

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 724 328**

51 Int. Cl.:

C12Q 1/68 (2008.01)

A61K 31/713 (2006.01)

A61K 48/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **20.06.2014 PCT/CN2014/000608**

87 Fecha y número de publicación internacional: **29.01.2015 WO15010442**

96 Fecha de presentación y número de la solicitud europea: **20.06.2014 E 14830344 (9)**

97 Fecha y número de publicación de la concesión europea: **06.03.2019 EP 3024949**

54 Título: **B3GNT5 para su uso como biomarcador para nacimiento prematuro**

30 Prioridad:

24.07.2013 US 201361857975 P

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

10.09.2019

73 Titular/es:

**THE CHINESE UNIVERSITY OF HONG KONG
(100.0%)**

**Knowledge Transfer Office (KTO) Room 328, Pi
Ch'iu Building Sha Tin, N.T.
Hong Kong, CN**

72 Inventor/es:

**CHIM, SIU-CHUNG STEPHEN;
CHEUNG, CHEE YIN;
CHEUNG, WAN CHEE;
LEUNG, TAK YEUNG y
LEE, WING SHAN**

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 724 328 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

B3GNT5 para su uso como biomarcador para nacimiento prematuro.

5 Solicitudes relacionadas

Esta solicitud reivindica prioridad de la solicitud de patente provisional U. S. No. 61/857.975, presentada el 24 de julio de 2013.

10 Antecedentes de la invención

En los seres humanos, el nacimiento prematuro o el parto prematuro se refieren al nacimiento a una edad gestacional de menos de 37 semanas. El nacimiento prematuro es una de las principales causas de muerte infantil en todo el mundo. Los bebés que nacen prematuramente también tienen más probabilidades de sufrir diversas complicaciones tanto a corto como a largo plazo, incluidas las discapacidades y los impedimentos en el crecimiento y en el desarrollo mental. Si bien se han logrado avances sustanciales para mejorar la tasa de supervivencia y el desarrollo posterior de los bebés que nacieron prematuramente, la causa precisa de los nacimientos prematuros aún no se ha comprendido por completo. Dada la prevalencia y las implicaciones del parto prematuro, existe la necesidad de nuevos métodos para detectar con mayor precisión un mayor riesgo de parto prematuro en mujeres embarazadas, de modo que se puedan tomar medidas preventivas oportunamente para reducir o eliminar las posibilidades de un parto prematuro. Esta invención satisface esta y otras necesidades relacionadas.

Breve compendio de la invención

25 Los presentes inventores descubrieron que la transcripción de determinados genes marcadores, como se ve a nivel del ARNm, en los glóbulos sanguíneos de una mujer embarazada puede elevarse o suprimirse en correlación con la probabilidad de un nacimiento prematuro. Como tal, en un primer aspecto, la presente invención proporciona un método para determinar el riesgo que corre una mujer embarazada al parir al bebé prematuramente. El método incluye las etapas de: (a) medir el nivel de ARNm de un marcador, que puede ser uno de los genes enumerados en la Tabla 2 o CD16A o CD62L, en una muestra de sangre tomada de una mujer embarazada; y (b) comparar el nivel de ARNm obtenido en la etapa (a) con un control estándar. Cuando se detecta un aumento o una disminución en el nivel de ARNm en comparación con el control estándar, indica que la mujer tiene un mayor riesgo de un parto prematuro. Para cualquier marcador en particular, si un aumento o disminución indica mayor riesgo será evidente según la información proporcionada en esta solicitud, por ejemplo, la Tabla 2. Por ejemplo, cuando el marcador es B3GNT5, CD16A o CD62L, un aumento en el nivel de ARNm cuando se compara con el control estándar indica que la mujer tiene un mayor riesgo de parto prematuro, mientras que, cuando el marcador es CLC o GBP3, una disminución en el nivel de ARNm cuando se compara con el control estándar indica que la mujer tiene un mayor riesgo de parto prematuro. En este método, se pueden usar sangre entera y diversas fracciones de sangre, tales como suero o plasma o glóbulos aislados.

40 En algunas realizaciones, el nivel de ARNm se normaliza sobre el nivel de ARNm de un gen de referencia en la misma muestra antes de la etapa (b). Por ejemplo, el nivel de ARNm de un gen marcador puede expresarse como una relación sobre el nivel de ARNm de un gen de referencia. Un gen de referencia de ejemplo es GAPDH. En algunos casos, el nivel de ARNm de más de un marcador de genes se mide y se compara con sus respectivos controles estándar para determinar el riesgo de nacimiento prematuro

50 En algunas realizaciones, la etapa (a) comprende espectrometría de masas o hibridación con una micromatriz, sonda de fluorescencia o baliza molecular. En algunas realizaciones, la etapa (a) comprende una reacción de amplificación, como una reacción en cadena de polimerasa (PCR), especialmente una reacción en cadena de polimerasa con transcriptasa inversa (RT-PCR) que incluye RT-PCT cuantitativa (qRT-PCR). En algunas realizaciones, la etapa (a) comprende un ensayo de hibridación de polinucleótidos que utiliza una sonda de polinucleótidos que comprende una unidad estructural detectable. Por ejemplo, el ensayo de hibridación de polinucleótidos puede ser un análisis de transferencia Southern, un análisis de transferencia Northern o un ensayo de hibridación in situ.

55 En determinadas realizaciones, cuando se ha indicado que una mujer embarazada tiene un mayor riesgo de parto prematuro, el método puede incluir, además, una etapa terapéutica para reducir o eliminar el riesgo de parto prematuro.

60 En un segundo aspecto, la presente invención proporciona un kit para determinar el riesgo de parto prematuro en una mujer embarazada. El kit incluye estos componentes: (1) un control estándar que proporciona un nivel promedio de ARNm de un gen marcador; y (2) un agente que identifica específica y cuantitativamente el ARNm del gen marcador. El gen marcador se selecciona del grupo que consiste en los genes de la Tabla 2, CD16A y CD62L. En algunas realizaciones, el agente es una sonda polinucleotídica que hibrida con el ARNm del gen marcador. La sonda polinucleotídica incluye opcionalmente una unidad estructural detectable. En algunas realizaciones, el kit incluye, además, dos cebadores de oligonucleótidos para amplificar específicamente, en una reacción de amplificación, al

menos un segmento del ADNc del gen marcador o al menos un segmento del complemento del ADNc del gen marcador. A menudo, el kit contiene, además, un manual de instrucciones.

5 En un tercer aspecto, la presente descripción proporciona un método para reducir el riesgo de parto prematuro o prevenir el parto prematuro. El método incluye la etapa de administrar a la mujer una cantidad eficaz de (1) una secuencia de polinucleótidos antisentido o un ARNsi contra un gen marcador en la Tabla 2, o contra CD16A o CD62L; o (2) un casete de expresión que comprende la secuencia de ADNc de un gen marcador en la Tabla 2 y que dirige la transcripción del gen marcador. En algunas realizaciones, el casete de expresión comprende un promotor unido operativamente a la secuencia de ADNc marcador. La selección de (1) o (2) se basa en si se encuentra que un
10 marcador ARN particular está elevado en relación con el nacimiento prematuro.

Breve descripción de los dibujos

15 Figura 1: Diagramas de caja de las concentraciones del ARNm de B3GA/T5 en sangre de mujeres sintomáticas que dieron a luz antes de las 34 semanas de gestación (grupo de prueba) y las que dieron a luz a las 37 semanas o más tarde (grupo de referencia). La caja se reduce al percentil 25 y hasta el percentil 75. La línea dentro de la caja se dibuja como la mediana. Los bigotes se reducen al percentil 10 y hasta el 90. Los puntos por debajo y por encima de los bigotes se dibujan como puntos individuales.

20 Figura 2: Curva de características operativas del receptor del ARNm de B3GNT5 para predecir el nacimiento antes de las 34 semanas de gestación entre mujeres sintomáticas.

25 Figura 3: Diagramas de caja de concentraciones de ARNm de CLC en sangre de mujeres sintomáticas que dieron a luz antes de las 34 semanas de gestación (grupo de prueba) y aquellas que dieron a luz a las 37 semanas o más tarde (grupo de referencia). La caja se reduce al percentil 25 y hasta el percentil 75. La línea dentro de la caja se dibuja como la mediana. Los bigotes se reducen al percentil 10 y hasta el 90. Los puntos por debajo y por encima de los bigotes se dibujan como puntos individuales.

30 Figura 4: Curva de características operativas del receptor del ARNm de CLC para predecir el nacimiento antes de las 34 semanas de gestación entre mujeres sintomáticas.

35 Figura 5: Diagramas de caja de las concentraciones del ARNm de GBP3 en la sangre de mujeres sintomáticas que dieron a luz antes de las 34 semanas de gestación (grupo de prueba) y las que dieron a luz a las 37 semanas o más tarde (grupo de referencia). La caja se reduce al percentil 25 y hasta el percentil 75. La línea dentro de la caja se dibuja como la mediana. Los bigotes se reducen al percentil 10 y hasta el 90. Los puntos por debajo y por encima de los bigotes se dibujan como puntos individuales.

40 Figura 6: Curva de características operativas del receptor del ARNm de GBP3 para predecir el nacimiento antes de las 34 semanas de gestación entre mujeres sintomáticas.

45 Figura 7: Diagramas de caja de las concentraciones del ARNm de CD16A en la sangre de mujeres sintomáticas que dan a luz antes de las 34 semanas de gestación (grupo de prueba) y las que dan a luz a las 37 semanas o más tarde (grupo de referencia). La caja se reduce al percentil 25 y hasta el percentil 75. La línea dentro de la caja se dibuja como la mediana. Los bigotes se reducen al percentil 10 y hasta el 90. Los puntos por debajo y por encima de los bigotes se dibujan como puntos individuales.

Figura 8: Curva de características operativas del receptor del ARNm de CD16A para predecir el parto antes de las 34 semanas de gestación entre mujeres sintomáticas.

50 Figura 9: Diagramas de caja de las concentraciones del ARNm de CD62L en la sangre de mujeres sintomáticas que dan a luz antes de las 34 semanas de gestación (grupo de prueba) y las que dan a luz a las 37 semanas o después (grupo de referencia). La caja se reduce al percentil 25 y hasta el percentil 75. La línea dentro de la caja se dibuja como la mediana. Los bigotes se reducen al percentil 10 y hasta el 90. Los puntos por debajo y por encima de los bigotes se dibujan como puntos individuales.

55 Figura 10: Curva de características operativas del receptor del ARNm de CD62L para predecir el nacimiento antes de las 34 semanas gestacionales entre mujeres sintomáticas.

Definiciones

60 En esta descripción, las expresiones “nacimiento prematuro” y “parto prematuro” tienen el mismo significado y se refieren al nacimiento de un infante humano de menos de 37 semanas de edad gestacional, por ejemplo, a una edad gestacional de 34 semanas o menos.

65 En esta descripción, el término “o” se emplea generalmente en su sentido incluyendo “y/o”, a menos que el contenido indique claramente lo contrario.

El término “sangre”, como se usa en este documento, se refiere a una muestra o preparación de sangre de un sujeto que se está analizando. El término abarca sangre entera o cualquier fracción de sangre, que puede contener glóbulos o puede ser virtualmente acelular, como plasma o suero.

En esta descripción, el término molécula de ácido nucleico “aislada” significa una molécula de ácido nucleico que se separa de otras moléculas de ácido nucleico que normalmente están asociadas con la molécula de ácido nucleico aislada. Por lo tanto, una molécula de ácido nucleico “aislada” incluye, sin limitación, una molécula de ácido nucleico que está libre de secuencias de nucleótidos que flanquean naturalmente uno o ambos extremos del ácido nucleico en el genoma del organismo del cual se deriva el ácido nucleico aislado (por ejemplo, un ADNc o un fragmento de ADN genómico producido por PCR o digestión con endonucleasa de restricción). Dicha molécula de ácido nucleico aislada se introduce en general en un vector (por ejemplo, un vector de clonación o un vector de expresión) por comodidad de manipulación o para generar una molécula de ácido nucleico de fusión. Además, una molécula de ácido nucleico aislada puede incluir una molécula de ácido nucleico modificada por ingeniería genética, como una molécula de ácido nucleico recombinante o sintética. Una molécula de ácido nucleico que existe entre cientos de millones de otras moléculas de ácido nucleico, por ejemplo, dentro de una biblioteca de ácido nucleico (por ejemplo, un ADNc o una biblioteca genómica) o un gel (por ejemplo, agarosa o poli(acrilamida)) que contiene ADN genómico digerido por restricción, no es un ácido nucleico “aislado”.

La expresión “ácido nucleico” o “polinucleótido” se refiere a ácidos desoxirribonucleicos (ADN) o ácidos ribonucleicos (ARN) y sus polímeros, ya sea en forma de cadena simple o doble. A menos que se limite específicamente, la expresión abarca ácidos nucleicos que contienen análogos conocidos de nucleótidos naturales que tienen propiedades de unión similares a las del ácido nucleico de referencia y se metabolizan de manera similar a los nucleótidos naturales. A menos que se indique lo contrario, una secuencia particular de ácido nucleico también abarca implícitamente sus variantes modificadas conservativamente (por ejemplo, sustituciones de codones degenerados), alelos, ortólogos, polimorfismos de un solo nucleótido (SNP) y secuencias complementarias, así como la secuencia explícitamente indicada. Específicamente, las sustituciones de codones degenerados se pueden lograr generando secuencias en las que la tercera posición de uno o más codones (o todos) seleccionados están sustituidos con residuos de base mixta y/o de desoxinosina (Batzer et al., *Nucleic Acid Res.* 19: 5081 (1991); Ohtsuka et al., *J. Biol. Chem.* 260:2605-2608 (1985) y Rossolini et al., *Mol. Cell. Probes* 8:91-98 (1994)). La expresión ácido nucleico se usa indistintamente con gen, ADNc y ARNm codificados por un gen.

El término “gen” se usa para describir el segmento de ADN involucrado en la producción de una cadena polipeptídica; incluye las regiones que preceden y siguen a la región de codificación (líder y trasera) involucradas en la transcripción/traducción del producto génico y la regulación de la transcripción/traducción, así como las secuencias de intervención (intrones) entre los segmentos de codificación individuales (exones). Tal como se utiliza en esta solicitud, un gen, cuando se identifica específicamente por su nombre (por ejemplo, cualquiera de los enumerados en la Tabla 2 más CD16A y CD62L), abarca cualquier variante o mutante natural de ese gen. Por ejemplo, la secuencia de ADNc del gen B3GNT5 humano se expone en el número de acceso de GenBank NM_032047. Un “gen B3GNT5” en el sentido de esta solicitud incluye variantes que tienen una secuencia de polinucleótidos con al menos el 80%, 85%, 90%, 95%, 98%, 99% o mayor identidad de secuencia con la secuencia de ADNc de NM_032047. El porcentaje de identidad de secuencia para otros genes, incluidos los que se proporcionan en la Tabla 2, se expresa de manera similar. El número de acceso de GenBank para CD16A es NM_000569, NM_001127592, NM_001127593, NM_001127595 o NM_001127596 y para CD62L es NM_000655 o NR_029467.

En esta solicitud, los términos “polipéptido”, “péptido” y “proteína” se usan de manera indistinta en este documento para referirse a un polímero de residuos de aminoácidos. Los términos se aplican a polímeros de aminoácidos en los que uno o más residuos de aminoácidos son un mimético químico artificial de un aminoácido natural correspondiente, así como a polímeros de aminoácidos naturales y polímeros de aminoácidos no naturales. Como se usa en este documento, los términos abarcan cadenas de aminoácidos de cualquier longitud, incluidas proteínas de longitud completa (es decir, antígenos), en donde los residuos de aminoácidos están unidos por enlaces peptídicos covalentes.

El término “aminoácido” se refiere a los aminoácidos naturales y sintéticos, así como a los análogos de aminoácidos y los miméticos de aminoácidos que funcionan de una manera similar a los aminoácidos naturales. Los aminoácidos naturales son aquellos codificados por el código genético, así como aquellos aminoácidos que se modifican posteriormente, por ejemplo, hidroxiprolina, γ -carboxiglutamato y O-fosfoserina. Para los fines de esta solicitud, los análogos de aminoácidos se refieren a compuestos que tienen la misma estructura química básica que un aminoácido natural, es decir, un carbono que está unido a un hidrógeno, un grupo carboxilo, un grupo amino y un grupo R, por ejemplo, homoserina, norleucina, sulfóxido de metionina, metionina metil sulfonio. Dichos análogos tienen grupos R modificados (por ejemplo, norleucina) o esqueletos peptídicos modificados, pero conservan la misma estructura química básica que un aminoácido natural. Para los fines de esta solicitud, los compuestos miméticos de aminoácidos se refieren a compuestos químicos que tienen una estructura que es diferente de la estructura química general de un aminoácido, pero que funciona de una manera similar a un aminoácido natural.

Los aminoácidos pueden incluir aquellos que tienen quiralidad D no natural, como se describe en el documento WO 01/12654, que puede mejorar la estabilidad (por ejemplo, la vida media), la biodisponibilidad y otras características de un polipéptido que comprende una o más de tales D-aminoácidos. En algunos casos, uno o más, y potencialmente todos los aminoácidos de un polipéptido terapéutico tienen quiralidad D.

Se puede hacer referencia a los aminoácidos en el presente documento mediante los símbolos de tres letras comúnmente conocidos o mediante los símbolos de una letra recomendados por la Comisión de Nomenclatura Bioquímica IUPAC-IUB. Los nucleótidos, igualmente, pueden mencionarse por sus códigos de una sola letra comúnmente aceptados.

Como se usa en el presente documento, las expresiones "idéntico" o porcentaje de "identidad", en el contexto de describir dos o más secuencias de polinucleótidos o aminoácidos, se refieren a dos o más secuencias o subsecuencias que son iguales o tienen una porcentaje especificado de residuos de aminoácidos o nucleótidos que son iguales (por ejemplo, un gen B3GNT5 variante utilizado en el método de esta invención tiene al menos el 80% de identidad de secuencia, preferiblemente el 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad, con una secuencia de referencia, por ejemplo, un ADNc de B3GNT5 humano de tipo salvaje establecido en GenBank No. de registro NM_032047), cuando se comparan y alinean para obtener la máxima correspondencia sobre una ventana de comparación, o una región designada según lo medido utilizando uno de los siguientes algoritmos de comparación de secuencias o mediante alineación manual e inspección visual. Se dice entonces que tales secuencias son "sustancialmente idénticas". Con respecto a las secuencias de polinucleótidos, esta definición también se refiere al complemento de una secuencia de prueba. Preferiblemente, la identidad existe sobre una región que tiene una longitud de al menos aproximadamente 50 aminoácidos o nucleótidos, o más preferiblemente sobre una región que tiene una longitud de 75-100 aminoácidos o nucleótidos.

Para la comparación de secuencias, normalmente una secuencia actúa como una secuencia de referencia, con la que se comparan las secuencias de prueba. Cuando se utiliza un algoritmo de comparación de secuencias, las secuencias de prueba y de referencia se ingresan en un ordenador, las coordenadas de la subsecuencia se designan, si es necesario, y se designan los parámetros del programa de algoritmos de secuencia. Se pueden usar los parámetros predeterminados del programa, o se pueden designar parámetros alternativos. El algoritmo de comparación de secuencias luego calcula el porcentaje de identidades de secuencia para las secuencias de prueba en relación con la secuencia de referencia, en función de los parámetros del programa. Para la comparación de secuencias de ácidos nucleicos y proteínas, se utilizan los algoritmos BLAST y BLAST 2.0 y los parámetros predeterminados que se analizan a continuación.

Una "ventana de comparación", como se usa en este documento, incluye una referencia a un segmento de cualquiera de las posiciones contiguas seleccionadas del grupo que consiste en 20 a 600, en general, de aproximadamente 50 a aproximadamente 200, más en general, de aproximadamente 100 a aproximadamente 150, en las que una secuencia puede compararse con una secuencia de referencia del mismo número de posiciones contiguas después de que las dos secuencias estén alineadas de manera óptima. Los métodos de alineación de secuencias para comparación son bien conocidos en la técnica. La alineación óptima de las secuencias para la comparación se puede realizar, por ejemplo, mediante el algoritmo de homología local de Smith & Waterman, *Adv. Appl. Maths.* 2:482 (1981), por el algoritmo de alineación de homología de Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), por el método de búsqueda de similitud de Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85: 2444 (1988), mediante implementaciones computarizadas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en el paquete de software de genética de Wisconsin, Genetics Computer Group, 575 Science Dr., Madison, WI), o por alineación manual e inspección visual (ver, por ejemplo, *Current Protocols in Molecular Biology* (Ausubel et al., eds. 1995 suplemento)).

Los ejemplos de algoritmos que son adecuados para determinar el porcentaje de identidad de secuencia y la similitud de secuencia son los algoritmos BLAST y BLAST 2.0, que se describen en Altschul et al., (1990) *J. Mol. Biol.* 215:403-410 y Altschul et al. (1977) *Nucleic Acids Res.* 25: 3389-3402, respectivamente. El software para realizar los análisis de BLAST está disponible públicamente en el sitio web del Centro Nacional de Información Biotecnológica, ncbi.nlm.nih.gov. El algoritmo implica primero la identificación de pares de secuencias de alta puntuación (HSP) mediante la identificación de palabras cortas de longitud W en la secuencia de consulta, que coinciden o satisfacen alguna puntuación de umbral de valor positivo T cuando se alinean con una palabra de la misma longitud en una secuencia de base de datos. T se conoce como el umbral de puntuación de la palabra vecina (Altschul et al., supra). Estos aciertos de palabras vecinas iniciales actúan como semillas para iniciar búsquedas para encontrar HSP más largos que los contienen. Los aciertos de palabras se extienden en ambas direcciones a lo largo de cada secuencia hasta donde se puede aumentar la puntuación de alineación acumulativa. Las puntuaciones acumuladas se calculan utilizando, para secuencias de nucleótidos, los parámetros M (puntuación de recompensa para un par de residuos coincidentes; siempre >0) y N (puntuación de penalización para residuos no coincidentes; siempre <0). Para las secuencias de aminoácidos, se utiliza una matriz de puntuación para calcular la puntuación acumulativa. La extensión de los aciertos de palabra en cada dirección se detiene cuando: la puntuación de alineación acumulada cae en la cantidad X desde su valor máximo alcanzado; el puntaje acumulativo llega a cero o por debajo, debido a la acumulación de una o más alineaciones de residuos de puntaje negativo; o se llega al final de cualquiera de las secuencias. Los parámetros del algoritmo BLAST W, T y X determinan la sensibilidad y la

velocidad de la alineación. El programa BLASTN (para secuencias de nucleótidos) usa como predeterminado un tamaño de palabra (W) de 28, una expectativa (E) de 10, M=1, N=-2 y una comparación de ambas cadenas. Para las secuencias de aminoácidos, el programa BLASTP utiliza como valor predeterminado un tamaño de palabra (W) de 3, una expectativa (E) de 10, y la matriz de puntuación BLOSUM62 (véase Henikoff y Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 (1989)).

El algoritmo BLAST también realiza un análisis estadístico de la similitud entre dos secuencias (véase, por ejemplo, Karlin y Altschul, Proc. Nat'l. Acad. Sci. USA 90:5873-5787 (1993)). Una medida de similitud proporcionada por el algoritmo BLAST es la probabilidad de suma más pequeña (P(N)), que proporciona una indicación de la probabilidad por la cual una coincidencia entre dos secuencias de nucleótidos o aminoácidos se produciría por casualidad. Por ejemplo, un ácido nucleico se considera similar a una secuencia de referencia si la probabilidad de la suma más pequeña en una comparación del ácido nucleico de prueba con el ácido nucleico de referencia es menor que aproximadamente 0,2, más preferiblemente, menor que aproximadamente 0,01, y más preferiblemente, menor que aproximadamente 0,001.

La frase "se une específicamente con" cuando se usa en el contexto de describir una relación de unión de una molécula particular a una proteína o péptido, se refiere a una reacción de unión que es determinante de la presencia de la proteína en una población heterogénea de proteínas y otras biológicas. Por lo tanto, en condiciones de ensayo de unión designadas, el agente de unión especificado (por ejemplo, un anticuerpo) se une a una proteína particular al menos dos veces el fondo y no se une sustancialmente en una cantidad significativa a otras proteínas presentes en la muestra. La unión específica de un anticuerpo en tales condiciones puede requerir un anticuerpo que se seleccione por su especificidad para una proteína particular o una proteína, pero no sus proteínas "hermanas" similares. Se puede usar una variedad de formatos de inmunoensayo para seleccionar anticuerpos específicamente inmunorreactivos con una proteína particular o en una forma particular. Por ejemplo, los inmunoensayos ELISA en fase sólida se usan de manera rutinaria para seleccionar anticuerpos específicamente inmunorreactivos con una proteína (véase, por ejemplo, Harlow & Lane, Antibodies, Laboratory Manual (1988) para una descripción de los formatos de inmunoensayo y las condiciones que se pueden usar para determinar inmunorreactividad específica). Típicamente, una reacción de unión específica o selectiva será al menos el doble de la señal o ruido de fondo y, más típicamente, más de 10 a 100 veces el fondo. Por otro lado, la expresión "unirse específicamente con" cuando se usa en el contexto de referirse a una secuencia de polinucleótidos que forma un complejo de doble cadena con otra secuencia de polinucleótidos describe "hibridación de polinucleótidos" basada en el emparejamiento de bases de Watson-Crick, tal como se proporciona en la definición de la expresión "método de hibridación de polinucleótidos".

Tal como se utiliza en esta solicitud, un "aumento" o una "disminución" se refieren a un cambio positivo o negativo detectable en la cantidad de un control de comparación, por ejemplo, un control estándar establecido (como un nivel promedio de un ARNm de gen marcador encontrado en la sangre de una mujer embarazada que da a luz al bebé en un período de tiempo normal de su embarazo). Un aumento es un cambio positivo que suele ser de al menos el 10%, o al menos el 20%, o el 50%, o el 100%, y puede ser tan alto como al menos 2 veces, 3 veces, 4 veces, 5 veces, 6 veces, 7 veces, 8 veces, 9 veces o 10 veces el valor de control. De manera similar, una disminución es un cambio negativo que suele ser de al menos el 10%, o al menos el 20%, 30% o 50%, o incluso tan alto como al menos el 80% o el 90% del valor de control. Otros términos que indican cambios cuantitativos o diferencias con respecto a una base comparativa, como "más", "menos", "más alto" y "más bajo", se usan en esta solicitud de la misma manera que se describió anteriormente. Por el contrario, la expresión "sustancialmente igual" o "sustancialmente falta de cambio" indica poco o ningún cambio en la cantidad con respecto al valor de control estándar, típicamente dentro de $\pm 10\%$ del control estándar, o dentro de $\pm 5\%$, 2%, o incluso menos variación del control estándar.

Un "método de hibridación de polinucleótidos", como se usa en el presente documento, se refiere a un método para detectar la presencia y/o la cantidad de una secuencia de polinucleótidos predeterminada basada en su capacidad para formar el emparejamiento de bases de Watson-Crick, en condiciones de hibridación apropiadas, con una sonda polinucleotídica de una secuencia conocida. Los ejemplos de tales métodos de hibridación incluyen transferencia Southern, transferencia Northern e hibridación in situ.

"Cebadores", como se usan en este documento, se refieren a oligonucleótidos que pueden usarse en un método de amplificación, como una reacción en cadena de polimerasa (PCR), para amplificar una secuencia de nucleótidos basada en la secuencia de polinucleótidos correspondiente a un gen marcador (cualquiera en la Tabla 2 y que incluye además CD16A y CD62L), por ejemplo, el ADNc o la secuencia genómica para el gen B3GNT5 humano o una porción del mismo. Típicamente, al menos uno de los cebadores de PCR para la amplificación de una secuencia de polinucleótidos es específico de la secuencia para esa secuencia de polinucleótidos. La longitud exacta del cebador dependerá de muchos factores, incluida la temperatura, la fuente del cebador y el método utilizado. Por ejemplo, para aplicaciones de diagnóstico y pronóstico, dependiendo de la complejidad de la secuencia diana, el cebador oligonucleotídico contiene típicamente al menos 10 o 15 o 20 o 25 o más nucleótidos, aunque puede contener menos nucleótidos o más nucleótidos. Los factores implicados en la determinación de la longitud apropiada del cebador son conocidos fácilmente por un experto en la técnica. En esta descripción, la expresión "par de cebadores" significa un par de cebadores que hibridan con cadenas opuestas de una molécula de ADN diana o con regiones del ADN diana que flanquean una secuencia de nucleótidos que se amplificará. En esta descripción, la expresión "sitio

de cebador” significa el área del ADN objetivo u otro ácido nucleico con el que se hibrida un cebador.

Una “marca”, “marca detectable” o “unidad estructural detectable” es una composición detectable por medios espectroscópicos, fotoquímicos, bioquímicos, inmunoquímicos, químicos u otros medios físicos. Por ejemplo, los marcadores útiles incluyen ³²P, colorantes fluorescentes, reactivos densos en electrones, enzimas (por ejemplo, como se usan comúnmente en un ELISA), biotina, digoxigenina o haptenos y proteínas que pueden hacerse detectables, por ejemplo, incorporando un componente radioactivo en el péptido o utilizado para detectar anticuerpos específicamente reactivos con el péptido. Típicamente, un marcador detectable se une a una sonda o una molécula con características de unión definidas (por ejemplo, un polipéptido con una especificidad de unión conocida o un polinucleótido), para permitir que la presencia de la sonda (y por lo tanto su objetivo de unión) sea fácilmente detectable.

“Control estándar” como se usa en el presente documento se refiere a una cantidad o concentración predeterminada de una secuencia de polinucleótidos, por ejemplo, ARNm de uno de los genes marcadores enumerados en la Tabla 2 más CD16A y CD62L, que está presente en una muestra de sangre tomada de una mujer embarazada sana o una mujer embarazada con contracciones uterinas frecuentes antes de las 37 semanas de gestación, y que da a luz al bebé dentro del marco de tiempo normal de su embarazo. El valor de control estándar es adecuado para el uso de un método de la presente invención, para servir como base para comparar la cantidad de ARNm de un gen marcador que está presente en una muestra de prueba. Una muestra establecida que sirve como control estándar proporciona una cantidad promedio de un ARNm marcador que es típico de una muestra de sangre de una mujer embarazada saludable promedio o una mujer embarazada con contracciones uterinas frecuentes antes de las 37 semanas de gestación y que da a luz a su bebé dentro del marco temporal normal de su embarazo como se define convencionalmente. Un valor de control estándar puede variar dependiendo de la naturaleza de la muestra (por ejemplo, cómo se procesó después de la recolección), si el nivel de ARNm se normaliza con respecto al nivel de ARNm de otro gen de referencia (por ejemplo, la relación entre el nivel de ARNm marcador y el ARNm de referencia), así como otros factores tales como la edad, la edad gestacional y el origen étnico de los sujetos en función de los cuales se establece dicho valor de control.

La expresión “gen de referencia”, como se usa en el presente documento, se refiere a un gen “constitutivo” que se sabe que se expresa de manera consistente a un nivel fácilmente detectable y sustancialmente constante en las muestras de sangre. Los ejemplos de dichos genes “constitutivos” para muestras de sangre incluyen GAPDH (gliceraldehído 3-fosfato deshidrogenasa), SDHA (succinato deshidrogenasa), HPRT1 (hipoxantina fosforribosil transferasa 1), HBS1L (proteína similar a HBS1), AHSP (proteína estabilizadora de hemoglobina alfa), ACTB (beta-actina), RNA18S5 (ARN, 18S ribosomal 5), FCGR3A (el fragmento Fc de IgG, IIIa de baja afinidad, receptor), FCGR3B (el fragmento Fc de IgG, IIIa de baja afinidad, receptor), B2M (beta-2-microglobulina), HUWE1 (dominio HECT, UBA y WWE que contiene 1, proteína ligasa de ubiquitina E3), TPT1 (proteína tumoral, control de la traducción 1), MYL12B (miosina, cadena liviana 12B, regulador), SKP1 (proteína 1 asociada a quinasas de fase S) y cualquier gen identificado como adecuado para la normalización de los datos de expresión de muestras de sangre (Chang et al., 2011; Cheng et al., 2011). También se pueden usar múltiples “genes constitutivos”.

El término “promedio”, tal como se usa en el contexto de describir a una mujer embarazada que está sana, o una mujer embarazada con contracciones uterinas frecuentes antes de las 37 semanas de gestación, y que luego se confirma que da a luz dentro del marco temporal normal de embarazo, se refiere a ciertas características, especialmente la cantidad de cierto ARNm de gen marcador que se encuentra en la muestra de sangre de la mujer que es representativa de un grupo seleccionado al azar de mujeres embarazadas sanas, o mujeres embarazadas con contracciones uterinas frecuentes antes de las 37 semanas de gestación, y que más adelante se confirman para dar a luz dentro del marco temporal normal del embarazo. Este grupo seleccionado debe comprender un número suficiente de mujeres de manera que la cantidad promedio del ARNm marcador en la sangre de estos individuos refleje, con una precisión razonable, la cantidad correspondiente del ARNm marcador en la población general de mujeres embarazadas sanas o mujeres embarazadas con contracciones uterinas frecuentes antes de las 37 semanas de gestación, y que dan a luz a sus bebés en el marco temporal normal de embarazo. Además, el grupo seleccionado de mujeres generalmente tiene una edad gestacional similar a la de una mujer cuya muestra se analiza para determinar el riesgo de parto prematuro. Además, otros factores como la edad, el origen étnico, el historial médico también se consideran y, de preferencia, coinciden estrechamente entre los perfiles del sujeto de prueba y el grupo seleccionado de individuos que establecen el valor “promedio”.

El término “cantidad”, como se usa en esta solicitud, se refiere a la cantidad de un polinucleótido de interés, por ejemplo, un ARNm del gen marcador, presente en una muestra. Dicha cantidad puede expresarse en términos absolutos, es decir, la cantidad total del polinucleótido en la muestra, o en términos relativos, es decir, la concentración del polinucleótido en la muestra, incluida la expresada en forma de una relación entre el nivel de ARNm marcador y un nivel de ARNm de referencia producido por el llamado proceso de normalización.

El término “tratar” o “tratamiento”, como se usa en esta solicitud, describe un acto que conduce a la eliminación, reducción, alivio, reversión o prevención o retraso de la aparición o recurrencia de cualquier síntoma de una condición relevante. En otras palabras, “tratar” una condición abarca tanto la intervención terapéutica como la profiláctica contra la condición.

La expresión “cantidad eficaz” como se usa en este documento se refiere a una cantidad de una sustancia dada que es suficiente en cantidad para producir un efecto deseado, por ejemplo, para reducir el riesgo de nacimiento prematuro de un bebé antes de las 37 semanas o 34 semanas de edad gestacional o para prevenir dicho parto prematuro.

Descripción detallada de la invención

I. Introducción

El nacimiento prematuro en seres humanos, especialmente el parto antes de las 37 semanas o incluso antes de las 34 semanas de edad gestacional, conduce a un aumento de la mortalidad infantil y problemas de desarrollo. Hasta la fecha, la fibronectina fetal en los fluidos cervicovaginales y la longitud cervical mediante ecografía transvaginal son los marcadores clínicamente útiles para predecir el nacimiento humano antes de las 37 semanas de gestación. Por ejemplo, Lockwood et al. midieron las concentraciones de fibronectina fetal en fluido cervical o vaginal en 117 mujeres con síntomas de contracciones uterinas y membrana intacta. Utilizando >50 ng/ml como umbral, Lockwood et al. predijeron mujeres que dieron a luz antes de las 37 semanas de gestación con una sensibilidad del 81,7% y una especificidad del 82,5% (Lockwood et al., 1991). Recientemente, una revisión sistemática en múltiples estudios similares estimó que esta prueba podría predecir el nacimiento antes de las 34 semanas entre mujeres sintomáticas con una sensibilidad del 74,6% y una especificidad del 79,5% (Honest et al., 2003).

Por otro lado, Murakawa et al. midieron la longitud cervical mediante ecografía transvaginal en 32 mujeres con síntomas de contracciones uterinas antes de las 37 semanas. Usando <25 mm como umbral, Murakawa et al. predijeron mujeres que dieron a luz antes de las 37 semanas con una sensibilidad del 63,6% y una especificidad del 85,7% (Murakawa et al., 1993). Recientemente, una revisión sistemática en múltiples estudios similares estimó que esta prueba podría predecir el parto antes de las 34 semanas entre mujeres sintomáticas con una sensibilidad del 46,2% y una especificidad del 93,7% (Sotiriadis et al., 2010).

Además de la fibronectina fetal y la longitud cervical, 319 estudios que incluían 22 pruebas han sido revisados sistemáticamente por su desempeño en la predicción de nacimientos antes de las 37 semanas, y ninguno tiene una precisión excepcional (Honest et al. 2012)., Por lo tanto, los marcadores nuevos que podrían predecir el nacimiento humano antes de las 37 semanas con mayor sensibilidad y mayor especificidad son muy necesarios, por lo que la prevención y la intervención pueden dirigirse a aquellos que tienen más probabilidades de beneficiarse.

Los presentes inventores descubrieron por primera vez que el ARNm de varios biomarcadores encontrados en la sangre de una mujer embarazada puede servir como marcadores precisos para indicar la probabilidad de un parto prematuro. Este descubrimiento proporciona medios importantes para determinar el riesgo de parto prematuro y para el tratamiento profiláctico del parto prematuro. Este método para predecir el parto prematuro se puede aplicar a mujeres embarazadas con o sin síntomas que se sabe que están asociados con el parto prematuro, como contracciones uterinas antes de las 37 semanas de gestación o ruptura de la membrana antes del parto, incluso a mujeres que no han experimentado contracciones uterinas regulares pero tienen (1) antecedentes previos de partos antes de las 37 semanas en embarazos anteriores; (2) una longitud cervical acortada, cuña cervical o barros intraamnióticos en el cuello uterino; (3) signos de infección o inflamación del tracto reproductivo; (4) hemorragia anteparto; o (5) embarazos múltiples.

II. Metodología general

La práctica de esta invención utiliza técnicas de rutina en el campo de la biología molecular. Los textos básicos que describen los métodos generales de uso en esta invención incluyen Sambrook y Russell, *Molecular Cloning, A Laboratory Manual* (3ª ed. 2001); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); y *Current Protocols in Molecular Biology* (Ausubel et al., eds., 1994)).

Para los ácidos nucleicos, los tamaños se dan en kilobases (kb) o pares de bases (pb). Estas son estimaciones derivadas de electroforesis en gel de agarosa o acrilamida, de ácidos nucleicos secuenciados o de secuencias de ADN publicadas. Para las proteínas, los tamaños se dan en kilodaltons (kDa) o números de residuos de aminoácidos. Los tamaños de proteínas se estiman a partir de electroforesis en gel, de proteínas secuenciadas, de secuencias de aminoácidos derivadas o de secuencias de proteínas publicadas.

Los oligonucleótidos que no están disponibles en comercios se pueden sintetizar químicamente, por ejemplo, según el método del triéster de fosfoamidita en fase sólida descrito por primera vez por Beaucage y Caruthers, *Tetrahedron Lett.* 22: 1859-1862 (1981), usando un sintetizador automático, como se describe en Van Devanter et. al., *Nucleic Acids Res.* 12:6159-6168 (1984). La purificación de oligonucleótidos se realiza mediante cualquier estrategia reconocida en la técnica, por ejemplo, electroforesis en gel de acrilamida nativa o cromatografía líquida de alto rendimiento de intercambio aniónico (HPLC) como se describe en Pearson y Reanier, *J. Chrom.* 255:137-149 (1983).

La secuencia de interés utilizada en esta invención, por ejemplo, la secuencia de polinucleótidos del gen B3GNT5, CLC, CD16A o CD62L humano, y oligonucleótidos sintéticos (por ejemplo, cebadores) se puede verificar utilizando, por ejemplo, el método de terminación de cadena para secuenciación de plantillas de doble cadena de Wallace et al., Gene 16:21-26 (1981).

5

III. Adquisición de muestras de sangre y análisis de ARNm marcador

La presente invención se refiere a la medición de la cantidad de ARNm transcrito de al menos uno de los genes marcadores encontrados en la sangre de una mujer embarazada, especialmente en una muestra de plasma o suero, como un medio para evaluar el riesgo de que una mujer dé a luz al bebé en forma prematura. Los genes marcadores incluyen aquellos identificados en la Tabla 2, así como CD16A y CD62L. Por lo tanto, las primeras etapas para poner en práctica esta invención son obtener una muestra de sangre de una mujer embarazada que está siendo analizada y extraer el ARNm de la muestra.

10

A. Adquisición y preparación de muestras de sangre

Se obtiene una muestra de sangre de una mujer embarazada en una edad gestacional adecuada para su análisis utilizando un método de la presente invención. La edad gestacional adecuada puede variar según el trastorno examinado, como se explica a continuación. La recolección de sangre de una mujer se realiza de acuerdo con el protocolo estándar que generalmente siguen los hospitales o clínicas. Una cantidad apropiada de sangre periférica, por ejemplo, típicamente de entre 5 y 50 ml, se recolecta y se puede almacenar de acuerdo con el procedimiento estándar antes de la preparación adicional.

20

El análisis del ARNm transcrito de uno o más genes marcadores encontrados en la sangre materna según la presente invención se puede realizar utilizando, por ejemplo, sangre entera o cualquier preparación de sangre entera que contenga los glóbulos sanguíneos. Las preparaciones de sangre que no contienen los glóbulos sanguíneos, como plasma o suero, también son útiles con el fin de poner en práctica la presente invención, debido a que los glóbulos sanguíneos son la fuente predominante de ácidos nucleicos presentes en el plasma o suero (Lui et al., 2002). Para preparar glóbulos sanguíneos a partir de una muestra, los métodos para eliminar la porción acelular, como el plasma o el suero, de la sangre materna son bien conocidos entre los expertos en la técnica. Por ejemplo, la sangre de una mujer embarazada se puede colocar en un tubo con EDTA o un producto comercial especializado como Vacutainer SST (Becton Dickinson, Franklin Lakes, NJ) para evitar la coagulación de la sangre, y el plasma se puede separar y luego extraer de la fracción celular de la sangre entera mediante centrifugación o sedimentación por gravedad solo durante un período de tiempo apropiado. Si se usa centrifugación, entonces, típica, aunque no exclusivamente, se realiza a una velocidad apropiada, por ejemplo, 1,500-3,000 x g. Para la preparación de plasma/suero a partir de una muestra de sangre, los métodos también son bien conocidos entre los expertos en la técnica. Por ejemplo, una muestra de sangre recogida en tubos con EDTA se centrifuga a 16.000 x g durante 10 minutos a 4°C para eliminar el plasma y se vuelve a centrifugar a 1.600 x g durante 10 minutos a 4°C para eliminar el plasma residual. Además, después de que se haya recolectado o preparado la sangre entera, los glóbulos sanguíneos, el plasma o el suero, se pueden agregar aditivos para preservar el ARN. Estas aditividades pueden incluir una solución monofásica de fenol e isotiocianato de guanidinio, o los reactivos comercialmente disponibles, incluidos Trizol (Life Technologies), Trizol LS (Life Technologies), RNA Later (Life Technologies - Ambion), RNA Later ICE (Life Technologies - Ambion) y tubos de extracción de sangre asociados con el sistema PreamAlytix (Qiagen/Beckton Dickson). A menos que se indique lo contrario, estos reactivos disponibles en comercios para preservar y/o extraer el ARN se realizan de acuerdo con las recomendaciones del fabricante.

25

30

35

40

45

B. Extracción y cuantificación del ARN

Existen numerosos métodos para extraer ARNm de una muestra biológica. Se pueden seguir los métodos generales de preparación de ARNm (por ejemplo, descritos por Sambrook y Russell, Molecular Cloning: A Laboratory Manual 3d ed., 2001); diversos reactivos o kits disponibles comercialmente, como el reactivo Trizol (Invitrogen, Carlsbad, CA), los kits Oligotex Direct mRNA (Qiagen, Valencia, CA), los kits RNeasy Mini (Qiagen, Hilden, Alemania) y Polyattract® Series 9600™ (Promega, Madison, WI), también se puede usar para obtener ARNm de una muestra biológica de un sujeto de prueba. También se pueden usar combinaciones de más de uno de estos métodos.

50

55

Es esencial que todo el ADN contaminante se elimine de las preparaciones de ARN. Por lo tanto, se debe utilizar un manejo cuidadoso de las muestras, un tratamiento completo con ADNasa y controles negativos apropiados en las etapas de amplificación y cuantificación.

60

1. Determinación cuantitativa basada en PCR del nivel de ARNm

Una vez que se extrae el ARNm de una muestra, se puede cuantificar la cantidad de ARNm transcrito de uno o más de los genes marcadores identificados en la Tabla 2. El método preferido para determinar el nivel de ARNm es un método basado en la amplificación, por ejemplo, mediante la reacción en cadena de polimerasa (PCR), especialmente la reacción en cadena de polimerasa con transcripción inversa (RT-PCR).

65

Antes de la etapa de amplificación, se debe sintetizar una copia de ADN (ADNc) del ARNm del gen marcador. Esto se logra mediante transcripción inversa, que puede llevarse a cabo como una etapa separada, o en una reacción en cadena de polimerasa con transcripción inversa (RT-PCR), una modificación de la reacción en cadena de polimerasa para amplificar el ARN. Los métodos adecuados para la amplificación por PCR de los ácidos ribonucleicos están descritos por Romero y Rotbart en *Diagnostic Molecular Biology: Principles and Applications*, pp. 401-406; Persing et al., eds., Mayo Foundation, Rochester, MN, 1993; Egger et al, *J. Clin. Microbiol.* 33:1442-1447, 1995; y la patente U. S. N° 5.075.212.

Los métodos generales de PCR son bien conocidos en la técnica y, por lo tanto, no se describen detalladamente en este documento. Para una revisión de los métodos, protocolos y principios de PCR en el diseño de cebadores, véase, por ejemplo, Innis, et al., *PCR Protocols: A Guide to Methods and Applications*, Academic Press, Inc. N.Y., 1990. Los reactivos y protocolos de PCR también están disponibles de proveedores comerciales, como Roche Molecular Systems.

La PCR se realiza generalmente como un proceso automatizado con una enzima termoestable. En este proceso, la temperatura de la mezcla de reacción se cicla a través de una región desnaturalizante, una región de fusión del cebador y una región de reacción de extensión en forma automática. Las máquinas específicamente adaptadas para este propósito están disponibles en comercios.

Aunque la amplificación por PCR del ARNm diana se usa típicamente en la práctica de la presente invención, un experto en la técnica reconocerá que la amplificación de estas especies de ARNm en una muestra de sangre materna puede realizarse mediante cualquier método conocido, como la reacción en cadena de ligasa (LCR), amplificación mediada por transcripción y replicación de secuencia autosostenida o amplificación basada en secuencia de ácido nucleico (NASBA), amplificación dependiente de helicasa (HDA), amplificación de círculo rodante (RCA) y amplificación isotérmica mediada por bucle (LAMP), cada una de las cuales proporciona suficiente amplificación. La tecnología de ADN ramificado desarrollada más recientemente también se puede usar para determinar cuantitativamente la cantidad de marcadores de ARNm en la sangre materna. Para una revisión de la amplificación de la señal de ADN ramificado para la cuantificación directa de secuencias de ácido nucleico en muestras clínicas, véase Nolte, *Adv. Clinica Chem.* 33:201-235, 1998.

2. Otros métodos cuantitativos

El ARNm del gen marcador también puede detectarse utilizando otras técnicas estándar, bien conocidas por los expertos en la técnica. Aunque la etapa de detección está precedida típicamente por una etapa de amplificación, la amplificación no se requiere en los métodos de la invención. Por ejemplo, el ARNm puede identificarse por fraccionamiento de tamaño (por ejemplo, electroforesis en gel), sea o no precedida por una etapa de amplificación. Después de analizar una muestra en un gel de agarosa o poliacrilamida y marcar con bromuro de etidio según técnicas bien conocidas (véase, por ejemplo, Sambrook y Russell, *supra*), la presencia de una banda del mismo tamaño que la comparación estándar es una indicación de la presencia de un ARNm diana, cuya cantidad se puede comparar con el control en función de la intensidad de la banda. Alternativamente, se pueden usar sondas de oligonucleótidos específicas para el ARNm del gen marcador para detectar la presencia de tales especies de ARNm e indicar la cantidad de ARNm en comparación con la comparación estándar, según la intensidad de la señal impartida por la sonda.

La hibridación de la sonda específica de la secuencia es un método bien conocido para detectar un ácido nucleico particular que comprende otras especies de ácidos nucleicos. En condiciones de hibridación suficientemente estrictas, las sondas hibridan específicamente solo con secuencias sustancialmente complementarias. La rigurosidad de las condiciones de hibridación se puede relajar para tolerar cantidades variables de desajuste de secuencia.

Una serie de formatos de hibridación bien conocidos en la técnica, que incluyen, entre otros, ensayos de hibridación en fase de solución, fase sólida o fase mixta. Los siguientes artículos proporcionan una visión general de los diversos formatos de ensayos de hibridación: Singer et al., *Biotechniques* 4: 230, 1986; Haase et al., *Methods in Virology*, pp. 189-226, 1984; Wilkinson, *In situ Hybridization*, Wilkinson ed., IRL Press, Oxford University Press, Oxford; y Hames y Higgins eds., *Nucleic Acid Hybridization: A Practical Approach*, IRL Press, 1987.

Los complejos de hibridación se detectan de acuerdo con técnicas bien conocidas. Las sondas de ácido nucleico capaces de hibridar específicamente con un ácido nucleico diana, es decir, el ARNm o el ADN amplificado, pueden marcarse por uno cualquiera de diversos métodos típicamente utilizados para detectar la presencia de ácidos nucleicos hibridados. Un método común de detección es el uso de autorradiografía usando sondas marcadas con ^3H , ^{125}I , ^{35}S , ^{14}C o ^{32}P , o similares. La elección del isótopo radiactivo depende de las preferencias de investigación debido a la facilidad de síntesis, estabilidad y vida media de los isótopos seleccionados. Otros marcadores incluyen compuestos (por ejemplo, biotina y digoxigenina), que se unen a los antilíngandos o anticuerpos marcados con fluoróforos, agentes quimioluminiscentes y enzimas. Alternativamente, las sondas pueden conjugarse directamente con marcadores tales como fluoróforos, agentes quimioluminiscentes o enzimas. La elección de la marca depende de la sensibilidad requerida, la facilidad de conjugación con la sonda, los requisitos de estabilidad y la

instrumentación disponible.

Las sondas y los cebadores necesarios para poner en práctica la presente invención se pueden sintetizar y marcar usando técnicas bien conocidas. Los oligonucleótidos utilizados como sondas y cebadores pueden sintetizarse químicamente de acuerdo con el método del triéster de fosforamidita en fase sólida descrito por primera vez por Beaucage y Caruthers, *Tetrahedron Letts.*, 22:1859-1862, 1981, usando un sintetizador automático, como se describe en Needham-Van Deventer et al. *Nucleic Acids Res.* 12:6159-6168, 1984. La purificación de los oligonucleótidos se realiza mediante electroforesis en gel de acrilamida nativa o mediante HPLC de intercambio aniónico como se describe en Pearson y Regnier, *J. Chrom.*, 255:137-149, 1983.

IV. Establecimiento de un control estándar

A fin de establecer un control estándar para la práctica del método de esta invención, se selecciona primero un grupo de mujeres embarazadas "sanas" o mujeres embarazadas con contracciones uterinas frecuentes antes de las 37 semanas de gestación, y que luego se confirma que dan a luz dentro del marco temporal normal de su embarazo según lo determinado por métodos convencionales. Estas mujeres están dentro de los parámetros apropiados, como una edad gestacional particular y un estado de salud comparable. Opcionalmente, las mujeres se agrupan según la edad o el origen étnico similar.

El período de alumbramiento normal de las mujeres seleccionadas se confirmará más adelante, y cualquier persona entre las personas seleccionadas que paren antes o después de la fecha normal de parto será excluida del grupo para proporcionar datos como "control estándar".

Además, el grupo seleccionado de mujeres debe ser de un tamaño razonable, de modo que la cantidad/concentración promedio de ARNm marcador en la muestra de sangre obtenida del grupo pueda considerarse razonablemente como representativa del nivel normal o promedio entre la población general de mujeres embarazadas sanas, o mujeres embarazadas con contracciones uterinas frecuentes antes de las 37 semanas de gestación, y que darán a luz dentro del marco temporal normal. Preferiblemente, el grupo seleccionado comprende al menos 10 sujetos humanos.

De acuerdo con el principio científico fundamental de establecimiento de un valor de control, el nivel de ARNm en el grupo de control está determinado por el mismo método utilizado para determinar el nivel de ARNm en los individuos de prueba. Por ejemplo, si el nivel de ARNm de un gen marcador se determina en un tipo particular de muestra (por ejemplo, plasma) tomado de una mujer que se está analizando, el control también debe obtenerse del mismo tipo de muestra. Si el nivel de mRNA de un gen marcador se determina después de normalizarse sobre el nivel de mRNA de un gen de referencia (por ejemplo, representado por la relación de un mRNA de gen marcador a un mRNA de gen de referencia), entonces el control estándar también debe representarse en la forma de un valor normalizado sobre el mismo nivel de ARNm del gen de referencia.

Una vez que se establece un valor promedio para cualquiera de los marcadores de ARNm basados en los valores individuales encontrados en cada sujeto del grupo de control seleccionado, este promedio o mediana se considera un control estándar. Una desviación estándar también se determina durante el mismo proceso. En algunos casos, se pueden establecer controles estándar separados para grupos definidos por separado que tienen características distintas, como edad, género u origen étnico.

V. Métodos terapéuticos para prevenir el parto prematuro

Al ilustrar la correlación entre el nivel de ARNm de los genes marcadores identificados en la Tabla 2 y el riesgo de parto prematuro, la presente descripción proporciona, además, un medio para tratar profilácticamente a las mujeres embarazadas que, de lo contrario, podrían experimentar un parto prematuro y dar a luz a sus bebés mucho antes de que alcancen el término completo: una vez que se detecta un aumento o disminución del nivel de ARNm del marcador y se determina un mayor riesgo de parto prematuro, el médico tiene la opción de tratar a la mujer de manera profiláctica, por ejemplo, con corticosteroides prenatales, que se demostró que reduce la morbilidad y la mortalidad neonatales por dificultad respiratoria, hemorragia intraventricular, enterocolitis necrotizante y conducto arterioso persistente (Roberts y Dalziel, 2006, *Cochrane Database Syst Rev* (3): CD004454; Wapner et al. 2006, *Am J Obstet Gynecol* 195(3): 633-42). Además, los fármacos tocolíticos se pueden usar para prolongar el embarazo en mujeres con alto riesgo de dar a luz demasiado pronto. El uso de estos fármacos proporciona un parto retrasado de 48 horas, lo que permite la transferencia a una unidad especializada y la administración de corticosteroides para reducir la morbilidad y la mortalidad neonatales (Iams et al., 2008 *Lancet* 371(9607): 164-75). Además, se ha informado que el tratamiento con trinitrato de glicerilo transdérmico disminuye efectivamente la morbilidad neonatal (Smith et al. 2007, *Am J Obstet Gynecol* 196(1): 37 e1-8). Tal como se usa en el presente documento, el tratamiento del parto prematuro incluye reducir o eliminar la posibilidad de que una mujer embarazada dé a luz en cualquier momento antes de las 37 semanas de edad gestacional, por ejemplo, antes de las 34 semanas de edad gestacional.

Otra posibilidad de tratar el nacimiento prematuro es mediante la regulación directa del nivel de ARNm de los genes marcadores que se ha demostrado que se desvían de un valor de control estándar. Por ejemplo, cuando se halla

que un gen marcador ha aumentado con respecto al valor de control estándar, se pueden tomar medidas para reducir específicamente el nivel de ARNm de este gen. Se pueden administrar secuencias de polinucleótidos antisentido y ARNs a la mujer embarazada para este fin. Por otro lado, cuando se halla que un gen marcador ha disminuido con respecto al valor de control estándar, se pueden tomar medidas para aumentar específicamente el nivel de ARNm de este gen. Se puede administrar un ácido nucleico aislado, como un casete de expresión, que contiene la secuencia codificante del gen marcador y que dirige la transcripción de la secuencia a la mujer embarazada con este propósito.

VI. Kits y dispositivos

La invención proporciona composiciones y kits para poner en práctica los métodos descritos en el presente documento para evaluar el nivel de ARNm de cualquiera de los genes marcadores (como los enumerados en la Tabla 2, así como CD16A y CD62L) para determinar el riesgo de parto prematuro en una mujer embarazada.

Los kits para realizar ensayos para determinar el nivel del ARNm del gen marcador incluyen típicamente al menos una sonda de oligonucleótido útil para la hibridación específica con al menos un segmento de la secuencia codificante del gen marcador o su secuencia complementaria. Opcionalmente, esta sonda de oligonucleótido se marca con una unidad estructural detectable. En algunos casos, los kits pueden incluir al menos dos cebadores oligonucleotídicos que pueden usarse en la amplificación de al menos un segmento del ADN o ARNm del gen marcador por PCR, particularmente por RT-PCR. En algunos casos, los kits pueden contener múltiples conjuntos de la sonda y/o cebadores descritos anteriormente, de manera que se pueden probar y cuantificar más de un ARNm de gen marcador. En algunos casos, los kits pueden contener, además, la sonda y/o cebadores descritos anteriormente que se pueden usar para determinar el nivel de ARNm de un gen de referencia.

A menudo, los kits también incluyen un control estándar apropiado. Los controles estándar indican el valor promedio de uno o más ARNm del gen marcador en un tipo particular de muestra de sangre. En algunos casos, dicho control estándar puede proporcionarse en forma de un valor establecido. Además, los kits de esta invención pueden proporcionar manuales de instrucciones para guiar a los usuarios en el análisis de muestras de prueba y evaluar el riesgo de nacimiento prematuro en un sujeto de prueba.

En un aspecto adicional, la presente invención también puede realizarse en un dispositivo o un sistema que comprende uno o más de dichos dispositivos, que es capaz de llevar a cabo todos o algunos de las etapas del método descritos en el presente documento. Por ejemplo, en algunos casos, el dispositivo o el sistema realiza los siguientes pasos al recibir una muestra de sangre, por ejemplo, una muestra de plasma tomada de una mujer embarazada que está siendo examinada para determinar el riesgo de parto prematuro, evaluando el riesgo de parto prematuro: (a) determinar en la muestra la cantidad o concentración de un ARNm de un gen marcador, que opcionalmente puede normalizarse sobre el nivel de ARNm de un gen de referencia; (b) comparar la cantidad o concentración con un valor de control estándar; y (c) proporcionar un resultado que indique si existe un mayor riesgo de nacimiento prematuro. En otros casos, el dispositivo o sistema de la invención realiza la tarea de las etapas (b) y (c), después de que se haya realizado la etapa (a) y la cantidad o concentración de (a) se haya ingresado en el dispositivo. Preferiblemente, el dispositivo o sistema está parcial o totalmente automatizado.

Ejemplos

Los siguientes ejemplos se proporcionan solo a modo de ilustración y no a modo de limitación. Los expertos en la técnica reconocerán fácilmente una variedad de parámetros no críticos que podrían cambiarse o modificarse para producir esencialmente los mismos resultados o resultados similares.

Introducción

Actualmente, la fibronectina fetal en el líquido cervicovaginal y la longitud cervical determinada por la ecografía transvaginal son los mejores fabricantes disponibles para predecir el nacimiento humano antes de las 37 semanas de gestación. Si bien estos métodos se utilizan por su alta especificidad, su sensibilidad es solo moderada.

Los presentes inventores tienen un interés de larga data en el descubrimiento sistemático de ARN, microARN asociados con el embarazo (Tsui et al. 2004; Chim et al., 2008) y marcadores de metilación del ADN (Chim et al. 2005; Chim et al. 2008; Tsui et al. 2010) circulando en plasma sanguíneo periférico materno. De interés clínico, los inventores han descubierto que ciertos transcritos de ARN circulantes se detectan con mayor frecuencia en mujeres que presentan contracciones uterinas y dan como resultado el nacimiento antes de las 34 semanas de gestación, pero no en mujeres de la misma edad gestacional (Chim et al., 2012). Los datos prometedores en el plasma materno libre de células llevaron a los inventores a investigar sistemáticamente si otros compartimentos sanguíneos, incluidas las glóbulos sanguíneos, contienen transcritos de ARN que pueden predecir el nacimiento antes de las 34 semanas.

Para identificar dichos marcadores en un enfoque sistemático y de genoma completo, los inventores han perfilado los niveles de ARN de casi todos los 30.000 genes humanos y sus variantes utilizando la tecnología de matriz de

exones. Esta tecnología ha permitido a los inventores generar el perfil global de expresión génica (ARN), o transcriptoma, de la sangre en alta resolución en el nivel del exón, que es más detallado que el nivel del gen. Los inventores han perfilado sistemáticamente los transcriptomas de sangre materna obtenidos de mujeres durante su presentación de contracciones uterinas regulares antes de las 34 semanas. De estos, se identificó que un panel de transcripciones de ARN de 32 genes se puede medir fácilmente y se expresa de manera diferente en la sangre de las mujeres que dan a luz antes de las 34 semanas, en comparación con las que dan a luz a las 37 semanas o después.

Dado que el nacimiento humano antes de las 34 semanas de gestación es más susceptible a la mortalidad y la morbilidad neonatales, el nacimiento antes de las 34 semanas se ha utilizado como medida de resultado en la mayoría de las partes de este estudio. Sin embargo, dicho diseño experimental y la medida de resultado no impiden el descubrimiento de marcadores para predecir el nacimiento antes de las 37 semanas, o dentro de los 2 a 7 días de la presentación de las contracciones uterinas.

Usando la reacción en cadena de polimerasa transcriptasa inversa cuantitativa (qRT-PCR), los inventores han demostrado que los marcadores de sangre identificados, cuando se usan solos, ya podrían usarse para predecir el nacimiento antes de las 34 semanas con alta sensibilidad y alta especificidad.

Métodos y resultados

Reclutamiento de participantes. Las mujeres embarazadas con contracciones uterinas regulares y frecuentes (>1 cada 10 minutos) antes de las 34 semanas de gestación fueron invitadas a participar en este estudio con consentimiento informado. Se obtuvo sangre periférica de cada participante durante la contracción uterina prematura. Se siguió el resultado del parto. Las mujeres que luego se confirmó que dieron a luz en el parto antes de las 34 semanas se clasificaron como el grupo de prueba, y las que luego se confirmó que dieron como resultado el nacimiento a las 37 semanas o después se clasificaron como el grupo de referencia. Se excluyeron los embarazos que incluyeron el parto prematuro indicado, preeclampsia, embarazos múltiples, sufrimiento fetal, restricción del crecimiento, anomalías cromosómicas o estructurales.

Procesamiento de sangre. Se recogieron 12 ml de sangre periférica en tubos con EDTA (Beckton Dickson) de mujeres embarazadas durante la presentación de la contracción uterina prematura, procesada en 6 horas. Brevemente, la sangre fue centrifugada 1.600 x g. Se eliminó el plasma. La muestra de sangre se centrifugó de nuevo a 5.000 x g para eliminar más plasma. Se mezclaron 0,3 ml de glóbulos sanguíneos recolectados con 0,9 ml de Trizol LS (Invitrogen, Life Technologies) y se almacenaron a -80°C hasta la extracción del ARN.

Perfil de los transcriptomas de sangre basados en la matriz de exones. Para cada muestra de sangre, el ARN se extrajo de los glóbulos sanguíneos Trizol LS y se trató con DNaseI (Invitrogen, Life Technologies) para eliminar la contaminación por ADN genómico. La cantidad y la calidad de las preparaciones de ARN a partir de tejido placentario se evaluaron mediante un espectrofotómetro y un bioanálisis (Agilent). Se analizaron seis muestras de ARN en sangre (3 del grupo de prueba y otras 3 del grupo de referencia, Tabla 1) utilizando la matriz de expresión génica Exon 1.0ST (Affymetrix), de acuerdo con las instrucciones del fabricante.

Preprocesamiento de datos de matriz de exones. Los datos de la señal de la sonda se analizaron luego con Partek Genomics Suite (versión 6.5, Partek Inc.). Para normalizar las señales de la sonda a partir de diferentes muestras de sangre, se realizó la normalización Robust Multi Array (RMA) (Irizarry et al., 2003). Cada matriz contiene más de 1.400.000 conjuntos de sondas que interrogan los niveles de expresión de ARN de esencialmente todos los >30.000 genes humanos y variantes de transcripto. Aunque la mayoría de las señales de la sonda no se modificaron en ningún dato de micromatriz, su número total obstaculizaría el análisis estadístico, incluidas las pruebas de hipótesis múltiples y, por lo tanto, debe eliminarse. Con este fin, los inventores primero realizaron una prueba T (sin ajuste para comparación de prueba múltiple) en todas las sondas y encontraron que solo se cambiaron 9.264 sondas entre la prueba y los grupos de referencia.

Minería de datos e identificación sistemática de marcadores. Para explicar la diferencia en la edad gestacional en la recolección de sangre, cada muestra en el grupo de prueba se emparejó con otra muestra en el grupo de referencia con edad gestacional compatible (en el lapso de 1 semana). Para identificar las sondas que se cambiaron entre los dos grupos, se realizó una prueba T pareada en las 9.264 sondas. Entre ellos, se cambiaron 3.778 sondas entre los dos grupos (valores p, rango $5,0 \times 10^{-6}$ a 0,049). Para realizar el ajuste para múltiples pruebas de hipótesis, los valores q se calcularon mediante el método de la tasa de descubrimiento falso (Storey 2002), y se seleccionaron 3500 sondas (valor q < 0,007639). Entre ellos, las señales medias de 153 sondas se cambiaron >2,6 veces entre los dos grupos (72 sondas y 81 sondas fueron reguladas hacia arriba y hacia abajo, respectivamente, en el grupo de prueba, en comparación con el grupo de referencia).

Para aislar aún más las sondas reguladas hacia arriba que podrían distinguir potencialmente los dos grupos, los inventores buscaron sondas cuyo primer cuartil de señales en el grupo de prueba era >2 veces más alto que el tercer cuartil de señales en el grupo de referencia, y seleccionó 52 sondas cumpliendo con este criterio. De manera similar, para aislar aún más las sondas reguladas a la baja que podrían distinguir a los dos grupos, los inventores

buscaron sondas cuyo primer percentil de las señales en el grupo de referencia era >2 veces más alto que el tercer percentil de las señales en la referencia grupo, y se seleccionaron 72 sondas que cumplen este criterio. Por lo tanto, se identificaron 124 sondas (=52+72) con potencial para distinguir los dos grupos.

5 Para refinar aún más los transcritos de ARN que serían fácilmente detectables en la muestra de sangre, los inventores seleccionaron solo los transcritos de ARN representados por >1 sonda con señal de expresión mediana >169 unidades ($=2^{7,4}$ unidades) en al menos un grupo, e identificó 48 sondas (14 sondas reguladas hacia arriba y 34 sondas reguladas hacia abajo). Estas señales de sonda se derivaron de los transcritos de ARN de 32 genes (13 genes regulados hacia arriba y 19 genes regulados hacia abajo). Se argumenta que este panel de 32 transcritos de
10 ARN se puede detectar fácilmente en sangre y se puede usar para predecir el resultado de un parto antes de las 34 semanas (Tabla 2).

15 Análisis de QRT-PCR de nuevos marcadores identificados. Para demostrar que los marcadores identificados mediante la tecnología de matriz de exones y la estrategia de extracción de datos anterior (Tabla 2) son útiles, los inventores realizaron qRT-PCR, el estándar de oro en el perfil de expresión génica. Se analizaron veinte muestras de ARN en sangre (10 de los grupos de prueba, otras 10 del grupo de referencia, Tabla 3) recolectadas de mujeres.

20 Las concentraciones de 3 transcritos de ARN marcador identificados en la Tabla 2 se determinaron mediante qRT-PCR. Es decir, se midieron las concentraciones de los ARNm que codifican UDP-GlcNAc:betaGal beta-1,3-N-acetilglucosaminiltransferasa 5 (B3GNT5), galectina de cristal de Charcot-Leyden (CLC) y proteína de unión a guanilato 3 (GBP3). Se ha demostrado que el transcrito del ARN del marcador anterior está regulado hacia arriba en la sangre de las mujeres que dan a luz antes de las 34 semanas, en comparación con las que dan como resultado el nacimiento a las 37 semanas o después, mientras que se ha demostrado que los dos últimos marcadores se regulan hacia abajo. Para controlar las variaciones en la entrada de ARN en cada qRT-PCR, las
25 concentraciones de ARN marcador se normalizaron frente al ARN de referencia, el ARNm de gliceraldehído-3-fosfato deshidrogenasa (GAPDH).

30 Para el ensayo por qRT-PCR de B3GNT5, el cebador directo fue 5'-TTG GGC TTG CTT TGT TTC CT-3', el cebador inverso fue 5'-GCC TGC CGA TCT GGT AGA AG-3', y la sonda de hidrólisis fue 5'(6FAM)-AGG CCC AGC ATT T-3' (MGB), donde 6FAM fue el colorante indicador de 6-carboxifluoresceína y MGB fue el inactivador no fluorescente de unión a la ranura menor. Para el ensayo por qRT-PCR de CLC, el cebador directo fue 5'-GCT GCC TCT TTG TCT ACT GGT TCT A-3', el cebador inverso fue 5'-GCA GAT ATG GTT CAT TCA AGA AAC A-3', y la sonda de hidrólisis fue 5(6FAM)'-AAT CAA AGG GCG ACC ACT-3' (MGB). Para el ensayo por qRT-PCR de GBP3, el cebador directo fue 5'-GGC CTC GTC TAG AGA GCC TAG TG-3', el cebador inverso fue 5'-TGC GTT CTC CAT GCA GGG-3', y la
35 sonda de hidrólisis fue 5'(6FAM)-TGA CCT ATA TCA ATG CTA TCA G-3' (MGB). Para el ensayo por qRT-PCR de GAPDH, las secuencias para los cebadores y la sonda se publicaron previamente (Chim et al., 2012).

40 Para minimizar el efecto de cualquier ADN genómico contaminante en las preparaciones de ARN, el ensayo por qRT-PCR para todas las dianas de ARNm, excepto el ARNm de B3GNT5, se diseñó para que se extendiera por intrones. Sin embargo, debido a ciertas restricciones de la secuencia del ARNm, el ensayo por RT-qPCR para el ARNm de B3GNT5 no abarcó ningún intrón.

45 Para todos los ensayos por qRT-PCR, excepto el ensayo de B3GNT5, cada qRT-PCR se configuró en un volumen de reacción de 25 µL utilizando componentes suministrados en un conjunto de reactivos de PCR EZ rTth RNA (Life Technologies). Cada reacción contenía 5 µL de 5x tampón EZ, y concentraciones finales de 3 mM de Mn(OAc)₂, 300 µM cada una de dATP, dCTP, dGTP, 600 µM de dUTP, 2,5 U de polimerasa rTth, 0,25 U de uracilo N-glicosilasa (UNG) y 5 µL de ARN extraído de una muestra de sangre. Para el ensayo de B3GNT5, cada qRT-PCR se configuró en un volumen de reacción de 12,5 µL que contenía 2,5 µL de 5x de tampón EZ y concentraciones finales de 3 mM de Mn(OAc)₂, 300 µM cada una de dATP, dCTP, dGTP, 600 µM de dUTP, 2,5 U de polimerasa rTth, 0,25 U de N-glicosilasa de uracilo (UNG) y 2,5 µL de ARN extraído de una muestra de sangre. Para todos los ensayos de qRT-PCR en este estudio, las concentraciones finales de cebadores directos, cebadores inversos y sondas de hidrólisis fueron 300 nM, 300 nM y 200 nM, respectivamente. Las condiciones de los ciclos térmicos fueron 50°C durante 2 minutos, 60°C durante 30 minutos, 95°C durante 5 minutos, seguidos de 45 ciclos de (94°C durante 20 segundos y 60°C durante 1 minuto).
50

55 La amplificación de la diana del ARNm fue monitoreada y analizada por un sistema de detección de secuencias ABI Prism 7900 (Life Technologies) y un software de detección de secuencias versión 2.1 (Life Technologies). Para cada ensayo, se preparó una calibración amplificando diluciones en serie de oligonucleótidos de ADN sintético purificados por HPLC (Sigma-Prologo) que representan el amplicón objetivo en concentraciones conocidas. Las concentraciones absolutas de dianas de ARNm se calcularon como el número de copias por ng de ARN total en los glóbulos sanguíneos. Para cada muestra de ARN en sangre, se calculó una concentración de ARN marcador normalizado dividiendo la concentración absoluta del ARN marcador (es decir, el ARNm de B3GNT5, el ARNm de CLC y el ARNm de GBP3) por la concentración absoluta del ARN de referencia (es decir, ARNm de GAPDH). No hubo unidad para esta concentración de ARN marcador normalizado.
60

65 Análisis de QRT-PCR de los transcritos de ARN no incluidos en la Tabla 2. Para comparar y contrastar el

desempeño predictivo de los transcritos de ARN marcadores identificados en este estudio (Tabla 2), se analizaron dos transcritos de ARN que no figuran en la Tabla 2 en paralelo, a saber, aquellos que codifican el fragmento Fc de IgG, IIIa de baja afinidad, receptor (FCGR3A, sinónimo: CD16A) y la selectina L (SELL, sinónimo: CD62L). Se ha demostrado que estos dos transcritos de ARN se expresan en gran medida en los glóbulos sanguíneos y solo se expresan en otras células humanas (Su et al. 2004). Se analizaron veinte muestras de ARN de sangre (10 de los grupos de prueba, otras 10 del grupo de referencia, Tabla 3) recolectadas de mujeres.

Para el ensayo por qRT-PCR de CD16A, el cebador directo fue 5'-ACC CGG TGC AGC TAG AAG TC-3', el cebador inverso fue 5'-GAA TAG GGT CTT CCT CCT TGA ACA-3' y la sonda de hidrólisis fue 5'(6FAM)-TTG CTC CAG GCC CCT-3' (MGB). Para el ensayo por qRT-PCR de CD62L, el cebador directo fue 5'-TTC AGC CTC CCC ACC TTC T-3', el cebador inverso fue 5'-GGT GTG GAA GTC AGC CAA CTG-3' y la sonda de hidrólisis fue 5'(6FAM)-CAG CCA CCT CTC TT-3' (MGB). Las condiciones de reacción y el perfil térmico fueron los mismos que en otros ensayos de qRT-PCR mencionados anteriormente. Para cada muestra de ARN en sangre, las concentraciones normalizadas del ARNm de CD16A y el ARNm de CD62L se calcularon como se indicó anteriormente.

Datos de la prueba qRT-PCR dirigidos al ARNm de B3GNT5 preseleccionados en este estudio (Tabla 2). Se recogieron muestras de sangre de 20 mujeres durante la presentación de contracciones uterinas regulares y frecuentes. Entre ellas, 10 mujeres dieron a luz antes de las 34 semanas (el grupo de prueba) y las 10 restantes dieron a luz a las 37 semanas o más tarde (el grupo de referencia). Las medianas (primer cuartil y tercer cuartil) de las concentraciones de ARNm de B3GNT5 en sangre normalizada con GAPDH fueron 1,06 (0,644-1,44) y 0,362 (0,211-0,391) en el grupo de prueba y el grupo de referencia, respectivamente (Fig. 1). Esta mediana de concentraciones de ARNm de B3GNT5 normalizadas en el grupo de prueba fue 2,99 veces mayor que en el grupo de referencia ($p=0,002$, prueba de suma de intervalos de Mann-Whitney).

Para determinar las concentraciones de umbral óptimas de este marcador para identificar el nacimiento antes de las 34 semanas, los inventores trazaron la curva de características operativas del receptor (ROC) (Fig. 2, área bajo la curva = 0,915, 95% de intervalo de confianza (CI) = 0,788-1,04, $p = 0,00172$). Al utilizar las concentraciones de ARNm de B3GNT5 en sangre normalizadas por GAPDH $>0,495$ como un umbral para definir a una mujer como positiva para este ensayo, los inventores pudieron identificar a las mujeres que dieron a luz antes de las 34 semanas con una sensibilidad del 90,0% y una especificidad del 90,0%. El valor predictivo positivo y el valor predictivo negativo para el ARNm de B3GNT5 son del 90,0% y del 90,0%, respectivamente.

Datos de la prueba qRT-PCR dirigidos al ARNm de CLC preseleccionados por este estudio (Tabla 2). Las medianas (primer cuartil-tercer cuartil) de las concentraciones de ARNm de CLC en sangre normalizadas con GAPDH fueron 2,40 (1,68-4,03) y 6,05 (4,43-15,6) en el grupo de prueba y el grupo de referencia, respectivamente (Fig. 3). Esta mediana de las concentraciones de ARNm de CLC normalizadas en el grupo de prueba fue 2,52 veces menor que en el grupo de referencia ($p=0,011$, prueba de suma de intervalos de Mann-Whitney).

La curva ROC se trazó (Fig. 4, área bajo la curva = 0,840, 95% de CI = 0,637 a 1,04, $p = 0,0102$). Al utilizar las concentraciones de ARNm de CLC en sangre normalizadas por GAPDH $<4,150$ como un umbral para definir a una mujer como positiva para este ensayo, los inventores pudieron identificar a las mujeres que dieron a luz antes de las 34 semanas con una sensibilidad del 80% y una especificidad del 90%. El valor predictivo positivo y el valor predictivo negativo para el ARNm de CLC son del 88,9% y del 81,8%, respectivamente.

Los datos del ensayo por qRT-PCR dirigidos al ARNm de GBP3 preseleccionados en este estudio (Tabla 2). Las medianas (primer cuartil-tercer cuartil) de las concentraciones de ARNm de GBP3 en sangre normalizadas con GAPDH fueron 0,0587 (0,0134-0,131) y 0,465 (0,107-0,867) en el grupo de prueba y el grupo de referencia, respectivamente (Fig. 5). Esta mediana de las concentraciones de ARNm de GBP3 normalizadas en el grupo de prueba fue 7,92 veces más baja que en el grupo de referencia ($p=0,0173$, prueba de suma de intervalos de Mann-Whitney).

La curva ROC se trazó (Fig. 6, área bajo la curva = 0,820, intervalo de confianza del 95% = 0,624 a 1,02, $p=0,0156$). Usando las concentraciones de ARNm de GBP3 en sangre normalizadas por GAPDH $<0,0914$ como umbral, los inventores pudieron identificar a las mujeres que dieron a luz antes de las 34 semanas con una sensibilidad del 60% y una especificidad del 90%. El valor predictivo positivo y el valor predictivo negativo para el ARNm de GBP3 son del 85,7% y 69,2%, respectivamente.

Los datos del ensayo qRT-PCR dirigidos al ARNm de CD16A no incluidos en la Tabla 2. Las medianas (primer cuartil-tercer cuartil) de las concentraciones de ARNm de CD16A en sangre normalizadas con GAPDH fueron 365 (267-541) y 341 (300-426) en el grupo de prueba y el grupo de referencia, respectivamente (Fig. 7). Estas concentraciones de ARNm de CD16A normalizadas en el grupo de prueba no fueron significativamente diferentes de las del grupo de referencia ($p=0,571$, prueba de suma de intervalos de Mann-Whitney).

Para visualizar el rendimiento predictivo del ARNm de CD16A, se representó la curva ROC (Fig. 8). No se observó una diferencia significativa entre el área de debajo de la curva ROC para el ARNm de CD16A (área bajo la curva = 0,580, 95% CI = 0,319 a 0,841) y el área debajo de la línea de identidad ($p=0,545$). Usando las concentraciones de

ARNm de CD16A en sangre normalizadas por GAPDH >438 como un umbral para definir a una mujer como positiva para esta prueba, los inventores pudieron identificar a las mujeres que dieron a luz antes de las 34 semanas con una sensibilidad del 30,0% y una especificidad del 90,0%. El valor predictivo positivo y el valor predictivo negativo para el ARNm de CD16A son del 75,0% y del 56,3%, respectivamente.

Datos del ensayo por qRT-PCR dirigidos al ARNm de CD62L no incluidos en la Tabla 2. Las medianas (primer cuartil-tercer cuartil) de las concentraciones de ARNm de CD62L en sangre normalizadas con GAPDH fueron 69,6 (59,7-112) y 76,7 (71,4-91,7) en el grupo de prueba y el grupo de referencia, respectivamente (Fig. 9). Estas concentraciones de ARNm de CD62L normalizadas en el grupo de prueba no fueron significativamente diferentes de las del grupo de referencia ($p=0,678$, prueba de suma de intervalos de Mann-Whitney).

Para visualizar el rendimiento predictivo del ARNm de CD62L, se representó su curva ROC (Fig. 10). No se observó una diferencia significativa entre el área de debajo de la curva ROC para el ARNm de CD62L (0,560, 95% CI = 0,288 a 0,832) y el área debajo de la línea de identidad (0,500, $p=0,650$). Usando las concentraciones de ARNm de CD62L en sangre normalizadas por GAPDH >62,4 como un umbral para definir a una mujer como positiva para esta prueba, los inventores pudieron identificar a las mujeres que dieron a luz antes de las 34 semanas con una sensibilidad del 30% y una especificidad del 90%. El valor predictivo positivo y el valor predictivo negativo para el ARNm de CD16A son del 75,0% y 56,3%, respectivamente.

20 Análisis

En este estudio, utilizando la tecnología de matriz de exones, los presentes inventores han perfilado en la resolución del nivel de exón los transcriptomas sanguíneos de las mujeres embarazadas durante la presentación de las contracciones uterinas. Los datos de expresión de ARN en todo el genoma de los glóbulos sanguíneos durante las contracciones uterinas nunca se han publicado en la literatura revisada por pares antes. Además, los inventores, por primera vez, compararon sistemáticamente los transcritos de ARN expresados diferencialmente entre las mujeres que dieron como resultado el parto antes de las 34 semanas de gestación y las que dieron como resultado el nacimiento o después de las 37 semanas. Además, se ha diseñado un método para la selección estratégica de transcripciones de ARN marcador útiles para predecir el nacimiento, entre >1,4 millones de puntos de datos de niveles de expresión de ARN. Estos nuevos datos y métodos han permitido a los investigadores en este campo hacer una lista corta de un panel de 32 transcritos de ARN (Tabla 2), entre los >30.000 genes humanos, que son útiles para predecir el nacimiento prematuro a través del análisis molecular de la sangre periférica materna.

Los inventores han demostrado la utilidad clínica de los ensayos de qRT-PCR dirigidos a nuestros transcritos de ARN marcadores de lista corta, a saber, el ARNm de B3GNT5, el ARNm de CLC y el ARNm de GBP3 (de la Tabla 2). Las concentraciones de ellas fueron significativamente diferentes entre las mujeres que dieron a luz antes de las 34 semanas (el grupo de prueba) y las que dieron a luz a las 37 semanas o más tarde (el grupo de referencia; prueba de Mann-Whitney, valores de $p = 0,002$, $0,011$ y $0,0173$ para el ARNm de B3GNT5, el ARNm de CLC y el ARNm de GBP3, respectivamente). No solo así, el intervalo intercuartil de las concentraciones dos de los tres marcadores probados, el ARNm de B3GNT5, el ARNm de CLC, no tuvo solapamiento entre los dos grupos.

En paralelo a los ensayos de qRT-PCR dirigidos a los transcritos de ARN marcador preseleccionados en este estudio (de la Tabla 2), los inventores también analizaron las muestras de ARN en sangre de los dos grupos de participantes mediante ensayos dirigidos a transcritos de ARN no enumerados en la Tabla 2. Por el contrario, las concentraciones normalizadas de GAPDH entre los dos grupos no fueron significativamente diferentes (prueba de Mann-Whitney, valores de $p = 0,571$ y $0,678$ para el ARNm de CD16A y el ARNm de CD62L, respectivamente).

Sobre la base de los resultados prometedores, el rendimiento predictivo de los marcadores se examinó mediante análisis ROC. Para cada marcador desarrollado en este estudio (como los que se enumeran en la Tabla 2), el área debajo de la curva ROC fue significativamente más grande que el área debajo de la línea de identidad ($x = y$), lo que demuestra un uso potencial para la predicción. Específicamente, las áreas bajo la curva del ARNm de B3GNT5, el ARNm de CLC y el ARNm de GBP3 fueron 0,915, 0,840 y 0,820, respectivamente. Esto indica que las probabilidades de predecir con precisión el nacimiento antes de las 34 semanas usando los tres marcadores son del 91,5%, 84,0% y 82,0%, respectivamente.

Por el contrario, las áreas bajo las curvas ROC del ARNm de CD16A y del ARNm de CD62L fueron del 0,580 y 0,560, respectivamente. Estas áreas no fueron significativamente diferentes de 0,500, que es el área debajo de la línea de identidad ($p = 0,545$ y $0,650$, respectivamente). Esto implica que estos análisis, que no se desarrollaron a partir de la Tabla 2, no tienen potencial para predecir el nacimiento.

Lo más importante es que se ha demostrado que dos de los tres marcadores identificados por las estrategias anteriores predicen el nacimiento a alta sensibilidad y alta especificidad. En particular, la sensibilidad y la especificidad para el ARNm de B3GNT5 fueron del 90,0% y 90,0%, respectivamente, y para el ARNm de CLC fueron del 88,9% y 81,8%, respectivamente. El rendimiento de estos dos marcadores novedosos se comparó favorablemente con el de la longitud cervical transvaginal (la sensibilidad y la especificidad fueron del 63,6% y del 85,7%, respectivamente), o fueron al menos a la par con la de la fibronectina fetal (la sensibilidad y la especificidad

fueron del 81,7% y 82,5%, respectivamente (Lockwood et al., 1991).

Otra ventaja importante de los marcadores sanguíneos preseleccionados en este estudio sobre la ecografía transvaginal y la fibronectina fetal es que no requieren un examen pélvico, lo que no siempre es tolerable por las mujeres embarazadas que necesitan ser examinadas. Para resumir, los presentes inventores han generado a través de este estudio un panel de 32 transcritos de ARN de sangre periférica, que es útil para la predicción del nacimiento en humanos antes de las 34 semanas con un rendimiento mejor o al mismo nivel que los marcadores actuales.

10 Listado de referencias

- Chang CW, Cheng WC, Chen CR, Shu WY, Tsai ML, Huang CL, Hsu IC. (2011). "Identification of human housekeeping genes and tissue-selective genes by microarray meta-analysis". *PLoS One*. 6(7):e22859.
- 15 Cheng WC, Chang CW, Chen CR, Tsai ML, Shu WY, Li CY, Hsu IC. (2011). "Identification of reference genes across physiological states for qRT-PCR through microarray meta-analysis". *PLoS One*. 6(2):e17347.
- Chim SSC, Jin S, Lee TYH, Lun FMF, Lee WS, Chan LYS, Jin YY, Yang N, Tong YK, Leung TY, Lau TK, Ding C, Chiu RWK and Lo YMD (2008). "Systematic search for placental DNA-methylation markers on chromosome 21: toward a maternal plasma-based epigenetic test for fetal trisomy 21". *Clin Chem* 54(3): 500-11.
- 20 Chim SSC, Lee WS, Ting YH, Chan OK, Lee SWY and Leung TY (2012). "Systematic identification of spontaneous preterm birth-associated RNA transcripts in maternal plasma". *PLoS One* 7(4): e34328.
- 25 Chim SSC, Shing TKF, Hung ECW, Leung TY, Lau TK, Chiu RWK and Lo YMD (2008). "Detection and characterization of placental microRNAs in maternal plasma". *Clin Chem* 54(3): 482-90.
- Chim SSC, Tong YK, Chiu RWK, Lau TK, Leung TN, Chan LYS, Oudejans CBM, Ding C and Lo YMD (2005). "Detection of the placental epigenetic signature of the maspin gene in maternal plasma". *Proc Natl Acad Sci U S A* 102(41): 14753-8.
- 30 Honest H, Bachmann LM, Coomarasamy A, Gupta JK, Kleijnen J and Khan KS (2003). "Accuracy of cervical transvaginal sonography in predicting preterm birth: a systematic review". *Ultrasound Obstet Gynecol* 22(3): 305-22.
- 35 Honest H, Hyde CJ and Khan KS (2012). "Prediction of spontaneous preterm birth: no good test for predicting a spontaneous preterm birth". *Curr Opin Obstet Gynecol* 24(6): 422-33.
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U and Speed TP (2003). "Exploration, normalization, and summaries of high density oligonucleotide array probe level data". *Biostatistics* 4(2): 249-64.
- 40 Lockwood CJ, Senyei AE, Dische MR, Casal D, Shah KD, Thung SN, Jones L, Deligdisch L and Garite TJ (1991). "Fetal fibronectin in cervical and vaginal secretions as a predictor of preterm delivery". *N Engl J Med* 325(10): 669-74.
- Lui YYN, Chik KW, Chiu RWK, Ho CY, Lam CWK, Lo YMD. 2002. "Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. *Clin Chem*. 48(3):421-7.
- 45 Murakawa H, Utumi T, Hasegawa I, Tanaka K and Fuzimori R (1993). "Evaluation of threatened preterm delivery by transvaginal ultrasonographic measurement of cervical length". *Obstetrics & Gynecology* 82(5): 829-832.
- 50 Sotiriadis A, Papatheodorou S, Kavvadias A and Makrydimas G (2010). "Transvaginal cervical length measurement for prediction of preterm birth in women with threatened preterm labor: a meta-analysis". *Ultrasound Obstet Gynecol* 35(1): 54-64.
- Storey JD (2002). "A direct approach to false discovery rates". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(3): 479-498.
- 55 Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR and Hogenesch JB (2004). "A gene atlas of the mouse and human protein-encoding transcriptomes". *Proc Natl Acad Sci U S A* 101(16): 6062-7.
- 60 Tsui DWY, Lam YMD, Lee WS, Leung TY, Lau TK, Lau ET, Tang MH, Akolekar R, Nicolaides KH, Chiu RWK, Lo YMD and Chim SSC (2010). "Systematic identification of placental epigenetic signatures for the noninvasive prenatal detection of Edwards syndrome". *PLoS One* 5(11): e15069.
- 65 Tsui NBY, Chim SSC, Chiu RWK, LauTK, Ng EKO, Leung TN, Tong YK, Chan KCA and Lo YMD (2004). "Systematic micro-array based identification of placental mRNA in maternal plasma: towards non-invasive prenatal gene ex-

pression profiling". J Med Genet 41(6): 461-7.

Tabla 1. Características de los participantes en el análisis de la matriz de exones de todo el genoma.

	Resultado en nacimientos prematuros	Resultado en nacimientos a término	Valor P ^a
Número de participantes (n)	3	3	-
Edad de la madre en años (media +/- desviación estándar (DE))	28 +/- 8,5	30 +/- 2,6	0,7621
Nulíparas (n, %), semanas gestacionales al momento de la extracción de sangre (media +/- DE)	0 (0%) 31,5 +/- 0,436	1 (33%) 31,9 +/- 0,869	1,0000 0,5846
Semanas gestacionales al momento del parto (media +/- DE)	31,6 +/- 0,515	39,3 +/- 0,218	<= 0,0001
Peso al nacer en gramos (media +/- DE)	1619 +/- 254,7	3095 +/- 172,8	0,0011
Recuento de glóbulos blancos maternos (media +/- DE)	9,47 +/- 2,59	12,1 +/- 0,808	0,1638
Hemorragia anteparto (n, %)	1 (33%)	0 (0%)	1.00

^a Prueba T para variables continuas. (Los datos pasaron la prueba de normalidad y la prueba de varianza igual). Prueba exacta de Fisher para variables nominales.

5

ES 2 724 328 T3

Tabla 2. Transcritos génicos y de ARN identificados como útiles como marcadores de sangre para predecir un nacimiento humano antes de las 37 semanas

Símbolo del gen	Nombre del gen	HGNC ID	Acceso a RefSeq	Primer cuartil de señal de sonda en grupo de prueba+	Tercer cuartil de señal de sonda en grupo de prueba	Primer cuartil de señal de sonda en grupo de referencia**	Tercer cuartil de señal de sonda en grupo de referencia	Dirección de cambio	Cambio en veces de señal de sonda mediana	Valor q
<i>B3GN75</i>	UDP-GlcNAc-betaGal beta-1,3-N-acetilglucosaminotransferasa 5	HGNC: 15684	NM_032047	1533	1855	627	745	Aumentado en grupo de prueba	2.64	0.00740
<i>EFCAB13</i>	Dominio de unión de calcio de mano EF 13	HGNC: 26664	NM_152347	106	206	457	618	Reducido en grupo de prueba	2.67	0.00758
<i>TREML4</i>	receptor de acionamiento expresado en células mieloides de tipo 4	HGNC: 30607	NM_198163	561	624	206	261	Aumentado en grupo de prueba	2.78	0.00707
<i>ADORA3</i>	receptor de adenosina A3	HGNC: 268	NM_020683	118	142	349	354	Reducido en grupo de prueba	2.88	0.00708
<i>PDE6D</i>	fosfodiesterasa 6D, específico de cGMP, rod., delta	HGNC: 8788	NM_002601	334	388	120	152	Aumentado en grupo de prueba	2.88	0.00772
<i>CD177</i>	molécula CD177	HGNC: 30072	NM_020406	4884	5518	1387	2127	Aumentado en grupo de prueba	2.90	0.00735
<i>SCMH1</i>	comb sex. en homólogo 1 en media pierna (drosophila)	HGNC: 19003	NM_001031694	89.9	105	230	307	Reducido en grupo de prueba	2.93	0.00707
<i>ATP2B4</i>	ATPasa, transporte Ca ^{***} membrana plasmática 4	HGNC: 817	NM_001001396	782	952	284	342	Aumentado en grupo de prueba	2.97	0.00769
<i>ALDH1A1</i>	aldehído deshidrogenasa 1 miembro de la familia A1	HGNC: 402	NM_000689	48.8	68.1	179	196	Reducido en grupo de prueba	3.03	0.00708
<i>GPR56</i>	receptor 56 acoplado con la proteína G	HGNC: 4512	NM_005682	606	1025	223	297	Aumentado en grupo de prueba	3.14	0.00708

ES 2 724 328 T3

Tabla 2. Transcritos génicos y de ARN identificados como útiles como marcadores de sangre para predecir un nacimiento humano antes de las 37 semanas

Símbolo del gen	Nombre del gen	HGNC ID	Acceso a RefSeq	Primer cuartil de señal de sonda en grupo de prueba*	Tercer cuartil de señal de sonda en grupo de prueba	Primer cuartil de señal de sonda en grupo de referencia**	Tercer cuartil de señal de sonda en grupo de referencia	Dirección de cambio	Cambio en veces de señal de sonda mediana	Valor q
<i>FAH</i>	fumarilacetato hidrolasa (fumarilacetacetasa)	HGNC: 3579 NM_000137		1365	1571	403	541	Aumentado en grupo de prueba	3.18	0.00764
<i>GPR34</i>	receptor 34 acoplado a la proteína G	HGNC: 4490 NM_005300		154	216	663	852	Reducido en grupo de prueba	3.20	0.00708
<i>CLK4</i>	quinasa de litio CDC 4	HGNC: 13659 NM_020666		59.3	67.0	204	211	Reducido en grupo de prueba	3.20	0.00684
<i>PTGDR</i>	receptor de prostaglandina D2 (DP)	HGNC: 9591 NM_000953		542	808	208	249	Aumentado en grupo de prueba	3.22	0.00724
<i>FNTA</i>	farnesiltransferasa, caja CAAX, alfa	HGNC: 3782 NM_002027 / NR_033698 / AB209689		77.7	109	228	272	Reducido en grupo de prueba	3.32	0.00740
<i>CTSG</i>	catepsina G	HGNC: 2532 NM_001911		58.2	69.0	188	248	Reducido en grupo de prueba	3.34	0.00679
<i>MPO</i>	mieloperoxidasa	HGNC: 7218 NM_000250		33.8	76.0	203	272	Reducido en grupo de prueba	3.59	0.00758
<i>CPA3</i>	carboxipeptidasa A3 (mastocito)	HGNC: 2298 NM_001870		59.1	95	272	377	Reducido en grupo de prueba	4.23	0.00684
<i>LILRA3</i>	receptor de tipo inmunoglobulina leucocítica, subfamilia A (sin dominio TM), miembro 3	HGNC: 6604 NM_006885		212	417	1145	1733	Reducido en grupo de prueba	4.26	0.00707
<i>AK5</i>	adenilato quinasa 5	HGNC: 365 NM_174858		47.2	61.1	179	229	Reducido en grupo de prueba	4.33	0.00684

ES 2 724 328 T3

Tabla 2. Transcritos génicos y de ARN identificados como útiles como marcadores de sangre para predecir un nacimiento humano antes de las 37 semanas

Simbolo del gen	Nombre del gen	HGNC ID	Acceso a RefSeq	Primer cuartil de señal de sonda en grupo de prueba*	Tercer cuartil de señal de sonda en grupo de prueba	Primer cuartil de señal de sonda en grupo de referencia**	Tercer cuartil de señal de sonda en grupo de referencia	Dirección de cambio	Cambio en veces de señal de sonda mediana	Valor q
<i>KLRD1</i>	receptor de tipo lectina de células asesinas subfamilia D, miembro 1	HGNC: 6378	NM_002262	453	747	140	204	Aumentado en grupo de prueba	4.43	0.00735
<i>YPEL1</i>	yippee de tipo 1 (Drosophila)	HGNC: 12845	NM_013313	219	296	51.1	70.4	Aumentado en grupo de prueba	4.59	0.00746
<i>NR4A3</i>	receptor nuclear subfamilia 4, grupo A, miembro 3	HGNC: 7962	NM_006981	234	450	44.7	94.3	Aumentado en grupo de prueba	4.77	0.00740
<i>CCR3</i>	receptor 3 de quimioquina (motivo C-C)	HGNC: 1604	NM_001837	32.6	45.7	138	245	Reducido en grupo de prueba	4.95	0.00712
<i>THEM5</i>	superfamilia de tioesterasa miembro 5	HGNC: 26755	NM_182578	297	429	61.1	103	Aumentado en grupo de prueba	5.01	0.00684
<i>KLRG1</i>	receptor tipo lectina de células asesinas, subfamilia C, miembro 1	HGNC: 6374	NM_002259	153	172	29.9	49.4	Aumentado en grupo de prueba	5.15	0.00748
<i>CLC</i>	galectina de cristal de Charcot-Leyden	HGNC: 2014	NM_001828	790	1049	3510	4871	Reducido en grupo de prueba	5.55	0.00737
<i>GBP3</i>	proteína de unión 3 de guanilato	HGNC: 4184	NM_018284	12.9	23.5	133	266	Reducido en grupo de prueba	7.86	0.00764
<i>HSD17B4</i>	hidroxiesteroide (17-beta) deshidrogenasa 4	HGNC: 5213	NM_000414	36.0	47.6	302	339	Reducido en grupo de prueba	8.81	0.00684
<i>IL5RA</i>	receptor 5 de interleuquina, alfa	HGNC: 6017	NM_000564	21.3	32.9	216	310	Reducido en grupo de prueba	9.64	0.00740

Tabla 2. Transcritos génicos y de ARN identificados como útiles como marcadores de sangre para predecir un nacimiento humano antes de las 37 semanas

Símbolo del gen	Nombre del gen	HGNC ID	Acceso a RefSeq	Primer cuartil de señal de sonda en grupo de prueba*	Tercer cuartil de señal de sonda en grupo de prueba	Primer cuartil de señal de sonda en grupo de referencia**	Tercer cuartil de señal de sonda en grupo de referencia	Dirección de cambio	Cambio en veces de señal de sonda mediana	Valor q
<i>HRH4</i>	Receptor de histamina H4	HGNC: 17383	NM_021624	14.9	19.7	127	191	Reducido en grupo de prueba	9.71	0.00740
<i>EDIL3</i>	Repeticiones similares a EGF Y dominios 3 similares a discodina I	HGNC: 3173	NM_005711	10.8	12.4	208	363	Reducido en grupo de prueba	29.0	0.00708

Legenda:

HGNC: Comité de Nomenclatura Genética HUGO en el European Bioinformatics Institute (<http://www.genenames.org/>).

Ref Seq: Base de Datos de Secuencia de Referencia en el National Center for Biotechnology Information, National Library of Medicine (<http://ftp://www.ncbi.nlm.nih.gov/refseq/>)

valor q: valores q ajustados por Tasa de Descubrimiento Falso

* Grupo de prueba compuesto por mujeres con contracciones uterinas (< 34 semanas gestacionales) y dando como resultado el nacimiento antes de las 34 semanas.

** Grupo de prueba compuesto por mujeres con contracciones uterinas (< 34 semanas gestacionales) y dando como resultado el nacimiento en o después de las 37 semanas.

Tabla 3. Características de los participantes para el análisis de qRT-PCR

	Resultado en nacimientos prematuros	Resultado en nacimientos a término	Valor P ^a
Número de participantes con contracciones uterinas en la extracción de sangre (n)	10	10	-
Edad de la madre en años (media +/- desviación estándar (DE))	30 +/- 7,4	33 +/- 5,8	0,3088
Nulíparas (n,%) semanas gestacionales al momento de extraer sangre (media +/- DE)	4 (40%) 31,7 +/- 2,34	4 (40%) 29,8 +/- 2,59	1,00 0,0976
Semanas gestacionales al momento del parto (media +/- DE)	31,9 +/- 2,13	39,4 +/- 1,06	< 0,0001
Peso al nacer en gramos (media +/- DE)	1735 +/- 470,2	3519 +/- 432,6	< 0,0001
Recuento de glóbulos blancos maternos en 10 ⁹ por L (media +/- DE)	13,4 +/- 4,33	10,4 +/- 1,75	0,0565
Hemorragia anteparto (n, %)	3 (30%)	1 (10%)	0,582
Rotura preparto de la membrana (n, %)	7 (70%)	4 (40%)	0,370

5 ^a Prueba T para variables continuas. (Los datos pasaron la prueba de normalidad y la prueba de varianza igual). Prueba exacta de Fisher para variables nominales.

REIVINDICACIONES

1. Un método para determinar el riesgo de nacimiento prematuro, que comprende las etapas de:
- 5 (a) medir el nivel de ARNm de un gen marcador en una muestra de sangre tomada de una mujer embarazada, en donde el gen marcador es B3GNT5; y
- (b) comparar el nivel de ARNm obtenido en la etapa (a) con un control estándar, en donde un aumento en el nivel de ARNm de B3GNT5 cuando se compara con el control estándar indica que la mujer tiene un riesgo mayor de parto prematuro.
- 10 2. El método de acuerdo con la reivindicación 1, en el que la muestra de sangre es
- 15 sangre entera,
glóbulos sanguíneos aislados,
plasma o
suero.
3. El método de acuerdo con la reivindicación 1, en donde el nivel de ARNm se normaliza sobre el nivel de ARNm de un gen de referencia en la misma muestra antes de la etapa (b).
- 20 4. El método de acuerdo con la reivindicación 1, en donde la etapa (a) comprende
- 25 espectrometría de masas o hibridación con una micromatriz, sonda de fluorescencia o baliza molecular, una reacción de amplificación o
- un ensayo de hibridación de polinucleótidos que utiliza una sonda de polinucleótidos que comprende una unidad estructural detectable.
- 30 5. El método de acuerdo con la reivindicación 4, en el que
- la reacción de amplificación es una reacción en cadena de polimerasa (PCR), preferiblemente PCR de transcriptasa inversa (RT-PCR), o el ensayo de hibridación de polinucleótidos es un análisis de transferencia Southern, análisis de transferencia Northern o ensayo de hibridación in situ.
- 35 6. El método de acuerdo con la reivindicación 1, cuando se indica que la mujer tiene un mayor riesgo de parto prematuro, que comprende además una etapa para tratar a la mujer por parto prematuro.
7. El método de acuerdo con la reivindicación 1, en donde el nivel de ARNm de más de un gen marcador se mide y se compara con sus respectivos controles estándar para determinar el riesgo de nacimiento prematuro.
- 40 8. Un uso ex vivo de un kit para determinar el riesgo de nacimiento prematuro, que comprende (1) un control estándar que proporciona un nivel promedio de un ARN de gen marcador; y (2) un agente que identifica específica y cuantitativamente el ARNm del gen marcador, en donde el gen marcador es B3GNT5.
- 45 9. El uso de acuerdo con la reivindicación 8, en donde el agente es una sonda polinucleotídica que se hibrida con el ARNm del gen marcador.
10. El uso de acuerdo con la reivindicación 9, en donde la sonda de polinucleótido comprende una unidad estructural detectable.
- 50 11. El uso de acuerdo con la reivindicación 8, que comprende además dos cebadores oligonucleotídicos para amplificar específicamente al menos un segmento del ADNc del gen marcador o su complemento en una reacción de amplificación.
- 55 12. El uso de acuerdo con la reivindicación 8, que comprende, además, un manual de instrucciones.

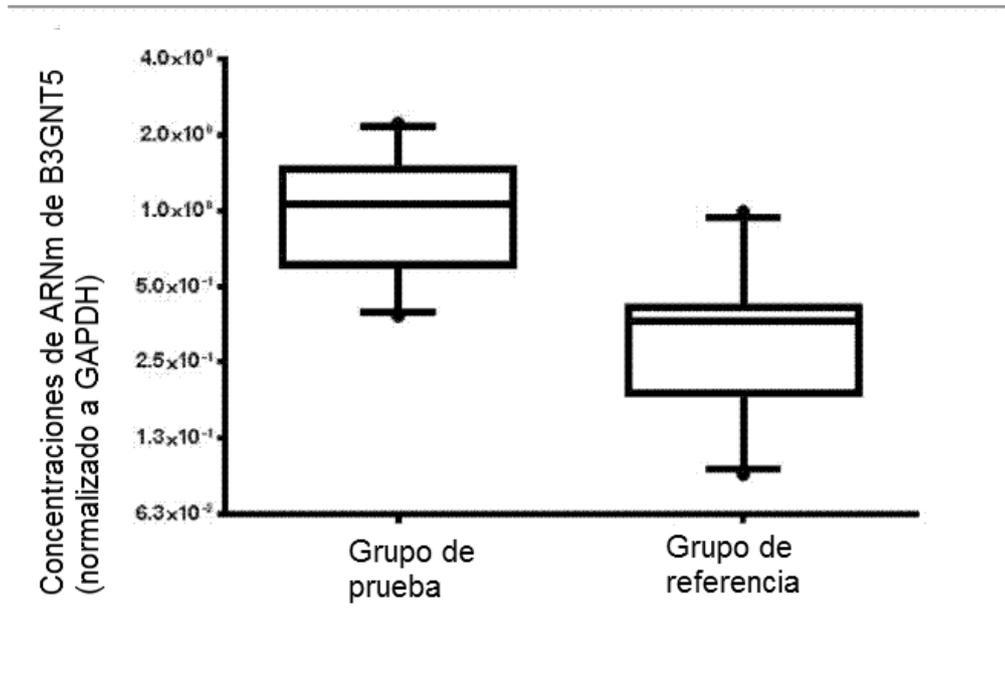


FIG. 1

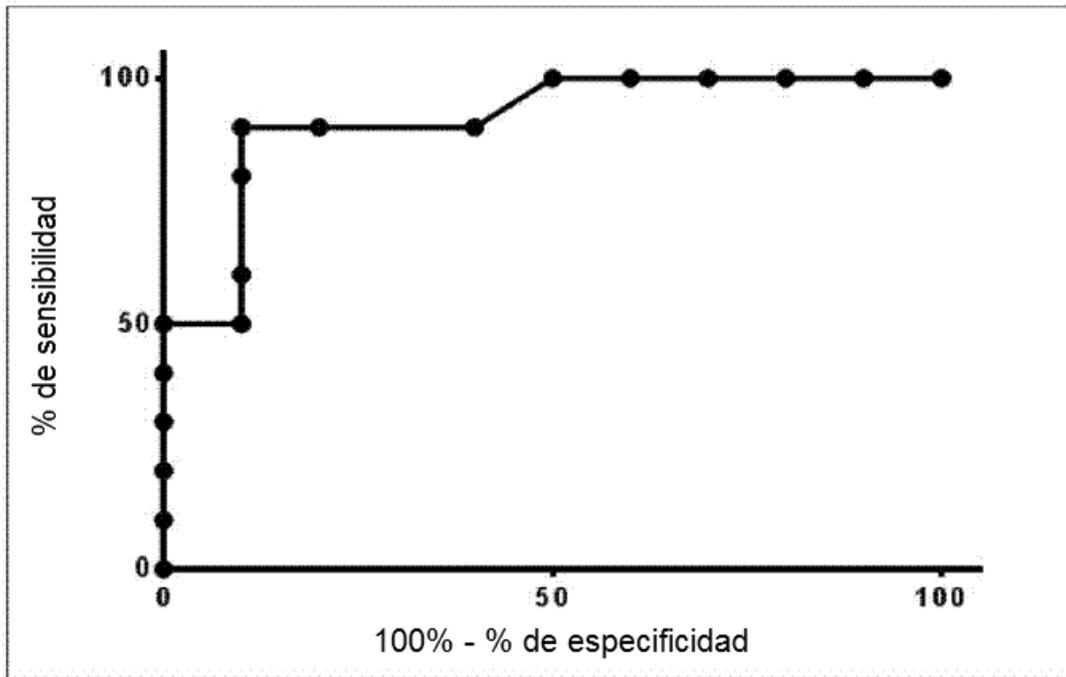


FIG. 2

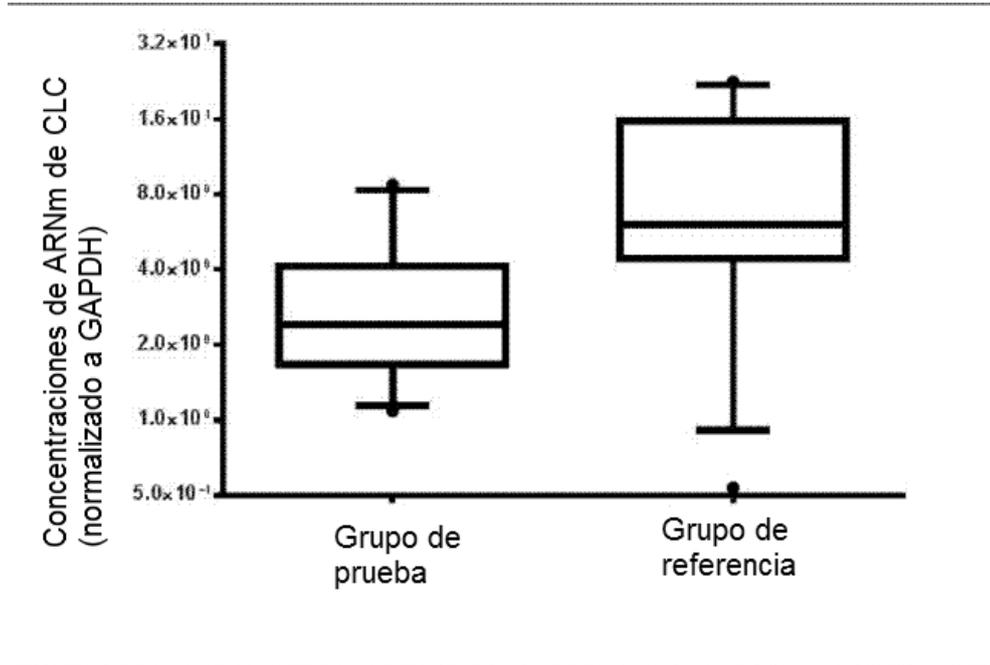


FIG. 3

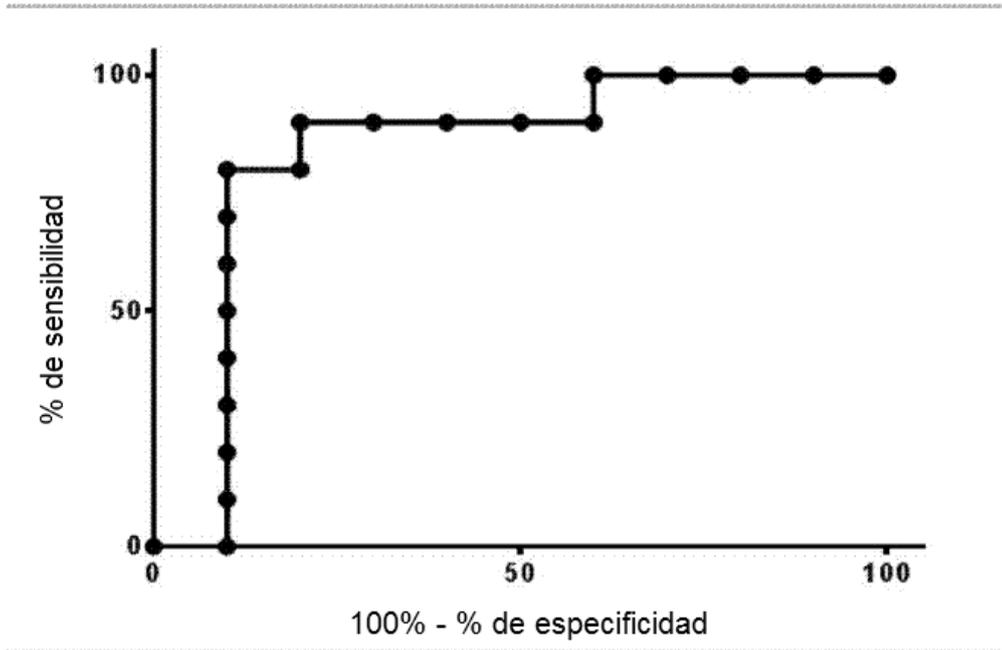


FIG. 4

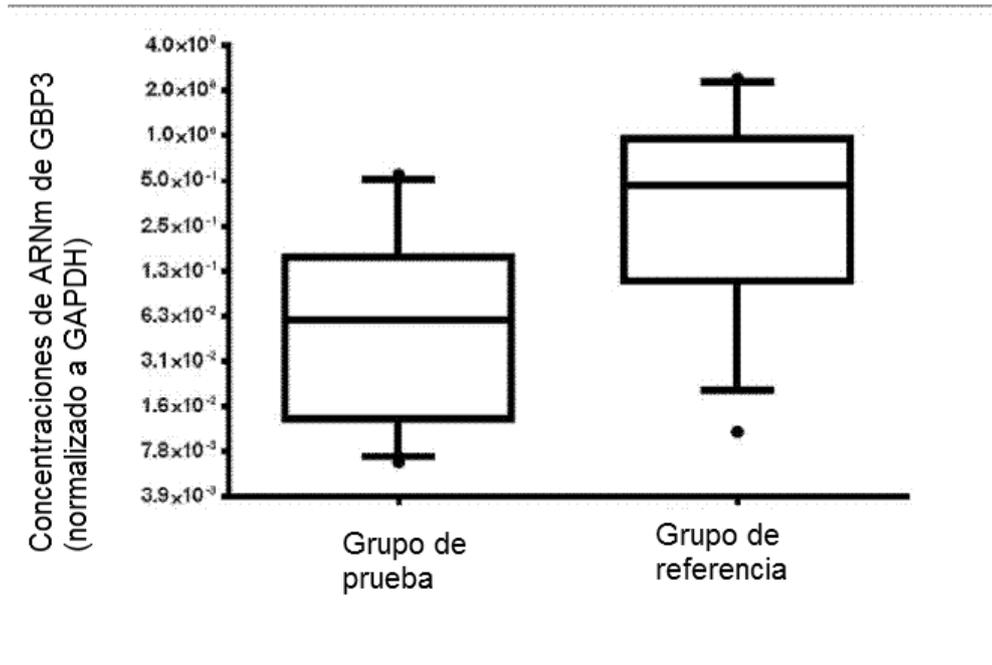


FIG. 5

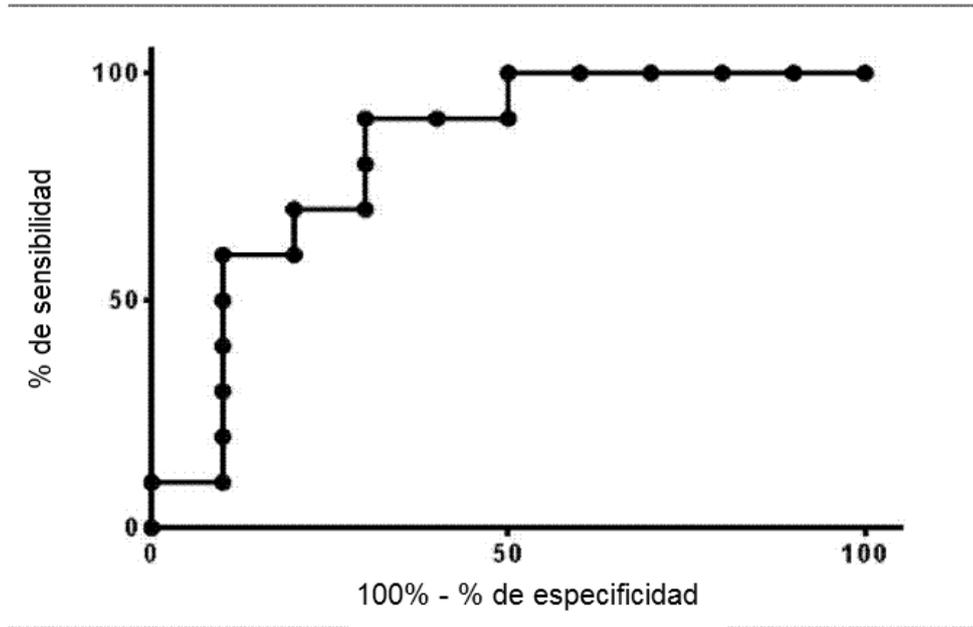


FIG. 6

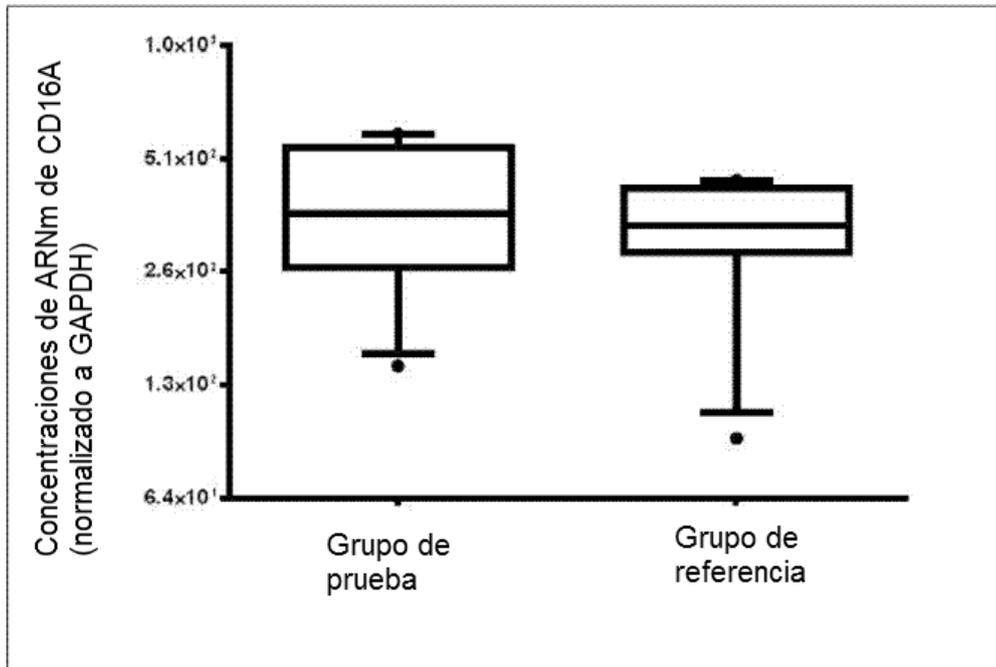


FIG. 7

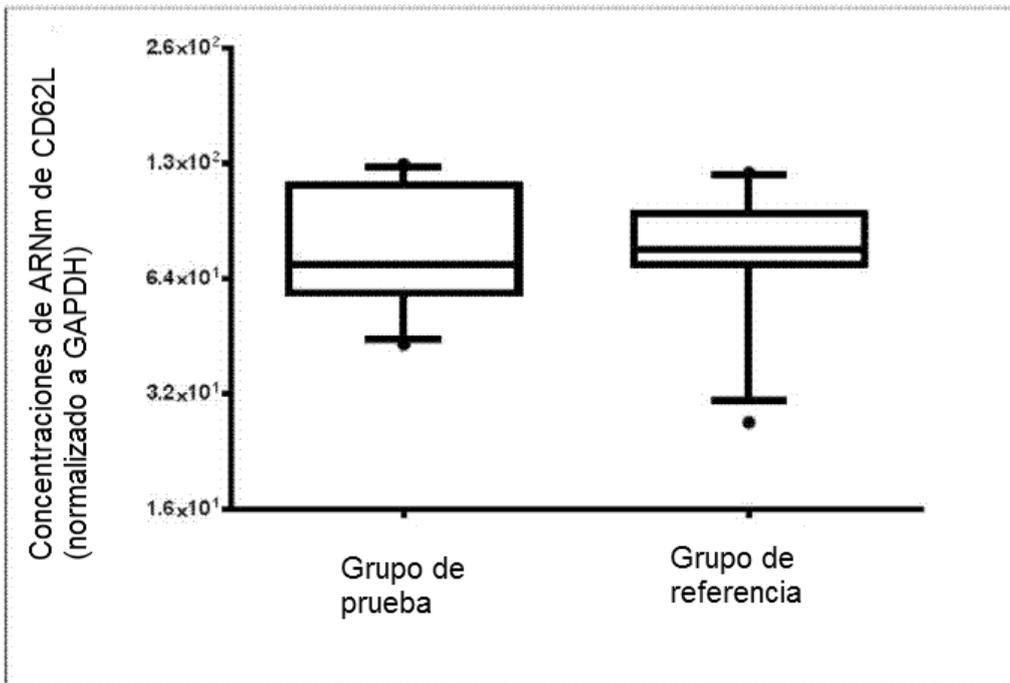


FIG. 9

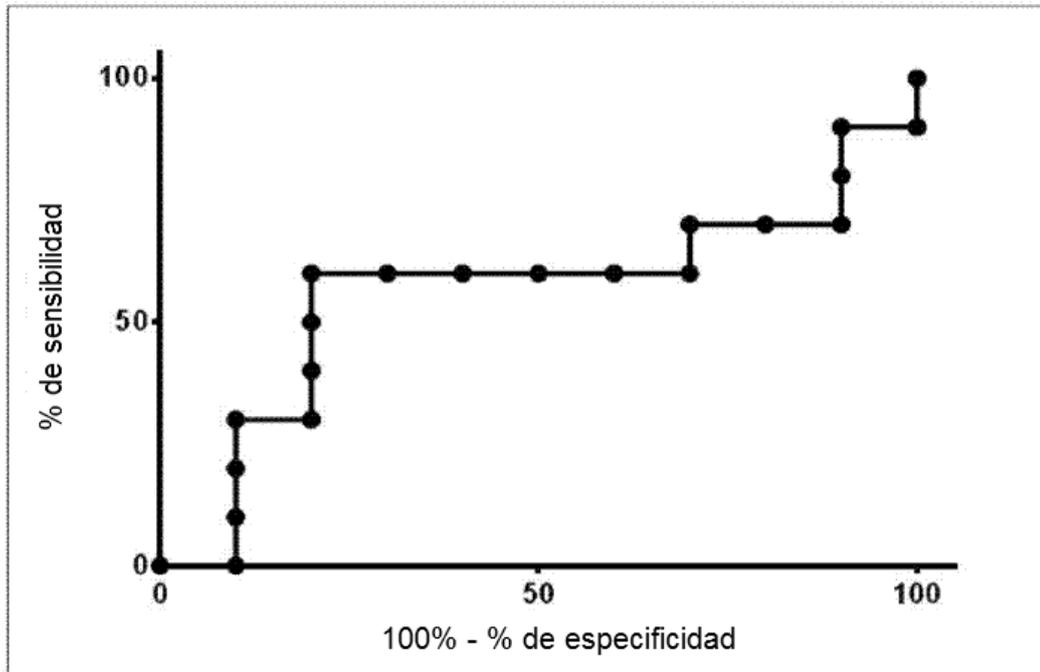


FIG. 10