dependiente de la metilación, la situación es a la inversa.

## 10 11. Ventajas del ensayo de relación de señal

El ensayo de la relación de señal de la presente invención tiene varias ventajas y beneficios sobre otros enfoques para analizar la metilación y, por lo tanto, el experto apreciaría los diversos usos ventajosos a los que se puede poner este ensayo de relación de señal. Por ejemplo, este ensayo es insensible a diversos factores de "ruido" inherentes cuando se basa en el nivel de intensidad absoluto de una señal de amplicón que fluctúa como consecuencia de los cambios en la concentración de ADN molde, el fabricante del termociclador, y las condiciones de PCR, y la presencia de inhibidores. En cambio, la relación de metilación calculada actualmente no depende de tales factores, ya que los dos loci analizados están co-amplificados en la misma reacción y, por lo tanto, están sujetos a los efectos de tales disparidades. Por lo tanto, el presente método inventivo no requiere una cuantificación absoluta de dianas genómicas o amplicones; y el ensayo es una única reacción independiente sin necesidad de una curva estándar o ningún control externo.

La relación de señal se puede realizar en cantidades muy pequeñas de ADN en una única reacción bioquímica y, por lo tanto, es un método barato, rápido y potente para establecer, por ejemplo, si una muestra contiene ADN natural o artificial. Una característica importante del diseño de los métodos inventivos es que se puede combinar con otros procedimientos basados en PCR, tales como el perfil de ADN, en una única reacción bioquímica.

## 12. Electroforesis capilar y multiplexación

La rapidez del análisis es evidente en consideración del uso de, por ejemplo, electroforesis capilar para separar numerosos productos de amplificación producidos a partir de la amplificación de múltiples pares de loci diana. Como se ha descrito anteriormente, el presente ensayo de relación de señal se puede realizar en múltiples loci, y en cada caso se calcula una relación de señal para cada par de loci por separado. Por ejemplo, si cuatro loci (A, B, C, D) están coamplificados en la reacción, se pueden calcular seis relaciones de señal diferentes, es decir: A/B, A/C, A/D, B/C, B/D, C/D.

Por consiguiente, si los "n" loci están co-amplificados, entonces pueden calcularse  $(n^2-n)/2$  relaciones diferentes. Por lo tanto, la cantidad de información que se proporciona por el presente ensayo de metilación aumenta exponencialmente con el número de loci analizados. La electroforesis capilar, a diferencia de los métodos de amplificación por PCR en tiempo real, puede distinguir entre una gran cantidad de loci en una sola ejecución. Por ejemplo, para el perfil de ADN, 17 loci genómicos se co-amplifican de forma rutinaria a partir de una muestra de ADN particular, y se analizan juntos. Como consecuencia, el rendimiento del presente ensayo de la relación de señal en todos los 17 loci produce 136 relaciones de señal independientes. La PCR en tiempo real no puede distinguir simultáneamente en una sola reacción los números de productos de amplificación discretos necesarios para producir 136 relaciones. Aproximadamente cuatro loci pueden distinguirse por PCR en tiempo real, lo que corresponde al cálculo de solamente seis relaciones.

Por el contrario, la electroforesis capilar puede separar fácilmente los productos de amplificación de todas las permutaciones pareadas de 17 loci y, por lo tanto, puede producir fácilmente datos para calcular simultáneamente las 136 relaciones de señal en una única reacción. En teoría, pueden ejecutarse juntos cientos de loci y separarse en una sola ejecución de electroforesis capilar. Para facilitar el análisis de relaciones múltiples, puede ser ventajoso calcular una puntuación única, denominada en el presente documento "puntuación combinada de señal" (CSS), a partir de estas relaciones. El CSS se puede utilizar, por ejemplo, para determinar si la muestra de ADN de origen es natural o artificial.

55

## 13. Secuencias

Las secuencias proporcionadas en el presente documento para los diversos CODIS, PowerPlex® 16, y otros loci usados comúnmente para los perfiles, es decir, las SEQ ID NOs. 1-25 son de loci genómicos usados comúnmente

para el perfil de STR, se han analizado en el presente documento para determinar (1) la posición de cada dinucleótido de citosina-guanina (CG), (2) el nivel de metilación de todos los dinucleótidos CG en el locus en ADN natural, y (3) el perfil de enzima de restricción dependiente de metilación y sensible a la metilación para ese locus particular. La lista de secuencias incluida en el texto de esta solicitud proporciona, por lo tanto, una guía para el experto en la identificación de endonucleasas de restricción dependientes de la metilación y sensibles a la metilación que pueden usarse de acuerdo con los métodos de ensayo generadores de la relación de la invención. Por ejemplo, la SEQ ID NO. 8, que es la secuencia de nucleótidos de 751 de longitud para el locus del CODIS "vWA" y el locus amplificado por PowerPlex® 16, comprende solamente un dinucleótido CG en la posición 571. Ese CG y los nucleótidos que lo flanquean inmediatamente, proporcionan la secuencia de reconocimiento para las enzimas BssKI, HpaII, Nt.CviPII, y las SEQ ID NOs 26-32 son loci genómicos que los inventores han encontrado para producir relaciones de señal con valores significativamente diferentes entre muestras de ADN natural y artificial. Las SEQ ID NOs 33-54 son loci genómicos que los inventores han encontrado para producir relaciones de señal con valores significativamente diferentes (basados en la prueba de Kolmogorov-Smirnov usando un umbral de  $p = 0,05$ ) entre muestras de ADN extraídas de diferentes tejidos.

La información de secuencia proporcionada en el presente documento también permite que el experto en la técnica diseñe cebadores de amplificación directos e inversos que se hibriden con regiones de un locus seleccionado que flanquean el CG y el sitio de restricción. Por lo tanto, la presente invención no se limita a la amplificación, por ejemplo, de los loci del CODIS, usando solamente los cebadores disponibles comercialmente, aunque el uso y la disponibilidad de cebadores disponibles comercialmente pueden ser una opción más conveniente y rentable para realizar los presentes ensayos de autenticación de la invención.

#### 14. Algoritmo y software

En una realización, el cálculo de las relaciones de señal y/o de representación (por ejemplo, para determinar si un perfil de ADN representa el ADN natural) se realiza en base al análisis de las intensidades de señal de los productos de amplificación de la PCR fluorescente que se ejecutan en una máquina de electroforesis capilar. La salida de la máquina de electroforesis capilar es un archivo informático binario (por ejemplo, un archivo FSA en el caso de las máquinas de electroforesis capilar de Applied Biosystems). Este archivo incluye información sobre la ejecución de la electroforesis capilar, incluyendo los datos de canal, que son las unidades fluorescentes relativas (rfu) de cada fluoróforo en función de cada punto de tiempo de muestreo (llamado punto de datos).

La presente descripción también describe un programa de software que acepta como entrada un archivo que es la salida de una máquina de electroforesis capilar ejecutada, y calcula las intensidades de fluorescencia de un conjunto de loci cuyos productos de amplificación se ejecutaron en la máquina de electroforesis capilar. En base a estas intensidades, el software calcula las relaciones de señal y/o las relaciones de representación (o puntuaciones combinadas basadas en estas relaciones), en base a un conjunto de pares de loci predefinidos para los cuales se definen las relaciones a calcular. Finalmente, el software produce una decisión basada en la comparación de las relaciones calculadas con respecto a los umbrales predefinidos.

A continuación se muestra un esquema de este análisis realizado por el programa de software:

1. Restar de la señal de cada canal de tinte el nivel inicial, definido como el nivel medio de rfu de ese canal en el intervalo de 400-500 pb, y eliminar la superposición espectral entre los fluoróforos basados en la matriz obtenida del procedimiento de calibración espectral de la máquina de electroforesis capilar.
2. Obtener el nivel de rfu de cada lugar analizado, definido como el nivel máximo de rfu en el intervalo de -0,5 pb a +0,5 pb del tamaño esperado del locus.
3. Calcular las relaciones de señal entre un conjunto de pares de loci, definidos como la relación entre los niveles de rfu del primer y segundo loci en el par (para evitar la división por cero, asignar un nivel de 50 rfu a todos los loci con niveles de rfu inferiores a 50).
4. Para cada relación calculada en la etapa 3 y para cada tipo de tejido, calcular una puntuación de probabilidad, definida como el valor de la función de probabilidad de referencia (véase a continuación) en la relación observada.
5. Para cada tipo de tejido, calcular una puntuación de probabilidad combinada, definida como el producto de las puntuaciones de probabilidad por separado para ese tejido obtenido en la etapa 4 (donde  $n$  es el número de relaciones de señal).
6. Para cada tipo de tejido, calcular una puntuación de probabilidad (que representa la probabilidad de que la muestra de ADN se originase a partir de ese tejido), definida como la relación entre la puntuación combinada de probabilidad del tejido y la suma de las puntuaciones combinadas de probabilidad de todos

los tejidos. El tejido con la mayor puntuación de probabilidad se considera el tejido de origen de la muestra.

#### 15. Funciones de probabilidad de referencia

- 5 La función de probabilidad de referencia de un cierto par de loci en un cierto tejido es la función de distribución gamma ajustada al conjunto de relaciones de señal observadas en muestras de ADN natural de ese tipo de tejido.

#### 16. Corrección para la digestión incompleta

- 10 Debido a que la digestión incompleta del ADN molde puede ocurrir como resultado de la presencia de inhibidores en la muestra, es útil incorporar en el ensayo un análisis del nivel real de digestión. Esto puede conseguirse mediante la incorporación de al menos un "locus de control digerido" y al menos un "locus de control no digerido". Un locus de control digerido es un locus genómico que se sabe que no está metilado en el ADN natural en todos los tejidos potenciales. Un locus de control no digerido es un locus que se sabe que está metilado en el ADN natural en todos los tejidos potenciales, o es un locus que carece de la secuencia de reconocimiento de la endonucleasa utilizada en el ensayo. La relación de señal entre un locus de control digerido y un locus de control no digerido se correlaciona con el nivel real de digestión en el ensayo. El algoritmo puede aprovechar tales loci comprobando si se produjo una digestión adecuada (y de no ser así, abortando el análisis), y modificando los niveles de rfu reales de loci a valores que representan los valores de rfu que se habrían obtenido si se hubiera producido una digestión completa. Si se realiza esta modificación, entonces las relaciones de señal se calculan a partir de los valores de rfu modificados.

#### Ejemplos

#### MATERIALES Y MÉTODOS

25

##### A. Recogida de tejidos biológicos

Se recogieron de voluntarios sangre (venosa/menstrual), saliva, semen, epidermis de la piel, orina y secreción vaginal. Se obtuvo el consentimiento informado de todos los participantes reclutados en el estudio.

30

##### B. Extracción y cuantificación de ADN

- El ADN se extrajo de todas las muestras usando extracción orgánica. La cantidad de ADN recuperado se determinó utilizando el kit de cuantificación de ADN humano Quantifiler® (Applied Biosystems) y el sistema de PCR en tiempo real 7300 (Applied Biosystems).

35

##### C. Selección de loci genéticos y diseño de imprimación

- Se buscaron islas CpG aleatorias (definidas como una región  $\geq 200$  pb y con un contenido de GC  $\geq 0,5$ , y con al menos 8 CG, correspondientes a la relación CpG observada/esperada  $>0,6$ ) utilizando un programa de software desarrollado para esta tarea. De estos, se seleccionaron islas CpG que contenían una secuencia de reconocimiento HhaI para el cribado inicial. Los cebadores (18-28 pb) que flanquean la secuencia de reconocimiento de HhaI se diseñaron con una  $T_m$  de 64-66 °C y para tamaños de amplicón de 66-150 pb. Se adquirieron los cebadores directos marcados por fluorescencia (FAM o JOE) de Integrated DNA Technologies. Los loci se seleccionaron por pares para la metilación diferencial mediante la comparación de las señales obtenidas a partir de la amplificación de muestras de ADN agrupadas (conteniendo cada una ADN de 10 individuos) de cada tipo de tejido. Los loci que mostraron patrones de amplificación diferenciales significativos (definidos como la relación de señal más grande observada en todos los tipos de tejido que fueron mayores de tres veces la relación más pequeña observada) se seleccionaron para la multiplexación. En total, se cribaron 205 loci genómicos de los cuales 38 mostraron patrones de amplificación diferencial significativos. De estos, se usaron 16 loci en los experimentos. Para el ensayo de identificación de tejido independiente, se usaron 15 loci que pudieron coamplificarse sin ruido significativo. Para la detección combinada de semen y el ensayo de perfil, se utilizaron dos loci (L68346 que también se usó para el ensayo independiente y L16264).

50

##### 55 D. Digestión por endonucleasa, PCR y electroforesis capilar

En experimentos de ensayo independientes, cada muestra de ADN se digirió por HhaI y se sometió a amplificación por PCR en la misma reacción, que consistía en 30 U de HhaI (New England Biolabs), 2,5 U de AmpliTaq Gold (Applied Biosystems), 2,5  $\mu$ g de BSA (New England Biolabs), cada dNTP 0,2 mM, cada cebador 0,1-0,3  $\mu$ M, 2,5  $\mu$ l

de tampón de reacción (Tris-HCl 150 mM, MgCl<sub>2</sub> 15 mM) y DDW hasta un volumen total de 25 µl. La composición de reacción en los experimentos combinados de identificación-análisis de tejido fue la misma que en el ensayo independiente, excepto para los cebadores, que fueron 0,2 µM de cada cebador de identificación de semen junto con 5 µl de Primer Set de ProfilerPlus (Applied Biosystems). Las reacciones (tanto para el ensayo independiente como para el combinado) se realizaron en un sistema GeneAmp® PCR 9700 (Applied Biosystems), y el programa de termociclado utilizado fue: 37 °C durante 15 min, 95 °C durante 11 min, seguido de 28 ciclos de 94 °C durante 1 min, 59 °C durante 1 min, 72 °C durante 1 min, y seguido de una etapa de extensión final de 60 °C durante 45 min. Una mezcla que contenía 1,5 µl de productos de amplificación, se desnaturalizaron (95 °C durante 3 min inmediatamente seguido de 3 min sobre hielo) 24,5 µl de HiDi formamida (Applied Biosystems) y 0,5 µl de tamaño estándar de ROX (Applied Biosystems), y se realizó en un analizador ABI 310 Genetic Analyzer (Applied Biosystems) de acuerdo con las instrucciones del fabricante. Todos los electroferogramas resultantes se analizaron usando software desarrollado internamente, y los electroferogramas resultantes de los experimentos combinados de identificación-análisis de tejido también se analizaron usando el software de análisis GeneMapper ID-X (Applied Biosystems).

#### 15 E. Ensayo de identificación de tejidos - Análisis de datos

El análisis de datos para la identificación del tejido se realizó mediante un software desarrollado internamente que acepta como entrada archivos fsa y saca el tejido de origen más probable. La derivación del porcentaje de semen en las muestras de mezcla se realizó en base a los loci de análisis y los loci de identificación tisular. La derivación basada en los perfiles de loci se realizó dividiendo la suma de los valores de rfu de los alelos correspondientes al contribuyente del semen por la suma total de los valores de rfu de todos los alelos en los loci STR que tenían 4 alelos genotipados. La derivación basada en los loci de identificación tisular se realizó en primer lugar calculando la relación de señal esperada para todos los porcentajes posibles de mezclas (del 0 al 100 % con aumentos del 1 %, en base a la relación de señal de ADN no digerido), y el algoritmo proporcionó el porcentaje de semen para el que la relación de señal esperada es la más cercana a la relación de señal observada.

#### F. Simulación de muestras degradadas

Se digirieron veinte nanogramos de cada muestra durante 10 min a 37 °C con 0,01 U de DNaseI (Ambion), 2 µl de tampón 10 x DNaseI (Ambion) en un volumen de reacción total de 20 µl seguido de inactivación por calor a 75 °C durante 10 min. Un nanogramo de cada muestra de ADN digerido se sometió entonces al ensayo combinado de identificación-análisis tisular.

#### Ejemplo 1

#### 35 Categorización de ADN de acuerdo con el ensayo independiente del tipo de tejido

En este ejemplo, el procedimiento de categorización se denomina "ensayo de identificación tisular". El ensayo de identificación tisular utiliza un panel de loci que están metilados de forma diferente entre los tejidos para determinar el tejido de origen más probable de una muestra de ADN como se muestra en la Tabla 1 (figura 7). Un esquema del ensayo se presenta en la figura 1A. Un nanogramo de ADN de una muestra forense en cuestión se digiere con la enzima de restricción sensible a la metilación HhaI, que escinde ADN en su secuencia de reconocimiento GCGC solamente si no está metilado (dejando intactas las dianas metiladas). A continuación, se amplifica un panel de loci de identificación de tejido mediante PCR a partir del ADN digerido usando cebadores marcados por fluorescencia, y una alícuota de productos amplificados se separa mediante electroforesis capilar. Los loci con niveles de metilación más altos se amplifican con mayor eficiencia porque se protegen más moléculas de ADN de la digestión, produciendo una señal relativamente fuerte en el electroferograma (figura 1B, locus A). Por el contrario, los loci con un nivel de metilación inferior se amplifican con menor eficiencia, produciendo una señal relativamente débil en el electroferograma (figura 1B, locus B).

El software de análisis de señal automatizado permite el análisis de un electroferograma de salida, y asigna alturas (en rfu) a los amplicones correspondientes a los loci de identificación de tejido. Aunque la altura de un solo locus está correlacionada con su nivel de metilación, el nivel real de metilación no puede derivarse, ya que ese nivel también depende de la concentración precisa de la plantilla y de las condiciones específicas de la PCR. En cambio, las relaciones de los niveles de metilación entre los loci co-amplificados pueden calcularse, ya que todos los loci se someten a la misma concentración de plantilla y a las condiciones de reacción específicas. Por lo tanto, para cada par de loci se calcula una relación de señal, definida como la relación entre las alturas del primer y segundo loci, y esta relación refleja la relación entre los niveles de metilación de los loci correspondientes. Todas las relaciones de señal calculadas se combinan en un único vector (si, por ejemplo, en un panel de 15 loci, el vector contiene 105

relaciones, entre el locus 1 y 2, el locus 1 y 3, etc.), que después se compara con una base de datos de vectores de referencia obtenidos de un conjunto de datos de muestras de origen tisular conocido. El algoritmo de identificación tisular calcula para cada fuente de tejido potencial una puntuación de probabilidad, lo que refleja la probabilidad de que la muestra de ADN se originase a partir de ese tejido, y el resultado del algoritmo es el tejido más probable.

5

Se realizó un ensayo de identificación de tejido independiente en 50 muestras: 14 de sangre, 14 de saliva, 11 de semen y 11 epidermis usando un panel de 15 loci de identificación tisular, que varían en tamaño de amplicón de 66 a 150 pb. La figura 2 representa electroferogramas de ocho de estas muestras, dos de cada tipo de tejido. Aunque las muestras del mismo tipo de tejido de diferentes individuos no eran completamente idénticas (debido a la variabilidad natural de los niveles de metilación y los efectos de la PCR estocástica), fue evidente que cada tipo de tejido tenía un perfil de metilación distinto. Por ejemplo, la relación de L91762/L68346 en muestras de semen varío de 0,04-0,53 y fue mayor en todas las demás muestras de tejido (2,15-18,28; Tabla 2 en la figura 8) y, por lo tanto, una baja relación de L91762/L68346 fue distintiva de las muestras de semen. La relación de L76138/L26688 fue baja en sangre y saliva (0,08-1,54) y mayor en semen y piel epidérmica (2,04-19,56; Tabla 2 en la figura 8); por lo tanto, una relación alta de L91762/L68346 concomitante con una alta relación de L76138/L26688 fue distintiva de las muestras de epidermis de la piel. El algoritmo de identificación tisular identificó correctamente la verdadera fuente de tejido de las 50 muestras. El análisis adicional de los datos utilizando subconjuntos de loci de identificación de tejido reveló que el conjunto completo de 15 loci fue redundante y se logró una identificación del 100 % utilizando solamente 7 de los 15 loci (L91762, L68346, L50468, L14432, L30139, L15952, y L26688). Debido a que la identificación del tejido se basa en una digestión eficiente, se evaluó la digestión completa del ADN molde por HhaI en una única configuración de reacción de digestión-amplificación. Las 50 muestras de ADN se sometieron al mismo procedimiento pero usando un panel de dos loci diferentes como se muestra en la Tabla 3 (figura 9) - un locus de control sin un sitio HhaI (L98328) y un locus que previamente se demostró que estaba completamente sin metilar en estos tejidos (SW14; véase D. Frumkin et al., Authentication of forensic DNA samples, Forensic Sci. Int. Genet. 2009, doi: 10.1016/j.fsigen.2009.06.009, que se incorpora en el presente documento por referencia). Todas las muestras mostraron la presencia del locus de referencia y la ausencia completa de SW14, lo que indica la digestión completa por HhaI en contraste con la muestra no digerida que mostró la presencia de ambos loci (se muestran realizaciones ejemplares en la figura 4).

## 30 Ejemplo 2

### **Análisis de los factores de ruido que afectan a las relaciones de señal**

El ensayo descrito en el ejemplo 1 se basa en la utilización de loci que están metilados de forma diferente entre tejidos, para los cuales los valores de la relación señal son generalmente específicos de tejido (figura 1C). Sin embargo, incluso en los tejidos metilados de forma diferente, podría existir algún solapamiento en las relaciones de señal observadas como resultado de la variabilidad natural en los niveles de metilación entre diferentes individuos y/o artefactos asociados con la PCR, tales como diferencias en la concentración de plantilla de ADN (por ejemplo, como resultado de errores de pipeteo y efectos estocásticos). El efecto de cada uno de estos factores de "ruido" se analizó en relación con la "señal", que es la diferencia de relación de señal media entre los tejidos. Se seleccionaron dos loci con una diferencia promedio de 10 veces en los valores de relación señal entre el semen y la sangre. Se realizaron diferentes PCR en la misma muestra de semen usando 1 ng y 2 ng de ADN como plantilla. Las relaciones obtenidas a partir de estas amplificaciones se compararon entre sí y con las relaciones obtenidas a partir de la amplificación de diferentes muestras de sangre. Las diferencias entre las relaciones obtenidas en diferentes PCR y el uso de diferentes cantidades de ADN de entrada fueron más de un orden de magnitud menor que las diferencias entre las relaciones obtenidas de los diferentes tejidos (figura 1D).

## Ejemplo 3

### 50 Detección combinada de semen y perfil de ADN

En este ejemplo, la categorización del ADN se encuentra en una de dos categorías: semen y no seminal, donde no seminal se refiere a todos los tejidos, excepto el semen. La categorización se realiza junto con el perfil de ADN en la misma reacción.

55

El procedimiento de identificación de tejido utiliza las mismas plataformas utilizadas en el perfil de STR estándar (es decir, los instrumentos de electroforesis capilar y termociclador) y, por lo tanto, es fácilmente susceptible de integración con el perfil de ADN en una única reacción. Esta capacidad se demostró mediante la generación de un perfil de STR integrado y un ensayo de detección de semen. Una PCR del kit de perfil de STR (Profiler Plus) se

complementó con cebadores específicos para dos loci de identificación de semen (Tabla 4 en la figura 10). Estos loci consisten en un amplicón de 70 pb que se amplifica de manera eficiente solamente en el semen y un amplicón de 95 pb que se amplifica de manera eficiente en todos los tejidos excepto en el semen. El ensayo se ensayó en muestras puras de semen, orina (macho), sangre venosa, sangre menstrual, secreción vaginal y saliva. Se observó el patrón correcto de la amplificación de la identificación del semen y el algoritmo de identificación tisular identificó correctamente la presencia/ausencia de semen en todas las muestras. Las figuras 3A y B muestran electroferogramas completos de muestras de semen y orina (obtenidas del mismo individuo), y la figura 3C representa los datos del canal FAM de todas las muestras, incluidos los dos loci de identificación de semen y tres loci de perfil. La relación de señal observada fue de 25,04 en la muestra de semen, y 0,04-0,089 en las otras muestras. El ensayo combinado también se ensayó en mezclas de semen y saliva a varias relaciones (figura 3D). Para cada una de estas mezclas, el porcentaje del semen en la muestra se derivó de la relación de señal de los loci de identificación del semen y se comparó con el porcentaje correspondiente derivado de los loci de análisis (las muestras de semen y saliva se obtuvieron de diferentes individuos y, por lo tanto, podían diferenciarse en base a los loci STR). Los porcentajes derivados fueron comparables, con una diferencia máxima del 10 % (Tabla 5 en la figura 11). Además, la presencia frente a la ausencia de semen se determinó correctamente en todas las muestras (incluyendo una muestra que contenía solo el 13 % de semen; figura 3D). Los perfiles de STR obtenidos a partir de las muestras utilizando el ensayo integrado de análisis-detección de semen fueron idénticos a los perfiles obtenidos de las mismas muestras usando ProfilerPlus (sin detección de semen).

## 20 Ejemplo 4

### Categorización de muestras de ADN maduras y degradadas

El ensayo de identificación de tejido independiente (descrito en el ejemplo 1) se ensayó en dos muestras de ADN extraídas de manchas de 20 meses de antigüedad (sangre y semen en algodón) como prueba de concepto para la aplicación a muestras de tipo forense. Ambas muestras se amplificaron bien, de manera similar a las muestras frescas (figura 5), y el algoritmo identificó correctamente la fuente de tejido de ambas muestras. Con el fin de simular muestras de ADN degradadas, dos muestras adicionales de ADN (sangre y semen) se sometieron a digestión con DNaseI en condiciones que producen una digestión parcial. Estas muestras se analizaron después con la detección combinada de semen y el ensayo de perfiles. La digestión de DNaseI dio como resultado la desviación de los alelos en los STR más grandes en ambas muestras de modo que sus perfiles fueron parciales. En la muestra de sangre solamente se tipificaron los alelos D3S1358 y amelogenina, y en la muestra de semen se observó la desviación de alelos en FGA, D21S11, D18S51, D13S317 y D7S820 (figura 6). Los loci de identificación de tejido más pequeños se amplificaron con éxito con el patrón esperado, y el algoritmo identificó correctamente la presencia/ausencia de semen en estas muestras.

## Ejemplo 5

### Perfiles combinados y categorización de ADN

Este ensayo particular emplea tanto el perfil convencional de una muestra de ADN (con el kit Promega PowerPlex 16), como la categorización del ADN en categorías de natural frente a artificial.

Por lo tanto, un aspecto de este ensayo es verificar la naturaleza natural del ADN contra varios o todos los tipos de ADN artificial (véase la sección Definiciones) y, por lo tanto, contiene un análisis de varias relaciones diferentes. La combinación de relaciones hace posible, por lo tanto, comparar un único parámetro de diversos ADN sintetizados a partir de diversos métodos. Un único parámetro, por ejemplo, la relación de representación, puede distinguir entre una muestra que contiene el genoma completo (ADN natural pero también el ADN sintetizado MDA, como en este ejemplo) y una muestra que contiene solamente un subconjunto (por ejemplo, solamente los loci del CODIS).

Se analizaron dos muestras de ADN, una muestra de ADN natural extraída de la sangre del sujeto A, y una muestra de ADN artificial sintetizada mediante amplificación de desplazamiento múltiple *in vitro* usando un kit comercial (a partir de una cantidad mínima del ADN real del sujeto A como plantilla).

Las dos muestras se analizaron en primer lugar con un kit convencional (PowerPlex16) y se analizaron usando un software de análisis convencional (GeneMapperID-X). El esquema de este procedimiento se puede describir brevemente como realizar PCR múltiple utilizando la mezcla de cebadores PowerPlex16 en una muestra de ADN; separar los productos de amplificación en una máquina de electroforesis capilar; y después analizar los datos de salida con, por ejemplo, el software GeneMapperID-X.

Para cada muestra, el producto final de este ensayo es un perfil. Los perfiles de las dos muestras son idénticos, como se puede ver en la figura 9A,B. Además, el software GeneMapperID-X determinó que los perfiles de ambas muestras eran perfiles de un único contribuyente (es decir, de una única persona), sin anomalías. Esto muestra que un perfil obtenido a partir de ADN artificial puede ser idéntico a un perfil obtenido a partir de ADN natural. Véase D. Frumkin, et al., Authentication of forensic DNA samples, Forensic Sci. Int. Genet. (2009), doi:10.1016/j.fsigen.2009.06.009, que se incorpora en el presente documento por referencia.

Posteriormente, ambas muestras de ADN también se analizaron mediante un ensayo de análisis y categorización combinado basado en el siguiente esquema: digerir una muestra de ADN con HhaI y realizar una PCR múltiple usando la mezcla de cebadores PowerPlex16 con la adición de cebadores para amplificar un locus adicional, tal como Hypo23; separar los productos de amplificación en una máquina de electroforesis capilar; y analizar los datos de salida con el software del analizador capilar, tal como el algoritmo de software descrito en el presente documento.

Para cada muestra, el producto final de este ensayo es un perfil y una asignación de categoría. Como se puede observar en la figura 9C y D, el ADN natural se asignó como "natural", en contraste con la muestra artificial que se asignó como "artificial".

En este ejemplo, la asignación de categoría se realizó mediante el análisis de tres parámetros:

SR1 (relación de señal TPOX/D8S1179), SR2 (relación de señal de D3S1358/Hypo23), y RR (relación de representación de OCA2/D3S 1358).

Los valores de umbral para estos parámetros se proporcionan en la Tabla 6:

**Tabla 6**

	SR1	SR2	RR
Umbral para natural	$\geq 0,9$	$\geq 10$	$\geq 2$
Umbral para artificial	$<0,9$	$<10$	$<2$

Los valores observados para estos parámetros, y la categoría asignada se proporcionan en la Tabla 7 (los valores que están por encima del umbral para el ADN natural están sombreados).

**Tabla 7**

	SR1	SR2	RR	Categoría
Muestra de ADN natural	1,23	17,54	3,31	Natural
Muestra de ADN artificial	0,06	2,06	2,47	Artificial

La asignación de la categoría se realizó de acuerdo con la siguiente regla:

Si todos los valores de los parámetros observados están por encima de sus umbrales respectivos para el ADN natural, entonces se asigna una categoría "natural", de lo contrario se asigna una categoría "artificial".

#### 40 Ejemplo 6

##### **Perfiles combinados y categorización de ADN (MDA frente a no MDA)**

Este ensayo particular emplea tanto el perfil convencional de una muestra de ADN (con el kit Promega PowerPlex16), como la categorización del ADN en categorías de MDA (ADN artificial que se sintetizó por amplificación de desplazamiento múltiple frente a otros tipos (no MDA) de ADN).

Se analizaron dos muestras de ADN, una muestra de ADN natural extraída de la sangre del sujeto A, y una muestra

de ADN artificial sintetizada mediante amplificación de desplazamiento múltiple *in vitro* usando un kit comercial (a partir de una cantidad mínima del ADN real del sujeto A como plantilla).

Las dos muestras se analizaron en primer lugar con un kit convencional (PowerPlex16) y se analizaron usando un software de análisis convencional (GeneMapperID-X). El esquema de este procedimiento es como se indica a continuación:

10 ADN → PCR múltiple usando la mezcla de cebadores PowerPlex16 → separación de productos de amplificación en una máquina de electroforesis capilar → análisis de datos con el software GeneMapperID-X

Para cada muestra, el producto final de este ensayo es un perfil. Los perfiles de las dos muestras son idénticos, como puede observarse en la figura 15A-B. Además, el software GeneMapperID-X determinó que los perfiles de ambas muestras eran perfiles de un único contribuyente (es decir, de una única persona), sin anomalías.

15 Ambas muestras de ADN también se analizaron posteriormente por el software Capillary Analyzer. Para cada muestra, el producto final de este ensayo es un perfil y una asignación de categoría. Como se puede ver en la figura 15C-D, el ADN natural se asignó como "no MDA", a diferencia de la muestra de MDA que se asignó como "MDA".

20 En este ejemplo, la asignación de categoría se realizó mediante el análisis de tres parámetros:

RR1 (relación de representación de vWA/D18S51), RR2 (relación de representación de CSF1PO/Penta\_D), y RR3 (relación de representación de D21S11/D7S820).

25 Una puntuación de representación combinada se calculó basándose en estas tres relaciones como se indica a continuación:

$$\text{CRS} = \sqrt[3]{R1 * R2 * R3}$$

30 El valor umbral que se determinó en base al análisis de muestras de MDA múltiples y muestras diferentes de MDA fue 1,5, por lo tanto, una muestra desconocida con un valor de CRS superior a 1,5 se clasifica como "MDA", mientras que una muestra desconocida con un valor de CRS inferior a 1.5 se clasifica como "no MDA".

Los valores de CRS de la muestra de MDA y de la muestra diferente de MDA se calcularon, y son como se indica a continuación:

Valor de CRS de la muestra de MDA = 2,04  
Valor de CRS de la muestra diferente de MDA = 0,91

40 En función del valor umbral, ambas muestras se categorizaron correctamente.

### Ejemplo 7

#### Kit detector de semen independiente

45 El ensayo empleado por el kit acepta como entrada una muestra de ADN que luego se somete a un procedimiento bioquímico seguido de análisis de señal por un software dedicado. El resultado del ensayo es la categoría de la muestra de ADN (semen o no seminal), y un nivel de confianza estadística que representa la probabilidad de que la categoría emitida sea la verdadera categoría de la muestra de ADN.

50 Un kit de detección de semen de la presente invención puede comprender uno o más de los siguientes componentes:

1. Una caja que puede comprender uno o más de los siguientes componentes:

55 **Tubo 1** que comprende una mezcla de cebadores 5 x de:  
SD1f 0,6 μM (AAGAGCCCATCAGGCAGGTC);  
SD1r 0,6 μM (GTTTCTTGTCGAGCAGCACGTGGATGATG);  
SD2f 1,75 μM (CTCCAGAACTGGAACCTCCTG);

SD2r 1,75 µM (GTTTCTTAACCTTGGAGACGACGGCATC);  
 SD3f 1,25 µM (TGGAGGACAATGCCCTGGTG);  
 SD3r 1,25 µM (GTTTCTTGGCTTCACCTGCGACCGTCTC);  
 SD4f 1,75 µM (CCCTCCGAGTGGCCAGCAG);  
 SD4r 1,75 µM (GTTTCTGACCACTGCCGTGGGAATG);  
 SD5f 1,75 µM (CTTCTCAGCCAATGGGAAGAG);  
 SD5r 1,75 µM (ACGTAGAAGGACCCGAGGAC);  
 SD6f 0,9 µM (TACAGACAAATCACTCAGCAGC); y  
 SD6r 0,9 µM (GTTTCTTGTCTGACACTCGGTTGTAGGTATT)

Los cebadores directos (por ejemplo, SD\_f) están marcados por fluorescencia;

**Tubo 2** que comprende un tampón de reacción 10 x (TRIS-HCl 150 mM, MgCl<sub>2</sub> 15 mM, cada dntp 0,2 mM, BSA 2,5 µg);

**Tubo 3** que comprende la endonucleasa de restricción HhaI;

**Tubo 4** que comprende una muestra de ADN de semen de control;

**Tubo 5** que comprende una muestra de ADN diferente de semen de control;

**Tubo 6** que comprende una escalera de ADN; y

una ficha de datos de seguridad (MSDS).

2. Un documento que detalla el siguiente protocolo, o instrucciones similares:

(a) Para cada reacción, combinar en un tubo de 0,2 ml (no suministrado) los siguientes ingredientes: 5 µl de mezcla de cebador 5 x, 2,5 µl de mezcla de reacción 10 x, 0,5 µl de endonucleasa HhaI, 0,5 µl de ADN polimerasa, 0,5 ng de ADN, y DDW (agua destilada) hasta un volumen total de 25 µl.

(b) Además del ADN ensayado, establecer una reacción de semen de control positivo usando el ADN de semen proporcionado, una reacción de control diferente de semen usando el ADN de sangre suministrado, una reacción de control negativa usando DDW en lugar de ADN, y una reacción de control de digestión usando DDW en lugar de HhaI.

(c) Colocar los tubos de reacción en el termociclador y ejecutar el siguiente programa: 37 °C durante 15 min, 95 °C durante 11 min seguido de 30 ciclos de 94 °C durante 1 min, 59 °C durante 1 min, 72 °C durante 1 min seguido de una etapa de extensión final de 60 °C durante 45 min, y mantenimiento a 25 °C.

(d) Para cada reacción de post-amplificación, mezclar 1,5 µl de producto con 24,5 µl de formamida y 0,5 µl de tamaño de fluorescente estándar en un nuevo tubo de 0,2 ml.

(e) Desnaturalizar las muestras a 95 °C durante 3 minutos, después enfriar inmediatamente sobre hielo durante 3 minutos.

(f) Realizar muestras desnaturalizadas en la máquina de electroforesis capilar utilizando los siguientes parámetros:

Módulo = GS STR POP4 (1 ml) F;

S. de Iny. = 5;

kV de iny. = 15;

kV de realización = 15;

Realización C = 60;

Tiempo de realización = 24 min

(g) Analizar los ficheros .fsa de salida en el software de análisis Tissueldentifier.

3. Software de análisis "Tissueldentifier" que implementa el algoritmo de identificación tisular como se describe en la sección "Algoritmo y software" anterior, con las siguientes modificaciones:

En la etapa 3 del algoritmo, el conjunto de pares de loci usados son: {SD1,SD6}, {SD2,SD6}, {SD3,SD6}, {SD4,SD6}, y {SD5,SD6}.

Se realizó una corrección para la digestión incompleta usando la relación de señal de {SD5,SD6}. SD5 actuó como un locus de control digerido y SD6 actuó como un locus de control no digerido.

Se obtuvieron funciones de probabilidad de referencia de semen y no seminal para {SD1,SD6}, {SD2,SD6}, {SD3,SD6}, {SD4,SD6} usando un conjunto de 27 muestras de semen de referencia obtenidas de 27 individuos diferentes y 86 muestras de referencia diferentes de semen (sangre, saliva, orina, hisopo vaginal, sangre menstrual) 5 obtenidas de diferentes individuos.

El ensayo se basa en el análisis de 6 loci genómicos. De estos, los inventores encontraron 4 loci para ser metilados de forma diferente en el semen frente a tejidos no seminales. Dos de estos cuatro loci están metilados en todos los tejidos, excepto el semen, y los otros dos loci no están metilados en ninguno de los tejidos excepto en el semen. El 10 ensayo también incluye un locus de control no digerido que no contiene una secuencia de reconocimiento de HhaI (SD6) y, por lo tanto, se espera que se amplifique con éxito, independientemente del tipo de tejido, y un locus de control digerido (SD5) que los inventores descubrieron que no estaba metilado en ninguno de los tejidos potenciales.

El kit se probó en 27 muestras de ADN de semen y 86 muestras no seminales (sangre, saliva, orina, hisopo vaginal, 15 sangre menstrual) de 95 donantes diferentes de diferentes sexos, edades y etnias. El algoritmo identificó correctamente la fuente de todas las muestras con un nivel de confianza típico >99,999999 %. La Figura 12 muestra gráficos de muestra de 6 muestras. Las muestras ensayadas fueron las mismas muestras utilizadas para obtener funciones de probabilidad de referencia. Sin embargo, para evitar el sesgo, en cada análisis específico, la muestra 20 analizada no se utilizó para obtener las funciones de probabilidad de referencia.

**SECUENCIAS**

**Secuencia 1: D16S539**

```

1 CTCTTCTCAT TCCACAAGCT CTCCCCAAAA GACCCCATTC CTCCCCACCT TCAACCATCT
61 CTGGCAGGGA GGAGGGGGAA CTGAGAGGCT ACTTCTGAC CCAGGACCCCT AAGCCTGTGT
121 ACGGAGAGAG CATGAGCTGG GTGAGCTGCT TGCCAAGGAG TGGCATCTGC CCTCATCAGT
181 GGACACAAAA AGCCCCAGGG GTTAACTGGC CATGGCTGCC CTCATGGCTG CACCCGGAGG
241 ATGACTGTGT TCCCACCTCTC AGTCCCTGCC AGGTGCCTGA CAGCCCTGCA CCCAGGAGCT
301 GGGGGGTCTA AGAGCTTGTA AAAAGTGATC AAGTGCCAGA TGCTCGTTGT GCACAAATCT
361 AAATGCAGAA AAGCACTGAA AGAAGAATCC AGAAAACCAC AGTTCCCAT TTTATATGGG
421 AGCAAAACAAA GGCAGATCCC AAGCTCTTCC TCTTCCCTAG ATCAATACAG ACAGACAGAC
481 AGGTGGATAG ATAGATAGAT AGATAGATAG ATAGATAGAT AGATAGATAT CATGAAAGA
541 CAAAACAGAG ATGGATGATA GATACATGCT TACAGATGCA CACACAAACG CTAATGGTA
601 TAAAAATGGA ATCACTCTGT AGGCTGTTTT ACCACCTACT TTAATAAATT AATGAGTTAT
661 TGAGTATAAT TTAATTTTAT ATACTAATTT GAAACTGTGT CATTAGGTTT TTAAGTCTAT
721 GGCATCACTT TCGCTTGAT TTTTCTATTG ATTTCTTTTC TTTTCTTTTC TTTTTTGAGA
781 CAGAGTCTCA CTCTCACCCA GGCTGGAGTA CCGTGGCAGC ATCTTGGCTC ATTGCAACCA
841 CCACCTCCCG GGCTCAAGTG ATTATCTCTG CTCAGCCTCC CAAATAGCT
    
```

25

**Secuencia 2: D7S820**

```

1 ATATGCTAAC TGGATGTGAA CAATTGTGTT CTAATGAGCT TAATATGAGT TTCATAATTT
61 GTGCATTTTG CTGTTAAAAA GCCAGAAAAC AAAACAAAAC AAAAFACTGA AACCAGTGTG
121 AACAAAGAGTT ACACGATGGA AGGCATCAGT TTTACACACCA GAAGGAATAA AAACAGGCAA
181 AAATACCATA AGTTGATCCT CAAAAATGA TTGATTTTAA GCCTTATGAG ATAATTGTGA
241 GGTCTTAAA TCTGAGGTAT CAAAACTCA GAGGGAATAT ATATTCTTAA GAATTATAAC
301 GATTCCACAT TTATCCTCAT TGACAGAATT GCACCAAATA TTGGTAATTA AATGTTTACT
361 ATAGACTATT TAGTGAGATT AAAAAAACT ATCAATCTGT CTATCTATCT ATCTATCTAT
421 CTATCTATCT ATCTATCTAT CTATCTATCT ATCGTTAGTT CGTTCTAAAC TATGACAAGT
481 GTTCTATCAT ACCCTTTATA TATATTAACC TTTAAAATAAC TCCATAGTCA GCCTGACCAA
541 CATGGTAAAA CCCCGTCTCT AAAAAAATA CAAAAATTAG CTGGATGCAG TAGCACATGC
601 CTGTAGTCCC AGCTACTCAG GAGGCTGGGG CAGGAGAACC ACTTGACCCA AGAAGCGGAG
661 GTTGAGTGA GCCGAGATCG CACCACTGCA CTCCAGCCTG GGTGACAGAG TGAGACTCCA
721 TCTCAAGATA AAGAAATAAA TAAAAACAAA CAAACAAAAA AATTCCATAG GGGGTCAAGT
781 GCGGTGGCTC ATGCCTGTAA TCCCAGCACT TTGGGAGGCC GAAGCAGGTG GATCACTTGA
841 GGT
    
```

**Secuencia 3: D13S317**

1 AATATGAATT CAATGTATAC AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGACTTCTAC  
 61 AGAGCTCTAA GCATAATTGT GTRACTCCAA GCTCACAGTG CCTAAGACCA GTACCAGGCT  
 121 GACTCATTTGG AAAGCTGCCA TAGTAAGACT CTTCTGTTC A CTGCATTAT TATTGATGTA  
 181 TTGCAAGCAC TTAGTTACAT TTCTAGCATA TAACACATGA TCAATAAATA TTTTGACATG  
 241 AACAAATGGT AATTCTGCCT ACAGCCAATG TGAATATTGG GATGGGTTC TGGACATGGT  
 301 ATCACAGAAG TCTGGGATGT GGAGGAGAGT TCATTTCTTT AGTGGGCATC CGTGACTCTC  
 361 TGGACTCTGA CCCATCTAAC GCCTATCTGT ATTTACAAAT ACATTATCTA TCTATCTATC  
 421 TATCTATCTA TCTATCTATC TATCTATCAA TCAATCATCT ATCTATCTTT CTGTCTGTCT  
 481 TTTTGGGCTG CCTATGGCTC AACCCAAAGT GAAGGAGGAG ATTTGACCAA CAATTCAGGC  
 541 TCTCTGAATA TGTTTTGGAA ATAATGTATA TTAATGAATG TACAAATTTT CCCACTTGTA  
 601 CTTTCAGACT GTTATCTGTG AGTTAAAACCT CCTCCACTCT TTTTCCCTACC CAAATAATAG  
 661 CATACTTTT TCTGAGTATA TTTTGGGAG AAGAGTTATT CAGTTATTGT TATATTTTAA  
 721 AAAATTCCTT ATACCAAACT CTACTTGATC TAAGGCTATT CATTGAACT TTCAGCATGC  
 781 TTAATAGCAG TC

Secuencia 4: D5S818

1 CCCTCTGTGT AGCCTGGCTA TGTGCCACAT TGTGAGGTT TCTCCCTCTA GCTACTTCTT  
 61 CCAGTTTCCCT CTTTCAGGAT CCGAATTCCT TTCTCAAAGC ACTGGTGAAT AACTCCAAT  
 121 ACTCCATCAT TTCAATTATC AGAGTAATAT AAGTCTTCTT TTTCTAAACC TCTCCCATCT  
 181 GGATAGTGGG CCTCATATTT CAGATGCTAA TAGGCTGTG AGGTAGTTT CTAAGCAAAA  
 241 AAGTAATGGT CTCCTCAGA GGAATGCTTT AGTGCCTTTT AGCCAAGTGA TTCCAATCAT  
 301 AGCCACAGTT TACAACATTT GTATCTTTAT CTGTATCCTT ATTTATACCT CTATCTATCT  
 361 ATCTATCTAT CTATCTATCT ATCTATCTAT CTATCTTCAA AATATTCATC AAGGATACCA  
 421 AAGAGGAAAA TCACCCTTGT CACATACTTG CTATTAATAA ATACTTTTAT TAGTACAGAT  
 481 TATCTGGGAC ACCACTTTAA TTAGAAGCTT TAAAAGCATA TGCATGTCTC AGTATTTAAT  
 541 TTTAAAATTA TTACATAATT ATATACTCCT TTGAATTAGA AAATTACAAA TGTGGCTATG  
 601 TATTAATTTTC TCCCATGTAT TTTCAAATG AGGGGGTAAG CCAGACCCTC TCCCTCTCCC  
 661 ATGCCTAATA GCTCAAAGTT AAAGGTAAG AAACAAGAAA ATATGGTGAA AGTTAACCAG  
 721 CTTCACTTCA GAGGA

5

Secuencia 5: CSF1PO

1 ATTC AACACA TGAGGCACGG CCAGACCTAA ATGTCTCAGA GCCTGCTCCC ACTCCGATGA  
 61 GGTGCTGCCT TGCTTCAGGG TCTGAGTCCA GTGACTGCCA CTGCTGCAC CCAATCAGCA  
 121 TAGCCAGAGA CCTGGAGGTC ATCCTTATCT CTTTCTCTTT CCTCATCCCT GCATCTCAGA  
 181 CTCTTCCACA CACCACTGGC CATCTTCAGC CCATTCTCCA GCCTCCAGGT TCCCACCAA  
 241 CCCACATGGT GCCAGACTGA GCCTTCFCAG ATACTATCTC CTGGTGACA CTTGGACAGC  
 301 ATTTCTGTG TCAGACCTTG TTCTAAGTAC TTCTATCTA TCTATCTATC TATCTATCTA  
 361 TCTATCTATC TATCTATCTA TCTATCTAAT CTATCTATCT TCTATCTATG AAGGCAGTTA  
 421 CTGTTAATAT CTTCAATTTA CAGGTAGGAA AACTGAGACA CAGGGTGGTT AGCAACCTGC  
 481 TAGTCTTGG CAGCTCAGG TTGGAACACT GCCCTGGAGT GTGTGCTCCT GACCACCAG  
 541 AAGTGCTCC TCTGTACAAT CTGACCCCAT CACTCTCCTC TTTACAATGA CCTCCCAATA  
 601 GGTAAAGATG CAGTTATPCT TTCTCACTTT AAGACACCTT TACCTCCGGC TTCTGCCACC  
 661 TCCTCTGTCT CCCTGTGGCC ACTCCTCACA CCACTCCACA TCCCAGCTGT TGTCAAGTTC  
 721 TTTCAAGTGT CCAATGATC TATGTTCTCT TTGCCPTTGA GCCTTGCATA TGTTCCTCCC  
 781 TCTGCCAGAA GCGCTGTTC CTTTCCTTTT CCACCTTCT GCCCAGCCAA CTCACCTTCA  
 841 TTCTTCCCAT CCCAGTTTAG AGGCCACCTT CTCGAAGCCT GGGTTGGGGG GACTCTTCAG  
 901 TGTTCACAG ACACCTTGTG CTTCCCCAT AATCACTGGG TGATCATTTG

10

Secuencia 6: TPOX

1 CCCAGCACAC ACCTTGCCTC TGGCTGGGAC CCCCTTTGCT GCTGGCCCTG CTCAGGCCCC  
 61 ACAGCTTGAT CTCCTCATGT TCCCAC TGCT GACTTCCCCA AGCTAACTGT GCCACAGAGT  
 121 GGGGACCCCT CTCCTGCCTC TCACAACCCC CACTTCCCT TGCTTCACTT TTCACCAACT  
 181 GAAATATGGC CAAAGGCAAA AACCCATGTT CCCACTGGCC TGTGGGTCCC CCCATAGATC  
 241 GTAAGCCAG GAGGAAGGGC TGTGTTTCAG GGCTGTGATC ACTAGCACCC AGAACCTTCG  
 301 ACTGGCACAG AACAGGCACT TAGGGAACCC TCACTGAATG AATGAATGAA TGAATGAATG  
 361 AATGAATGTT TGGGCAATA AACGCTGACA AGGACAGAAG GGCTAGCGG GAAGGGAACA  
 421 GGAGTAAGAC CAGCGCACAG CCCGACTTGT GTTCAGAAAG CCTGGGATFG GACCTGAGGA  
 481 GTTCAATTTT GGATGAATCT CTTAATTAAC CTGTGGGGTT CCCAGTTCTT CCCCTGAGCG  
 541 CCCAGGACAG TAGAGTCAAC CTCACGTTTG AGCGTTGGGG ACGCAACAC GAGAGTGTCT  
 601 GGTGTGAGCA CACAGGAGGA GTCACGACAG AGCAGTGTA GAGCCGCCAC GTGGGTCCCA  
 661 CACAGGGGGA GTCACGACAG AGCAGTGTA GAGCCGCCAC GAGGGTCCCA CACAGGGGGA  
 721 GTCGCGACAG AGCAGTGTAA GAGCCGCCAC GAGGGTCCCA CACAGGGGGA GTCACGACAG  
 781 AGCAGTGTA GAGCCGCCAC GAGGGTCCCA CACAGGGGGA GTCACGACAG AG

Secuencia 7: TH01

1 TTACCCCTGG GGTGGGGTGG TAGGATGCAG CTGGGGCTGC AGTTCAGGC CACGGAGAGC  
 61 CTGTGAGGCT GGGCCCCGGG GCGCCCTGGG GAGGGGATGC CTGATGGGGA GCCTGGTGGG  
 121 GGAGGGTAGG GGAGGGCGGG GGAGGACGGG GGAGGGCGCC CTGTGTCCCT GAGAAGGTAC  
 181 CTGGAAATGA CACTGCTACA ACTCACACCA CATTTCAAATC AAGGTCCATA AATAAAAACC  
 241 CATTTTAAAT GTGCCAGGGA GCCCAAGGTT CTGAGTGCCC AAGGAGGCAC CGAAGACCCC  
 301 TCCTGTGGGC TGAAGAGCTC CCGATTATCC AGCCTGGCCC ACACAGTCCC CTGTACACAG  
 361 GGCTTCCGAG TGCAGGTCAC AGGGAACACA GACTCCATGG TGAATGAATG AATGAATGAA  
 421 TGAATGAATG AGGGAATAA GGGAGGAACA GGCCAATGGG AATCACCCCA GAGCCAGAT  
 481 ACCCTTTGAA TTTTGCCCCC TATTTGCCCA GGACCCCA CATGAGCTG CTGCTAGAGC  
 541 CTGGGAAGGG CCTTGGGGCT GCCTCCCCAA GCAGGCAGG TGGTTGGGGT GCTGACTAGG  
 601 GCAGCTGGGG CAGAGGGAGG CAGGGGCAGG TGGGAGTAGG CTGGGGGCTG GGTGCAGCAG  
 661 CCGGGGACCT CTGGCCATCT TGGATTTTTT GGATGGATT TTTCCACAT TCCGATCGTT  
 721 AAGATTCAAG ATGAAACAAG ACACAGAGAC CCACAGGACC CCCCAG

Secuencia 8: vWA

1 AGATGATAGA TACATATGTT AGACAGAGAT AGGATAGATG ATAGATACAT AGGTTAGATA  
 61 GAGATAGGAT AGATTATAAA TAGATACACA GGTAGATAG ATTAGACAGA CAGATAGATA  
 121 CATACATAGA TATAGGATAG ATAAGTAGAT ACAATAGAGA TAGATAGATA GATAGATAGA  
 181 TGATAGAGGA TAGATGATAA ATAGATATAT AGCTTAGATA GAGATAGGAT AGATGATAGA  
 241 TACATAGGAT AGATAGAGAC AGGATAGATG ATAAATAGAT ACATAGGTTA GATAGAGATA  
 301 GGACAGATGA TAAATACATA GGATGGATGG ATAGATGGAT AGATAGATAG ATAGATAGAT  
 361 AGATAGATAG ATAGATAGAT AGACAGACAG ACACACAGAC AGATAGATCA ATCCAAGTCA  
 421 CATACTGATT ATCTTATCA TCCACTAGGG CTTTCACATC TCAGCCAAGT CAACTTGGAT  
 481 CCTCTAGACC TGTTCCTTCT TCTGGAAGGT GGGAACTCTA CCTTATAGGA TCAGTCTGAG  
 541 GAGTTCACAA AATAATAAGG GCAAAGTGCC CGGCACATTTG TAGGAGACTA GTAATGTCTA  
 601 TAAAATGAGG GGCTTGAAGT AAATGATCCC TCTAGTTCTC TCTACTGCTA ACATTCTAAG  
 661 ACCTCCTTTA CATTAATGTG TCTCAAGCCA CATCTCCCTC CCCTACAGGA CTTCTATTTA  
 721 TTTCTGATCA ATTCATGAG TACAAAATAAG T

5

Secuencia 9: FGA

1 ACTGAACATT TGCTTTTGAA ATTFACATC TATGTACCGT TGGAAAATTT ACTTAATATC  
 61 TCTGAATTTT TTTTCTTCAA CTGTGGAGTG AGGAAAATAA TACCTACTTT TAGGTAGATG  
 121 ATGGATATAA CACTTTTCTC TGCATATAGT AGACACTCAG TGCATAACTA TCGCCTTCCCT  
 181 TTTCCCTCTA CTCAGAAACA AGGACATCTG GGACCACAGC CACATACTTA CCTCCAGTFCG  
 241 TTTTATATCA ACCAACTGAG CTCTAACATT TTTCTGCAGA AGCTGGATAT GCTGTACTTT  
 301 TTCTATGACT TTGGCGTTCA GGACTTCAAT TCTGTCTTC AGATCCTCTG ACACTCGGTT  
 361 GTAGTATATA TCAGGGTCTG AAATCGAAAA TATGGTTAT TGAAGTAGCTG CTGAGTGATT  
 421 TGCTGTGAAT TGCCAGCAAA AAAGAAAGGA AGAAAGGAAG GAAGGAGAAA GAAAGAAAGA  
 481 AAGAAAGAAA GAAAGAAAGA AAGAAAGAAA GAAAGAAAGA AAGAGAAAAA AGAAAGAAAG  
 541 AAACATAGCTT GTAATATGCT CTAATTTTAT TTTGGTTACA GTTTAATCTG TGAGTTCAAA  
 601 ACCATGGGGG CATTTGACTT TTGGATAATG TTATGCCCTG CAGCCTTCCA TGAATGCCAG  
 661 TTAAGATGTC CTAATAGCAA TTAGTAATCC CAAAGAAATA TAGAAGAAGA ACTTCTTTTG  
 721 GAATTTTAAA GGTGTAATTT GGAGTTAAAA TAGTTGGTTT GATTGCATTT CAATTTATTT  
 781 ATAACATCCT TAAACAAGGG ACTTGAACAT ATTTGATTTT CTTACTGATG AGCTTTTCTT  
 841 TTTAATCTAT AGATTTGAAA TGGTTCCTAA GCTGTTTTGG GTCAACAGGA TCACTCACTT  
 901 GCCAGCTAGT GTTGCAATCAG TGATTTTAAA TGTCAGTGT TTTGT

Secuencia 10: D21S11

1 GTTGGCTGGG GCTCAGAGAG AACAAAAAGG CAGAGGAAAA ACAAAATTTCC CCTTCACTT  
 61 CTGGAGATGG AACACTTTC TTCTGCTTTT GGACATCAGA AATCCAAGTT CTCTGGCCTT  
 121 TGGACTTTGG GACTTGTGCC AGCACCTTCC TGGGTTCCTT GCCTTTTGGC CTCAACTGA  
 181 AGGTTTACACT ATCAGCTTCC GTTGTCTTAA GGGCTTCAGA CTTGGACAGC CACTTCCCA  
 241 GCTTCCCTGA TTCTTCAGCT TGTAGATGGT CTGTTATGGG ACTTTTCTCA GTCTCCATAA  
 301 ATATGTGACT CAATTCCTCA AGTGAATGTC CTTCTATCTA TCTATCTATC TGTCTGTCTG  
 361 TCTGTCTGTC TGTCTATCTA TCTATATCTA TCTATCTATC ATCTATCTAT CCATATCTAT  
 421 CTATCTATCT ATCTATCTAT CTATCTATCT ATCTATCTAT CGTCTATCTA TCCAGTCTAT  
 481 CTACCTCCTA TTAGTCTGTC TCTGGAGAAC ATTGACTAAT ACAACATCTT TAATATATCA  
 541 CAGTTTAAAT TCAAGTTATA TCATACCACT TCATACATTA TATAAAACCT TACAGTGTTT  
 601 CTCCCTTCTC AGTGTTTATG GCTAGTAAT TTTTACTGGG TGCCAGACAC TAATTTTAT  
 661 TTTGCTAAGT GGTGAATATT TTTTATATCC TTTAAAATAT TTTTGAAGT TGAATGGGGT  
 721 AAAGTTAAGT TCAATATTGG AAAAATATTG ATTCTTTTGA GGATAGTTAT CTTCTAATTA  
 781 GTCTACCTGT TGCCCATATA ATGGCATGAT TTTCCACTCT GTG

10

Secuencia 11: D8S1179

1 TACTACAGCA AGAGCGCTTG AACAGATGT AGGGGAGATA GCAGCTGGAG AGCATAACAG  
 61 AGGCACTGAC ATGTGAGCAG CTAACGAGGC CTTTTACAAG ACATCTGTGA CCACACGGCC  
 121 AAGTAGAAGA AAGCCGTAA AAGCATCAAG GTAGTTAGGT AAAGCTGAGT CTGAAGTAA  
 181 TAAAACATTT TACAGGATC CTTGGGCTGT CGCTTTTCTG GCCAGAAACC TCTGTAGCCA  
 241 GTGGCGCCTT TGCTGAGTT TTGCTCAGGC CCACTGGGCT CTTTCTGCC ACACGGCCTG  
 301 GCAACTTATA TGTATTTTGT TATTTTATGT GTACATTCGT ATCTATCTGT CTATCTATCT

361 ATCTATCTAT CTATCTATCT ATCTATCTAT CTATTCCCCA CAGTGAAAAAT AATCTACAGG  
 421 ATAGGTAAT AAATTAAGGC ATATTACAGC AATGGGATAC GATACAGTGA TGAATAATGAA  
 481 CTAATTATAG CTACGTGAAA CTATACTCAT GAACACAATT TGGTAAAAGA AACTGGAAAC  
 541 AAGAATACAT ACGGTTTTTG ACAGCTGTAC TATTTTACAT TCCCAACAAC AATGCACAGG  
 601 GTTTCAGTTT CTCCACATCC TTGTCAACAT TTGTTATTTT CTGGGTTTTT GATAATAGCT  
 661 GTGAAAGGAA AATAAAAACT TGGGCCGGC GCGGTGGCTC ACGCCTGTAA TCCCAGCACT  
 721 TTGGGAGGCC AAGCGGGCA GATCTCAAG TCGGGAGATT GAGACCATCC TGGCTAACAT  
 781 GGTGAAACC CATCTCTACT AAAAAACAA AAACAAAAA TTAG

**Secuencia 12: D18S51**

1 CATGCCACTA AGCTGTACAC TGAAAAACGG TTAACATGAT AAATTTTATG TTACATACAT  
 61 TTTTACCACAA TTTAAAAAAA TTATTAATAA ATACTAACAA TAGGCCAAGC GTGATGGCTC  
 121 ACACCTGTAA TCCCAGCACT TTGGGAGGCT GAGACAGGTG GATCAATTGA GCTCAGGAGT  
 181 TTGAGACCAG CCTGGGTAAAC ACAGTGAGAC CCCTGTCTCT ACAAATAAAT ACAAATAAAT  
 241 GTTGGCAGTG GTGGCAGGTG CCTGTAGTCT CAGCTACTTG CAGGGCTGAG GCAGGAGGAG  
 301 TTCTTGAGCC CAGAAGGTTA AGGCTGCAGT GAGCCATGTT CATGCCACTG CACTTCACTC  
 361 TGAGTGACAA ATTTAGACCT TGCTCAGAA AGAAAGAAAG AAAGAAAGAA AGAAAGAAAG  
 421 AAAGAAAGAA AGAAAGAAAG AAAGAAAGAA AGAAAGAAAG AAAGAAAGAA AGAAAGAAAG  
 481 AAAGAAAG AAATAGTAGC AACTGTTATT GTAAGACATC TCCACACACC AGAGAAGTTA  
 541 ATTTTAAATTT TAACATGTTA AGAACAGAGA GAAGCCAACA TGTCACACTT AGGCTGACGG  
 601 TTTGTTTATT TCGCTTGTG CTGGTAGTCG GGTGTTTATT TTTTAAAGTA GCTTATCCAA  
 661 TACTTCATTA ACAATTTTCA TAAGTTATTT CATCTTTCAA CATAAATACG CACAAGGATT  
 721 TCTTCTGGTC AAGACCAAAC TAATATTAGT CCATAGTAGG AGCTAATACT ATCACAATTA  
 781 CTAAGTATTC TATTTGCAAT TTGACTGTAG CCCATAGCCT TTTGTCCGCT AAAGTGAGCT  
 841 TAATGCTGAT CAGGTAATTT AAAAAATATA GTTAATTAAG AGGGCATAAA TGTTACCTGA  
 901 CTCATTAAGT CATTTCAATT AGGTCTG

5

**Secuencia 13: D3S1358**

1 CTGGTTTTGG TGGAAATTGAC TCCCTCTGTC ACAAACCTCAG CTTCAGCCCA TACCCTGAGC  
 61 CATAGACCTA TCCCTCTAAT GCATTGTAAT AGTCTCAGGG CTAATAACAA GGGAGAGGTG  
 121 TCAAAGGGCC AGTTCCACCT CCACCACCAG TGGAAAAGCT ATTCCCAGGT GAGGACTGCA  
 181 GCTGCCAGGG CACTGCTCCA GAATGGGCAT GCTGGCCATA TTCACTTGCC CACTTCTGCC  
 241 CAGGATCTA TTTTCTGTG GTGTGTATTC CCTGTGCCCT TGGGGGCATC TCTTATACTC  
 301 ATGAAATCAA CAGAGGCTTG CATGTATCTA TCTGTCTATC TATCTATCTA TCTATCTATC  
 361 TATCTATCTA TCTATCTATC TATCTATCTA TGAGACAGGG TCTTGCTCTG TCACCCAGAT  
 421 TGGACTGCAG TGGGGGAATC ATAGCTCACT ACAGCCTCAA ACTCCTGGGC TCAAGCAGTC  
 481 CTCCAGCCTC AGCCTCCCAA GTACCTGGGA TTATAGGCAT GAGCCACCAT GTCCGGCTAA  
 541 TTTTTTTTTT TAAGAGATGG GGTCTCGCTG TGTCCCCAG CCTTGCTTTA AACTCCTGGC  
 601 CTAAGTGAT CCTCCATCT CAGCCTTCCA AAGTGCTGAG ATTACAGCAG AGGCTTTTAA  
 661 CTCAAAGCTT TCCCTGCTAG GACAAGCCCT AGTTAAAGTC CTGGAGCACT GGCACCTGCA  
 721 GCTGCACTTG G

**Secuencia 14: Penta D**

1 CCTACTCGGG AGGCTGAGGC AGGAGAATCG CTTGAACCCA GGAGGGGGCG ACTGCAGTGA  
 61 GCCGAGATCG TGCCACTGCA CTCCAGCCTG GGTGACAGAG CGAGACTCCA TCTCAAAAAA  
 121 AAAAAAATAA AAACAGAATC ATAGGCCAGG CACAGTGGCT AATTGTACCT TGGGAGGCTG  
 181 AGACGGGAGG ATCGAGACCA TCCTGGGCAC CATAGTGAGA CCCATCTCT ACAAATAAAT  
 241 AAAAAATTT TTTTAAATA GCCAGGCATG GTGAGGCTGA AGTAGGATCA CTTGAGCCTG  
 301 GAAGGTCGAA GCTGAAATGA GCCATGATCA CACCCTACA CTCCAGCCTA GGTGACAGAG  
 361 CAAGACACCA TCTCAAGAAA GAAAAAAG AAAGAAAAGA AAAGAAAAGA AAAGAAAAGA  
 421 AAAGAAAAGA AAAGAAAAGA AAAGAAAAGA AAAGAAAAGA CGAAGGGGAA AAAAGAGAA  
 481 TCATAAACAT AAATGTAAAA TTTCTCAAAA AAATCGTTAT GACCATAGGT TAGGCAATA  
 541 TTTCTTAGAT ATCACAATAA CATGACCTAT TAAAAATAA TAATAAAGTA AGTTTATCA  
 601 AAATTAATAA GTTCTACTCT TCAAAAGATA CCTTATAAAG AAAGTAAAAA GACACGCCAC  
 661 AGGCTAAGAG AAAGTACTTC TAATCACATA TCTAAAAAAG GACTTGTGTC CAGATTAAG  
 721 AATCTTTACA CATCAATAAG ACAACCCCAAT TAAAAATGGG CAAAAGATTT GAAGAGATAT  
 781 TTAACCAAAG AAACATATA AATGTGTCGG GCGCGATGG TAATCCAGC ACTTTGAGAG  
 841 GCCGAGCAG GCGGATCACT TGAGGTCAGG AGTTTAGGAC CAGTCTGGCC AACATGGTGA  
 901 AACCTGTCT CTAATAAAAA TACAAAAAT AGCTGGGTGT GGTGGCGTAA GCCTGTAATC  
 961 CCACTGCTC AGGAGGCTGA GGCAGAAAG TGTCTTGAAC CTGGGAGGTG GAGGCTGCAG  
 1021 TAAGCC

10

**Secuencia 15: Penta E**

1 CACATGTGGA CATTCTTAT TTTCTCATAT TGGTGGTATG GCTCATTAT GAAGTTAATA  
 61 CTGGACATTG TGGGGAGGCT GTGTAAGAAG TGTTAAAGGG GATCAGGGAT ACATTCACCT  
 121 CTCTTTTCTT TFGCTAGTTC TGTGGTCTTA AGCAAAGTAG CCTCAAACAT CAGTTTCCTC  
 181 TTTTATAAAA TGAGGAAAAT AATACTCATT ACCTTGCAATG CATGATATAA TGATTACATA  
 241 ACATACATGT GTGTAAAGTG CTTAGTATCA TGATTGATAC ATGGAAAAGAA TTCCTTATT  
 301 TGGGTATTA ATTGAGAAAA CTCCTTACAA TTTTCTTTTC TTTTCTTTTC TTTTCTTTGA  
 361 GACTGAGTCT TGCTCAGTFCG CCCAGGCTGG AGTGCAATGG CGTGATCTCG GCTCACTTCA  
 421 AFCTCCACCT CCTGGGTTC AAGTATTCTC CTGTTTCAGC CTCCAGAGTA GCTGGGATTA  
 481 CAGGTGCCFA CCACCACACC CAGCTAATTT TTTGTATTTT AGTAGAGACG GGGTTTCACC  
 541 ATGTTGCCCA GGCTGGTCTT GATCTCCTGA GCTCAGGTAA TACACCTGCA TCGGCCTCCC  
 601 AAAGTGCTAG GATTGCAGGC GTGAATCACC GCACCTGTCC ACAAATTTCT TGTATTGGT  
 661 ACCCTTTTAT GTTGGTAAAA TGTATTTTAT TTTCTCTTAT CAAATAATTT TCAATGCAAT  
 721 GAGACGTCAA CTTTAAGCCC AAAGTAGACC AGTAGTAAAA CTAAGGCTGA AACCATTGAT  
 781 TGATTAATAC CATATAATGT CCTAAAATAA TCGGCCTTTA AAACATTTGG TTTCAATTTT  
 841 CATGATAAAA ATATGTAGCA TTTTTCACAT TTTAATTCAC TTTGTAGAGT TCTCAATCAT  
 901 TTCTAACACA TGCTTGGCAA TGACAAGCCA TTTGTGAAAG AGTTTTCGCT GCTTTAAAAA  
 961 ATATGCAAAAT GTAATAA

**Secuencia 16: AMEL X**

1 AGGTCTCCTC TTCATACAG CACATTGTT CAAACTAAAA ACAGACCTCA AGTATATTC  
 61 GCACTATATA GATTTTTTTA AAGTAGCTTC AGTCTCCTTT AATGTGAACA ATTGCATACT  
 121 CATTTAATCT CTTCTCTCTT CTTCTCTTCC TTCACTCTCT CCCTTCCTCT CTCTTTCTAT  
 181 TCTCCTCCCG TCCCTCCCTGT AAAAGCTACC ACCTCATCCT GGGCACCTCG GTTATATCAA  
 241 CTTCAGCTAT GAGGTAATTT TTTCTTTTAC TAATTTTGAC CATTGTTTGC GTTAACAATG  
 301 CCCTGGGCTC TGTAAGAAT AGTGTGTTGA TTTTTFATCC CAGATGTTTC TCAAGTGGTC  
 361 CTGATTTTAC AGTTCCTACC ACCAGCTTCC CAGTTTAAGC TCTGATGGTT GGCCTCAAGC  
 421 CTGTGTCTGC CCAGCAGCCT CCCGCCTGGC CACTCTGACT CAGTCTGTCC TCCTAAATAT  
 481 GGCCGTAAGC TTACCCATCA TGAACCACTA CTCAGGGAGG CTCCATGATA GGGCAAAAAG  
 541 TAAACTCTGA CCAGCTTGGT TCTAACCCAG CTAGTAAAAA GTAAGGATTA GGTAAAGATG  
 601 TATTTAAAC TCTTTCCAGC TCAAAAACCT CCTGATTCCTA AGATAGTCAC ACTCTATGTG  
 661 TGTCTCTTGC TTGGCTCTGC TGAATATTA GTGACTAAGT GGTATA

5

**Secuencia 17: AMEL Y**

1 TTATTCTCCA ATATTTTGAA ATGTGAATAT TACAGTAAT TCCCTTGTCC AAATGAGAAA  
 61 ACCAGGGTTC CAAAGAGAGG AAATTAATTTG CCCAAGTTA GTAATTTTAC CTAATCTTTA  
 121 CATTTTACCG GATGGGATAG AACCAGCTG GTCAGTCAGA GTTGACTTTT TGCCCTTTCA  
 181 TGGAACCTTC CTGAGCAGTG GTTCATGAAT GAATAAACIT ACAGCCATAT TTAGGAGGAA  
 241 AGAGTCAATC CGAATGGTCA GGCAGGAGGG TGCTGGAGCA ACACAGGCTT GAGGCCAACC  
 301 ATCAGAGCTT AAACCTGGGAA GCTGATGGTA GGAACGTGTA AATTGGGACC ACTTGAGAAA  
 361 CCACITTTAT TGGGATGAAG AATCCACCCA CTATTCTTTA CAGAGCCAG GGGACTGCTA  
 421 ATGCAACACG TGATCAAAAT TAGTAAAGAG AAAAATPACC TCATAGCTGA AGTTGATATA  
 481 ACCAGGGTGC CCAGGATGAG GTGGTAGCTT TTATAGGGAG GAGGGGAGGA GAAGAGAAAAG  
 541 AGAGAGGAGG GGAGAGTGTG AAGGAAGGGA AGAGAGAGTA AGAGATTAAG TCAATATGCA  
 601 ATTTGTAACA TTAAGAGAGA CTAATAATAC TTTTAAAAA TCTATATAGT ACAGAAATATA  
 661 TTTGAGGTTCT GTTTTCTGTT AAAACAAGTG TGCTATGTAG GAGAGGAGAC TT

**Secuencia 18: D2S1338**

1 ACAAGGCACG GAACTCACAC CCAGCCTCTC TCCATACAAC AGAATATGGG TTCTTGCGGA  
 61 GCTGGACTCT GCAGGAGTCT ATCTAATATG GACTCTGTGT CAATGACTCC TGGGCCTCCT  
 121 CTGATCACCC CATFAAAGTC CTTTCGATTGC TTTGAGCCTC AAATCTATGT GACATCAATA  
 181 GCTTCTATTC TCCCTAGCAC TTAGAAGTGT TTCTTGTGA TACATTTGCT GGCTCTTCC  
 241 CTGTCTCACC CCTTTTCTTA CCAGAAAGCC AGTCCAGAG GCCCTTGTCA GTGTTTATGC  
 301 CTACATCCCT AGTACCTAGC ATGGTACCTG CAGGTGGCCC ATAAATCATGA GTTATTCAGT  
 361 AAGTTAAAGG ATTTCCAGGAG GGAAGGAAGG ACGGAAGGAA GGAAGGAAGG AAGGAAGGAA  
 421 GGAAGGAAGG AAGGAAGGAA GGAAGGACAG CAGGCAGGCA GGCAGGACAG CAAGGCCAAG  
 481 CCATTTCTGT TTCCAAATCC ACTGGCTCCC TCCCACAGCT GGATATATGG CCAGTAGGAA  
 541 TTTGCCATTT CAGGGTTTGT CTGTCACTGT AGTCAGGACC ATGAAGTCTT TAGGCACCTC  
 601 CACTCCACAC ACCCCCTGGT GAGAGCTCCC ATCTCCCTGT TCTGAAACAG CTCCCAATA  
 661 TAGTACTGAT TCCGGTTAAA CTTGAACCCC TGCCCCCTGC CCTGCCCTGT ATTTACATGA  
 721 GGACACTGAG CCCAGAGGGG GTAAAGTGAC TGCCAGGGGT CACACAGCTA GAAAGTGGCG  
 781 GTGCCAGAAC TGAAGGAGG CCCTCATTC TGAGTCACGG CTTTCCATA GCACAGCCTT

10

**Secuencia 19: D19S433**

1 ATGAACTGG ACACAGAAAC CAGACCCAG AGCACATACC GTATGAGTCC ATTTGGTATGA  
 61 AGTTTAAAAA CAGATGGCAC TAGTCCAAG GATTGGAAGT TGAATAGTGC GTTACCAGGA  
 121 CTGGGGGGAG GAAGGGATGG TGGATGGTGA ACAAAGGAC CTTGGAGGGC TCTGGGGT  
 181 CTAGGAATCA ATCTTCTTTC TTTCTTCTCT TCTCTCTTTC CTCTTCTCTCT CTTCTTCTCT  
 241 GTTTTATTTT CAAATAGGTTT TTAAGGAACA GGTTGGTGTG GTTACATGAA TAAGTTCTTT  
 301 AGCAGTGATT TCTGATATTT TGGTGCACCC ATTACCCGAA TAAAAATCTT CTCTCTTCTT  
 361 TCTCTCTTCC CTTCTTCTTC CTTCTCTTCC TTTCTTCTCT CTTCTCTCTT TTTCTTCTTA  
 421 CTTCTTCTTC TTTAACAGAA TCTTATTTCTG TTGCCAGGC TGGAGTGCAG TGGTACAATT  
 481 ATAGCTTTTTC CGAGCCTCAA CCTCCTGGGC TCAAGTGATC TTCTGCCCCC AGCCTCCTGA  
 541 GTAGCCAGGA GTACAGGAAT GTGCCAACAT GCCTGGCTAA TTTTAAAAAA TTTTATATAG  
 601 AGAAGAGGTC TCACTAATGT GCCCAGACTA GACTTGAACCT CCTTCCCTCA AGTGATCTTT  
 661 CTGCATCAGT TTTCCAAAGT GCTGGGATGG CAGGCATGAG CCACCTCACC CAGCCTTAGA  
 721 AATGTTCTGT TTTCTGACCT GAGAGCTGGA TATACAGGAT TGCTCACTTT GTGAAAAATC

Secuencia 20: ACTBP2SE33

1 GTACTTCAGA GTCAGGATGC CTCCTTTGCT CTGGGCTCC TTGCCACAT AGGAGTCTTT  
 61 CTGACCCATG CCCACCATCA CTCCTGGTG CCTAGGGTGC CCCACAATGG AGGGGAAGAC  
 121 GGCCCTGGGA GCCTTGCAGC TGCTGGAGCA GTTGTGAGC ACGACGAGCG CGGTGATAGC  
 181 ATCATPCATG GTGAGCTGGC GCGGGTGGC GAGCAAGGC GCAGCGGCAA GGCAAGGTT  
 241 CTGTGCTCGC TGGGCTGAGC CCGTCTCCGC GGTGTAAGGA GGTTTATATA TATTTCTACA  
 301 ACATCTCCCC TACCGCTATA GTAACCTGCT CTTTCTTTCC TTCTTTTCTT TCTTTCTTTC  
 361 TTTCTTTCTT TCTTCTTCTT TTTCTTTCTT TCTTTCTTTC TTTCTTTTTC TTTCTTTCTT  
 421 TCTTTCTTTC TTTCTTTCTT TCTTCTCTT TCTTTCTTTC TCTTTCTTTC TTTTCTTTTC  
 481 TTTTCTTTCC TTCTTCTCTT TCTCTCTCTC TCTTCTTCTT TCTTTCTAAC TCTCTTTGTC  
 541 TCTTTCTTTC TTTCTTTTGA CCGAGTTTCA CTCTTGTGCG CCAGATTGGA GTGCAATGGC  
 601 ATGACCTCGG CTCACCTTAG CCTCCACCTC CCAGGTTCAA GCGATTATCC TGCCTCAGCC  
 661 TCCCTAGGAG CTGGAATTAC AGACGTGCAC CACCAAGCCT GGCTAATTTT TGTATTATTA  
 721 TGGACTGGAA ATATGACATT CATATTGGAA TCCAGTGTCT TTTTCTGAAA AAGAGAGTTC  
 781 CGACCTGGCT TAGGCTCCCA AAGTCTGGG ATTATAGGCA TGAGCCACAG TGCCACGCTT  
 841 TCTTTTCATT TAATACTATA GTAGTGTGAT CCTCTCTACC TATTACA

5

Secuencia 21: D10S1248

1 TTCTGTTTTC CCGTGGTTC TAGTATGGTA CCTGGCCAAG GGCACACTAG ATCTTTGTCA  
 61 AGGTAATGAC TACTTTTAT TAAATGCTTT CCATGTATCA AGTTCGTGTC CAAGCACTTG  
 121 ACATATATCA TTTTATTTTA TCCCCTGAAG TAGTTATTGG TATCTTCATT TACAAATAAA  
 181 AAAACAAGCT TAGTACTTAA CTCACCTGCC TGAACATAAT TATTGCTTTA AAGGTAGCTA  
 241 GGATTCTTAA TAGCTATTAT TACCAAAGCA TGAACAATCA GTAAAAAGCA AACCTGAGCA  
 301 TTAGCCCCAG GACCAATCTG GTCACAAACA TATTAATGAA TTGAACAAAT GAGTGAAGTG  
 361 AAGGAAGGAA GGAAGGAAG AAGGAAGGAA GGAAGGAAG AAGGAAGGAA ATGAAGACAA  
 421 TACAACCAGA GTTGTCTCTT TAATAACAAG ACAAGGAAA AAGAGAACCTG TCAGAAATAAG  
 481 TCTTAATTAT AATATCCAGG GGTGGGATAC AGAGGTTTTA GCATCTGCTC TTTGCCAAGC  
 541 ACTGCACTTA TTCTTGAAGG ATACCTGAGG GAAAAAGTAT GGTTTCTCAC AGGATCTAGT  
 601 TGGACTGGAA ATATGACATT CATATTGGAA TCCAGTGTCT TTTTCTGAAA AAGAGAGTTC  
 661 GTTCCAAGCT TAGCTCACAT GCAAGCTAAG ACAACCACTA GAAATTACTC TCCCAGGGC

Secuencia 22: D1S1656

1 GTCATGCCA CAGTGTAAACG GGAATTGACC AGGTAGGCGA CTTGAACCTCC AACTGCAGGC  
 61 TATGGGGAGA CATGTGACAA TGCTAATCCC TTAGGCATTT ATTCAGTGCA TTGCAGTTTA  
 121 AATGTCTGCC TTTTCCAGCAT TTCAGAGATT ATGTACCTTA AAGAGGCAGG CTGGAATTTA  
 181 AAACGGCAAG CCAGGAAAGA GAGAAACCAT GTGATCCAC CGCAGCACAA AACCTCGTTA  
 241 GCAGCTGTAA GCGCCTGGTC TTTGTTTATT TTTAATTTCC TTTCTTTCCC AATTCTCCTT  
 301 CAGTCTGTG TTAGTCAAGG TTCTTCAAGG AAATAGAAATC ACTAGGGAAC CAAATATATA  
 361 TACATACAAT TAAACACACA CACACCTATC TATCTATCTA TCTATCTATC TATCTATCTA  
 421 TCTATCTATC TATCTATCTA TCTATCTATC TAGATCACAC AGTTGACCTC TGAGCAACAC  
 481 AGGCTTGAAC TTATATGGGG ATTTTCTTCC ATCTCTACCA CCCCTGAGAC AGCAAGACCA  
 541 ACTCCTCCTC CTCCTTCTCA GCCTACTCAA CATGAAGATA ATAAGGATGA AGACCTTTAC  
 601 AATGACCAG TTCCACTTAA TAAATAGTAA ATGTATTTCC TCTTCCCTAT GATTTTCTTG  
 661 ATAACATTTT TTTTCTCTGG CTTATTATTAT STAAGAATAC AGTATATAAT ATAAATAATT  
 721 ATAAAACATG TTAATTGGTT CTTTACGTTA TCGATAAGAC TTCTGGTCAA TGGTAGGCTA

10

Tscl McrBC (semisito)

752: BspDI Clal parcialmente metilado

15

Secuencia 23: D22S1045

1 GAGCCCAAGT TTAAACCCAG GCCCTCTGTG TCCCCTTACA GGGTGACTGC ATCTCCGAGT  
 61 CCTGGCTTGT CATGCCTGAC AGAGGGCTGC CGAGTGAGCA GCTTAAGGCA TCCTGCCACT  
 121 GTGCAGCTGC CAACCTTACA GCCCGGCAGC CTTGCGGGAG GAAGCTCTAG TGCAGGCCCTC  
 181 TTAGGATCTG GGGTCCAGGA TGCTGATTTT AGGGCCGGGA CCTTGGGCAC CGTCCCTCTG  
 241 GTCTGCATAA GACCCACTAT GGGCAAACCT TAAACCTGAT CGTTGGAATT CCCCAACTG  
 301 GCCAGTTCCT CTCCACCCTA TAGACCCTGT CTTAGCCTTC TTATAGCTGC TATGGGGGCT  
 361 AGATTTTCCC CGATGATAGT AGTCTCATTA TTATTATTA TATTATTATT ATTATTATTA  
 421 TTATTATTAC TATTATTGTT ATAAAAATAT TGCCAAATCA ACATTCCGCT GATCACTCAC  
 481 ACTGTGCCGG GCACTCTTGA GAGCACTTTA CATATATTGT CTCATTTAAT TCTCTCAACT  
 541 TGGGCACAGG CACTGTCACT ATTTCCATTC TACAGCTGAG GAGACTGAAG CACAGAGAGC  
 601 CTTAGGGACT TGCCCTGAGT CACACAGCTA AGAAATGGTG GAGCCAGGAT CAGAAACCAG  
 661 GCCACCTACA GAGCTCCCTG CAAGGGGAAC AGCATCCGCT TCCAGAGGCT GTGATTTTAT  
 721 CAGTACACT GTGTGACTCC ATCTTACAC TCTCCTGCC CTCAGAAGA CATATAACCT

Secuencia 24: D2S441

1 ATGAAGAGAT GGTGAGGCGA GGTATGGGGG AAGGGGCGTG GAGCTTCCAT GTCTCCCTG  
 61 GGCGCCACCC FCCAGGAACC TCCACGTGTT CAGCTATACA GAAGCTTCCCT GAACCCAGTC  
 121 CTCTTGGGGT TTAGGGGAAG CTTTCATGACA TCAGCATTCCT TTCTCCAGG GTATTAATGG  
 181 GACCCTCTCT GAAGAGATTC TTAAGACCCA CGGCCAGAAA CTTGGGTAAA GACTAGAGTC  
 241 CTGCCTTGGG GCAGGTGAAA GGAGTGCAAG AGAAGGTAAG AGAGATTCTG TTCCTGAGCC  
 301 CTAATGCACC CAACATTTCTA AAAAAAGGCT GTAACAAGGG CTACAGGAAT CATGAGCCAG  
 361 GAACTGTGGC TCATCTATGA AAACCTCTAT CTATCTATCT ATCTATCTAT CTATCTATCT  
 421 ATCTATCTAT CTATATCATA ACACCACAGC CACTTAGCTC CAATTTAAAA GATTAATCAT  
 481 AAACATTTGG GAAGGAGAGT GAAGATTTTT GTGATGTTAA ATAAGAATGA TTATACTAAA  
 541 AACCAAAATA ATATGTTATT TATGGCTGGG TGTGGTGGCT TAAGCCTGTA ATCCAGAAC  
 601 TTTGGGAGGC CAAGCCTTGT GGATCACTTG AGCCAGAAAG TTCAGAGCCA GCTTGGGCAA  
 661 CATAGGGAGA CCGTGTCTCT AAAAAAATT TTAATAATTAG CTGGACATGA TGGCAGGCAC  
 721 CCGTAGTCTC AGCTACTCAG GAGGCTCACG CCACTGCATT CCAGTCTGGG TAAGGCACAC

Secuencia 25: D12S391

1 GTCAGGAGTT CGAGACCAGC CTGGCCAACA TGGCGAAACC CTGTCTCTAC TAAAAATACA  
 61 AAAAAATTAG CTGGGCATGG TGGTGTGTTT CTGTAACCC AGCTACTCAG GAGCTGAGG  
 121 CAAGAGAATC GCTGGAACCC AGGAGGTGGA AGTTGCAGTG AGCTGAGATT GCACCACTGC  
 181 ACTCCAGTGT GGGCAACAGA GCGAGACTCT GTCTCAGAAA AAAAAAAGAA TACATGAAAT  
 241 CAGAGAACT CAAATTTGTA TAGTAGTTTC TTCTGGTGAA GGAAGAAAAG AGAATGATAT  
 301 CAGGGAAGAT GAAAAAAGAG ACTGTATTAG TAAGGCTTCT CCAGAGAGAA AGAATCAACA  
 361 GGATCAATGG ATGCATAGGT AGATAGATAG ATAGATAGAT AGATAGATAG ATAGATAGAT  
 421 AGATAGACAG ACAGACAGAC AGACAGACAG ACAGATGAGA GGGGATTAT TAGAGGAATT  
 481 AGCTCAAGTG ATATGGAGGC TGA AAAAATCT CATGACAGTC CATCTGCAAG CTGGAGACCC  
 541 AGGGACACTA GAGCAGTGGC TCAGTCCAGG TCTAAAAGCC AAAAAACAGG GGAACCTGAT  
 601 GGTGTAATTA FCCATCCAG GTGGAAGGCC TGAGAACCTG GAGTGCCCTT GGTATAAGTC  
 661 CCAGAGTACA AAGACAGGAG AGCCTGGAGT TCTGACTTCC AAGGGCAGAA GAATGTGTCT  
 721 CAGCTCCAGG AGAGAGAGAG AAAGAATTTT TTTCTCCGC CTTTGTATTG TATCTGGGGG

5

Secuencia 26: ADD6 (del ncbi acceso NT\_022135)

16664701 cttgaacctg agaggcagag gttgcagtga gccgagacca tgccattgca ctccagcctg  
 16664761 gqcaatagag taaactcca tcctcccgc ccaaaaaagt agacaacgct catgaggatga  
 16664821 tgaggaagg gttatcgtgt gttgcttgc gagaacagga ccccagaact caccgtgtcg  
 16664881 acgcccggca gtagcatctc agtcaacttg gcgtagatct cctgcagcgt cagagcctgg  
 16664941 ctaaggaaga ggtatgtgag aagtccccgc ctcaacctcc ggctcggct catttggtac  
 16665001 tgatgtccc tcaacttgtt gtcaacatga atttggcctg ttgaaaaaca gtatttcttt  
 16665061 tgaaggaggt ttgggttgag aatcatcttt tcagtctcaa agcctctgt cctcccagta  
 16665121 gcttaactaa accagtgcca ggtgacagag ggttaaggaa cccaatttat ctaacgctaa  
 16665181 cctgggaggt tcaactcatac acttgcttat gtaaatgaa gaaaagttaa aagacaagct

Secuencia 27: ADD10 (del cromosoma 17:3477839-3478292)

1 AGAGAGCCCA GGAGACAGGC AGAAAGGAAG GCATGTGACC GGATCACAAT CATCAGCTCT  
 61 CTGCTGTCTT CTTTGGGAAG GGTTTTAGTA TTAAAAGGAC ATTTATTCTC ATTAATGCAA  
 121 AATTAAGGAG TTTTAAAAGC TTTTACAACC TAGACTCCCT CTGAGAGGTT AGCCTTGACA  
 181 CCCTAATCGC TTCTGTCTCC CGCCACTGCT CGGTGCCAAG CAGCTCCAC GGCCCGGGC  
 241 GGTCTGATGA TAGCCGACAC GGAGGGAGGA AGGGGAGGAG GAAGAGCCTG CATCAGCTCC  
 301 TACGATTGCC CAGCCCAATC CTGGGAGTGA TTAACCGGTG CATACCAAAA TGCCAGTCCC  
 361 ACTGACAGGC AGGTACCCGT GCACCTCAGG GCACCTTAAA TTGCCGACTC TCCATGTAGA  
 421 GAGGGATGAA TCCAATATTG AAATCCTCAT AACTACAGCC CCCCAAAAGTA GCCGTCCATC  
 481 TTCTGCTTAA AATGTTGATC TGTAGTAAAA TGTGATTTT GTTGAAGCTG AGTGATG

10

Secuencia 28: ADD17 (del cromosoma 1: 50149332-50149574)

1 TTGAACCTAG GAGATGGAGG TTGCAGTGAG CTGCGATCAT GCCACTTCCC TCCAGGCTGG  
 61 GCAACAGAGC GAGACTCCAT CTCAAAAAAA CAAAAAAGAA AACCAACCTT TTGAATGTAG  
 121 GGAAACTTT TCAAAGGATA TCTAGTTTTT AATTACAGTA AACTTGTGGA AGGGAGGTTT  
 181 AGAGTTGAGA TTGAGATTAT AGATTTTGTCT GATGATAAAC CATGAGTTC AGAGGACATA  
 241 CTAGACTATT CTGGGCAGTT ATACAGGGGT GGATGGAATG TGGGAGTGGG GTTGTATAGT  
 301 GCCATAAAGA AATGAGAGTC CCGATTAAAA ATAATGAGCT GGACTCGCGA GCCTTTGTGTA  
 361 ACTGAAATTA ATAGAAAAT AAGAAATACA TTATTTCTGT GATTGTTGAG AGGAAGAAAT  
 421 GGTGGAATC TTGTGAGAAG CACACTGAGC TCTAGCACCA CCTCTTCACT CCTACAGATG  
 481 GTGGAATAAA CGGCAGGCAA GTTCAAAATC ACATATAGTC ATTATTGCAA GATAGTTCTA  
 541 TGGATATAGA TACTACATAC AATATAAATC ATGCTCATTG AATGTTTTCAG TGGAACTAC  
 601 TCTGAACCT

15

Secuencia 29: Hypo23 (de ncbi acceso NM\_004907)

123161 GAGTTTGGGA AGGGTATTTT AGGGGGGAAT AACTTTTGTAG TTCCAGCGT GCGGGGGAAG  
 123221 GGCGGGACGG GAGGGTGTCC CAAGGCCTGA GAAGATCAGT GTGGGGCAGG GGTGAGGAAT  
 123281 AACCTGGGAG GGGGCCTTGT ATGGGGGAAA TAATTGGGAA GAGGAGAGAT GGGATGAAGG  
 123341 GGCCCTCAGC GGGTCGTCTC CTGTGTATGC AGGGTCGTTC TGCAGCGTCT CTGGGAGATG  
 123401 GCGTCCCTGG GAGCCCTCAG GTCGCCCTA CCGCTGCGG GGTGCTTTCC TGGCGTCAG  
 123461 CCTTCTGGC CCCTGGAGG AAGGAAGTGA AACTCTCTC TTCCCCACC CGGCTGGAAT  
 123521 GCGAGTCAGG AAGCCTGGGG CTCAGCCTG CTCGGCTGC CCGGGTCGGG GATGGGGAGG  
 123581 GGCGTGGCCG GAGCGAAAG CCCC GCCCT CCGCGCCCC CCCC GGAAG CCCC CGCC  
 123641 GGCGCTAAG GCGATCAGG GCCCTGTCTT AATATGGGCA ACCGGAAGCG GCCCGCGCA  
 123701 CTGCCCTACG TCACTCCGTC CAAATTTAGT TGTGGAAGTC AGCGGGCGCT GGTGGCGGGA  
 123761 AGGCGCCCGG AGCCAGTGG GCGGAAAGG GGGCGGGGG CGCACCACC CTTAAAGGGC  
 123821 CCGCACAGG AATGAATGGA GCCATTGCA CAAATTCGCA TCCTATTTT GGAGGAAGTG  
 123881 GAATTAGTAT TT

**Secuencia 30: Hypo28 (del cromosoma 5: 85949232-85949719)**

949232 TTTATTTTAA AAAAGAAAG AAAGAAGAGA AAAGGGATGG GTTTATTGTC CTTTCAACA  
 949292 GACTAGAGTA TACGGGGTGA AACTGCTTCA CTGATTCAA TAAATCGTT TCCGGTAACA  
 949352 GGCCCCAGGA ATCCTAGACC TAAGCCTGGC GCGAAACTAC ATTTCCACA ATCCTTCGGG  
 949412 GGCTGATAAG GCTCCGCAAT GGTCGAAC ACAAATCCA CAATCCAGGG CGATTTCCGC  
 949472 TTTGTCGGT TTCTCAAG CTCGCCCA TTCCCATCT TTCTTTTTCG TCCTTGGCA  
 949532 CCGGGGAAAG AGGTCGTGAA AAAAAAGTCT TTGGTGAGGT GCCGCCATTT CATCTGTCTT  
 949592 CATCTCTGCG GCCTTTCGCA GAGCTCCAG CAGCGTATG TTGGCCAGA GCATCCGGAG  
 949652 GTTCAACCC TCTGTGTCC GTAGGAGCA CTATGAGGAG GGCCCTGGGA AGGTTAGTGT  
 949712 GTAAGGGG

5

**Secuencia 31: Hypo33 (del cromosoma 6: 34211067-34211314)**

211067 ACCCTCATTT CACATTTTAC CCCTTCTCA AAATGCTCC TTCATATTAC CTCCTAGAA  
 211127 ACCAAGAAATA TGGCTACTAA TTCTCCCTGG CCCCATGCTG CAGGTGAACC GGTAGCCAG  
 211187 AGGTATCACA TAATTTCTCC AAAGTCACAC AGCAAATCAA GATGCATCCA GGACTAGAAG  
 211247 CCATGTCAGC CACTGTTGGA AGCCCCAGCG AAGCTGACAG AAAGTTTCAT AATACCACC  
 211307 TCTCCCTT

10

**Secuencia 32: secuencia OCA2: (del cromosoma 15: 25276967-25277446)**

6967 aaacacccca gtctgaaat aacctagatt tgttgctctt acgagtgaat atgctatttc  
 7027 atacacgaag ctttctctt cagcaccaca gatttaagga taattatgga tgaatattat  
 7087 ggattcattt taaatccttt ggcaaatctg ctctgggggc ttctctgtca gaaggtctct  
 7147 ccttcccaac tctaaagaac gttattccta tgcaaatgct gctgagtcac gacggggagg  
 7207 gaagtgcaga gagaagggt ggtgcatgg tcagtaagtc atgagggtga gattaggggt  
 7267 gacacactgc ttgccaacgt aggagaaggc tctgccctca cctagcaagt ctgatggaag  
 7327 ccccttattc cgtcctctct gccgggttcc accgagatcc aaaaaggaat gctgtgtagg  
 7387 agcacatgat atgtgataaa tgagagaaag gtcaaacatt taaggaacgc ccagagaaag

**Secuencia 33: SD1 (cromosoma 11: 72085699 - 72086161)**

72085699 CGCTGGGCGC GAGTGAAGAG CCCATCAGGC AGGTCGGCGCA GGAAGCGCTT GAGCGCCGAG  
 72085759 GAAACATCAT CCACGTGCTG CTCGCCCTCC TTGAGGTGCA CAGAGCGGC ATCCTGCCGC  
 72085819 AGGCTCTCCA GCAGCCGCTG TGTCTTCGAT GTCGCCCCAC ACTTGGCGTA GATGCCCTCG  
 72085879 GAGGTCAGG CTAGGGAGGG GCGGGGCCAA GCGTTCGGGG CCTGAGGCAT AGAGTCATGG  
 72085939 GCGGGGGCCG CGCAGCTCTG GGGCGGGAGG CCGCTCTCCG GGAGGGGCGG GGCTGGCAC  
 72085999 CTAGGGGCGG GGCTCACCGC ACTGCTGAT GTAGTCCACA CAGCGGTACA CGATCACCGG  
 72086059 GATATCCGAG TCCCCAAGCT GCTGCTCCGA CAGCGTGTCC CCCATGCTGG CGGCTGCTTT  
 72086119 CTGGATGGCC CCCAGCCAAC CCATGAAGTC CAGCCCGCCG TCG

15

**Secuencia 34: SD2 (cromosoma 4: 25286907 - 25287230)**

25286907 CGATCCCGTT CACTCGCCTG CCCATCCGCA TGGCCAAGGG GCTGGGCAAC ATCTCTGCCA  
 25286967 AGTATCGCTG GTTCGCCGTC TTCTACCTGA TCATCTTCTT CTTCCGTATC CCGCTGACGG  
 25287027 TGTTTGGCCT CTCGCTGGCC GGCTGGCGGG TGCTGGTTGG TGTGCGGGTT CCGTCTGCT  
 25287087 TCATCATCAT CCTGGTACTG TGCTCCGAC TCCTGCGATC TCGCTGCCCA CGGCTCTGC  
 25287147 CGAAGAAACT CCAGAACTGG AACTTCCCTG CGCTGTGGAT GCGCTCGCTG AAGCCCTGGG  
 25287207 ATGCCGTCGT CTCCAAGTTC ACCG

**Secuencia 35: SD3 (cromosoma 11: 57170977 - 57171240)**

57170977 CGCGAGAGGA GGGGAATGGC AGGTCCCCGC TTCGCTGACC TTGGAATAA GCTCATCGTG  
 57171037 TTTGGCCAGG TGTGCACGGC AGTGGACACA GCTGTAAGTG CCGTGACAGC GGGGAGGATA  
 57171097 GCTGCGGAAA GTCTTGGTGG GGAGGCAGGC AGGGCCCGGA CCGGGGTCAC AGCTGGGAT  
 57171157 GACGGGCTGG AGGACAATGC CCTGGTGGGC TGGAGGGGCT GGGCCGCTGC AGCCCCCGCA  
 57171217 GAGACGGTGC CAGGTGAAGC AGCG

20

**Secuencia 36: SD4 (cromosoma 11: 1493306 - 1493605)**

1493306 CGGTGCTCAC CGGCTCGACC CCCCAGCAAC ACGAGAGACC TCACAGAGGG AGTCACACTA  
 1493366 ACCTGGTCCG GGCTCCAGAG CGAAACCCCA ACCACTATGC TCACAGCCAG GACCGAGCAG  
 1493426 GCTGGGCCAA CGGCAGTCCC TGCCAGCGC CCGGCTCCCT CCGAGTGGCC AGCAGCGCCC  
 1493486 TCTGGTGGAG ACTGGCTCGG CTTCCGCGGC ACTGCAATCC CACGGCAGTG GTCCATCTAG  
 1493546 TCCCCAAGTC CTAGAGGAG CCCCCTCTCT TCCTCAGCC CTGGCAGGTT CTTGGCGCC

**Secuencia 37: SD5 (cromosoma 22: 16491391 - 16491872)**

16491391 CTGAGAGCCA TGGCGAGAGC AATGCTAGGC CGGTGAACAG TAGGCTCGAG TTTAGGTTTG  
 16491451 AAAGGTGAGG TGAGAGAAAT CGGCAAAGGG AACCCCTGCG CAGATCTCGG GTTCCTTTAC  
 16491511 TTTATAACCG CGGGTCCCG TTCTCTGCCAG GTGACTGCAC AGTTCATCCT CATGACCCTT  
 16491571 CTCAGCCAAT GGAAGAGAG CGACGCCAG GAAGTCCCAG CCTGTCCCGG CCTGTGGGGC  
 16491631 CGTCTCGGG TCCTTCTACG TCGCTGACTC GTGACCTGAC CGGTATTTT TCCCTAAAC  
 16491691 GGGATCTTGG GTAGGAGGAA GAAAAGATAA GGAGTTCCTC TATCTGAAAT TAGTCGGGCT  
 16491751 GTTTTGAGGA GTACTGGTTA GGTATATTGG AGAATGTGCC TTTATTGGAA TTTTTCGGAT  
 16491811 AATTACACAG GGTATGGGC TTAGGGGTAC ATGATTTGTTG GCCGGCGTG GTGGCTCAGC  
 16491871 CC

5

**Secuencia 38: SD6 (cromosoma 4: 155727706 - 155729006)**

155727706 TTGCAGGGCT AGCCAGGATA AGCAATGGAT GCTGTTTAA TCACTTAGC CTCAGAATGG  
 155727766 ATCAAGAGTA AGGTGCTATT CCCTCTCAA GGAAACGCTA ATTAATGAT CCTTGGTGAA  
 155727826 ACTCAGGCTG AGGGCTCAGA GGTGTGGTGA TGTAATGGGC TTTGGAAATA GACAGGGACT  
 155727886 GAACATTTGC TTTTGAAAT TACTATCTAT GTACCGTTGG AAAATTTACT TAATATCTCT  
 155727946 GAATTTTTTT TCTTCAACTG TGGAGTGAGG AAAATAATAC CTACTTTTAC GTAGATGATG  
 155728006 GATATAACAC TTTTCTCTGC ATATAGTAGA CACTCAGTGC ATAACATACG CCTTCTTTT  
 155728066 CCTCTACTC AGAAACAAGG ACATCTGGGA CCACAGCCAC ATACTTACCT CCAGTCGTTT  
 155728126 CATATCAACC AACTGAGCTC TAACATTTTT CTGCAGAAGC TGGATATGCT GTACTTTTTT  
 155728186 TATGACTTTG CGCTTCAGGA CTCAATTCT GCTTCTCAGA TCCTCTGACA CTCGGTTGTA  
 155728246 GGTATTATCA CGGCTGAAA TCGAAAATAT GGTATTGAA GTAGCTGCTG AGTGATTTGT  
 155728306 CTGTAATTGC CAGCAAAAA GAAAGGAAGA AAGGAAGGAA AGAAAAGAA AGAAAAGAA  
 155728366 AAAGAAAAG AAAGAAAAG AAAGAAAAG AAAGAAAAG AAAGAAAAG AAAGAAAAG  
 155728426 CTAGCTTGTA AATATGCCTA ATTTTATTTT GGTACAGTT TAATCTGTGA GTTCAAAACC  
 155728486 TATGGGGCAT TTGACTTTTG GATAATGTTA TGCCCTGCAG CCTTCCATGA ATGCCAGTTA  
 155728546 AGATGTCCTA ATAGCAATTA GTAATCCCAA AGAAATATAG AAGAAGAACT TTCCTTGAA  
 155728606 TTTTAAAGGT GTAATTTGGA GTTAAATAG TTGGTTGAT TGCATTTCAA TTATTTTATA  
 155728666 ACATCCTTAA TCAAGGGACT TGAACATATT GGATTTTCTT ACTGATGAGC TTTTCTTTTT  
 155728726 AATCTATAGA TTTGAAATGG TTCCTAAGCT GTTTTGGGTC AACAGGATCA CTCACTTGCC  
 155728786 AGCTAGTGT GCATCACTGA TTTTAAATGT CAAGTGTTTG TGAAGGTGTA AAAAGGCAA  
 155728846 GCAAAACCTG AGAACTGAG GACTCCCTAG ACTCGCTGTC CATGCCAGA GTGAATGCCA  
 155728906 TGTTTCTTAA CCTAATGAG TAGACTGTGA GAATGACGTA GCTTGACCCT ATATTTTAAA  
 155728966 TTTAAAAATC TACCTAATCA GCTCAGTGA GTCTGGGAGT T

**Secuencia 39: L91762 (cromosoma 12: 73985133 - 73985792)**

73985133 GAACGAAAT TTCGGAGGGA GCTAAGAGCT CCATGTTTTT ATTTTACTAT TCTTTAAACA  
 73985193 AGCAGACACA GTGCTGGTTT CTGATCATGT GGGAGGTGTG TAGGAGAGAA CACAGGTAAA  
 73985253 GTAAAAATAA AATATAAAGC CATAAATGCC ACAACAGAAA TCAATGAAC AAACATAGAA  
 73985313 GAGAATACCT TAATTCCTAG AAGACAGAGA GAAGAGTTTA TGAAGGCAT CCTAGAAGAG  
 73985373 GAGATATTG CACTGATTTCT TAAAAGGTGA GAGTACAAGA CTTCCTCTCC AGGAGTCGGC  
 73985433 GGAGGGAGGG CAGACTCAA GTGCTCCTGG ACGCAGAGGC TCTTGTGAGA GGGCAGAGC  
 73985493 CGGTTGGAAC AAAGTAAGGA TCTGAGCGAC CCAACTTTG CAGCCGAGGC CTCCAGCTCC  
 73985553 GAGGTGCATA GCAACCTTAG GGTCCGGTA GGGCTTCTC CGTCCGCGAC CCTGGCAGAG  
 73985613 CCGCAGAGCC CTCTCCAGA CCGACCACT GCCCAGGCC CCGAGCGACC GGAAGGCTCA  
 73985673 GCAGCAGGCC GCGGAGAAG GCGGCGCCA CCAGGCCCT AGCACTGCC AGTCTGGCCC  
 73985733 GGCACAGCTG CTGCAAGAG GCACAGACA GCTCTGTGG TACATGGCG TCAAGGACCA

10

**Secuencia 40: L68346 (cromosoma 3: 12175184 - 12175532)**

12175184 CGCGGACCTC GGACTCACCA AGTGCCGAGT GGCAGATGTG CTGCTGAGGG TGCGCCGGGG  
 12175244 CGCAGCTGCA TGCCCTCACC AGCCCCGGG GCCCAGCAA CGCCAGCAGC CGCAGCAACA  
 12175304 GCACCCAGCT TGGCGCGGGC CGAGGGCTCC CAGGCATGAC ACTGCAGATC CGCGACTGAG  
 12175364 CCTGTGAGGT CTGGGGGACT GGACGGCCCC AGCAGGGCTC CTTCCTCAAG CCGTTGTGCC  
 12175424 CCTCGCCCCA GGCTTTATGA GGTGGCCCTA AGGCCAATC CCGCCCCGAC GGGCTCCGCC  
 12175484 TTCTTTGGC TCTAGGCCAC CCGCGGAGC AGGACCCGAG GCTCTTTCCG

15

**Secuencia 41: L50468 (cromosoma 3: 55492699 - 55493120)**

55492699 CGCTGCCAC GGGGAGGAG ACTGGGGGCG TTTGAGGGGC TCAGCGCAC AGAGGAGTGA  
 55492759 GGTGGAGGAG GCGTTCCTC CTCTCTCTC TTCAATCCAG AGCAGCTCAA CGAGTGGCT  
 55492819 CCTTTCTAT GTATCCCTCA AAGCCTTCGC GTCGGATTAA AGGTGTCTT GATCCTTCTT  
 55492879 TACCAACCAC GGTGCGGGCC AGGCGTGATG AGGATGAGG GAGAGGAAAC CTCAGTAGCA  
 55492939 AAATTTGTTCA GAGGACGTTT GGAGGGCGCG GGGAGCAGCC GGATGCACAC CTAAGTCTC  
 55492999 GCAACACCCA ACTCCTCCTC CGCAGGCAGT CCTTAAAGAG AATAGATAAA AAGGCCAAGC  
 55493059 AAAGATCTC TCCCCTCGGC CCGGGGCGC TGCCACATTC AGTCTAGGCA TCCCCTTCTG  
 55493119 CT

**Secuencia 42: L14432 (cromosoma 22: 29866031 - 29866650)**

29866031 GCTCGGGTCC GGGGGTGTGG ATGAAGGAGC CCCAGGCAGT CTCATGAGCA CAGAGAGCAC  
 29866091 CGTAGGCTGC GGTGAGCTCC GGCTCATCCT CCCAGCTACA TGACTGCAGC CTCCTATGGG  
 29866151 CATCCCAGCG GGCATGGATC AGGGCCAGTC CCTGAGCATC CTTGGCCAGG AAGCTCAGGT  
 29866211 ACCCCAGTGG GTTGCACAGG ACGGCCTTGG TCAAGTGGCA GGAGGTCCTG TACCAGCGGA  
 29866271 GGGCAGGGGA GCCCCCAGG GCCACCCCA GGAAGCCAG CATCCCAAAC AGCCCTGCCT  
 29866331 GAACCCCAT TCTGCACTGG CCCAGTCCAG TCAGACAAAG CCCCTGGGAT GCCTGCCCTT  
 29866391 GGTGGCCCA CCAGGCGGCA GCTGAGCAGT GGAACGGAA CGGAGCCAG CAGGCCCGGT  
 29866451 GCGCGGGAC CAATGAATGG AGCTGCGGGA GGAGGAGAA AGAGGCTGGA GTATGGGGTG  
 29866511 ATCTTGGGCT TGTAACCGAA TCCACCAGCC GGGCAGGAGC CTGATTGGCT GAGGAGTGCA  
 29866571 CTTGCCAGG CCCACGCCCT CATCTGCCT GCCTTCTCAG CACCATCCAG TCACCTGCTG  
 29866631 CCAGCCCTGC CTAGATTGGG

**Secuencia 43: L4648 (cromosoma 1: 35815020 - 35816031)**

35815020 CGGCGGCTG GCTCCCTGGA GCGGGGCGGG ACGGCGCCGC GCGGACTCAC GTGCACAACC  
 35815080 GCGCGGGACG GGGCCACGCG GACTCACGTG CACAACCCGC GGACCCAGC GCCAGCGGGA  
 35815140 CCCCAGCGCC AGCGGGACCC CAGCGCCAGC GGGACCCAG CGCCAGCGGG ACCCCAGCGC  
 35815200 CAGCGGACC CCAGCCGCA CGGACCCCA GCGCCAGCG GTCTGTGGCC CAGTGGAGCG  
 35815260 AGTGGAGGCA TGGCGACCTG AGCGGAGACT GCGCCCTGGA CGCCCCAGC TAGACGTCAA  
 35815320 GTTACAGCCC GCGCAGCAGC AGCAAAGGGG AAGGGGCAGG AGCCGGGCAC AGTTGGATCC  
 35815380 GGAGTCTGT ACCCAGGGA AAGCGTGGC GGTGACCCA GGGCAGTGC GCGGGCGAGG  
 35815440 CAGTGGGCT CCTTGCTCCC TGGAGCCGC CCTCCCAACA CCTGCCCTCG GCGCCCCAG  
 35815500 CAGTTTTCAC CTTGGCCCTC CCGGTCAC TGGGGATTG GCGTTGCCG CAGCCAGTGT  
 35815560 GGGAGTGAAT TAGCGCCCTC CTTGCTCCTC GGCCCTTCG ACGGCACGAG GAACCTCTGT  
 35815620 CCTGCCCCAC AGACCTTCGG CCTCCGCCA GTGCGTACT GGAGCCTGCC CCGCCAGGGC  
 35815680 CCTGGAATCA GAGAAAGTC CTTTGGGCC ACCTGAAGCG TCGGATCCCT ACAGTGCCTC  
 35815740 CCAGCCTGG CGGGAGCGGC GGCTGCGTC CTGAAGGTTG GGTCTCTGG TCGGAAAGGG  
 35815800 AGGCAGTGC AGCCTCAGCC CCACCCAGA AGCGGCCTTC GCATCGCTGC GGTGGGCGTT  
 35815860 CTCGGGCTFC GACTTCGCCA GCGCCGCGG GCAGAGGCAC CTGGAGCTCG CAGGGCCAG  
 35815920 ACCTGGGTTG GAAAAGTTC GCTGACTGCA GGCAAGCGTC CCGGAGGGGC GGCCAGCGCA  
 35815980 AGCCCGGCG CTTTACCACA CACTTCCGG TCCCATGCCA GTTGATCCG CG

5

**Secuencia 44: L39664 (cromosoma 17: 53710207 - 53710501)**

53710207 CGCGCTGAGC GGTGGTGGAG GAGACAGGGG TCATCGTGCA GGTGTGTCAA GGGCAGCAGG  
 53710267 GCCCGGCCGT TGCTTGGAA GCGTGGTTG ACGGCCAGCA GCCCAGCTG GTTGGACATG  
 53710327 TTGCGCAGGT TCCTGGCCAG GGGCTCCTCG CTGCCGTACA CCATGCTGGC GTCCACGAAG  
 53710387 GAAGTGAGCG CGTTGATCTG GTTGGCGATG GTGATGTTGC TCCCGGGCA AGCCGGCAG  
 53710447 GAGCGGAAGA ACGGGATGCA GTGCGCTGG TTCTTGATGC GGGGGTCATT GGGCG

10

**Secuencia 45: L30139 (cromosoma 17: 36996273 - 36997073)**

36996273 TGCCACCAG AAGCCCATCA CCACCAGCAA AGCCACCACC AAAGCCACCA CCCAAGCCAG  
 36996333 CACCAAGGCC ACCACCATAT CCTCCCCCAA AGCCACTACC AAAGCTGCTG CTGCTGCTGC  
 36996393 TGAAGCCACC GCCATAGCCG CCCCCAGCC CGCAGGCTCC CCCAGAGGAG AAGCGGGAGG  
 36996453 ATGAGACAGA CAGGCCGCC CCGTAGGTGC TGGGGGCGCG GCAGGACCT CCGGCCAGGA  
 36996513 CCGAGGAGAT GCGCTGGAG CCGCCCCCA TGCCGCCCCG GATGCCCGAG GAGCCCTTCA  
 36996573 TGGAGCTGGA GGAGGTGAAC TGGCGGCTGC AGGTGGTCA TGGTCAGAGG AAGGAGGTGA  
 36996633 GCGAGCGAGC AGTTGGCTGA GTGAAGAGAA GGTGCTCGG TAAATTTGAA AAGGATGGCA  
 36996693 GTGCTTTATA CTCGTGGGTA GGGGGCGGGT CTGGCACTTT CCATCCCTT TGGCTTTTAT  
 36996753 CACCCAGAGG CTAGCGCCAA CTCCAGCCA GGTCCCTCCT CTCATCCGCC TCATCATGTC  
 36996813 TGTCATATTT TACTGGAAAC TCATTGTTTG GAGTGTFTTG GGCTTCTTTG TCCCGCCAGG  
 36996873 GCGTATTAC AGGGGGAGGT GTGGGCTGC AGGCTACACT TTCCCATCG ACCCTGGGAG  
 36996933 TCCAGCCCT CAGGAACCCG CGCCTGGGC TTAGCCAGGG TGACAGAGAG CAGGCCCTCT  
 36996993 GCACCTTAA CCTAGTTACC TGCTAGCTCT CCATGAAGT AATGGCCCT TAACATCCAC  
 36997053 GTAGAGGAAG CCTGCTGCTG C

**Secuencia 46: L55429 (cromosoma 5: 1547854 - 1548287)**

1547854 CGATGGAGGA CATCGTCATG GTCACAGGGA TGGCGTCGAA GTAGGACGAG TGAGGCGCGA  
 1547914 GCGTGAGGAT GGCCGCTCG GTGGGACGCG CCTGCGGCC CTTACGGCC ACCCGGTGGA  
 1547974 AGCCGCCGGC GAACCATATG GTGCGCATGA TGGCCTTCAG CAGGAAGTCC ACAACCTTGC  
 1548034 AAAAGAGGGC GCTGCGTCAC GCGGGCACAC GTCGCGAGTC TCGGAGTCTG TGTGAGGCAC  
 1548094 AGGGGCGGTC CCACGGGAGA GCCCTCCAG GCGCAGTCCA GGCCAGGGG TFCCTGTGGT  
 1548154 CGCCGCCGCT GGGACATCTG CTTGAGGGAA GAAAAGACGC CGCGCCTCC TGGCTCCGC  
 1548214 CTGTGCTCTC GCTGGCTGCG CTCTACCTA CGAAGGAGG CGCGGGGCC TGTGTGGTGG  
 1548274 TCACGGGCA CGC

15

**Secuencia 47: L62086 (cromosoma 19: 40478392 - 40478802)**

40478392 CGTCCATCTC GGCCTTCGAA GGCAGTGC GTCCTCATCC CTGCCGCTTT GACTTCCCGG  
 40478452 ATGAGCTGCG GCCCGTGTG GTGCATGGTG TCTGGTACTT CAATAGCCCC TACCCAAAGA  
 40478512 ACTACCCCCC GGTGGTCTTC AAGTCGCGCA CCAAAGTAGT CCACGAGAGC TTCCAGGGCC  
 40478572 GCAGCCGCT CTTGGGGGAC CTGGGCTGC GAAACTGCAC CCTCTGCTC AGCAACGTCA  
 40478632 GCCCGAGCT GGGCGGGAAG TACTACTTCC GTGGGGACCT GGGCGGCTAC AACCAATACA  
 40478692 CCTTCTCAGA GCACAGCTC CTGGATATCG TCAGTGTGTC CCCAGCGGTT GTGCAGGCAC  
 40478752 CGGGAGCTGG GGCAGCGGG CCGGAAGGAG TGTGGCCGGA AGGCCTCCCC G

**Secuencia 48: L76138 (cromosoma 19: 3129742 - 3130341)**

3129742 CGGGTGGCTC CACCCTGCGT CGGGCCTCAG TCAGCCCCCG GGGGAGGCCA TGAACGCCAC  
 3129802 GGGGACCCCG GTGGCCCCCG AGTCCCTGCCA ACAGCTGGCG GCCGGCGGGC ACAGCCGGCT  
 3129862 CATTGTTCTG CACTACAACC ACTCGGGCCG GCTGGCCGGG CGCGGGGGGC CGGAGGATGG  
 3129922 CGGCCTGGGG GCCCTGCGGG GGC'TGTGGT GGC'CGCCAGC TGCCCTGGTGG TGCTGGAGAA  
 3129982 CTTGCTGGTG CTGGCGGCCA TCACCAGCCA CATGCGGTGCG CGACGCTGGG TCTACTATTG  
 3130042 CCTGGTGAAC ATCACGCTGA GTGACCTGCT CACGGGCGCG GCCTACCTGG CCAACGTGCT  
 3130102 GCTGTGCGGG GCCCGCACCT TCCGTCTGGC GCCCGCCAG TGGTTCCTAC GGGAGGGCCT  
 3130162 GCTCTTCACC GCCCTGGCCG CCTCCACCTT CAGCCTGCTC T'CACTGCAG GGGAGCGCTT  
 3130222 TGCCACCATG GTGCGGCCCG TGGCCGAGAG CGGGGCCACC AAGACCAGCC GCGTCTACGG  
 3130282 CTTTCATCGC CTCTGCTGGC TGCTGGCCCG GCTGCTGGGG ATGCTGCCTT TGCTGGGCTG

**Secuencia 49: L15952 (cromosoma 7: 2740971 - 2741510)**

2740971 CGGGGCGAGC ACGTGCACCT CGTGACCACC GGCTCTTAG AGCCCCAGGC AGAGGCCAG  
 2741031 CTCTGCAGTA GGGAAAGGCAT CGGGAGTGCC AGGGTGAACG TACCCCAAGG GCCCGGCAGC  
 2741091 ACTGACCTCT CGTGCCTGCT CTCTCCCTTC CTCGCCAAGC CCCCGTGATG TGGGAAGCCA  
 2741151 GCGTGAGGCC GGTGGGGCAG CCGCCTTCCC GTGGCTGTGC CAAGTCCCC CGGTCTCTG  
 2741211 CACATCATGC CTCCTTCCAC ACCCTGACAG GAAGCAGCTG GGAGAAGCCG TTGGGTGCAC  
 2741271 TCAC'TCCCTG ATTTACGACA AGTTCCTTCC TCAGCGCCTC TCTCTCCTGC CTCCTCCTGC  
 2741331 TCTCTGCCCT TCCCTTGGCG CTCGGGAGGT GCCACGCAAG CCAAGAAGC ATCAGCATAC  
 2741391 TGTCCTTCCC TCTCCTGTGG CCACGGGCTC CCCAGGGAG CTGAGAGTAG CAGCAGCTCA  
 2741451 CAGCCCAAGC CACCCTTGCC CGT'TTCTAGG CAGGTGGTGG CACCAGGCAC GAAGGAAGCA

5

**Secuencia 50: L36599 (cromosoma 19: 4867561 - 4867784)**

4867561 CGGTGACCCG GCCGGTCTGA AGGGCAGAGT TCCGCTGTCA CTAGCCCTCC ACCCGTCTG  
 4867621 TGTGCTGGGA TGCCCTCGCG GCGCCGTCCA CGCCACCGCC GCCCCCTCTT GTGGGTCTG  
 4867681 TCTCTCCCGT GTCTAGGATC CTCTGCATC CGTTT'TCCT TCCTCCCTTC TCTCCCTCCG  
 4867741 TCTGTCTTGC CCGCACCTGA GGT'TGTGCA GAGGCGCTGA GACC

**Secuencia 51: L26688 (cromosoma 17: 77844647 - 77844962)**

77844647 CGCGGGTGC GGGGCTGCGG GGAGGGAGGG GCGGGTTCAG GGCTGTGGGC CCCGCCGGG  
 77844707 GCACCCGCTG GATCAGGCC'T GAGGCCGGCT' GCAGAGAACT TGGGCACCGC GGACCAGATT  
 77844767 GCTGGTCACC AGGATGAAAA ATAAAAAGA CCCAAACAGC TCCCCCGACC CCGCTTTCGC  
 77844827 GCAGGCCCTT GCGGT'TCTG TCAGGGGTTT T'TGTT'FAATG AAGAACGTCA AACATCTGGC  
 77844887 AAGGTCCGAA TGATTTTGGC ATGAGCACCG CCTGGAATTC GCGGTGAAAG CGACTGTGTG  
 77844947 CGCGCTCACG GCCTCG

10

**Secuencia 52: L81528 (cromosoma 19: 47395445 - 47395816)**

47395445 CGGGCGCCGC CCCCTT'PCT C'TCCATCAG CAACAGGCGG CGCCGGCCAG CCTCATAGTC  
 47395505 AGCCTCATCC AACTGACCA GCAGGCGAAC AGCCT'CCCG CCCACAGCCT CTGCGAGGGC  
 47395565 CTCAGTCAAG AACACGCCCC GCAGGGCCTG CAGCAGGGCG CCAC'TCAGGT AGTCGCCCCA  
 47395625 GAAGGCGTCC AGATAGGAGA GCTCTGAGAA CTTGATGTCA CAAACCACAG AGCCAGGTC  
 47395685 CCTTGAGCGC AGCACTGCGG TGGCT'GCCC AAACACGTCC AGCTGCCGCG CCAGGCGCTG  
 47395745 GGGCCGCCCG GATGCCACGC CCTGCT'CCAA GGCTGGCCCA TGCTCGCAGT ACTCTGCTCG  
 47395805 AACCCGAGC CG

15

**Secuencia 53: L36556 (cromosoma 19: 50962003 - 50962260)**

50962003 CGCCAGGGC CAGGCCGTTG ATGATGACGG GGC'CCCGTT GAGGAGCACT GCTGGGGAGC  
 50962063 CGCTGGCTGC CAGGAAGCTC CCGTTCACCA GGATGGAGGA GGAAGCCGGG CAAGGCGCGG  
 50962123 GAGGGCCGGT CCCTGCCAGG AATATGGAGC C'TGGGCAGC GGCCTCGGCG GACTCTGGGG  
 50962183 CCGCCCTCT' CTCCAGGTCC TCAGGACTTC GGCTGGACTC GTCC'FCAGTC GTGGGATTC  
 50962243 CATCAGACTC GCTGGCCG

**Secuencia 54: L16264 (cromosoma 2: 2733750 - 2734147)**

2733750 CGCCTTTTCC TTCGGTCACT CAGACCGCGC CCGAGGTCCC TGGTGCCCC GCGGGGAGC  
 2733810 TGAGGTTTGC GGTCCCTCCT GGGGTTCGCT TCCCGCGGGG CTA'AAACCGG GCGACCAGGG  
 2733870 CCCCTTTCTC CACTGGTGC TTTTCCGGGA AACGCTGCTC CTTAGATGGA CGAATACGTA  
 2733930 CTCGGTACCC AGCACATCCT GGACGAGTTA ACTTCCTTAA TTTCCATA'TI TGCCGGAGAA  
 2733990 CCAGCTGTT CAGAGCCAC AGCAAACGGG AGAAGTTAAA CCAGGTTCCG CGACCCAGAG  
 2734050 CCCAGGGGTG GCCCCGGGCA CACCACCTG ACTCCGACC CCCCACGAGA GGGGAGGATC  
 2734110 CTTGCAGACC TCACCTT'GCG TGCAACGCT' GCGGCCCG

20

## REIVINDICACIONES

1. Un método para categorizar una muestra de ADN mediante la determinación de relaciones de señal entre loci genómicos co-amplificados en la misma reacción, que comprende:
- 5 (A) digerir una muestra de ADN con una endonucleasa de restricción sensible a la metilación o una endonucleasa de restricción dependiente de la metilación;
- (B) amplificar al menos dos loci genómicos del ADN digerido, en el que al menos uno de los loci es un locus de restricción;
- 10 (C) determinar la intensidad de la señal del producto de amplificación de cada locus;
- (D) calcular relaciones de señal entre las intensidades de señal de los productos de amplificación; y
- (E) comparar las relaciones de señal con valores de referencia de diferentes categorías de ADN,
- en el que la aproximación del valor de una relación de señal con respecto a los valores de las relaciones de referencia de una categoría de ADN particular indica la fuente categórica de la muestra de ADN en la que los loci genómicos que se amplifican se eligen para producir distintas relaciones de señal para diferentes categorías potenciales, en el que no se necesita ninguna curva estándar o un ADN de referencia y en el que no se indica ningún nivel real de metilación en ningún locus genómico por dicho método.
- 20 2. El método de la reivindicación 1, en el que los valores de referencia con los que se comparan las relaciones de señal son de al menos una categoría de ADN seleccionada del grupo que consiste en tipo de tejido, tipo de célula, condición fisiológica, condición patológica, edad, etnia, sexo, nivel de metilación, especie, líneas celulares, ADN natural y ADN sintetizado artificialmente, riesgo de desarrollar una condición patológica, condiciones prenatales, pronóstico, propensión a responder al tratamiento, efectos del subcultivo de la línea celular, respuesta a la medicación.
3. El método de la reivindicación 2, en el que el tipo de tejido o tipo de célula se selecciona del grupo que consiste en sangre, saliva, semen, epidermis, orina, plasma, cabello, sangre menstrual, células y/o secreción vaginal, sudor, heces, cerebro, esófago, pulmón, estómago, corazón, duodeno, hígado, vesícula biliar, intestino, riñón, glándula suprarrenal, vejiga urinaria, uretra, colon, testículo, ovario, útero, vagina, músculo, tendón, ligamento, grasa, cartílago, hueso, células endoteliales, cuello uterino, linfoma, tiroides, glándula pituitaria, cerebelo y mama.
- 30 4. El método de la reivindicación 2, en el que la afección patológica es cáncer, inflamación, trastorno autoinmune, trastorno metabólico, infección, enfermedad degenerativa, desequilibrios hormonales, un trastorno **caracterizado por** metilación anormal del ADN, trastorno del desarrollo neurológico de ICF (inmunodeficiencia, inestabilidad centromérica y anomalías faciales), síndrome de Rett y síndrome de X frágil; y en el que la afección prenatal es síndrome de Prader-Willi, síndrome de Angelman, síndrome de Beckwith-Wiedemann, síndrome de X frágil, síndrome de Russell-Silver, diabetes mellitus neonatal transitoria, osteodistrofia hereditaria de Albright, síndrome de McCune-Albright, paraganglioma familiar no cromafínico, síndromes UPD14 materno y paterno.
- 40 5. El método de la reivindicación 1, en el que los valores de referencia contra los cuales se comparan las relaciones de señal representan categorías que son mezclas de otras categorías.
6. El método de la reivindicación 5, en el que los valores de referencia contra los que se comparan las relaciones de señal son de mezclas de semen y diferentes de semen en diferentes relaciones.
- 45 7. El método de la reivindicación 1, en el que la endonucleasa de restricción sensible a la metilación se selecciona del grupo que consiste en AatII, Acc65I, Accl, Acil, ACII, Afel, Agel, Apal, ApaLI, AscI, AsiSI, Aval, Avall, Bael, BanI, Bbel, BceAI, Bcgl, BfuCI, BgII, BmgBI, BsaAI, BsaBI, BsaHI, Bsal, BseYI, BsiEI, BsiWI, BslI, BsmAI, BsmBI, BsmFI, BspDI, BsrBI, BsrFI, BsshII, BssKI, BstAPI, BstBI, BstUI, BstZ17I, Cac8I, ClaI, DpnI, DrdI, Eael, EagI, EagI-HF, EciI, EcoRI, EcoRI-HF, Faul, Fnu4HI, Fsel, FspI, Haell, Hgal, Hhal, HincII, HincII, HinfI, HinPII, Hpal, HpaII, Hpy166ii, Hpy188iii, Hpy99I, HpyCH4IV, KasI, MluI, Mmel, MspAII, MwoI, NaeI, NarI, NgoNIV, Nhe-HFI, NheI, NlaIV, NotI, NotI-HF, Nrul, Nt.BbvCI, Nt.BsmAI, Nt.CviPII, PaeR7I, PteI, PmeI, PmlI, PshAI, PspOMI, PvuI, RsaI, RsrII, SacII, Sall, Sall-HF, Sau3AI, Sau96I, ScrFI, Sfil, SfoI, SgrAI, SmaI, SnaBI, TfiI, TscI, TseI, TspMI, y ZraI; y en el que la endonucleasa de restricción dependiente de la metilación se selecciona del grupo que consiste en McrBC, McrA, y MrrA.
- 55 8. El método de la reivindicación 1, en el que la endonucleasa de restricción sensible a la metilación es Hhal.

9. El método de la reivindicación 2, en el que la categoría de ADN de referencia es ADN natural o ADN sintetizado artificialmente, en el que al menos un locus genómico en el ADN natural o ADN sintetizado artificialmente comprende un locus de repetición corta en tándem (STR) de núcleo usado para el perfil de ADN.
- 5
10. El método de la reivindicación 9, en el que el locus genómico comprende un locus humano seleccionado del grupo que consiste en D16S539 (SEQ ID NO. 1), D7S820 (SEQ ID NO. 2), D13S317 (SEQ ID NO. 3), D5S818 (SEQ ID NO. 4), CSF1PO (SEQ ID NO. 5), TPOX (SEQ ID NO. 6); TH01 (SEQ ID NO. 7), vWA (SEQ ID NO. 8), FGA (SEQ ID NO. 9), D21S11 (SEQ ID NO. 10), D8S1179 (SEQ ID NO. 11), D18S51 (SEQ ID NO. 12), y
- 10 D3S1358 (SEQ ID NO. 13), Penta D (SEQ ID NO. 14), Penta E (SEQ ID NO. 15), y Amelogenina (SEQ ID NOs. 16 y 17), D2S1338 (SEQ ID No. 18), D19S433 (SEQ ID No. 19), ACTBP2SE33 (SEQ ID No. 20), D10S1248 (SEQ ID No. 21), D1S1656 (SEQ ID No. 22), D22S1045 (SEQ ID No. 23), D2S441 (SEQ ID No. 24), y D12S391 (SEQ ID No. 25).
11. El método de la reivindicación 2, en el que la categoría de ADN de referencia es ADN natural o ADN
- 15 sintetizado artificialmente, en el que se sabe que al menos un locus genómico está metilado en todos los tejidos de ADN natural, o se sabe que no está metilado en todos los tejidos de ADN natural.
12. El método de la reivindicación 11, en el que se sabe que el locus genómico está metilado en todos los tejidos de ADN natural es TPOX (SEQ ID NO. 6) o FGA (SEQ ID NO. 9), y en el que se sabe que el locus genómico
- 20 está sin metilar en todos los tejidos de ADN natural es Hypo23 (SEQ ID NO. 29), Hypo28 (SEQ ID NO. 30) o Hypo33 (SEQ ID NO. 31).
13. El método de la reivindicación 9, en el que los cebadores para la amplificación de al menos uno de los loci genómicos están disponibles en:
- 25
- (1) un kit comercial de Promega Corporation seleccionado del grupo que consiste en PowerPlex® 16 HS (Cat. N.º DC2100, DC2101), PowerPlex® 16 (Cat. N.º DC6530, DC6531), PowerPlex® 2.1 (Cat. N.º DC6470, DC6471), PowerPlex® 16 BIO (Cat. N.º DC6540, DC6541), y PowerPlex® ES Systems (Cat. N.º DC6730, DC6731);
- 30 (2) un kit comercial de Applied Biosystems seleccionado del grupo que consiste en SGM, SGM+, AmpFISTR Identifiler; AmpFISTR Profiler, AmpFISTR ProfilerPlus, AmpFISTR ProfilerPlusID, AmpFISTR SEfiler, AmpFISTR SEfiler Plus, AmpFISTR Cofiler, AmpFISTR Identifiler Direct, AmpFISTR Identifiler Plus, AmpFISTR NGM, AmpFISTR Y-filer, y AmpFISTR Minifiler; o
- 35 (3) Investigator ESSPlex, Investigator ESSPlex SE, Investigator Nonaplex ESSPlex, Investigator Hexaplex ESSPlex, Investigator Triplex AFS QS, Investigator Triplex DSF, Investigator IDplex, Investigator Decaplex SE, Investigator HDplex, Investigator Argus X-12, Investigator Y-12 QS, Investigator DIPplex.
14. El método de la reivindicación 9, que comprende además calcular al menos una de las siguientes relaciones:
- 40
- (1) D3S1358/D18S51,  
 (2) D3S1358/D7S820,  
 (3) D3S1358/Penta\_D,  
 (4) D3S1358/TPOX,
- 45 (5) D3S1358/FGA,  
 (6) TH01/Penta\_D,  
 (7) D21S11/D18S51,  
 (8) D21S11/D7S820,  
 (9) D21S11/Penta\_D,
- 50 (10) D21S11/AMEL,  
 (11) D21S11/TPOX,  
 (12) D21S11/FGA,  
 (13) D5S818/D18S51,  
 (14) vWA/D18S51,
- 55 (15) D5S818/Penta\_E,  
 (16) vWA/Penta\_E,  
 (17) D5S818/D7S820,  
 (18) D5S818/Penta\_D,  
 (19) D5S818/TPOX,

- (20) D5S818/FGA,  
 (21) D13S317/D7S820,  
 (22) D13S317/Penta\_D,  
 (23) D13S317/TPOX,  
 5 (24) D13S317/FGA,  
 (25) D16S539/D7S820,  
 (26) CSF1PO/D7S820,  
 (27) vWA/D7S820,  
 10 (28) D8S1179/D7S820,  
 (29) D16S539/TPOX,  
 (30) D16S539/FGA,  
 (31) CSF1PO/Penta\_D,  
 (32) CSF1PO/TPOX,  
 15 (33) vWA/Penta\_D,  
 (34) AMEL/TPOX,  
 (35) AMEL/FGA,  
 (36) vWA/D8S1179,  
 (37) vWA/TPOX,  
 20 (38) vWA/FGA,  
 (39) D8S1179/TPOX, y  
 (40) D8S1179/FGA.
15. El método de la reivindicación 1, en el que la intensidad de la señal es el nivel de fluorescencia del producto de amplificación medido durante la electroforesis capilar.  
 25 16. El método de la reivindicación 1, en el que las etapas (B) y (C) se realizan por PCR en tiempo real.
17. El método de la reivindicación 16, en el que la intensidad de señal de cada producto de amplificación es el nivel de fluorescencia de cada producto de amplificación.  
 30 18. El método de la reivindicación 1, que comprende además asignar niveles de confianza a categorías potenciales de una muestra de ADN, en el que el nivel de confianza refleja la probabilidad de que una categoría de ADN particular indique la fuente categórica de la muestra de ADN; en el que la probabilidad se calcula:
- 35 (A) asignando una puntuación de probabilidad a cada comparación de una relación de señal a los valores de referencia que corresponden a las diversas categorías de ADN, en el que la puntuación de probabilidad se obtiene (i) asignando una función de distribución de probabilidad gamma a los valores de referencia de cada categoría potencial; en el que (ii) la puntuación de probabilidad es igual al valor de la función de distribución de probabilidad gamma asignada en la etapa (A) a la relación de señal observada;  
 40 (B) para cada categoría, calcular la puntuación de probabilidad de la categoría, que es el producto de las puntuaciones de probabilidad obtenidas en la etapa (A); y  
 (C) normalizar las puntuaciones de probabilidad dividiendo la puntuación de probabilidad de categoría de cada categoría potencial por la suma de todas las puntuaciones de probabilidad de categoría;
- 45 en el que el nivel de confianza de cada categoría es la puntuación de probabilidad normalizada de esa categoría.
19. El método de la reivindicación 1, que comprende además realizar simultáneamente el perfil de ADN con la categorización de ADN.
- 50 20. Un kit para categorizar una muestra de ADN como semen o diferente de semen, que comprende:
- (a) una mezcla de cebadores, que comprende los siguientes cebadores:  
 SD1f (AAGAGCCCATCAGGCAGGTC);  
 SD1r (GTTTCTTGTGCGAGCAGCACGTGGATGATG);  
 55 SD2f (CTCCAGAACTGGAACCTTCCTG);  
 SD2r (GTTTCTTAACTTGGAGACGACGGCATC);  
 SD3f (TGGAGGACAATGCCCTGGTG);  
 SD3r (GTTTCTTGGCTTCACCTGCGACCGTCTC);  
 SD4f (CCCTCCGAGTGGCCAGCAG);

SD4r (GTTTCTGACCACTGCCGTGGGAATG);  
 SD5f (CTTCTCAGCCAATGGGAAGAG);  
 SD5r (ACGTAGAAGGACCCGAGGAC);  
 (b) tampón de reacción;  
 (c) endonucleasa de restricción HhaI; y  
 (d) ADN polimerasa;

5

en el que al menos un par de combinaciones de par de cebadores directos (f) e inversos (r) se selecciona del grupo que consiste en:

10

(1) SD1f (AAGAGCCCATCAGGCAGGTC) y SD1r (GTTTCTTGTCGAGCAGCACGTGGATGATG);  
 (2) SD2f (CTCCAGAAGTGGAACTTCCTG) y SD2r (GTTTCTTAACTTGGAGACGACGGCATC);  
 (3) SD3f (TGGAGGACAATGCCCTGGTG) y SD3r (GTTTCTTGCTTACCTGCGACCGTCTC);  
 (4) SD4f (CCCTCCGAGTGGCCAGCAG) y SD4r (GTTTCTGACCACTGCCGTGGGAATG);  
 (5) SD5f (CTTCTCAGCCAATGGGAAGAG) y SD5r (ACGTAGAAGGACCCGAGGAC).

15

21. El kit de la reivindicación 20, en el que la concentración de los cebadores en la mezcla de cebadores son: SD1f 0,6 µM, SD1r 0,6 µM, SD2f 1,75 µM, SD2r 1,75 µM, SD3f 1,25 µM, SD3r 1,25 µM, SD4f 1,75 µM, SD4r 1,75 µM, SDSf 1,75 µM y SD5r 1,75 µM.

20

22. Un kit para analizar, categorizar una muestra de ADN como semen o diferente de semen, y obtener un nivel de confianza de categorización asociado, que comprende:

(a) cebadores para amplificar al menos un locus específico de semen;  
 (b) cebadores para amplificar al menos un locus usado para el perfil de ADN;  
 (c) tampón de reacción;  
 (d) endonucleasa de restricción HhaI; y  
 (e) ADN polimerasa, en el que

25

el al menos un locus específico de semen es el locus L68346, y en el que los cebadores para amplificar un producto de amplificación específico de semen de 70 pb de L68346 son un cebador directo que comprende la secuencia de CAGCAACAGCACCCAGCTTG (FAM) y un cebador inverso que comprende la secuencia de CACAGGCTCAGTCGCGGATC; o en el que el al menos un locus específico de semen es el locus L16264, y en el que los cebadores para amplificar un producto de amplificación específico de semen de 95 pb de L16264 son un cebador directo que comprende la secuencia de GGACGAGTTAACTTCCTTAATTTT (FAM) y un cebador inverso que comprende la secuencia de GTTTCTTCGCGGAACCTGGTTAACTTC.

35

23. Un kit para categorizar una muestra de ADN como sangre, saliva, semen o epidermis de la piel, y para obtener un nivel de confianza de categorización asociado, que comprende:

40

(a) mezcla de cebadores que comprende cebadores directo e inverso para amplificar los loci representados como se indica a continuación:

1. L91762 (GCAGCAGGCCGCGGAGAAG directo (FAM); AGCAGCTGTGCCGGGCCAG inverso)
2. L68346 (CAGCAACAGCACCCAGCTTG directo (JOE); CACAGGCTCAGTCGCGGATC inverso)
3. L50468 (AGGAAACCTCAGTAGCAAAATTG directo (JOE); GCGAGACTTTAGGTGTGCATC inverso)
4. L14432 (CGTAGGCTGCGGTGAGCTC directo (FAM); GATCCATGCCCGCTGGGATG inverso)
5. L4648 (CAGCCTAGACGTCAAGTTACAG directo (JOE); ACGACCTCCGGATCCAACCTG inverso)
6. L39664 (CCCAGCTGGTTGGACATGTTG directo (FAM); CACTTCTTCGTGGACGCC inverso)
7. L30139 (GAGAAGCGGGAGGATGAGAC directo (FAM); CCGCATCTCCTCCGTCCTG inverso)
8. L55429 (GCCTTCAGCAGGAAGTCCAC directo (JOE); CCTGTGCCTCACACAGACTC inverso)
9. L62086 (GTGCATGGTGTCTGGTACTTC directo (FAM); GAAGCTCTCGTGGACTACTTG inverso)
10. L76138 (CAGCCTGCTCTTCACTGCAG directo (JOE); AGAGGCCGATGAAGCCGTAG inverso)
11. L15952 (CTCCCTGATTTACGACAAGTTC directo (FAM); GACAGTATGCTGATGCTTCTTG inverso)
12. L36599 (AAGGGCAGAGTTCGCTGTC directo (FAM); CGGATGCAGGAGGATCCTAG inverso)
13. L26688 (CGACCAAGATTGCTGGTCCAC directo (JOE); CGACCTTGCCAGATGTTTGAC inverso)
14. L81528 (AGCCTCATCCACACTGACCAG directo (JOE); TCAGAGCTCTCCTATCTGGAC inverso)
15. L36556 (GCCAGGCCGTTGATGATGAC directo (JOE); GAATATGGAGCCCTGGGCAG inverso)

55

(b) tampón de reacción;  
 (c) endonucleasa de restricción HhaI; y

(d) ADN polimerasa.

Figura 1

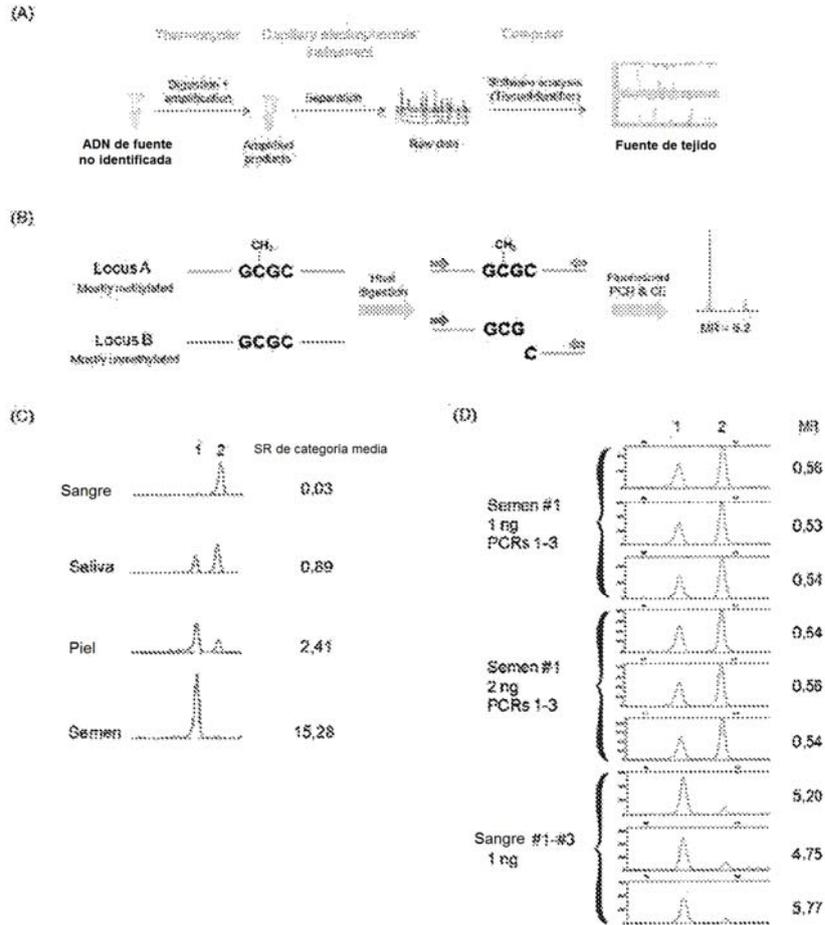


Figura 2

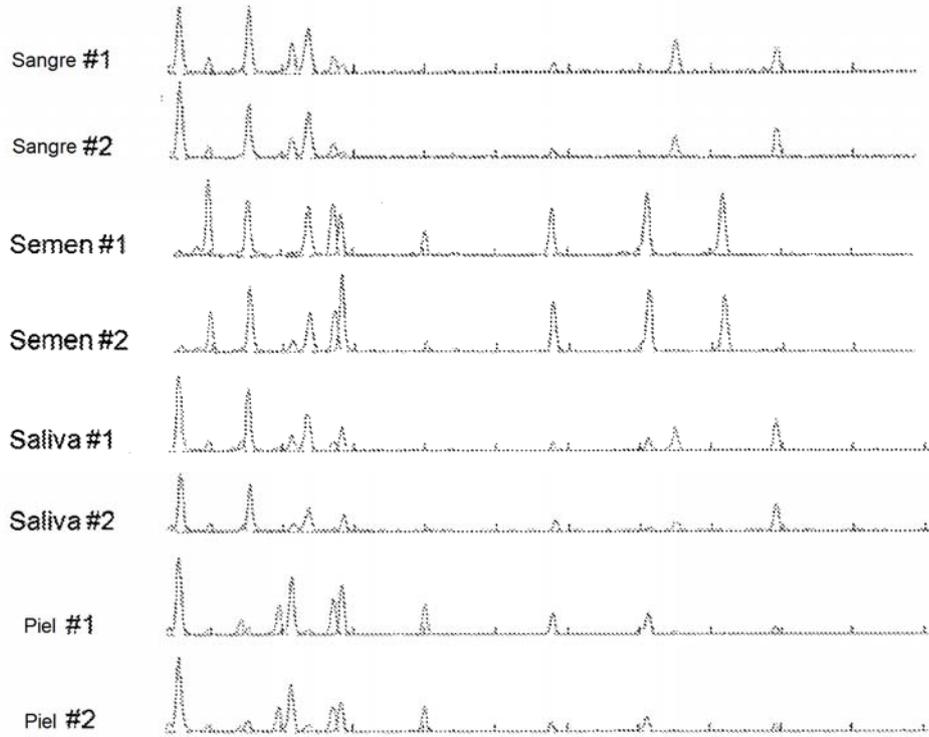


Figura 3

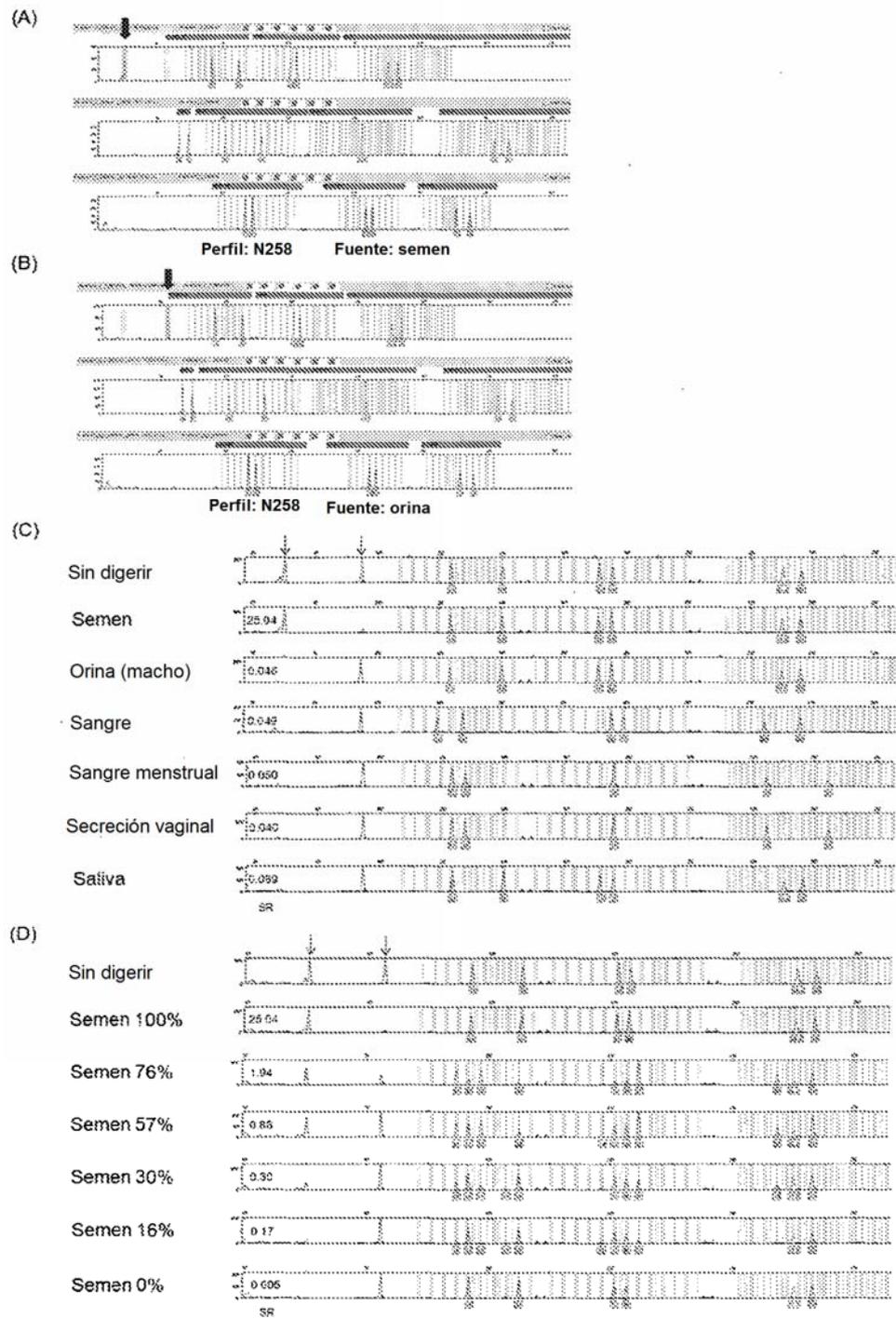


Figura 4

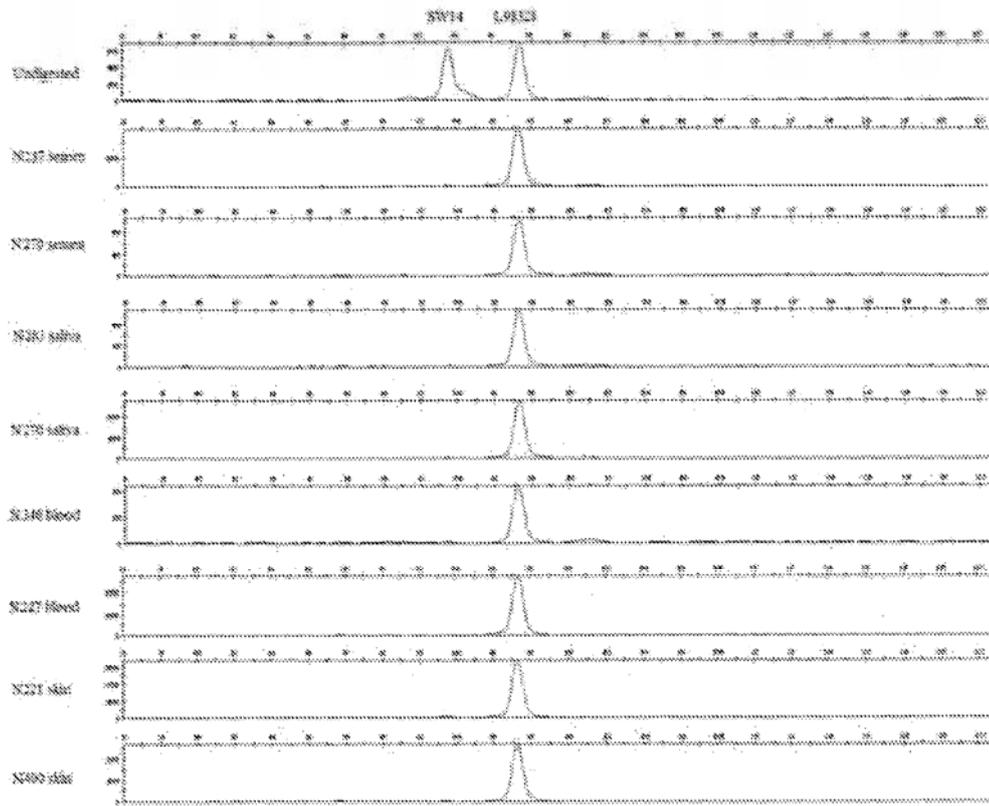


Figura 5

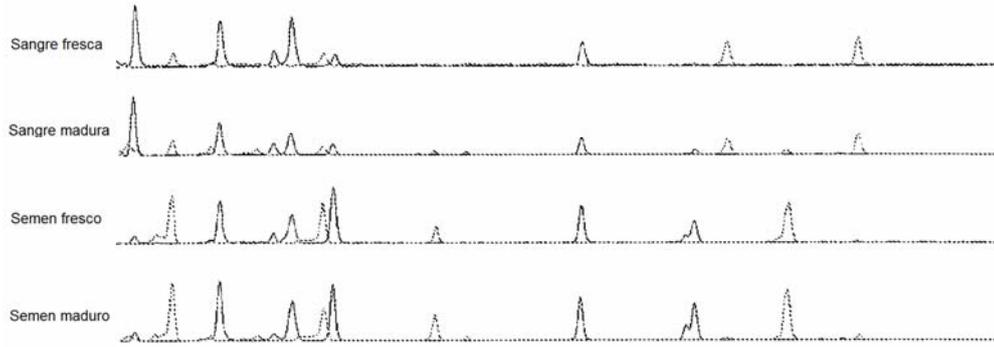
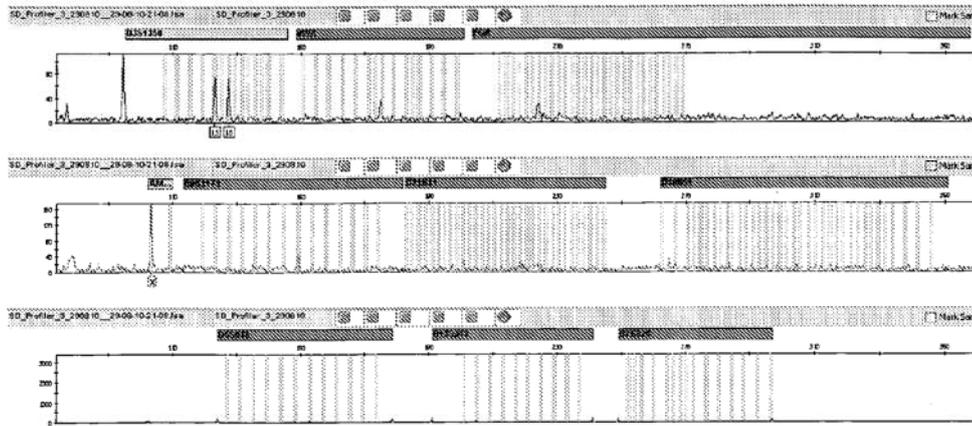


Figura 6

Muestra de sangre tratada con DNaseI



Muestra de semen tratada con DNaseI

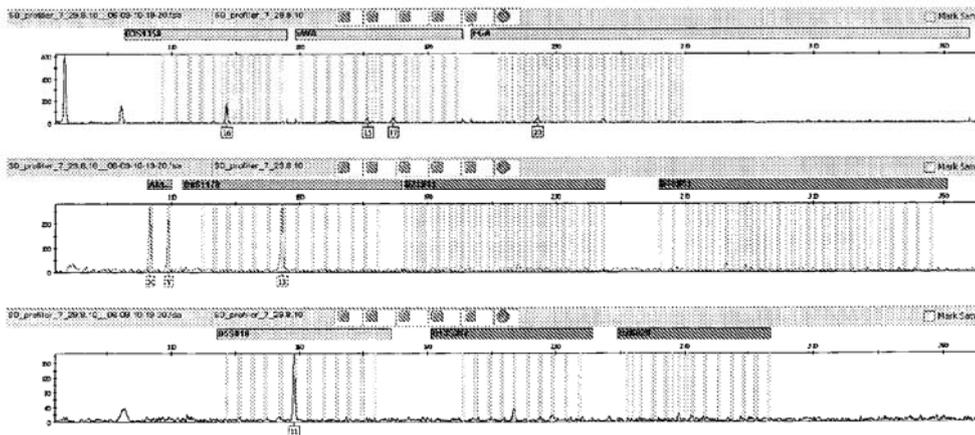


Figura 7

**Tabla 1 - Panel de loci usado para el ensayo de identificación tisular**

#	Locus	Ubicación <sup>a</sup>	Cebador directo	Cebador inverso	Tamaño <sup>b</sup>	Conc. <sup>c</sup>
1	L91762	Chr12 (73985697)	GCAGCAGGCCGCGGAGAAG (FAM)	AGCAGCTGTGCCGGCCAG	66	0.2µM
2	L68346	Chr3 (12175317)	CAGCAACAGCACCCAGCTTG (JOE)	CACAGGCTCAGTCGGGATC	70	0.2µM
3	L50468	Chr3 (55492965)	AGGAAACCTCAGTAGCAAAATTG (JOE)	GCGAGACTTTAGGTGTGCATC	75	0.2µM
4	L14432	Chr22 (29866149)	CGTAGGCTGCGGTGAGCTC (FAM)	GATCCATGCCCGCTGGGATG	75	0.2µM
5	L4648	Chr1 (35815331)	CAGCCTAGACGTCAAGTTACAG (JOE)	ACGACCTCCGGATCCAACG	80	0.2µM
6	L39664	Chr17 (53710330)	CCCAGCTGGTTGGACATGTTG (FAM)	CACCTCCTCGTGGACGCC	82	0.2µM
7	L30139	Chr17 (36996489)	GAGAAGCGGGAGGATGAGAC (FAM)	CCGCATCTCCTCCGTCCTG	84	0.2µM
8	L55429	Chr5 (1548043)	GCCTTCAGCAGGAAGTCCAC (JOE)	CCTGTGCCTCACACAGACTC	88	0.2µM
9	L62086	Chr19 (40478538)	GTGCATGGTGTCTGGTACTTC (FAM)	GAAGCTCTCGTGGACTACTTG	89	0.1µM
10	L76138	Chr19 (3130217)	CAGCCTGCTCTTCACTGCAG (JOE)	AGAGGCCGATGAAGCCGTAG	100	0.2µM
11	L15952	Chr7 (2741305)	CTCCCTGATTTACGACAAGTTC (FAM)	GACAGTATGCTGATGCTTCTTG	117	0.2µM
12	L36599	Chr19 (4867642)	AAGGGCAGAGTCCGCTGTC (FAM)	CGGATGCAGGAGGATCCTAG	130	0.2µM
13	L26688	Chr17 (77844826)	CGGACCAGATTGCTGGTAC (JOE)	CGACCTTGCCAGATGTTGAC	134	0.2µM
14	L81528	Chr19 (47395603)	AGCCTCATCCACACTGACCAG (JOE)	TCAGAGCTCTCTATCTGGAC	141	0.2µM
15	L36556	Chr19 (50962118)	GCCAGGCCGTTGATGATGAC (JOE)	GAATATGGAGCCCTGGGCAG	149	0.3µM

<sup>a</sup> La ubicación genómica del CG examinado de acuerdo con la versión UCSC hg18 se proporciona entre paréntesis tras el número de cromosoma

<sup>b</sup> en pb

<sup>c</sup> La concentración se refiere tanto a cebadores directos como inversos

Figura 8

Tabla 2. Relaciones de señal de L91762/L68346 y L76138/L26688				
	L91762/L68346		L76138/L26688	
	media±est.	intervalo	media±est.	intervalo
Semen	0,16±0,14	0,04-0,53	6,50±5,80	2,04-19,00
Sangre	4,36±1,69	2,15-7,73	0,30±0,20	0,08-0,76
Saliva	7,10±3,30	3,36-15,38	0,56±0,41	0,16-1,54
Epidermis	11,40±4,20	4,60-18,28	8,41±5,58	2,1-19,56

Figura 9

<b>Tabla 3. Loci usados para evaluar la digestión de HhaI</b>			
<b>Localización*</b>	<b>Cebador directo</b>	<b>Cebador inverso</b>	<b>Tamaño (pb)</b>
Chr7 (141084738)	GCGAAGGAAGTGCTGGAGTC (FAM)	GTTTCTTGAAAGGGCCAGACAC	94
Chr11 (86241803)	CAAAGTACTGGGGTTACAGGTG (FAM)	GGATGAACCTTTAAGACATCATC	97

\* La ubicación genómica del CG examinado de acuerdo con la versión UCSC hg18 se proporciona entre paréntesis tras el número de cromosoma (en el locus L98328, que no tiene un sitio de reconocimiento de HhaI, la ubicación es del comienzo del cebador directo)

Figura 10

Locus	Localización <sup>a</sup>	Cebador directo	Cebador inverso	Tamaño <sup>b</sup>	Conc. <sup>c</sup>
L68346	Chr3 (12175317)	CAGCAACAGCACCCAGCTTG (FAM)	CACAGGCTCAGTCGCGGATC	70	0,2µM
L16264	Chr2 (2734005)	GGACGAGTAACTTCCTTAATTTC (FAM)	GTTTCTTCGCGGAACCTGGTTTAACTTC	95	0,2µM

<sup>a</sup> La ubicación genómica del CG examinado de acuerdo con la versión UCSC hg18 se proporciona entre paréntesis tras el número de cromosoma

<sup>b</sup> en pb

<sup>c</sup> La concentración se refiere tanto a los cebadores directos como inversos

Figura 11

Tabla 5. Detección y cuantificación del componente de semen en mezclas de semen-saliva			
Número de muestra	% de semen en la muestra (según la tipificación de STR)	% de semen en la muestra (según este ensayo)	¿Semen detectado?
1	100	100	Sí
2	100	100	Sí
3	100	100	Sí
4	76	66	Sí
5	57	47	Sí
6	30	23	Sí
7	16	15	Sí
8	0	0	No
9	0	0	No
10	0	0	No

Cada fila representa una única reacción

Figura 12

