(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 670 544**

(51) Int. CI.:

C12Q 1/68 (2008.01)
G06F 19/00 (2008.01)
G06F 19/22 (2011.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **28.02.2014 PCT/IB2014/059351**

(87) Fecha y número de publicación internacional: **04.09.2014 WO14132244**

(96) Fecha de presentación y número de la solicitud europea: **28.02.2014 E 14757525 (2)**

(97) Fecha y número de publicación de la concesión europea: **25.04.2018 EP 2961851**

(54) Título: **Análisis de transcriptoma de plasma materno por secuenciación masiva en paralelo de ARN**

(30) Prioridad:

28.02.2013 US 201361770985 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
30.05.2018

(73) Titular/es:

The Chinese University Of Hong Kong (100.0%)
Knowledge Transfer Office Room 301 Pi Ch'iu
Building Shatin New Territories
Hong Kong, CN

(72) Inventor/es:

LO, YUK MING DENNIS;
CHIU, ROSSA WAI KWUN;
CHAN, KWAN CHEE;
JIANG, PEIYONG y
TSUI, BO YIN

(74) Agente/Representante:

ELZABURU, S.L.P

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

Análisis de transcriptoma de plasma materno por secuenciación masiva en paralelo de ARN

Se han descrito moléculas de ARN extracelular que son específicas del embarazo en el plasma materno[1]. Han proporcionado una herramienta de pruebas no invasivas para la evaluación fetal y seguimiento del embarazo utilizando simplemente una muestra de sangre periférica de la madre.

Hasta la fecha, se han desarrollado una serie de aplicaciones de diagnóstico prenatal prometedoras utilizando ARN de plasma materno[2-8]. Los investigadores han estado buscando activamente más marcadores de ARN con el fin de ampliar las aplicaciones en diferentes contextos de trastornos fetales y patologías del embarazo. Por ejemplo, el documento WO2004065629A1 proporciona métodos y equipos para diagnóstico, seguimiento o predicción de trastornos relacionados con el embarazo tales como preclamsia o embarazo en fase inicial, que comprende cuantificar los ARNm circulantes que codifican marcadores tales como hCG-beta.

En teoría, el método de identificación de ARN marcador más sencillo consiste en describir directamente moléculas de ARN extracelular en plasma materno. Este método sin embargo no ha sido fácil porque las tecnologías de selección de alto rendimiento convencionales, tales como el análisis de chips y análisis en serie de la expresión génica (SAGE), tienen una capacidad limitada para detectar los concentraciones normalmente bajas de ARN extracelular parcialmente degradado en el plasma materno[9] . En cambio, la mayoría de las estrategias de detección del marcador de ARN descritas operan indirectamente comparando las características de la expresión de la placenta y las células sanguíneas maternas[10]. Sólo las transcripciones que se expresan mucho más altas en la placenta que en las células de sangre materna se estudian más en muestras de plasma materno por tecnologías de alta sensibilidad pero de bajo rendimiento, tal como la reacción en cadena de la polimerasa de la transcriptasa inversa en tiempo real (RT-PCR). Hasta ahora los marcadores de ARN del plasma identificados por este método indirecto son relativamente limitados. Esto es posiblemente debido a que la estrategia de extracción en base al tejido no ha tenido totalmente en cuenta todos los factores biológicos que influyen en las concentraciones de ARN de la placenta en el plasma materno. Además, las transcripciones que se expresan y liberan por tejidos no placentarios en respuesta al embarazo no pudieron identificarse por este método. Una metodología sensible y de alto rendimiento que permite la caracterización directa del transcriptoma del plasma materno sería por lo tanto muy deseable.

Asimismo, la caracterización directa del ARN del plasma puede ser útil en otros escenarios en los que hay una mezcla de moléculas de ARN de dos individuos, tales como para trasplante de órganos. La circulación de las receptoras de trasplante contiene moléculas de ácido nucleico tanto de la donante como de la receptora. Un cambio en las características relativas de las moléculas de ARN aportadas por la donante o la receptora puede revelar patologías en el órgano trasplantado o la receptora, tales como el rechazo del trasplante.

El documento US2010311046A1 describe métodos para el diagnóstico prenatal de trisomía 21, que comprende analizar SNP informativos o transcritos de ARN que proceden de locus en el cromosoma 21. La trisomía 21 fetal se determina a continuación comparando las relaciones entre los alelos de un sitio polimórfico en transcritos de ARN específicos de locus y de tejidos desde un feto aneuploide a un feto euploide. El estado aneuploide viene determinado por las relaciones anormales entre los SNP informativos en los transcritos de ARN. Se seleccionan SNP que se encuentran en el ARN con expresión de tejido placentario.

Según la presente invención, se proporciona un método de diagnóstico de un trastorno relacionado con el embarazo utilizando una muestra de una hembra embarazada con un feto, comprendiendo el método: recibir un gran número de lecturas de secuencias, en donde las lecturas de secuencias se obtienen a partir de un análisis de moléculas de ARN obtenidas de la muestra, muestra que contiene una mezcla de moléculas de ARN de la madre y del feto; identificar, mediante un sistema informático, ubicaciones de las lecturas de secuencias en una secuencia de referencia; identificar uno o más locus informativos, cada uno de los cuales es homocigótico en una primera entidad de un primer alelo correspondiente y un segundo alelo correspondiente, en donde la primera entidad es la mujer embarazada o el feto, y la segunda entidad es distinta de la mujer embarazada y del feto; filtrar uno o más locus informativos para identificar uno o más locus informativos filtrados: que están situados dentro de una región expresada de la secuencia de referencia, en la que está situado al menos un primer número determinado de lecturas de secuencias en el gran número de lecturas de secuencias que contienen el segundo alelo correspondiente, para cada uno de los locus informativos filtrados: determinar un primer número de lecturas de secuencias situadas en el locus informativo filtrado y que contienen el primer alelo correspondiente, determinar un segundo número de lecturas de secuencias situadas en el locus informativo filtrado y que contienen el segundo alelo correspondiente, calcular una primera suma de los primeros números y la segunda suma de los segundos números; calcular la relación de la primera suma y la segunda suma, y comparar la relación con un criterio de valoración para determinar si el feto, la madre o el embarazo tiene un trastorno relacionado con el embarazo, determinándose el valor discriminatorio a partir de una o más muestras obtenidas de las mujeres embarazadas sin el trastorno relacionado con el embarazo.

Los presentes métodos no están limitados a diagnósticos prenatales y pueden aplicarse a cualquier muestra biológica que contenga una mezcla de moléculas de ARN procedentes de dos individuos.

Por ejemplo, puede utilizarse una muestra de plasma sanguíneo obtenido de una receptora de trasplante de

órganos. Los transcritos expresados en el órgano trasplantado reflejan genotipo de la donante y se producen a niveles detectables en la sangre de la receptora. Mediante la lectura de estas transcripciones, pueden ser identificados locus informativos dentro de los genes expresados tanto por la donante como por la receptora, y se puede medir los niveles de expresión relativos de los alelos de cada individuo. Niveles de expresión anormales del alelo aportado únicamente por la donante o la receptora se pueden utilizar para diagnosticar trastornos relacionados con el trasplante.

Las muestras biológicas que pueden utilizarse en los presentes métodos incluyen muestras de sangre, plasma, suero, orina, saliva y de tejido.

Por ejemplo, se han detectado ácidos nucleicos fetales en la orina de mujeres embarazadas. Se ha demostrado que la orina de los receptores de trasplante renal contiene ácidos nucleicos sin células y células del órgano trasplantado. Se ha observado microhibridismo en muchas enfermedades. Microhibridismo se refiere a la presencia de una fuente de células o ácidos nucleicos de otra persona en el cuerpo, incluidos órganos y tejidos de un individuo en particular. Se ha observado microhibridismo en biopsias de la tiroides, hígado, bazo, piel, médula ósea y otros tejidos. El microhibridismo puede producirse como resultado de embarazos anteriores o transfusiones de sangre.

La Fig. 1 muestra una relación positiva entre la relación alélica B/A y los niveles relativos de expresión de los respectivos genes en la placenta y células sanguíneas maternas.

La Fig. 2 muestra una relación entre el nivel de la expresión génica total y los niveles relativos de expresión de los genes en la placenta y células sanguíneas maternas.

La Fig. 3 es un diagrama de flujo para métodos de diagnóstico de un trastorno relacionado con el embarazo,

Las Fig. 4 y 5 muestran resúmenes de resultados de la alineación de lecturas de secuencias de ARN.

La Fig. 6 muestra los datos de la secuenciación del ARN de plasma sanguíneo de dos individuos. La Fig. 6A muestra la distribución de GC% de lecturas secuenciadas de las bibliotecas de secuencias de ARN de M9356P (con pretratamiento Ribo-Zero Gold) y M9415P (sin pretratamiento Ribo-Zero Gold). La Fig. 6B muestra la correlación de las características de la expresión génica entre M9356P y M9415P.

La Fig. 7 muestra las relaciones alélicas ARN-SNP y las concentraciones plasmáticas totales de diferentes transcritos de ARN en el plasma materno representadas frente a los niveles de expresión del tejido relativos (células sanguíneas/placenta).

La Fig. 8 muestra relaciones alélicas ARN-SNP y los niveles plasmáticos totales de diferentes genes que contienen los SNP informativos representados frente al nivel de expresión relativo del tejido (células sanguíneas/placenta).

La Fig. 9 muestra relaciones alélicas ARN-SNP para alelos de SNP feto-específicos y los niveles de expresión de transcritos de ARN que contienen estos alelos en el plasma materno.

La Fig. 10 muestra relaciones alélicas ARN-SNP para alelos de SNP específicos de la madre y los niveles de expresión de transcritos de ARN que contienen estos alelos en el plasma materno.

La Fig. 11 muestra datos para alelos de SNP específicos de la madre en pacientes con preeclampsia y de referencia. La Fig. 11A muestra relaciones alélicas ARN-SNP del caso 5641 (preeclamsia de aparición tardía en el tercer trimestre) y del caso 7171 (referencia) para SNP informativos que comprenden alelos de SNP específicos de la madre. La fig. 11B muestra el número de veces de diferencias de relaciones alélicas ARN-SNP y de los niveles plasmáticos del caso 5641 respecto a las de 7171 (referencia) para SNP informativos que comprenden alelos de SNP específicos de la madre.

La Fig. 12 muestra datos para alelos de SNP específicos de la madre en pacientes preeclámpsicas y de referencia. La Fig. 12A muestra relaciones alélicas ARN-SNP del caso 5641 (preeclampsia de aparición tardía en el tercer trimestre) y del caso 9356 (referencia) para SNP informativos que comprenden alelos de SNP específicos de la madre. La Fig. 12B muestra el número veces de diferencias de relaciones alélicas ARN-SNP y de niveles plasmáticos de caso 5641 respecto a las de 9356 para SNP informativos que comprenden alelos de SNP específicos de la madre.

La Fig. 13 muestra los datos en la Fig. 11 en forma de tabla.

La Fig. 14 muestra los datos en la Fig. 12 en forma de tabla.

La Fig. 15 es un diagrama de flujo para procedimientos de determinación, en una muestra de una hembra embarazada con un feto, una parte de ARN que es de origen fetal.

La Fig. 16 muestra los resultados del análisis de SNP informativo del transcriptoma de plasma materno.

La Fig. 17 muestra contribuciones fetales y maternas relativas sin filtrar la expresión específica de alelo en las

muestras de plasma maternos de final del embarazo.

La Fig. 18 muestra contribuciones fetales y maternas en el transcriptoma de plasma materno. La Fig. 18A muestra proporciones de transcritos fetal y materno en el plasma materno de embarazos en las fases inicial y final. La Fig. 18B muestra los recuentos de alelos y proporciones fetal-alelo antes y después del parto.

La Fig. 19 compara contribuciones fetal y materna fraccionadas en un ARN-SNP para los casos de embarazo pre-eclámpsico y de referencia,

La Fig. 20 es un diagrama de flujo para los métodos de designación de un locus genómico como marcador materno o fetal.

La Fig. 21 es una lista de genes relacionados con el embarazo.

La Fig. 22 muestra niveles de expresión más bajos de genes relacionados con el embarazo en hembras después del parto en comparación con antes del parto.

La Fig. 23 muestra la agrupación jerárquica de las muestras de plasma que utilizan los 131 genes relacionados con el embarazo.

La Fig. 24 muestra los niveles de expresión de 131 genes relacionados con el embarazo en la placenta, células de sangre materna y el plasma materno. La Fig. 24A es un mapa de calor de la expresión génica (log_2 (nivel de transcripción)) de la placenta y de las células sanguíneas maternas, así como del plasma materno antes y después del parto de los dos casos de embarazo en la fase final. Los 131 genes relacionados con el embarazo se expresaron preferentemente en la placenta. La fig. 24B demuestra que los niveles de expresión de los 131 genes en la placenta y en el plasma se correlacionaron positivamente ($P < 0,05$, correlación de Spearman).

La Fig. 25 compara contribuciones fetales y maternas con ARN en el plasma para genes identificados con los métodos basados en el plasma y en tejidos.

La Fig. 26 lista 98 ARN relacionados con la preeclampsia en el plasma materno identificados por ARN-Seq. Se muestran los niveles de expresión en el plasma recogido de las mujeres embarazadas sin complicaciones y mujeres que han desarrollado preeclampsia.

La Fig. 27 muestra recuentos de lectura de secuenciación de ARN en puntos del polimorfismo de un solo nucleótido (SNP) en ARN con *PAPPA* para el caso de 9415. El alelo A es el alelo compartido y el alelo G es el alelo específico del feto.

La Fig. 28 muestra recuentos de lectura de la secuenciación de ARN en puntos del polimorfismo de un solo nucleótido (SNP) en ARN con *H19* para el caso 9415.

La Fig. 29 muestra la detección de estado de metilación *H19* materna por digestión de la enzima de restricción sensible a la metilación.

La Fig. 30 muestra un diagrama de bloques de un ejemplo de sistema de ordenador 3000 utilizable con el sistema y métodos según realizaciones de la presente invención.

I. Introducción

La presencia de ARN fetal sin células en el plasma materno se describió hace más de una década[11]. Después de este hallazgo, ya se han llevado a cabo muchos estudios para detectar ARN circulante de origen fetal y placentario en el plasma materno[1,10,12]. Curiosamente, se encontró que los niveles de expresión de los transcritos específicos de la placenta en el plasma se correlacionan positivamente con los de tejidos de la placenta[10], lo que subraya la utilidad clínica de análisis de ARN del plasma como una herramienta no invasiva para controlar la salud y desarrollo de la placenta o el feto. De hecho, el examen de ARN materno circulante ha encontrado aplicaciones clínicas para los trastornos relacionados con el embarazo o la placenta como la preeclampsia[2,13-15], retraso del crecimiento intrauterino[4] y parto prematuro[8] , así como para pruebas no invasivas de las aneuploidías cromosómicas fetales[5,7,16]. Dichos desarrollos destacan las posibles utilidades de biomarcadores de ARN para la evaluación molecular de trastornos prenatales.

A pesar de la prometedora perspectiva de análisis de ARN del plasma en las pruebas prenatales, sigue habiendo un número limitado de transcritos relacionadas con el embarazo o la placenta bien validados en el plasma materno hasta la fecha. En este sentido, el examen de los biomarcadores de ARN en el plasma había sido llevado a cabo utilizando la transcriptasa inversa (RT)-PCR[1,4,8,10,1,2,13], un método sensible que normalmente se dirigiría a un número relativamente pequeño de especies de ARN por análisis. A destacar que estos estudios se centraron principalmente en genes con niveles relativamente altos de expresión en la placenta, en comparación con las células de sangre materna, sobre la premisa de que los genes relacionados con el embarazo proceden en gran parte de la placenta. Dichos métodos han identificado dianas de ARN relacionadas con el embarazo que se expresan mucho en la placenta, pero podría haber perdido otras dianas importantes. Además, dadas las bajas concentraciones y la

escasa integridad de ARN del plasma[9,15], métodos de alto rendimiento convencionales, tales como el análisis en serie de la expresión génica y el análisis de chips no serían robusto para el examen directo del transcriptoma del plasma.

Las limitaciones técnicas mencionadas de análisis de ARN del plasma posiblemente podrían resolverse empleando secuenciación masiva en paralelo (SMP) para el análisis de ARN, es decir, secuenciación de ARN (RNA-seq)[17,18]. Dada la sensibilidad mejorada y amplio intervalo dinámico, se ha empleado RNA-seq para examinar la expresión de genes en muchos tejidos humanos, incluida la placenta[19]. La superioridad de la SMP se ha demostrado además por su viabilidad en la caracterización directa de ARNmi en el plasma en individuos sanos[20] y en mujeres embarazadas[21,22]. Sin embargo, el espectro completo de transcriptoma de plasma sigue siendo difícil de alcanzar, lo que podría deberse a la menor estabilidad de las especies de ARN largos en comparación con los ARNmi cortos en el plasma[23].

En este estudio, se ha demostrado que los transcritos fetales y maternos pueden ser detectados en el plasma materno y que sus contribuciones relativas pueden estimarse usando RNA-seq, mediante el examen de polimorfismos de un solo nucleótido (SNP) específicos fetal y materno. Además, los patrones de expresión específicos de alelos de la placenta pueden controlarse en el plasma materno. También se ha demostrado que los transcritos relacionados con el embarazo podrían ser identificados por examen directo del plasma materno antes y después del parto.

II. Método de diagnóstico de un trastorno relacionado con el embarazo utilizando relaciones alélicas de SNP

El análisis de las características de la expresión génica es útil para la detección del estado de enfermedad de un individuo. Las realizaciones de la presente invención incluyen métodos para analizar las características de la expresión de un individuo mediante el análisis de una mezcla de moléculas de ARN de dos individuos diferentes. Los métodos hacen uso de las abundancias relativas de alelos que son específicos de un individuo y los alelos compartidos entre los dos individuos. En base a las abundancias relativas de los alelos compartidos e individuales específicos, pueden determinarse las características de expresión génica de los dos individuos. Una posible aplicación de los presentes métodos es el análisis de las características de la expresión génica de un feto mediante el análisis del ARN en una muestra de plasma materno que contiene ARN tanto del feto como de la mujer embarazada.

En una mezcla que contiene ARN de dos individuos, las características de la expresión de cada individuo no pueden determinarse en base al análisis de la cantidad total de los diferentes transcritos de ARN en la mezcla porque es difícil determinar las contribuciones relativas de cada individuo en todo el ARN. La abundancia relativa de los diferentes transcritos de ARN puede ser útil para el seguimiento del bienestar de un individuo o para detectar una enfermedad.

En este método, el genotipo de los dos individuos puede determinarse en primer lugar ya sea por genotipado directo de los individuos o por análisis familiar. Por ejemplo, si los padres tienen el genotipo de AA y TT, entonces el genotipo del feto sería AT.

El siguiente ejemplo hipotético ilustra el principio de este método. Para cada uno de los genes de interés, hay un SNP en la región de codificación de modo que el polimorfismo se puede observar en las transcripciones de ARN de los diferentes genes. Suponer que los genotipos del feto y de la mujer embarazada son AB y AA, respectivamente. Por lo tanto, el alelo B sería específico para el feto y el alelo A sería compartido entre el feto y la madre. Los niveles relativos de expresión de los diferentes genes en la placenta y las células de sangre materna son como se muestran en la Tabla 1.

En este ejemplo, se supone que la placenta y las células sanguíneas maternas contribuirían con una proporción relativa igual del 2% de cada uno de sus transcritos de ARN en el plasma materno. En otras palabras, si el nivel de expresión en la placenta de un gen es 10.000, 200 transcripciones de ARN de ese gen contribuirían en el plasma materno. Como se muestra en la Fig. 1, la relación alélica B/A muestra una buena relación positiva con los niveles relativos de expresión de los respectivos genes en la placenta y en las células sanguíneas maternas. Por el contrario, el nivel de expresión génica total se ve afectado por la fluctuación de la expresión de las células de la sangre materna y por lo tanto, se correlacionan peor con la expresión de la placenta (Fig. 2).

Tabla 1

Gen	Genotipo materno	Genotipo del feto	Nivel de expresión relativo		Expresión relativa (placenta /células sanguíneas maternas	Plasma materno				Relación B/A
			en células sanguíneas maternas	en la placenta		Abundancia relativa de ARN materno a fetal	Conc. de ARN total	Conc. de ARN		
								con el alelo A	con el alelo B	
1	AA	AB	8.000	20.000	2,5	1:1	560	360	200	0,56
2	AA	AB	6.000	10.000	1,67	1:1	320	220	100	0,45
3	AA	AB	10.000	5.000	0,5	1:1	300	250	50	0,2
4	AA	AB	14.000	1.000	0,071	1:1	300	290	10	0,03

Otra ventaja de utilizar el análisis de la relación alélica es que el nivel de expresión de un gen concreto en la placenta está normalizado al nivel de expresión de las células de la sangre materna. Ya que el nivel de expresión puede variar en gran medida a través de diferentes genes y diferentes muestras, esta normalización haría más robusta la comparación con diferentes genes y evitaría la necesidad de comparar con un gen de referencia, por ejemplo, un gen constitutivo. En otras palabras, en el análisis de la expresión génica convencional, el nivel de expresión de un gen en una muestra se mide con referencia a un gen constitutivo o como una cantidad para todo el ARN de la muestra. Para identificar la expresión génica aberrante, convencionalmente, se podría comparar entonces esos valores relativos de la muestra de ensayo con las muestras de referencia.

En la presente memoria se propone un nuevo método en el que se utilizan las cantidades relativas aportadas por el feto normalizadas a la contribución de la madre del mismo gen como medio para determinar las características de la expresión génica. En los estados patológicos, por ejemplo preeclampsia, parto prematuro, enfermedades maternas tales como lupus eritematoso diseminado, mediante los cuales se altera la expresión del gen en la placenta o los órganos maternos, la relación fetal a materna de ese gen se alteraría en comparación con embarazos sin dichas enfermedades. Este método puede ponerse en práctica también utilizando el alelo SNP específico del feto en relación con el alelo compartido entre los transcritos de ARN en el plasma materno.

El método también se puede poner en práctica utilizando el alelo del SNP específico para la madre con respecto al alelo compartido entre los transcritos de ARN en el plasma materno. Pueden identificarse enfermedades cuando uno o más de dichas relaciones alélicas para uno o más transcritos de genes de ARN están alterados en comparación con lo que cabe esperar para un estado no patológico. El patrón de dichas relaciones alélicas a través de un locus del gen o múltiples locus de genes podría utilizarse para identificar enfermedades. Los estados no patológicos podrían representarse mediante relaciones alélicas en embarazos normales, en muestras obtenidas antes o después de la prueba cuando el embarazo ya no está afectado por el estado patológico o a partir de datos existentes previamente obtenidos a partir de embarazos normales, es decir, un intervalo de referencia obtenido previamente.

Los trastornos relacionados con el embarazo que pueden diagnosticarse empleando los presentes métodos incluyen cualquiera de los trastornos caracterizados por niveles de expresión relativos anormales de genes en tejido materno y fetal. Estos trastornos incluyen, pero no se limitan a, preeclampsia, restricción del crecimiento intrauterino, placentación invasiva, nacimiento prematuro, la enfermedad hemolítica del recién nacido, insuficiencia placentaria, hidropesía fetal, malformación fetal, síndrome de HELLP, lupus eritematoso diseminado y otras enfermedades inmunológicas de la madre. Los métodos distinguen entre moléculas de ARN aportadas por la madre y el feto en una muestra que contiene una mezcla de dichas moléculas. Por lo tanto, los métodos pueden identificar cambios en la contribución de un individuo (es decir, la madre o el feto) a la mezcla en un locus determinado o para un gen determinado, incluso si la contribución del otro individuo no cambia o se mueve en la dirección opuesta. Dichos cambios no se pueden detectar fácilmente cuando se mide el nivel de expresión global del gen sin tener en cuenta el tejido o el individuo de origen. Los trastornos relacionados con el embarazo, como se expone en la presente memoria, no se caracterizan por anomalías cromosómicas en el feto, por ejemplo aneuploidía.

Los presentes métodos también se pueden realizar sin información previa de genotipado de la placenta. Por ejemplo, el genotipo de la mujer embarazada puede determinarse por genotipo directo de sus células sanguíneas. A continuación, los transcritos de ARN de las muestras de plasma pueden analizarse, por ejemplo, pero no limitarse a la secuenciación masiva en paralelo. Pueden identificarse transcritos de ARN que muestran dos alelos diferentes. En este conjunto de transcritos, si la madre es homocigótica, indicaría que el feto es heterocigótico para el alelo específico del feto y el alelo materno. En este escenario, el análisis de la relación alélica de ARN se puede realizar sin información previa del genotipo del feto (o la placenta). Además, los casos en que la madre es heterocigótica pero el feto es homocigótico pueden ser identificados por una desviación de una relación 1:1. En la patente de EE.UU. nº 8.467.976 se describen más detalles de dichas técnicas.

A. Método de diagnóstico

Un método 300 para el diagnóstico de un trastorno relacionado con el embarazo según algunas realizaciones se muestra en la Fig. 3. El método utiliza una muestra de una hembra embarazada con un feto. La muestra puede ser de plasma sanguíneo materno y contiene una mezcla de moléculas de ARN materno y fetal. La muestra puede obtenerse según se desee, de una hembra en cualquier fase del embarazo. Por ejemplo, mujeres embarazadas en el primer, segundo y tercer trimestre con embarazos de un solo feto pueden servir como pacientes. Las mujeres embarazadas del primer y segundo trimestres se pueden clasificar como "casos en la fase inicial del embarazo " y las mujeres embarazadas del tercer trimestre, se pueden clasificar como "casos en la fase final del embarazo". En algunas realizaciones, las muestras también se pueden recoger después del parto del feto, o de la misma persona antes y después del parto, y las muestras de las hembras no embarazadas pueden servir como referencias. Las muestras se pueden obtener en cualquier momento después del parto (por ejemplo, a las 24 horas).

La muestra puede obtenerse de la sangre periférica. Una muestra de sangre materna puede tratarse como se desee, por ejemplo, por centrifugación para separar las células sanguíneas del plasma. Pueden añadirse estabilizadores a la muestra de sangre o partes de la misma, y la muestra se puede almacenar antes de su uso. En algunas realizaciones, también se obtienen muestras de tejido fetal, por ejemplo vellosidades coriónicas, líquido

amniótico, o tejido placentario. El tejido fetal se puede utilizar para determinar genotipos fetales, como se expone más adelante. El tejido fetal se puede obtener antes o después del parto.

Una vez se obtiene una muestra, el ARN o el ADN puede entonces extraerse de las células de la sangre materna y el plasma, así como de cualquier tejido fetal tal como una muestra de la placenta. Pueden pretratarse muestras de ARN placentario y de células sanguíneas, tal como con el Ribo-Zero Gold Kit (Epicenter), para eliminar el ARN ribosómico (ARNr) antes de la preparación de la biblioteca de secuenciación.

En la etapa 301 del método, se recibe un gran número de lecturas. Las lecturas se obtienen de un análisis de moléculas de ARN obtenidas de la muestra. En diversas realizaciones, las lecturas pueden obtenerse por secuenciación, PCR digital, RT-PCR y espectrometría de masas. Aunque la exposición se centra en la secuenciación, los aspectos de la descripción son también aplicables a otras técnicas de obtención de las lecturas. Por ejemplo, la PCR digital (incluidas PCR microfluida y en gotas) puede proporcionar conocimiento de diferentes alelos en un locus determinado utilizando sondas (incluidos los cebadores) dirigidos a cada alelo. Las sondas pueden llevar marcadores para que sean distinguibles entre sí, por ejemplo presentando diferentes colores. En el mismo experimento o en un experimento diferente (por ejemplo, en un portaobjetos o chip separado), sondas con diferentes marcadores pueden dirigirse a diferentes locus. La detección de los marcadores refleja la presencia y/o cantidad de moléculas de ARN o de ADNc que están amplificándose y corresponde con las lecturas de las moléculas de ARN o ADNc, donde tales lecturas proporcionan información sobre una secuencia de las moléculas de ARN o ADNc en los locus correspondientes a las sondas. La descripción en relación con las "lecturas de secuencias" se aplica igualmente a lecturas obtenidas mediante cualquier técnica adecuada, incluidas las técnicas sin secuenciación.

La secuenciación se puede realizar como se desee, utilizando cualquier tecnología disponible. Ejemplos de tecnologías y métodos de secuenciación de ácidos nucleicos incluyen la secuenciación masiva en paralelo, la secuenciación de la siguiente generación, la secuenciación de todo el genoma, la secuenciación del exoma, secuenciación con o sin enriquecimiento de objetivo, la secuenciación por síntesis (p. ej., secuenciación por amplificación, amplificación clonal, amplificación puente, secuenciación con terminadores reversibles), la secuenciación por ligadura, la secuenciación por hibridación (p. ej., la secuenciación de chips), secuenciación de una sola molécula, de secuenciación en tiempo real, la secuenciación por nanoporos, pirosecuenciación, la secuenciación por semiconductores, la secuenciación por espectroscopia de masas, la secuenciación por perdigonada y la secuenciación de Sanger. En algunas realizaciones, la secuenciación se realiza utilizando técnicas masivas en paralelo sobre una biblioteca de ADNc preparada a partir de ARN en la muestra. Las bibliotecas de ADNc se pueden sintetizar como se desee, por ejemplo usando el mRNA-Seq Sample Preparation Kit (Illumina) siguiendo las instrucciones del fabricante o con ligeras modificaciones. En algunas realizaciones, la ADN de Klenow polimerasa diluida 5 veces se utiliza para la etapa de reparación en el extremo de ADNc del plasma. El equipo de purificación por PCR QIAquick y el equipo QIAquick MinElute (Qiagen) se pueden utilizar para purificar productos reparados en el extremo y adenilados, respectivamente. En algunas realizaciones, se utilizan adaptadores emparejado en el extremo diluidos 10 veces para la muestra de ADNc del plasma, o se llevan a cabo dos rondas de purificación para los productos ligados al adaptador utilizando las perlas AMPure XP (Ageticourt). Una biblioteca de ADNc puede secuenciarse para 75 pares de bases en un formato de extremos emparejados en un instrumento HiSeq 2000 (Illumina), o usar otros formatos o instrumentos.

En la etapa 302, se determinan ubicaciones de las lecturas en una secuencia de referencia. Por técnicas basadas en PCR, puede determinarse una ubicación, por ejemplo, haciendo coincidir el color de un marcador detectado en una sonda o cebador para la secuencia para la que la sonda o cebador es específico. Por técnicas de secuenciación, la ubicación puede determinarse alineando la secuencia de lectura con la secuencia de referencia. La alineación de secuencias puede realizarse mediante un sistema informático. Puede utilizarse cualquier tubería bioinformática para el preprocesamiento de datos en bruto (tal como la eliminación de lecturas muy duplicadas y lecturas de baja calidad), alineación de datos y/o normalización de datos. Los niveles de transcripción de ARN en las células de sangre materna, tejidos de la placenta y plasma materno pueden calcularse como fragmentos por kilobase por millonésima de lecturas exónicas (FPKM). Para la determinación de relaciones alélicas, el preprocesamiento de datos y la alineación se pueden realizar, pero no se requiere normalización de datos. Para la alineación puede utilizarse cualquier secuencia de referencia (por ejemplo, el genoma humano de referencia hg19).

En un ejemplo de realización de las etapas 301 y 302, después de la eliminación de lecturas duplicadas y lecturas de ARNr, se obtuvieron un promedio de 3 millones de lecturas analizables para cada muestra de plasma de mujer no embarazada; un promedio de 12 millones de lecturas analizables se obtuvieron para cada muestra de plasma de mujeres embarazadas. Para RNA-seq en tejidos, se obtuvieron un promedio de 173 millones y 41 millones de lecturas analizables por muestra para la placenta y las células sanguíneas, respectivamente. Las estadísticas de alineación de RNA-seq se resumen en las Fig. 4 y 5. El contenido por GC de lecturas secuenciadas también se examinó en todas las muestras (Fig. 6A y tabla 2).

Tabla 2. Proporción de lecturas secuenciadas ricas en GC en bibliotecas de RNA-seq

Tipo de muestra	Muestra	Tratamiento	Lecturas ricas en GC (% de lecturas en bruto)
Plasma	M9356P	Disminución de ARNr	49,28%
	M9415P	Ninguno	0,00%
Placenta	N9356	Disminución de ARNr	0,15%
Células sanguíneas	M9356W	Disminución de ARNr	0,30%
a. Las lecturas en bruto se alinearon frente a la secuencia rica en GC montada de nuevo por BLAST (NCBI).			

En la etapa 303, se identifican uno o locus más informativos. Esta etapa puede determinar o hacer deducciones acerca de los genotipos de la hembra y el feto en el locus dentro del genoma, y comparar estos genotipos. Un locus se considera informativo si es homocigótico en una primera entidad para un primer alelo correspondiente (p. ej., AA), y es heterocigótico en una segunda entidad para el primer alelo correspondiente y un segundo alelo correspondiente (p. ej., AB). La primera entidad puede ser la hembra embarazada o el feto, y la segunda entidad es el otro de entre el la hembra embarazada y el feto. En otras palabras, un individuo (ya sea la madre o el feto) es homocigótico, y el otro individuo es heterocigótico, para cada locus informativo.

Los locus informativos pueden clasificarse además según qué individuo es heterocigótico y el único contribuyente de un alelo. Un locus se considera un locus materno informativa, o análogamente un locus materno específico, si el feto es homocigótico y la hembra embarazada es heterocigótica. Un locus se considera un locus fetal informativo, o análogamente un locus fetal específico, si la hembra embarazada es homocigótica y el feto es heterocigótico. En la etapa 303, todos los locus informativos identificados son materno específicos o fetal específicos, porque el mismo individuo sirve como segunda entidad para todos estos locus.

Un locus informativo puede representar un polimorfismo de un solo nucleótido o "SNP", en donde el primer alelo y el segundo alelo difieren en la identidad de un solo nucleótido. Un locus informativo también puede representar una corta inserción o eliminación, en donde uno o más nucleótidos se insertan o eliminan en un alelo en comparación con el otro alelo.

En algunas realizaciones, el genotipo de la hembra embarazada en uno o más locus informativos, o en cada locus informativo, se determina por secuenciación del ADN genómico obtenido del tejido materno. El tejido materno puede ser células de sangre materna o cualquier otro tipo de tejido. En algunas realizaciones, el genotipo del feto en uno o más locus informativos, o en cada locus informativo, se determina por secuenciación del ADN genómico obtenido del tejido fetal, tales como la placenta, vellosidades coriónicas, líquido amniótico. Si se prefiere un método menos invasivo, el ADN fetal puede obtenerse también para la secuenciación de una muestra de sangre materna. La misma muestra usada para obtener las moléculas de ARN para la secuenciación puede ser la fuente de dicho ADN fetal, o puede utilizarse una muestra diferente. Se valorará que los locus fetales informativos puedan ser identificados sin genotipar directamente el feto. Por ejemplo, si se determina que la hembra embarazada es homocigótica para un primer alelo en un locus, y la secuenciación de ARN indicar la presencia de un segundo alelo en la mezcla de RNA materno y fetal, entonces el feto se puede suponer que es heterocigótico en este locus.

En algunas realizaciones, tanto la madre como el feto se genotipan utilizando métodos masivamente paralelos para identificar locus informativos. La secuenciación se puede realizar en células de sangre materna enriquecidas con exoma y muestras de ADN genómico de placenta. El ADN genómico se puede extraer de los tejidos de la placenta y de células sanguíneas maternas utilizando el QIAamp DNA Kit y el QIAamp Blood Kit (ambos de Qiagen), respectivamente, siguiendo las instrucciones del fabricante. El enriquecimiento con exoma y la preparación de la biblioteca de secuenciación se pueden realizar utilizando el TruSeq Exoma Enrichment Kit (Illumina) siguiendo el protocolo del fabricante. Pueden secuenciarse las bibliotecas de 75 pb en un formato PE en un instrumento HiSeq 2000 (Illumina).

En la etapa 304, se filtran uno o más locus informativos. Como se emplea en la presente memoria, "filtrar" significa seleccionar determinados locus informativos, fuera de los identificados en la etapa 303, para su posterior análisis. Un locus informativo "filtrado" es un locus seleccionado de este modo. La selección puede hacerse sobre la base de uno o más criterios. Uno de dichos criterios es cuando los locus informativos se encuentran en el genoma o en una secuencia de referencia. En algunas realizaciones, solamente los locus situados dentro de un exón o la región expresada de la secuencia de referencia se analizan además.

Otro criterio para el filtrado es el número de lecturas de secuencias, fuera de todas las recibidas en la etapa 301 y alineadas con la secuencia de referencia en la etapa 302, que se alinean con el locus y contienen cada alelo. En algunas realizaciones, un locus informativo filtrado debe tener asociado a él al menos un primer número

predeterminado de lecturas de secuencias que contiene el primer alelo, y/o un segundo número predeterminado de lecturas de secuencias que contienen el segundo alelo. Los números predeterminados pueden ser 1, 2 o más, y pueden corresponder a una calidad de lectura o profundidad de secuenciación deseadas. Como resultado de la filtración, se identifican uno o más locus informativos filtrados.

En algunas realizaciones, solamente los locus informativos que representan SNP sone consideran para el filtrado. En un ejemplo del método, se examinaron aproximadamente un millón de SNP en la base de datos NCBI dbSNP Build 135, todos ellos situados en los exones, y se identificaron los SNP informativos de este millón utilizando genotipos maternos y fetales. Para ilustración de SNP informativos, Se asignó "A" como alelo compartido y "B" como alelo materno- o fetal-específico. Los SNP informativos incluían aquellos con un alelo de SNP específico de la madre, es decir, "AA" en el feto y "AB" en la madre en un locus dado, y aquellos con un alelo de SNP fetal-específico, es decir, "AA" en la madre y " AB" en el feto, en un locus dado. Los genotipos se designaron usando una tubería de bioinformática interna. Para realizar el filtrado, se incluyeron para análisis solamente los SNP informativos en los que tanto el alelo "A" como el alelo "B" presentaban al menos un recuento de lecturas, respectivamente.

En las etapas 305 y 306, las lecturas de secuencias que alinean a cada locus informativo filtrado se cuentan y clasifican dependiendo de qué alelo contienen. Para cada locus informativo filtrado, se determinan dos números: un primer número, que es el número de lecturas de secuencias que se alinean al locus y que contienen el correspondiente primer alelo (es decir, el alelo compartido); y un segundo número, que es el número de lecturas de secuencias que se alinean al locus y que contienen el segundo alelo correspondiente (es decir, el alelo materno- o fetal-específico). El primer número y el segundo número pueden calcularse como se desee. El primer número más el segundo número para cada locus proporciona un número total de lecturas de secuencias que se alinean al locus. Una relación del primer número y el segundo número para un determinado locus informativo filtrado puede denominarse una "relación alélica".

En las etapas 307 y 308, los números de los primeros alelos y los segundos alelos se suman en todos los locus informativos filtrados. En la etapa 307, se calcula una primera suma de los primeros números. La primera suma representa el número total de lecturas de secuencias que contiene los primeros alelos correspondientes (es decir, los alelos compartidos) para los locus informativos filtrados. En la etapa 308, se calcula una segunda suma de los segundos números. La segunda suma representa el número total de lecturas de secuencias que contiene los segundos alelos correspondientes (es decir, los alelos materno- o fetal-específicos) para los locus informativos filtrados. Las sumas se pueden calcular como se desee, utilizando todos los locus informativos filtrados identificados en la etapa 304 o uno de sus subconjuntos. En algunas realizaciones se ponderan las sumas. Por ejemplo, las contribuciones a las sumas de locus de un determinado cromosoma se pueden escalar hacia arriba o abajo multiplicando los primeros números y segundos números para esos locus por un escalar.

En la etapa 309, se calcula una relación de la primera suma y la segunda suma. Esta relación representa el número relativo de lecturas de secuencias para los primeros alelos y segundo alelos, agregados en todos los locus informativos filtrados. En algunas realizaciones, la relación se calcula simplemente como la primera suma dividida por la segunda suma. Dicho cálculo da el número de lecturas de secuencias para alelos compartidos como un múltiplo de las lecturas de secuencias para alelos materno- o fetal-específicos. Alternativamente, la relación se puede calcular como la segunda suma dividida por la suma de la primera suma y la segunda suma. Aquí la relación proporciona el número de lecturas de secuencias para los alelos específicos para la madre o el feto como una fracción de todas las lecturas de secuencias. Otros métodos de cálculo de la relación serán evidentes para el experto en la materia. La primera suma y la segunda suma, a partir de las cuales se calcula la relación, pueden servir como representantes de las cantidades de ARN (es decir, transcritos) presentes en la muestra que se originan a partir de los alelos compartidos y específicos, para los locus informativos filtrados.

En la etapa 310, la relación calculada en la etapa 309 se compara con un valor discriminatorio para determinar si el feto, la madre o el embarazo tienen un trastorno relacionado con el embarazo. En diversas realizaciones, un valor discriminatorio puede determinarse a partir de una o más muestras obtenidas de hembras embarazadas sin trastorno relacionado con el embarazo (es decir, pacientes de referencia) y/o determinarse a partir de una o más muestras obtenidas de hembras embarazadas con el trastorno relacionado con el embarazo. En algunas realizaciones, el valor discriminatorio se determina realizando el mismo método que se ha descrito anteriormente en muestras de pacientes de referencia, y examinando el mismo o un conjunto de la superposición de locus informativos filtrados. Un valor discriminatorio puede ajustarse a un valor entre los valores esperados para embarazos sin un trastorno y los valores esperados para embarazos con un trastorno. El valor discriminatorio puede basarse en una diferencia estadística a partir de un valor normal.

En algunas realizaciones, la relación calculada para la hembra embarazada (como en la etapa 309) usa locus informativos filtrados en un determinado conjunto de genes, y una relación (es decir, el valor discriminatorio) se calcula para los pacientes de referencia utilizando locus informativos filtrados en algunos o todos los genes iguales. Si la relación calculada para la hembra embarazada se basa en alelos específicos de la madre, entonces el valor discriminatorio calculado para las pacientes de referencia puede también basarse en alelos específicos de la madre. Del mismo modo, la relación y el valor discriminatorio pueden ambos estar basados en alelos fetal-específicos.

La comparación entre la relación y el valor discriminatorio puede realizarse como se desee. Por ejemplo, un

trastorno puede diagnosticarse si la relación excede el valor discriminatorio o cae por debajo del valor discriminatorio, en cualquier cantidad o en un determinado margen. La comparación puede implicar el cálculo de una diferencia o el cálculo de otra relación entre la relación para la hembra embarazada y el valor discriminatorio. En algunas realizaciones, la comparación implica evaluar si los niveles relativos de expresión de alelos compartidos y no compartidos para determinados genes difieren significativamente entre la hembra embarazada y las pacientes de referencia.

Si se desea, el método también se puede usar para estimar una parte de ARN en la muestra que es de origen materno o fetal. La estimación se hace multiplicando la relación calculada en la etapa 309 por un escalar. El escalar representa un nivel de expresión total en los locus informativos filtrados con respecto a la expresión de los segundos alelos en el individuo heterocigótico.

El ejemplo siguiente ilustra la aplicación del escalar. Si los segundos alelos contados en el método son fetal-específicos, entonces la relación calculada en la etapa 309 puede proporcionar el número de lecturas de secuencias para alelos fetal-específicos como una fracción de todas las lecturas de secuencias para los locus informativos filtrados. De las lecturas de secuencias que contienen el alelo compartido, algunos están aportados por la madre y algunos están aportados por el feto. Si se conocen los niveles relativos de expresión de los alelos específicos para el feto y compartidos en el feto, o pueden estimarse, entonces la relación se puede escalar para obtener una estimación de la contribución fetal fraccionada para todas las lecturas de secuencias para los locus informativos filtrados.

En algunas realizaciones, los alelos fetales y compartidos se expresan de forma simétrica en la mayoría de los locus y el escalar se supone que es aproximadamente dos. En otras realizaciones, el escalar se aparta de dos y tiene en cuenta la expresión génica asimétrica[24]. La contribución materna fraccionada a las lecturas de secuencias es entonces uno menos la contribución fetal fraccionada. Cálculos similares pueden hacerse cuando los segundos alelos en los locus informativos filtrados son específicos de la madre.

B. Ejemplos:

1. Correlación de relaciones alélicas y concentraciones plasmáticas

Se analizaron las características de la expresión de una muestra de plasma materno (edad de gestación 37 2/7) y se compararon con la placenta correspondiente y las células de sangre materna para una mujer embarazada. Se realizó la secuenciación masiva en paralelo de ARN para cada una de estas muestras utilizando el instrumento Illumina HiSeq 2000. La placenta y las células de sangre materna se genotiparon mediante secuenciación del exoma por secuenciación masiva en paralelo. Se examinaron aproximadamente un millón de SNP en la base de datos NCBI dbSNP Build 135 que se encuentran en los exones, y se clasificaron los SNP informativos cuando la madre es homocigótica (genotipo AA) y el feto es heterocigótico (genotipo AB). Por lo tanto, el alelo A sería el alelo compartido entre la madre y el feto y el alelo B es específico del feto.

Las relaciones alélicas ARN-SNP y las concentraciones plasmáticas totales de diferentes transcritos de ARN en el plasma materno se representan frente a los niveles relativos de expresión hística (placenta/células sanguíneas) (Fig. 7). Se puede observar que la relación alélica ARN-SNP en el plasma se correlaciona bien con el nivel relativo de expresión hística (R de Spearman = 0,9731679, P = 1,126e-09). Sin embargo, el nivel plasmático total no se correlacionó con la expresión tisular (R de Spearman = -0,7285714, P = 0,002927).

Además de los análisis de la expresión fetal, este método también puede utilizarse para la caracterización de la expresión materna. En este escenario, se pueden utilizar los SNP informativos que la madre es heterocigótica (genotipo AB) y el feto es homocigótico (genotipo AA). La relación alélica ARN-SNP se puede calcular entonces como el recuento alélico específico de la madre dividido por el recuento de los alelos compartidos. En la Fig. 8, las relaciones alélicas ARN-SNP y los niveles totales en el plasma de diferentes genes que contienen SNP informativos se representan frente al nivel de expresión hística relativo (células sanguíneas/placenta). Se puede observar una buena relación positiva entre el nivel de expresión hística con la relación alélica ARN-SNP en el plasma materno (R de Spearman = 0,9386, P <2,2e-16), pero no los niveles totales de transcripción en el plasma (R de Spearman = 0,0574, P = 0,7431).

Los transcritos de ARN, sus relaciones alélicas ARN-SNP y sus concentraciones en el plasma se tabulan en la fig. 9 y la fig. 10.

2. Comparación de la relación alélica-SNP ARN del plasma materno para los casos de preeclampsia y embarazo de referencia

Para los casos de dos hembras embarazadas, se obtuvieron un número medio de 117.901.334 fragmentos en bruto (TABLA 3A). La relación alélica ARN-SNP de un subconjunto de transcritos circulantes que llevan alelos SNP específicos de la madre se comparó entre un caso de preeclampsia de aparición tardía en el tercer trimestre (PET), 5641, y su caso de referencia de igual edad de gestación, 7171. Como se muestra en la Fig. 11A, las relaciones alélicas ARN-SNP de estos transcritos representaron unas características diferentes en PET y los casos de referencia. Es importante destacar que, las características de la relación alélica ARN-SNP son distintas de las

características del nivel de transcripción de plasma, como se muestra en la Fig. 11B. Esta sugiere que el análisis de la relación alélica ARN-SNP en el plasma de materna podría ser utilizado como otro indicador para distinguir mejor los casos PET de un casos de embarazo normal. El número de transcritos informativos podría aumentarse más con un mayor número de SNP informativos, ejemplificados por la comparación entre el caso de PET, 5641, y otro caso de referencia de embarazo normal, 9356 (Figs. 12A y 12B). Las relaciones alélicas ARN-SNP y el nivel en el plasma de los transcritos de ARN se muestran en la Fig. 11 y la Fig. 12 se tabulan en la Fig. 13 y Fig. 14, respectivamente.

Tabla 3A: Compendio de resultados de alineación de lecturas de ARN-seq

Tipo de muestra	Muestra	Lecturas en bruto	Lecturas preprocesadas[a]	(%)[b]	Lecturas cartografiables[c]	(%)[d]
Plasma materno	5641 (Preeclampsia)	119.902.178	30.106.800	25,11%	4.001.376	13,29%
Plasma materno	7171 (Referencia)	115.900.490	29.160.920	25,16%	1.397.345	4,79%

a. Lecturas conservadas tras la eliminación de lecturas muy repetitivas.

b. % de lecturas en bruto.

c. Para análisis detallado véanse Datos complementarios en tabla 2B.

d. % de las lecturas preprocesadas.

Tabla 3B: Compendio de resultados de alineación de lecturas de ARN-seq

Tipo de muestra	Muestra	Lecturas filtradas[a]	Lecturas analizables[b]	Exón (%)	Intrón (%)	Región intergénica (%)[c]
Plasma materno	5641 (Preeclampsia)	2.579.163	1.422.213	44,87%	12,32%	42,80%
Plasma materno	7171 (Referencia)	790.735	606.610	37,69%	16,09	46,22%

a. La mayoría de las lecturas filtradas son los ARNr y los ARNt nucleares y mitocondriales.

b. Lecturas analizables = lecturas cartografiables totales – lecturas filtradas.

c. % de lecturas que se alinean a regiones fuera de exones e intrones de los genes de referencia.

III. Método de determinación de la contribución del ARN fetal o materno

A. Método

En la Fig. 15 se muestra un método 1500 para determinar, en una muestra de una hembra embarazada con un feto, una parte de ARN que es de origen fetal. La muestra se puede preparar y obtener como se expuso anteriormente para el método 300.

En la etapa 1501, se reciben un gran número de lecturas de secuencias. En la etapa 1502, las lecturas de secuencias, están alineados con una secuencia de referencia. Estas etapas se pueden realizar como se describió anteriormente para las etapas 301 y 302.

En la etapa 1503, se identifican una o más locus informativos maternos. Cada locus informativo materno es homocigótico en el feto para un primer alelo correspondiente y heterocigótico en la hembra embarazada para un primer alelo correspondiente y un segundo alelo correspondiente. El primer alelo es compartido por lo tanto entre la madre y el feto, y el segundo alelo es específico de la madre. Los locus maternos informativos pueden identificarse como se describió anteriormente para la etapa 303.

En la etapa 1504, los locus informativos maternos se filtran, como en la etapa 304, y se identifican uno o más locus informativos maternos filtrados. Las lecturas de secuencias se manipulan entonces en cada uno de los locus informativos maternos filtrados en las etapas 1505-1511. En las etapas 1505 y 1506, se determinan un primer número y un segundo número para cada locus materno informativo filtrado. El primer número se determina en la etapa 1505 como el número de lecturas de secuencias que se alinean al locus y que contienen el correspondiente

primer alelo. El segundo número se determina en la etapa 1506 como el número de lecturas de secuencias que se alinean al locus y que contienen el correspondiente segundo alelo. Las etapas 1505 y 1506 se pueden realizar de manera similar a las etapas 305 y 306, expuestas anteriormente.

En la etapa 1507, para cada locus informativo materno se calcula una suma del primer número y el segundo número. La suma puede ser equivalente al número total de lecturas de secuencias que se alinean al locus. En la etapa 1508, se calcula una relación materna dividiendo el segundo número por la suma. La relación materna proporciona la fracción de todas las lecturas de secuencias que se alinean a cada locus que contiene el segundo alelo (específico de la madre).

En la etapa 1509, se determina un escalar. El escalar representa un nivel de expresión total en el locus informativo materno filtrado con relación a la expresión del segundo alelo correspondiente en la hembra embarazada. En algunas realizaciones, donde la expresión de los dos alelos en la hembra embarazada se sabe o se supone que son simétricos, el escalar es de aproximadamente dos. En otras realizaciones, el escalar puede apartarse de dos y tener en cuenta la expresión génica asimétrica. Como la mayoría de locus expresan de forma simétrica, la suposición de simetría es válida.

En la etapa 1510, la relación materna se multiplica por el escalar para obtener una contribución materna. La contribución materna representa, en el locus informativo materno filtrado, la fracción de lecturas de secuencias (o, por extensión, la fracción de ARN en la muestra) que aporta la madre. Uno menos la contribución materna es la contribución fetal, o la fracción de lecturas de secuencias aportadas por el feto. En la etapa 1511, se calcula la contribución del feto para el locus informativo materno filtrado.

En la etapa 1512, se determina una parte de ARN en la muestra que es de origen fetal. La parte es la media de las contribuciones fetales para los locus informativos maternos filtrados. En algunas realizaciones, este promedio se pondera, por ejemplo, por las sumas calculadas para los locus informativos maternos filtrados. Mediante el cálculo de una media ponderada, la parte determinada en la etapa 1512 puede reflejar los niveles de expresión relativos de diferentes locus o genes en la muestra.

La parte de ARN en la muestra que es de origen fetal puede calcularse para un locus informativo materna filtrado, muchos de dichos locus, o todos los dichos locus para los cuales se obtienen datos de secuenciación. Se valorará que esta parte refleje el transcriptoma circulante en la hembra embarazada concreta en el momento de obtención de la muestra. Si las muestras se obtienen de la persona en diferentes momentos durante el curso de la gestación, el conjunto de locus informativos maternos filtrados identificados en el método 1500 puede variar de una muestra a la siguiente, ya que pueden determinarse las porciones de ARN de origen fetal en estos locus. La parte de ARN en una muestra que es de origen fetal también puede variar de una persona a otra, incluso cuando se controla la etapa gestacional u otros factores, y puede diferir entre una paciente que tiene un trastorno relacionado con el embarazo y una persona sana. Por consiguiente, la parte se puede comparar con un valor discriminatorio para diagnosticar un trastorno de este tipo.

En algunas realizaciones, el valor discriminatorio se determina realizando el método 1500 que utiliza muestras procedentes de una o más personas sanas. En algunas realizaciones, se hace un diagnóstico si la parte excede del valor discriminatorio o cae por debajo del mismo, en cualquier cantidad o por un determinado margen. La comparación puede implicar el cálculo de una diferencia o el cálculo de una relación entre la parte del ARN de origen fetal en la hembra embarazada y el valor discriminatorio. En algunas realizaciones, la comparación implica evaluar si la parte de ARN de origen fetal en la hembra embarazada difiere significativamente de lo que se ve en las mujeres embarazadas sanas en períodos gestacionales similares.

La parte de ARN en la muestra que es de origen fetal también puede determinarse examinando locus informativos fetales. En cada locus informativo fetal, esta parte se puede estimar calculando en primer lugar una relación fetal, que es el número de lecturas de secuencias que contiene el segundo alelo (específico del feto) dividido por el número total de lecturas de secuencias. La contribución fetal es entonces la relación fetal multiplicada por un escalar que representa los niveles de expresión relativos de los dos alelos en el feto. En consecuencia, una variación del método 1500 puede llevarse a cabo identificando locus informativos fetales, filtrando estos locus, calculando una contribución fetal para cada locus filtrado y promediando las contribuciones fetales para los locus informativos filtrados. Si se desea, la parte de origen fetal del ARN determinada por el método 1500 puede aumentarse promediando las contribuciones fetales calculadas para locus informativos maternos filtrados con contribuciones fetales calculadas para locus informativos fetales filtrados. Los promedios calculados en la presente memoria pueden ponderarse, para reflejar la variación en el número de lecturas de secuencias en diferentes locus, para dar más peso a los locus específicos de la madre o del feto, o de otra manera según se desee.

B. Ejemplos

1. Identificación y estimación del transcripto fetal y materno en plasma materno

Genes informativos, definidos como genes con al menos un SNP informativo, se identificaron en primer lugar en base a los datos de genotipado. En los dos casos de embarazo en la fase inicial, un número total de 6.714 y 6.753 genes informativos estaban disponibles para examinar las proporciones relativas de las contribuciones fetales y

maternas, respectivamente (Fig. 16 y 17). En los dos casos de embarazo en la fase final, un número total de 7.788 y 7.761 genes informativos estaban disponibles para examinar las proporciones relativas de las contribuciones fetal y materna, respectivamente. Para medir la proporción relativa de contribución fetal en el plasma materno, se clasificaron transcritos de ARN, donde los alelos específicos del feto fueron cubiertos por al menos una lectura de ARN-seq en las muestras de plasma materno. Las proporciones relativas de dichos transcritos del feto fueron 3,70% y 11,28% durante gestaciones prematuras y tardías, respectivamente. Utilizando un método similar, las proporciones relativas de contribución materna en la circulación, examinadas usando los alelos SNP específicos de la madre, se estimó que eran 76,90% y 78,32% durante las gestaciones prematura y tardía, respectivamente (Fig. 18A).

2. Comparación de la contribución fetal y materna fraccionaria de ARN-SNP para PET y los casos de embarazo de referencia

Se ha examinado la contribución fetal y materna fraccionaria para el transcrito GNAS, que se detectó tanto en el caso de PET, 5641, como en su caso de referencia de igual edad de gestación, 7171. En el transcrito GNAS, había un sitio de SNP informativo que comprende el alelo SNP específico del feto en el locus rs7121 en el caso 5641, en el mismo sitio SNP en el caso 7171, había un alelo SNP específico de la madre. Las contribuciones fetales y maternas fraccionarias para este sitio SNP en el caso 5641 se calculó que eran 0,09 y 0,91, respectivamente. Por otra parte, las contribuciones fetales y maternas fraccionarias para el mismo sitio de SNP en el caso de 7171 fueron 0,08 y 0,92, respectivamente (Fig. 19). En comparación con el caso de referencia 7171, hubo un aumento de 12,5% en la contribución fetal fraccionada y una disminución 1,09% en la contribución maternal fraccionaria en este transcrito. Curiosamente, se detectó un aumento del 21% en FPKM para este transcrito en el caso 5641 en comparación con el 7171.

IV. Método de designar un locus genómico como marcador o fetal

A. Método

Un método 2000 de designación de un locus genómico como marcador materno o fetal se muestra en la Fig. 20. El método implica el análisis de una muestra de una hembra embarazada con un feto. La muestra puede ser de plasma de la sangre materna y contiene una mezcla de moléculas de ARN de la madre y del feto. La muestra se puede obtener y preparar como se expuso anteriormente para el método 300.

En la etapa 2001, se reciben un gran número de lecturas de secuencias. En la etapa 2002, las lecturas de secuencias, están alineadas con una secuencia de referencia. Estas etapas se pueden realizar como se describió anteriormente para las etapas 301 y 302.

En la etapa 2003, se identifican uno o más locus informativos. Cada locus es homocigótico en una primera entidad para un primer alelo correspondiente y heterocigótico en una segunda entidad para el correspondiente primer alelo y un correspondiente segundo alelo. La primera entidad es la hembra embarazada o el feto, y la segunda entidad es el otro de entre la hembra embarazada y el feto. Los locus informativos pueden identificarse como se describió anteriormente para la etapa 303.

En la etapa 2004, los locus informativos se filtran, como en la etapa 304, y se identifican uno o más locus informativos filtrados. En algunas realizaciones, solamente los locus informativos que representan los SNP se consideran para el filtrado. En un ejemplo del método, los locus informativos filtrados incluían SNP informativos con al menos una lectura de secuencia para el alelo "A" y una lectura de secuencia para el alelo "B" en el plasma materno.

Las lecturas de secuencias, se manipulan a continuación en cada uno de los locus informativos filtrados en las etapas 2005-2008, En las etapas 2005 y 2006, para cada locus informativo filtrada se determinan un primer número y un segundo número. El primer número se determina en la etapa 2005 como el número de lecturas de secuencias que se alinean al locus y que contienen el correspondiente primer alelo. El segundo número se determina en la etapa 2006 como el número de lecturas de secuencias que se alinean al locus y que contienen el segundo alelo correspondiente. Las etapas 2005 y 2006 pueden realizarse de manera similar a las etapas 305 y 306, expuestas anteriormente.

En la etapa 2007, se calcula una relación del primer número y el segundo número para cada locus informativo filtrado. Esta relación representa las abundancias relativas de lecturas de secuencias (y, por extensión, los transcritos en la muestra) que contienen el primer alelo y el segundo alelo. En algunas realizaciones, la relación se calcula simplemente como el primer número dividido por el segundo número. Dicho cálculo da el número de lecturas de secuencias para el alelo compartido como un múltiplo de las lecturas de secuencias para el alelo específico materno o fetal. Esta relación, o su recíproca, puede considerarse una relación alélica, y para los locus de SNP, una relación alélica ARN-SNP.

En algunas realizaciones, para los SNP informativos que incluyen alelos de SNP específicos para la madre, la relación alélica ARN-SNP para cada SNP se calcula como alelo específico para la madre:alelo compartido. Por otra parte, para los SNP informativos que incluyen alelos de SNP específicos para el feto, la relación alélica ARN-SNP puede calcularse como alelo específico para el feto: alelo compartido. A diferencia de análisis de expresión génica,

en el que la normalización de datos es un prerrequisito, el análisis de la relación alélica ARN-SNP no requiere normalización de datos y se introduce menos sesgo. Para los transcritos que contienen más de un SNP informativo, se puede calcular una relación alélica ARN-SNP para cada sitio SNP informativo y se puede calcular una relación alélica ARN-SNP media por transcrito.

En la etapa 2007, la relación puede calcularse alternativamente como el segundo número dividido por la suma del primer número y el segundo número. En la presente memoria la relación proporciona el número de lecturas de secuencias para el alelo específico para la madre o el feto como una fracción de todas las lecturas de secuencias en el locus. En términos de alelos "A" y "B", la relación puede expresarse como

Relación de alelo B = recuento de alelos B / (recuento de alelos A + recuento de alelos B)

Teóricamente, para una transcrito plasmático que es aportado únicamente por el feto o por la madre, la relación de alelo B debe ser 0,5, suponiendo que no hay expresión específica del alelo. Otros métodos de cálculo de la relación serán evidentes para el experto en la materia.

En la etapa 2008, el locus informativo filtrado se designa como un marcador cuando la relación excede de un valor discriminatorio. En algunas realizaciones, el valor discriminatorio es de aproximadamente 0,2 a aproximadamente 0,5. En algunas realizaciones, el valor discriminatorio es de 0,4. En un ejemplo del método, el valor discriminatorio de la relación de alelo B se definió como ≥ 0,4 para un transcrito de ARN con una alta contribución fetal o materna. Este valor discriminatorio tomó en consideración la distribución de Poisson y muestreo aleatorio de lecturas de RNA-seq. En algunas realizaciones, se dice que la contribución del alelo "B" es alta cuando la relación excede el valor discriminatorio.

Una alta relación alélica en un locus informativo concreto puede indicar que (i) el segundo o alelo "B" se expresa mucho más en el individuo heterocigótico que el alelo "A" (es decir, el alelo se expresa de forma asimétrica), (ii) una mayor participación de todo el ARN que se alinea al locus en el plasma materno es aportado por el individuo heterocigótico, o ambos (i) y (ii). Una alta relación alélica también puede indicar un trastorno relacionado con el embarazo si el gen relacionado con el locus informativo se está sobreexpresando patológicamente. Por consiguiente, algunas realizaciones del método también incluyen el diagnóstico de un trastorno relacionado con el embarazo en base a si un locus informativo filtrada se designa como marcador para la segunda entidad (es decir, el individuo heterocigótico). Al hacer dicho diagnóstico, el valor discriminatorio utilizado para la comparación con la relación alélica se puede basar en los niveles de transcripción en las muestras de plasma obtenidas de personas embarazadas sanas.

En algunas realizaciones, el método 2000 también se puede usar para estimar una parte de ARN en la muestra, en un locus informativo filtrado concreto, que es de origen materno o fetal. La estimación se realiza multiplicando la proporción calculada en la etapa 2007 por un escalar. El escalar representa un nivel de expresión total en el locus respecto a la expresión del segundo alelo en el individuo heterocigótico.

Para ilustrar, si el segundo alelo ("B") contado en el método es específico para el feto, entonces la relación B-alelo, calculada como anteriormente, representa el número de lecturas de secuencias para este alelo como una fracción de todas las lecturas de secuencias para el locus informativo filtrado. De las lecturas de secuencias que contiene el alelo ("A") compartido, algunas son aportadas por la madre y otras son aportados por el feto. Si se conocen o pueden estimarse los niveles relativos de expresión de los alelos específicos para el feto y compartidos en el feto, entonces la relación B-alelo se puede escalar para obtener una estimación de la contribución fetal fraccionada a todas las lecturas de secuencias para el locus. Si los alelos "A" y "B" se expresan igualmente en el feto, a continuación, el escalar es de aproximadamente dos. La contribución materna fraccionada a las lecturas de secuencias en el locus es entonces uno menos la contribución fetal fraccionada. Cálculos similares pueden realizarse cuando el alelo "B" en el locus informativa filtrada es específico de la madre.

B. Ejemplo: Grandes contribuciones fetal y materna

Utilizando un valor discriminatorio para el alelo B de 0,4 como se ha descrito anteriormente, se encontró que el 0,91% de los transcritos circulantes muestran gran contribución por el feto durante la gestación preliminar (es decir, el primer y segundo trimestres). Este porcentaje aumentó a 2,52% durante la gestación tardía (es decir, el tercer trimestre). Por otra parte, el 42,58% y el 50,98% presentaron una gran contribución de la madre durante las gestaciones preliminar y tardía, respectivamente (Fig. 16, 17 y 18A).

V. Uso de genes para el diagnóstico de trastornos relacionados con el embarazo

A. Genes relacionados con el embarazo

También se proporciona un método de identificación de genes relacionados con el embarazo. El método incluye recibir un gran número de primera lecturas de secuencias y gran número de segundas lecturas de secuencias. Las primeras lecturas de secuencias provienen de moléculas de ARN de secuenciación obtenidas a partir de una muestra de plasma de sangre de una mujer embarazada. Las segundas lecturas de secuencias provienen de la secuenciación de moléculas de ARN obtenidas a partir de una muestra de plasma de sangre de una mujer no

embarazada. Las primeras lecturas de secuencias y las segundas lecturas de secuencias se alinean con una secuencia de referencia, y se designa un conjunto de genes candidatos.

Según el método, las lecturas de secuencias se utilizan entonces para determinar los niveles de expresión para cada gen candidato en las muestras de la mujer embarazada y la mujer no embarazada. Específicamente, para cada gen candidato, se determina un primer número de transcritos correspondientes al gen candidato utilizando las primeras lecturas de secuencias; y se determina un segundo número de transcritos correspondientes al gen candidato usando las segundas lecturas de secuencias. El primer número de transcritos y el segundo número de transcritos pueden normalizarse. A continuación se calcula una relación de transcripción para el gen candidato, donde la relación de transcripción incluye el primer número de transcritos dividido por el segundo número de transcritos. La relación de transcripción se compara entonces con un valor discriminatorio. El gen candidato se identifica como un gen relacionado con el embarazo si la relación de transcripción excede del valor discriminatorio.

En algunas realizaciones del método, la normalización del primer número de transcripciones corresponde a ampliar el primer número de transcripciones por el número total de primera lecturas de secuencias. Del mismo modo, la normalización del segundo número de transcripciones puede corresponder a ampliar el segundo número de transcripciones por el número total de segundas lecturas de secuencias. En otras realizaciones, la normalización del primer número de transcripciones para cada gen candidato corresponde a la ampliación del primer número de transcripciones para el gen candidato por el número total de primeras transcripciones de todos los genes candidatos. La normalización de la segunda serie de transcripciones para cada gen candidato puede corresponder a ampliar el segundo número de transcripciones para el gen candidato por el número total de segundas transcripciones para todos los genes candidatos.

El método identifica generalmente genes como relacionados con el embarazo si se expresan a niveles más altos en una mujer embarazada, en comparación con una mujer no embarazada, siendo igual todo lo demás. Las mujeres embarazadas y no embarazadas pueden ser el mismo individuo; es decir, las muestras obtenidas de la persona antes y después de dar a luz pueden ser las fuentes de las primeras lecturas de secuencias y de las segundas lecturas de secuencias, respectivamente.

Se ha demostrado que un subconjunto de transcritos de ARN circulante que llevan alelos específicos para el feto desapareció completamente del plasma materno después del parto (Fig. 18B). Estos transcritos se consideraron específicos para el feto en el plasma materno. Por otro lado, una parte de los alelos específicos para la madre también eran indetectable después del parto. Por lo tanto, se han explorado genes que presentaban aumento durante el embarazo, denominados genes relacionados con el embarazo, comparando directamente su representación en el plasma materno antes y después del parto. Se definen genes relacionados con el embarazo los que se detectaron en el plasma antes del parto de mujeres embarazadas de tercer trimestre, en las que su nivel en el plasma después del parto se redujo en ≥ 2 veces en ambos casos. Empleando un algoritmo bioinformático para la normalización de datos y análisis de la expresión génica diferencial, se compilaron una lista de 131 genes relacionados con el embarazo (Fig. 21). Entre estos genes, se publicó previamente que 15 eran específicos del embarazo en el plasma materno[1,2,4,5,9,10,12]. Utilizando RT-PCR en tiempo real de una sola etapa, se ha validado aún más la relación con el embarazo de cinco transcritos recién identificados, que eran abundantes en el plasma materno antes del parto, es decir, *STAT1*, *GBP1* y *HSD17B1* en 10 muestras de plasma adicionales de mujeres embarazadas del tercer trimestre, así como KRT18 y *GADD45G* en 10 muestras de plasma de otra cohorte de mujeres embarazadas del tercer trimestre (Fig. 22).

Para evaluar la relación de estos 131 genes con el embarazo, se realizó la agrupación jerárquica para todas las muestras de plasma. Se observó una separación clara entre las muestras de plasma de mujeres embarazadas (es decir, en las fases inicial y final del embarazo) y los que no relacionados con el embarazo en curso (es decir, referencias de no embarazadas y después del parto) (Fig. 23).

Curiosamente, cuando los patrones de expresión de estos 131 genes se compararon entre las muestras de plasma de los dos casos de embarazo en fase final y sus correspondientes células de placenta y sanguíneas maternas, se observó una semejanza más estrecha entre las muestras de placenta y del plasma antes del parto, y entre las células sanguíneas maternas y las muestras de plasma después del parto (Fig. 24A). Esta observación apoya la tesis de que la mayoría de los genes relacionados con el embarazo se expresan preferentemente en la placenta más que en las células de la sangre materna. Además, los niveles de expresión de estos transcritos relacionados con el embarazo en las placentas y plasma materno se correlacionaron positivamente ($P < 0,05$, correlación de Spearman) (Fig. 24B).

B. Expresión génica diferencial en la placenta y la sangre materna

Mientras que hemos sido capaces de catalogar un grupo de genes relacionados con el embarazo mediante el examen directo de las muestras de plasma maternas antes y después del parto, también se han extraído los datos de RNA-seq y células de la placenta y de sangre para comparación. Suponiendo que los genes relacionados con el embarazo debe ser los expresados a un alto nivel en la placenta y a un bajo nivel en las células sanguíneas maternas según lo publicado antes[10,15], se fijó arbitrariamente una diferencia de 20 veces como un valor discriminatorio mínimo para el análisis basado en el tejido. Este análisis basado en el tejido produjo un total de 798

posibles genes candidatos, en el que se calcularon las proporciones de contribuciones fetales y maternas en el plasma materno. Se identificó una proporción relativamente alta de genes con contribución fetal predominante (Fig. 25) en comparación con la del transcriptoma completo (Fig. 18A). Sin embargo, la estrategia basada en plasma superó a la estrategia basada en el tejido en que es capaz de identificar una mayor proporción de genes con la contribución fetal predominante (Fig. 25).

C. Genes relacionados con la enfermedad o el trastorno

También se proporciona en la presente memoria un método de identificación de genes relacionado con un trastorno relacionado con el embarazo. El método incluye recibir un gran número de primeras lecturas de secuencias y gran número de segundas lecturas de secuencias. Las primeras lecturas de secuencias provienen de moléculas de ARN de secuenciación obtenidas de una muestra de plasma de sangre de una mujer embarazada sana. Las segundas lecturas de secuencias provienen de moléculas de ARN de secuenciación obtenidas de una muestra de plasma de sangre de una mujer embarazada que padece un trastorno relacionado con el embarazo, o que lleva un feto que padece un trastorno relacionado con el embarazo. Las primeras lecturas de secuencias y las segundas lecturas de secuencias están alineadas con una secuencia de referencia, y se designa un conjunto de genes candidatos.

Según el método, las lecturas de secuencias se utilizan entonces para determinar los niveles de expresión para cada gen candidato en las muestras para las dos mujeres embarazadas. Específicamente, para cada gen candidato, se determina un primer número de transcritos correspondientes al gen candidato usando las primeras lecturas de secuencias, y se determina un segundo número de transcritos correspondientes al gen candidato usando las segundas lecturas de secuencias. El primer número de transcritos y el segundo número de transcritos pueden normalizarse. A continuación, se calcula una relación de transcripción para el gen candidato, donde la relación de transcripción incluye el primer número de transcripciones dividido por el segundo número de transcripciones. La relación de transcripción se compara entonces con un valor de referencia. El gen candidato se identifica como relacionado con el trastorno si la relación de transcripción se desvía del valor de referencia.

En algunas realizaciones del método, la normalización de la primera serie de transcritos corresponde a la ampliación de la primera serie de transcritos por el número total de primeras lecturas de secuencias. Del mismo modo, la normalización de la segunda serie de transcritos puede corresponder a la ampliación del segundo número de transcritos por el número total de segundas lecturas de secuencias. En otras realizaciones, la normalización de la primera serie de transcritos para cada gen candidato corresponde a la ampliación de la primera serie de transcritos para el gen candidato por el número total de primeros transcritos de todos los genes candidatos. La normalización de la segunda serie de transcritos para cada gen candidato puede corresponder a la ampliación de la segunda serie de transcritos para el gen candidato por el número total de segundos transcritos para todos los genes candidatos.

Según el método de identificación de genes asociados a un trastorno relacionado con el embarazo, en algunas realizaciones, el valor de referencia es 1. En algunas realizaciones, la relación de transcripción se desvía del valor de referencia cuando la proporción de la relación transcripción y el valor de referencia excede o cae por debajo de un valor discriminatorio. En algunas realizaciones, la relación de transcripción se desvía del valor de referencia cuando la diferencia entre la relación de la transcripción y el valor de referencia excede un valor discriminatorio.

El método identifica generalmente un gen asociado a un trastorno relacionado con el embarazo si el nivel de expresión de ese gen difiere significativamente en los embarazos que demuestran y no demuestran el trastorno, siendo todo lo demás igual.

El diagnóstico y seguimiento de los trastornos fetales y patologías del embarazo se ha conseguido previamente utilizando ARN plasmáticos maternos que están relacionados con enfermedades. Por ejemplo, la concentración en el plasma materno de ARNm de la hormona liberadora de corticotropina (CRH) se ha demostrado que es útil para la detección no invasiva y la predicción de preeclampsia[2,3]. La detección de ARNm del receptor tipo 1 de interleucina 1 (IL1RL1) en el plasma materno se ha demostrado que es útil para identificar mujeres con parto prematuro espontáneo[8]. Un grupo de marcadores de ARN de plasma materno relacionados con el crecimiento también se ha investigado para la evaluación no invasiva de restricción del crecimiento fetal y del crecimiento intrauterino[4]. En este estudio, se considera que los nuevos marcadores de ARN circulantes relacionados con la enfermedad podrían ser identificados por los transcriptomas de plasma materno directamente comparados de mujeres embarazadas normales y mujeres con embarazos complicados, tales como la preeclampsia, restricción del crecimiento intrauterino, parto prematuro y aneuploidías fetales. Para demostrar la viabilidad de este método, se realizó RNA-Seq para muestras de plasma maternos obtenidas de tres mujeres embarazadas que desarrollaron preeclampsia y siete mujeres embarazadas sin complicaciones de gestaciones coincidentes. Se identificaron 98 transcritos que presentaban una elevación significativa en el plasma de mujeres embarazadas con preeclampsia (Fig, 26). Los transcritos relacionados con preeclampsia recién identificados son potencialmente útiles para la predicción, el pronóstico y el seguimiento de la enfermedad.

Esta tecnología también puede utilizarse para la predicción y seguimiento de parto prematuro. La tecnología también puede utilizarse para la predicción de muerte fetal inminente. La tecnología también puede usarse para la detección de enfermedades causadas por mutaciones génicas, siempre que el gen en cuestión se transcriba en tejidos fetales o placentarios y los transcritos sean detectables en el plasma materno.

La RNA-Seq del plasma se puede aplicar en otros escenarios clínicos. Por ejemplo, la metodología RNA-Seq del plasma desarrollada en este estudio sería posiblemente útil para el estudio de otros procesos patológicos, como el cáncer, en los que se han descrito[13,14] las concentraciones anormales de ARN plasmático. Por ejemplo, comparando los transcriptomas en el plasma de un paciente de cáncer antes y después del tratamiento, pueden identificarse marcadores de ARN circulante relacionados con tumores para la aplicación de diagnóstico no invasivo.

VI. Patrones de expresión alélica de genes específicos

Se ha empleado secuenciación de ARN para examinar el patrón de expresión alélica[25]. Se supone que el patrón de expresión alélica de un gen dado se conservaría cuando los transcritos de ARN se liberen de los tejidos en la circulación, por lo tanto, podrían detectarse en el plasma. En este estudio, se han examinado los recuentos alélicos de dos transcritos de ARN, a saber, *PAPPA*, gen específico del embarazo, y *H19*, gen impreso expresado por vía materna[26,27].

Para el gen *PAPPA*, se analizó un SNP, rs386088, que contiene un alelo SNP específico del feto (Fig. 27). La ausencia de lecturas de RNA-seq *PAPPA* en las muestras de plasma después del parto indicó que era de hecho específico del embarazo[4]. A destacar que no hubo diferencia estadísticamente significativa en las proporciones de recuentos de lecturas de alelos en el feto entre el plasma materno antes del parto y las muestras de ARN placentario ($P = 0{,}320$, prueba de la χ^2), lo que indica que los datos de plasma materno reflejaban el patrón de expresión bialélica de *PAPPA* en la placenta.

Recientemente se ha publicado de que el patrón de metilación del ADN del gen *H19* impreso expresado por vía materna en la placenta y las células de sangre materna se pudo detectar por secuenciación de ADN con bisulfito del ADN del plasma materno[28]. En la presente memoria, se ha examinado además si el estado de la impronta genómica del gen *H19* se podría explorar en el ARN. Se han centrado en primer lugar en un punto del SNP en el exón 1 del gen *H19*, rs2839698, que contiene un alelo específico materno, es decir, AA en el feto y AG en la madre. Como se muestra en la Fig. 28, sólo se detectó alelo G en el plasma materno en el puerperio (Fig. 28). Dicho patrón monoalélico estaba según su enlace al alelo G no metilado en el punto SNP rs4930098 en la región de control de impronta (Fig. 29). En el plasma materno antes del parto, mientras que el alelo G estaba presente, el alelo A, que fue aportado por la placenta, también fue detectado (Fig. 28). En otros tres puntos de SNP, es decir, rs2839701, rs2839702 y rs3741219, que llevan alelos específicos de la madre, se encontró un patrón alélico similar, es decir bialélico en el plasma materno antes del parto y monoalélico en el plasma materno en el puerperio (Fig. 28). Puede ser que los alelos específicos de la madre en estos puntos de SNP estuvieran en el mismo haplotipo materno, que estaba sin metilar y por lo tanto que estaban transcritos. Concretamente, *H19* ARN no se expresó en las células de la sangre materna (Fig. 28), lo que sugiere que las moléculas de *H19* ARN en el plasma procedían de tejidos/órganos maternos distintos de las células sanguíneas. Los tejidos no placentarios y no fetales que se han descrito que muestran la expresión *H19* incluyen las glándulas suprarrenales, los músculos esqueléticos, el útero, adipocitos, el hígado y el páncreas [29].

VII. Exposición

En este trabajo, se pretendía desarrollar una tecnología para proporcionar una visión global de las actividades transcriptómicas en el plasma materno utilizando ARN-seq. Se ha demostrado previamente que el ADN fetal fraccionado en el plasma materno puede calcularse eligiendo como objetivo a uno o varios locus específicos del feto porque todo el genoma fetal está representado de manera uniforme en el plasma materno[30]. A diferencia de ADN circulante, la medición de la proporción de transcritos de ARN fetal en el plasma materno es menos sencilla ya que se complica por la expresión génica diferencial en los tejidos fetales y maternos y tal vez por su liberación en la circulación. Al realizar la RNA-seq en el plasma materno y examinar las diferencias polimórficas entre el feto y la madre, fuimos capaces de estimar la proporción de transcritos en el plasma aportados por el feto. Aunque los transcritos maternos dominaban el transcriptoma en el plasma, como era de prever, 3,70% y 11,28% de los transcritos circulantes en el plasma materno procedían del feto durante las fases inicial y final del embarazo, respectivamente. Estos transcritos fetales incluyen las moléculas de ARN aportadas por el feto y la madre, así como las aportadas únicamente por el feto. Se encontró que esta última constituye el 0,90% y el 2,52% de las transcritos circulantes maternos, durante las fases inicial y final del embarazo, respectivamente. La mayor representación de dichos genes específicos del feto durante la fase final del embarazo se correlaciona quizás con un aumento en el tamaño del feto y la placenta a medida que evoluciona el embarazo.

En este estudio, se ha demostrado que una expresión alélica equilibrada de ARN del gen PAPPA específico del embarazo en la placenta y la expresión monoalélica del gen *H19* expresado por vía materna impreso se pudo observar en el plasma materno. Estos datos sugieren que el plasma materno podría utilizarse como una fuente de muestras no invasivas para el estudio de patrones de expresión alélica.

Por comparación cuantitativa de los transcritos de ARN en las muestras de plasma maternos antes y después del parto, se ha compilado una lista de 131 genes que habían aumentado durante el embarazo, como es evidente por su reducida representación en las muestras de plasma del puerperio. Como cabía esperar, las características de estos genes podrían utilizarse para diferenciar muestras de plasma de mujeres embarazadas de las no embarazadas. Dicha comparación directa de las muestras de plasma materno antes y después del parto nos ha permitido, en una

forma de alto rendimiento, clasificar los genes relacionados con el embarazo, que pueden no necesariamente expresarse a un nivel mucho más alto en la placenta que en las células sanguíneas maternas como se demuestra en el trabajo anterior[10,15]. En esencia, este método directo de examen del plasma presenta otra vía para el descubrimiento de transcritos circulantes de ARN relacionados con el embarazo, sin un conocimiento *a priori* de las características transcriptómicas de los tejidos de la placenta y las células sanguíneas.

Aunque se ha demostrado que la ARN-seq es un método viable para caracterizar el transcriptoma del plasma, se pueden mejorar más varias cuestiones técnicas. En primer lugar, el rendimiento de la información para la RNA-seq del plasma podría incrementarse por una mayor optimización del protocolo de secuenciación, en particular en la reducción del ARNr muy transcrito y de los genes de globina del plasma. En segundo lugar, se han centrado únicamente en las transcripciones de referencia y todavía no se han explorado isoformas individuales. Los estudios futuros podrían incluir la detección de nuevos transcritos y el análisis diferencial de las variantes de corte y empalme y sus isoformas[31-33] aumentando la profundidad de lectura de la secuenciación. En tercer lugar, se ha omitido el filtrado ASE en el análisis de las proporciones de los transcritos fetal y maternos para las muestras de embarazo en fase inicial a medida que las vellosidades coriónicas y el líquido amniótico se había agotado para el análisis de genotipado. No obstante, se ha demostrado en la fase final del embarazo que el filtrado ASE no tuvo un impacto pronunciado en la identificación de genes con aportes fetales y maternos predominantes en el plasma materno (Fig. 16 y 17).

En conclusión, se ha demostrado que la tecnología RNA-seq se puede utilizar para medir la proporción de transcritos fetales y para identificar genes circulantes relacionados con el embarazo en el plasma materno. Este estudio ha allanado el camino hacia una mejor comprensión del panorama transcriptómico del plasma materno, facilitando por tanto la identificación de candidatos a biomarcadores implicados en las enfermedades relacionadas con el embarazo. Se idea prevé que esta tecnología podría conducir a nuevas vías para el diagnóstico molecular de enfermedades relacionadas con el embarazo o la placenta, y también para otras enfermedades como el cáncer[34].

VIII. Trasplantes

Como se ha mencionado anteriormente, los métodos descritos en la presente memoria también pueden aplicarse a trasplantes. Los métodos para trasplantes pueden proceder de una manera similar al del análisis fetal. Por ejemplo, puede obtenerse un genotipo del tejido trasplantado. Las realizaciones pueden identificar locus donde el tejido trasplantado es heterocigótico, y donde el organismo anfitrión (p. ej., macho o hembra) es homocigótico. Y, las realizaciones pueden identificar locus donde el tejido trasplantado es homocigótico, y donde el organismo anfitrión (p. ej., macho o hembra) es heterocigótico. Las mismas relaciones pueden calcularse, y compararse con un valor discriminatorio para determinar si existe un trastorno.

IX. Sistemas informáticos

Cualquiera de los sistemas informáticos mencionados en la presente memoria puede utilizar cualquier número adecuado de subsistemas. Ejemplos de dichos subsistemas se muestran en la Fig. 30 en un ordenador 3000. En algunas realizaciones, un sistema informático incluye un único ordenador, donde los subsistemas pueden ser los componentes del ordenador. En otras realizaciones, un sistema informático puede incluir varios ordenadores, siendo cada uno un subsistema, con los componentes internos.

Los subsistemas mostrados en la fig. 30 están interconectados a través de un bus del sistema 3075. Se muestran subsistemas adicionales tales como una impresora 3074, el teclado 3078, dispositivo(s) de almacenamiento 3079, monitor 3076, que está acoplado al adaptador de pantalla 3082 y otros. Periféricos y dispositivos de entrada/salida (I/O), que se acoplan al controlador I/O 3071, se pueden conectar al sistema informático por cualquier número de medios conocidos en la técnica, tal como el puerto de serie 3077. Por ejemplo, el puerto de serie 3077 o la interfaz externa 3081 (p. ej., Ethernet, Wi-Fi, etc.) se puede utilizar para conectar el sistema informático 3000 a una red de área amplia tal como Internet, un dispositivo de entrada de ratón, o un escáner. La interconexión a través del bus del sistema 3075 permite al procesador central 3073 comunicarse con cada subsistema y controlar la ejecución de instrucciones desde la memoria del sistema 3072 o el/los dispositivo(s) de almacenamiento 3079 (p. ej., un disco fijo, tal como un disco duro o un disco óptico), así como el intercambio de información entre subsistemas. La memoria del sistema 3072 y/o el/los dispositivo(s) de almacenamiento 3079 pueden incorporar un medio legible por ordenador. Cualquiera de los datos mencionados en la presente memoria puede ser la salida de un componente a otro componente y puede ser salida para el usuario.

Un sistema informático puede incluir un gran número de los mismos componentes o subsistemas, p. ej., conectados entre sí por la interfaz externa 3081 o por una interfaz interna. En algunas realizaciones, los sistemas informáticos, el subsistema o aparatos pueden comunicarse a través de una red. En tales casos, un ordenador puede considerarse un cliente y otro ordenador un servidor, donde cada uno puede formar parte de un mismo sistema informático. Un cliente y un servidor pueden incluir cada uno múltiples sistemas, subsistemas o componentes.

Debe entenderse que cualquiera de las realizaciones de la presente invención se puede llevar a cabo en forma de lógica de control utilizando el equipo informático (*hardware*) (p. ej., un circuito integrado específico para aplicaciones o matriz de puertas programables en campo) y/o utilizando programas informáticos (software) con un procesador

generalmente programable de una manera modular o integrada. Como usuario en la presente memoria, un procesador incluye un procesador de varios núcleos en un mismo chip integrado, o múltiples unidades de procesamiento en una única placa de circuito o en red. Basándose en la descripción y las enseñanzas proporcionadas en la presente memoria, un experto en la técnica sabrá y valorar otras formas y/o métodos para poner en práctica realizaciones de la presente invención utilizando equipos informáticos y una combinación de equipos y programas informáticos.

[0137] Cualquiera de los componentes o funciones del programa informático descritas en esta aplicación de software pueden llevarse a cabo como código de programa informático a ejecutar por un procesador utilizando cualquier lenguaje de ordenador adecuado, tal como, por ejemplo, Java, C++ o Perl usando, por ejemplo, técnicas convencionales u orientadas al objeto. El código programa informático puede ser almacenado como una serie de instrucciones o comandos en un medio legible por ordenador para su almacenamiento y/o transmisión, medios adecuados incluyen memoria de acceso aleatorio (RAM), una memoria de sólo lectura (ROM), un medio magnético tal como un disco duro o un disquete, o un medio óptico, tal como un disco compacto (CD) o DVD (disco versátil digital), memoria flash y similares. El medio legible por ordenador puede ser cualquier combinación de dichos dispositivos de almacenamiento o transmisión.

Dichos programas también pueden codificarse y transmitirse utilizando señales portadoras adaptadas para la transmisión a través de redes cableadas, ópticas, y/o inalámbricas adecuadas a una variedad de protocolos, incluido el Internet. Como tal, puede crearse un medio legible por ordenador según una realización de la presente invención usando una señal de datos codificada con dichos programas. Medios legibles por ordenador codificados con el código del programa pueden empaquetarse con un dispositivo compatible o proporcionarse por separado de otros dispositivos (p. ej., mediante descarga por Internet). Cualquier medio legible por ordenador tal puede residir en o dentro de un único ordenador (p. ej., un disco duro, un CD, o un sistema informático completo), y puede estar presente en o dentro de diferentes ordenadores dentro de un sistema o red. Un sistema informático puede incluir un monitor, impresora u otra pantalla adecuada para proporcionar cualquiera de los resultados mencionados en la presente memoria a un usuario.

Cualquiera de los métodos descritos en la presente memoria puede ser total o parcialmente realizado con un sistema informático incluidos uno o más procesadores, que pueden configurarse para realizar las etapas. Por lo tanto, las realizaciones pueden dirigirse a sistemas informáticos configurados para realizar las etapas de cualquiera de los métodos descritos en la presente memoria, potencialmente con diferentes componentes que realizan unas etapas respectivas etapas o un grupo respectivo de etapas. Aunque presentadas como etapas numeradas, las etapas de los métodos en la presente memoria se pueden realizar al mismo tiempo o en un orden diferente. Además, partes de estas etapas se pueden utilizar con partes de otras etapas de otros métodos. También, toda o partes de una etapa pueden ser opcionales. Además, cualquiera de las etapas de cualquiera de los métodos se pueden realizar con módulos, circuitos, u otros medios para realizar estas etapas.

Los detalles específicos de realizaciones particulares se pueden combinar de cualquier manera adecuada sin apartarse del espíritu y alcance de las realizaciones de la invención. Sin embargo, otras realizaciones de la invención pueden dirigirse a realizaciones específicas relativas a cada aspecto individual, o combinaciones específicas de cada uno de estos aspectos.

La descripción anterior de realizaciones ejemplares de la invención se ha presentado a efectos de ilustración y descripción. No se pretende que sea exhaustiva o que limite la invención a la forma precisa descrita, y son posibles muchas modificaciones y variaciones a la luz de las enseñanzas anteriores. Las realizaciones se eligieron y describieron con el fin de explicar mejor los principios de la invención y sus aplicaciones prácticas para permitir por ello a otros expertos en la técnica utilizar mejor la invención en varias realizaciones y con varias modificaciones que son adecuadas al uso particular contemplado.

A relación de "un", "una" o "el", "la", "los", "las" quiere decir "uno o más" a menos que se indique específicamente lo contrario.

X. Referencias

1. Ng EKO, Tsui NBY, Lau TK, Leung TN, Chiu RWK, Panesar NS, Lit LC, Chan KW, Lo YMD. mRNA of placental origin is readily detectable in maternal plasma. *Proc. Natl. Acad. Sci. U S A* 2003; 100: 4748-53.

2. Ng EKO, Leung TN, Tsui NBY, Lau TK, Panesar NS, Chiu RWK, Lo YMD. The concentration of circulating corticotropin-releasing hormone mRNA in maternal plasma is increased in preeclampsia. *Clin. Chem*. 2003;49: 727-31.

3. Farina A, Sekizawa A, Sugito Y, Iwasaki M, Jimbo M, Saito H, Okai T. Fetal DNA in maternal plasma as a screening variable for preeclampsia. A preliminary nonparametric analysis of detection rate in low-risk nonsymptomatic patients. *Prenat. Diagn*. 2004;24: 83-6.

4. Pang WW, Tsui MH, Sahota D, Leung TY, Lau TK, Lo YM, Chiu RW. A strategy for identifying circulating placental RNA markers for fetal growth assessment, *Prenat. Diagn*. 2009:29: 495-504.

5. Lo YMD, Tsui NBY, Chiu RWK, Lau TK, Leung TN, Heung MM, Gerovassili A, Jin Y, Nicolaides KH, Cantor CR, Ding C. Plasma placental RNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection. *Nat. Med*. 2007; 13: 218-23.

6. Tsui NBY, Akolekar R, Chiu RWK, Chow KCK, Leung TY, Lau TK, Nicolaides KH, Lo YMD. Synergy of total PLAC4 RN A concentration and measurement of the RNA single-nucieotide polymorphism allelic ratio for the noninvasive prenatal detection of trisomy 21. *Clin. Chem*. 2010;56: 73-81.

7. Tsui NBY, Wong BCK, Leung TY, Lau TK, Chiu RWK, Lo YMD. Non-invasive prenatal detection of fetal trisomy 18 by RNA-SNP allelic ratio analysis using maternal plasma SERP1NB2 mRNA: a feasibility study. *Prenat. Diagn*. 2009; 29: 1031-7.

8. Chim SS, Lee WS, Ting YH, Chan OK, Lee SW, Leung TY. Systematic identification of spontaneous preterm birth-associated RNA transcripts in maternal plasma. PLoS One 2012:7: e34328.

9. Wong BCK, Chiu RWK, Tsui NBY, Chan KCA, Chan LW, Lau TK, Leung TN, Lo YMD. Circulating placental RNA in maternal plasma is associated with a preponderance of 5' mRNA fragments: implications for noninvasive prenatal diagnosis and monitoring. *Clin. Chem*. 2005;51 : 1786-95.

10. Tsui NBY, Chim SSC, Chiu RWK, Lau TK, Ng EKO, Leung TN, Tong YK, Chan KCA, Lo YMD. Systematic micro-array based identification of placental mRNA in maternal plasma: towards non-invasive prenatal gene expression profiling. *J. Med. Genet*. 2004;41: 461-7.

11. Poon LL, Leung TN, Lau TK, Lo YMD. Presence of fetal RNA in maternal plasma. *Clin. Chem*. 2000;46: 1832-4.

12. Go AT, Visser A, Mulders MA, Blankenstein MA, Van Vugt JM, Oudejans CB. Detection of placental transcription factor mRNA in maternal plasma. *Clin. Chem*. 2004;50:1413-4.

13. Smets EM, Visser A, Go AT, van Vugt JM, Oudejans CB. Novel biomarkers in preeclampsia. *Clin. Chim. Acta* 2006;364:22-32.

14. Purwosunu Y, Sekizawa A, Koide K, Farina A, Wibowo N, Wiknjosastro GH, *et al*. Cell-free mRNA concentrations of plasminogen activator inhibitor-1 and tissue-type plasminogen activator are increased in the plasma of pregnant women with preeclampsia. *Clin. Chem*. 2007;53 :399-404.

15. Miura K, Miura S, Yamasaki K, Shimada T, Kinoshita A, Niikawa N, *et al*. The possibility of microarray -based analysis using cell-free placental mRNA in maternal plasma, *Prenat. Diagn*. 2010;30:849-61.

16. Ng EKO, El-Sheikhah A, Chiu RWK, Chan KC, Hogg M, Bindra R, *et al*. Evaluation of human chorionic gonadotropin beta-subunit mRNA concentrations in maternal serum in aneuploid pregnancies: a feasibility study. *Clin. Chem*. 2004;50: 1055-7.

17. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptornes by RNA-Seq. *Nat. Methods* 2008;5:621-8.

18. Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, *et al*. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 2008;321 :956-60.

19. Kim J, Zhao K, Jiang P, Lu ZX, Wang J, Murray JC, Xing Y. Transcriptome landscape of the human placenta. *BMC Genomics* 2012; 13 :115.

20. Wang K, Li H, Yuan Y, Etheridge A, Zhou Y, Huang D, et al. The complex exogenous RNA spectra in human plasma: an interface with human gut biota? PLoS One 2012;7:e51009.

21. Li H, Guo L, Wu Q, Lu J, Ge Q, Lu Z. A comprehensive survey of maternal plasma miRNAs expression profiles using high-throughput sequencing. *Clin. Chim. Acta* 2012;413 :568-76.

22. Williams Z, Ben-Dov IZ, Elias R, Mihailovic A, Brown M, Rosenwaks Z, Tuschl T. Comprehensive profiling of circulating microRNA via small RNA sequencing of cDNA libraries reveals biomarker potential and limitations. *Proc. Natl. Acad. Sci. U S A* 2013,110:4255-60,

23. Chim SSC, Shing TK, Hung EC, Leung TY, Lau TK, Chiu RWK, Lo YMD. Detection and characterization of placental microRNAs in maternal plasma. *Clin. Chem*. 2008;54:482-90.

24. Pickrell JK *et al*. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* 2010;464:768-772.

25. Smith RM, Webb A, Papp AC, Newman LC, Handelman SK, Suhy A, *et al*. Whole transcriptome RNA-Seq allelic expression in human brain. BMC Genomics 2013; 14:571.

26. Frost JM, Monk D, Stojilkovic-Mikic T, Woodfine K, Chitty LS, Murrell A, *et al*. Evaluation of allelic expression of imprinted genes in adult human blood, PLoS One 2010:5:e13556.

27. Daelemans C, Ritchie ME, Smits G, Abu-Amero S, Sudbery IM, Forrest MS, *et al*. High-throughput analysis of candidate imprinted genes and allele-specific gene expression in the human term placenta. *BMC Genet*. 2010; 11:25.

28. Lun FMF, Chiu RWK, Sun K, Leung TY, Jiang P, Chan KCA, *et al*. Noninvasive prenatal methylomic analysis by genomewide bisulfite sequencing of maternal plasma DNA. *Clin. Chem*. 2013;59: 1583-94.

29. Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, *et al*. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol*. 2009; 10:R130.

30. Lo YMD, Chan KCA, Sun H, Chen EZ, Jiang P, Lun FMF, *et al*. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci. Transl. Med*. 2010;2:61 ra91.

31. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 2011;25:1915-27.

32. St Laurent G, Shtokalo D, Tackett MR, Yang Z, Eremina T, Wahlestedt C, *et al*. Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells. *BMC Genomics* 2012;13:504.

33. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* 2012; 22:2008-17.

34. Fleischhacker M, Schmidt B. Circulating nucleic acids (CNAs) and cancer— a survey. Biochim. Biophys. Acta 2007; 1775:181-232.

REIVINDICACIONES

1. Un método para diagnosticar un trastorno relacionado con el embarazo usando una muestra de una hembra embarazada con un feto, comprendiendo el método:

recibir un gran número de lecturas de secuencias, en donde las lecturas de secuencias se obtienen de un análisis de moléculas de ARN obtenidos de la muestra, muestra que contiene una mezcla de moléculas de ARN maternos y fetales;

identificar, mediante un sistema informático, las ubicaciones de las lecturas de secuencias en una secuencia de referencia;

identificar uno o más locus informativos, cada uno de los cuales es homocigótico en una primera entidad para un primer alelo correspondiente y que es heterocigótico en una segunda entidad para el correspondiente primer alelo y un correspondiente segundo alelo, en donde la primera entidad es la hembra embarazada o el feto, y la segunda entidad es el otro de entre la hembra embarazada y el feto;

filtrar uno o más locus informativos para identificar uno o más locus informativos filtrados:

que se encuentran dentro de una región expresada de la secuencia de referencia, en la que se encuentran al menos un primer número predeterminado de lecturas de secuencias en el gran número de lecturas de secuencias que contienen el correspondiente primer alelo, y

en el que se encuentra al menos un segundo número predeterminado de lecturas de secuencias en el gran número de lecturas de secuencias que contienen el segundo alelo correspondiente,

para cada uno de los locus informativos filtrados:

determinar un primer número de lecturas de secuencias situadas en el locus informativo filtrado y que contiene el primer alelo correspondiente,

determinar un segundo número de lecturas de secuencias situadas en el locus informativo filtrado y que contiene el segundo alelo correspondiente,

calcular una primera suma de los primeros números y la segunda suma de los segundos números;

calcular una relación entre la primera suma y la segunda suma, y

comparar la relación con un valor discriminatorio para determinar si el feto, la madre o el embarazo tiene un trastorno relacionado con el embarazo, determinándose el valor discriminatorio a partir de una o más muestras obtenidas de hembras embarazadas sin el trastorno relacionado con el embarazo.

2. El método de la reivindicación 1, en donde la muestra de la mujer embarazada es una muestra de plasma sanguíneo.

3. El método de la reivindicación 1 o 2, en donde la segunda entidad es la hembra embarazada.

4. El método de la reivindicación 1, en donde la segunda entidad es el feto.

5. El método de una cualquiera de las reivindicaciones precedentes, en donde el cálculo de una relación de la primera suma y la segunda suma comprende dividir la segunda suma por una suma de la primera suma y la segunda suma.

6. El método de una cualquiera de las reivindicaciones precedentes, en donde el análisis de moléculas de ARN obtenidas de la muestra comprende la secuenciación de las moléculas de ARN o de una de las copias de ADNc.

7. El método de una cualquiera de las reivindicaciones 1 a 5, en donde el análisis de moléculas de ARN obtenidas de la muestra comprende realizar la PCR digital.

8. El método de una cualquiera de las reivindicaciones precedentes, en donde la identificación de las ubicaciones de las lecturas de secuencias en una secuencia de referencia comprende la alineación de las lecturas de secuencias con la secuencia de referencia.

9. El método de una cualquiera de las reivindicaciones precedentes, en donde el primer número predeterminado de lecturas de secuencias es 1, y en donde el segundo número predeterminado de lecturas de secuencias es 1.

10. El método de una cualquiera de las reivindicaciones precedentes, que comprende además ADN genómico de secuenciación obtenido de tejido materno para determinar un genotipo de la hembra embarazada en cada locus informativo.

11. El método de una cualquiera de las reivindicaciones precedentes, que comprende además ADN genómico de secuenciación obtenido de la placenta, de vellosidades coriónicas, del líquido amniótico o del plasma materno para determinar un genotipo del feto en cada locus informativo.

12. El método de una cualquiera de las reivindicaciones precedentes, que comprende además la estimación de una parte de ARN en la muestra que es de origen materno o fetal, en donde la parte es la relación multiplicada por un escalar.

13. Un programa informático que comprende un gran número de instrucciones que pueden ser ejecutadas por un sistema informático, que cuando se ejecutan controlan el sistema informático para llevar a cabo el método de una cualquiera de las reivindicaciones precedentes.

14. Un producto del programa informático que comprende un medio legible por ordenador que almacena un programa informático según la reivindicación 13.

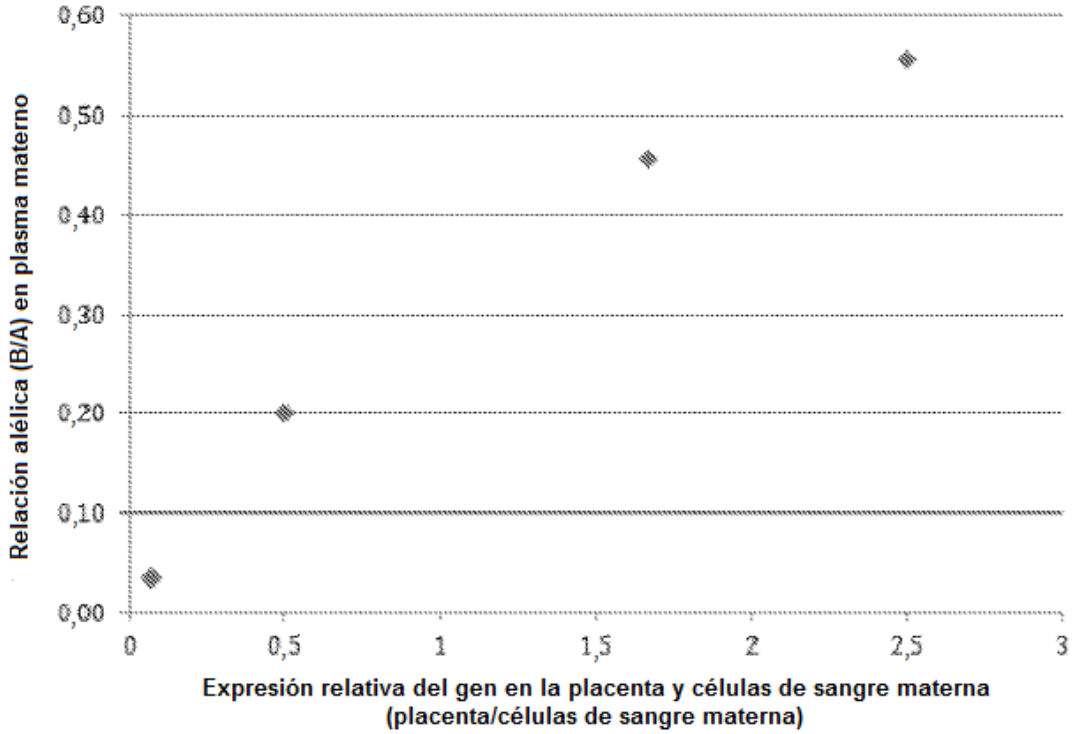
0,60
0,50
0,40
0,30
0,20
0,10
0,00
Relación alélica (B/A) en plasma materno
0
0,5
1
1,5
2
2,5
3
Expresión relativa del gen en la placenta y células de sangre materna
(placenta/células de sangre materna)

FIG. 1

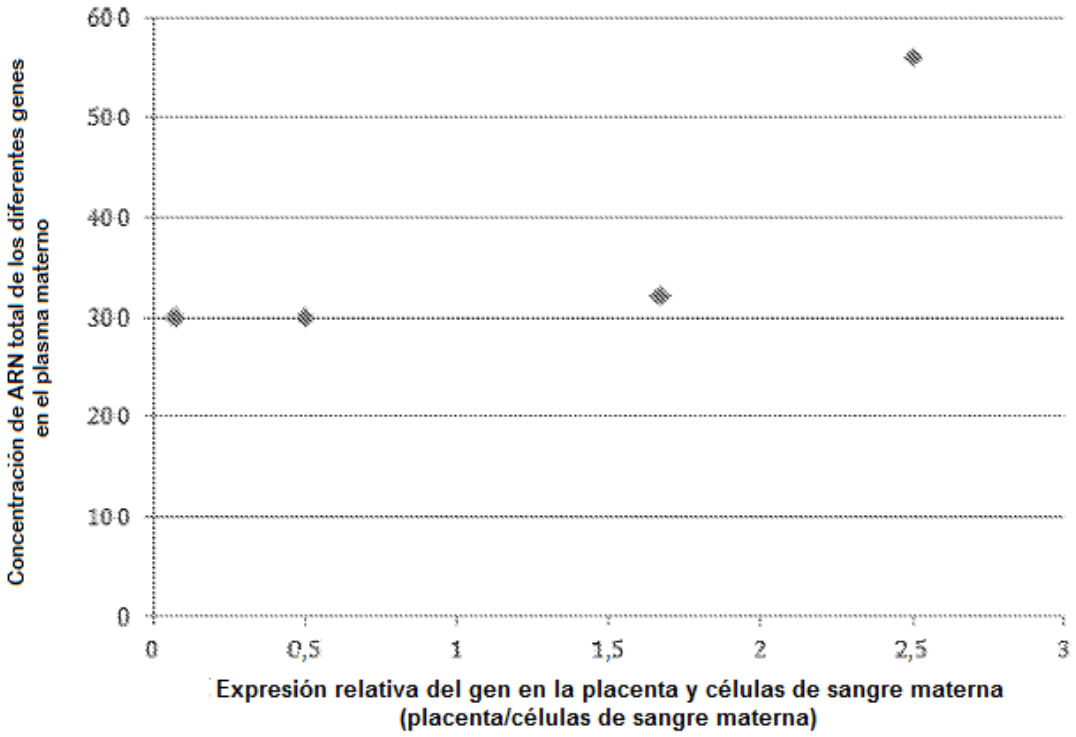
600
500
400
300
200
100
0
Concentración de ARN total de los diferentes genes en el plasma materno
0
0,5
1
1,5
2
2,5
3
Expresión relativa del gen en la placenta y células de sangre materna
(placenta/células de sangre materna)

FIG. 2

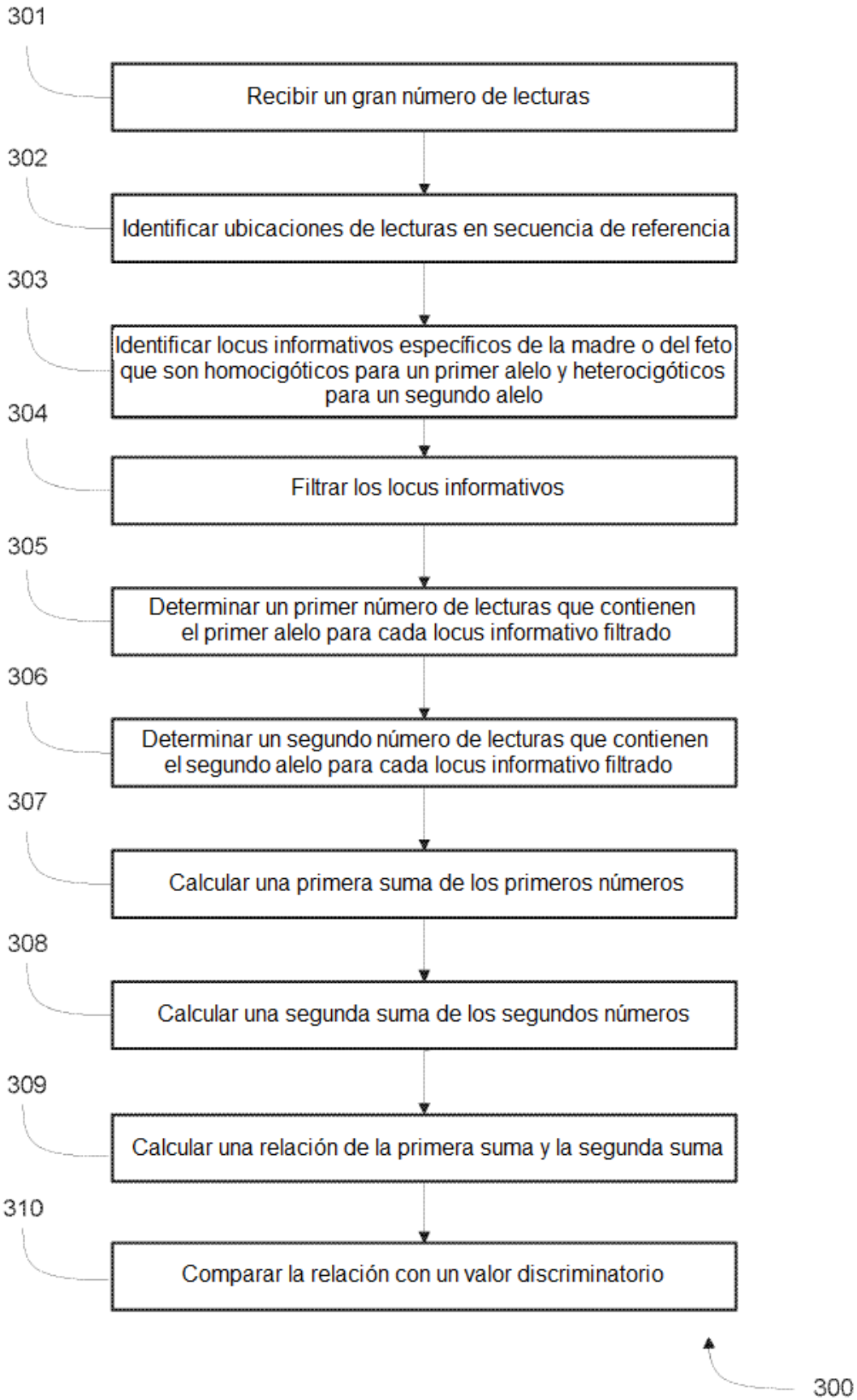
301
Recibir un gran número de lecturas
302
Identificar ubicaciones de lecturas en secuencia de referencia
303
Identificar locus informativos específicos de la madre o del feto que son homocigóticos para un primer alelo y heterocigóticos para un segundo alelo
304
Filtrar los locus informativos
305
Determinar un primer número de lecturas que contienen el primer alelo para cada locus informativo filtrado
306
Determinar un segundo número de lecturas que contienen el segundo alelo para cada locus informativo filtrado
307
Calcular una primera suma de los primeros números
308
Calcular una segunda suma de los segundos números
309
Calcular una relación de la primera suma y la segunda suma
310
Comparar la relación con un valor discriminatorio
300

FIG. 3

Tipo de muestra	Muestra	Lecturas en bruto	Lecturas preprocesadas[a]	(%)[b]	Lecturas cartografiables[c]	(%)[d]
Plasma de hembra no embarazada	CTR081	180.080.810	23.012.332	12,78%	7.708.011	33,50%
	CTR082	132.567.446	20.559.146	15,51%	6.848.249	33,31%
Plasma materno antes del parto	M9600P	367.301.939	45.255.199	12,32%	16.849.461	37,23%
	M9549P	597.724.438	151.243.195	25,30%	34.627.313	22,90%
	M9356P	820.742.404	190.030.278	23,15%	14.091.631	7,42%
	M9415P	970.145.241	143.615.377	14,80%	41.517.504	28,91%
Plasma materno después del parto	M9356BP	587.698.232	132.012.632	22,46%	7.828.124	5,93%
	M9415BP	853.401.365	101.070.496	11,84%	26.707.674	26,42%
Placenta	N9356	202.711.386	169.500.629	83,62%	160.417.620	94,64%
	N9415	257.219.381	203.094.642	78,96%	193.126.555	95,09%
Células de sangre materna	M9600W	163.832.386	38.707.658	23,63%	11.348.398	29,32%
	M9549W	240.383.043	83.922.150	34,91%	73.117.147	87,12%
	M9356W	208.400.902	60.851.642	29,20%	54.974.158	90,34%
	M9415W	265.636.207	49.943.628	18,80%	42.670.781	85,44%

a. Lecturas conservadas tras la eliminación de lecturas muy repetitivas.

b. % de lecturas en bruto.

c. Para desglose detallado véase Datos complementarios en la tabla 2B.

d. % de lecturas preprocesadas.

FIG. 4

ES 2 670 544 T3

Tipo de muestra	Muestra	Lecturas filtradas[a]	Lecturas analizables[b]	Exón (%)	Intrón (%)	Región intergénica (%)[c]
Plasma de hembra no embarazada	CTR081	4.186.990	3.521.021	64,00%	6,47%	29,53%
	CTR082	3.594.966	3.253.283	61,93%	7,29%	30,78%
Plasma materno antes del parto	M9600P	7.711.430	9.138.031	64,42%	6,47%	29,10%
	M9549P	20.481.388	14.145.925	41,84%	13,16%	45,00%
	M9356P	1.683.896	12.407.735	56,80%	30,76%	12,45%
	M9415P	23.164.778	18.352.726	53,61%	10,80%	35,59%
Plasma materno después del parto	M9356BP	988.066	6.840.058	50,41%	38,15%	11,44%
	M9415BP	15.853.377	10.854.297	50,12%	13,68%	36,19%
Placenta	N9356	3.204.907	157.212.713	31,18%	55,89%	12,93%
	N9415	4.186.711	188.939.844	26,51%	59,59%	13,90%
Células de sangre materna	M9600W	854.869	10.493.529	69,71%	23,50%	6,79%
	M9549W	8.120.655	64.996.492	66,98%	26,67%	6,36%
	M9356W	4.255.081	50.719.077	69,13%	24,70%	6,16%
	M9415W	5.318.314	37.352.467	85,68%	10,74%	3,58%

a. La mayoría de las lecturas filtradas son ARNr y ARNt nucleares y mitocondriales.

b. Lecturas analizables = lecturas cartografiables totales – lecturas filtradas.

c. % de lecturas que se alinean a regiones fuera de exones e intrones de los genes de referencia.

FIG. 5

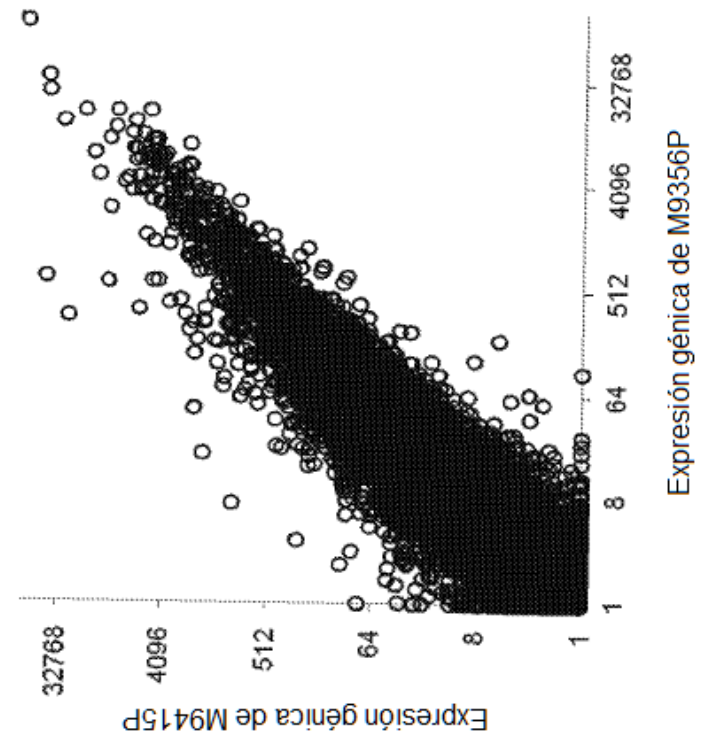
Expresión génica de M9356P
1
8
64
512
4096
32768
Expresión génica de M9415P
1
8
64
512
4096
32768

FIG. 6B

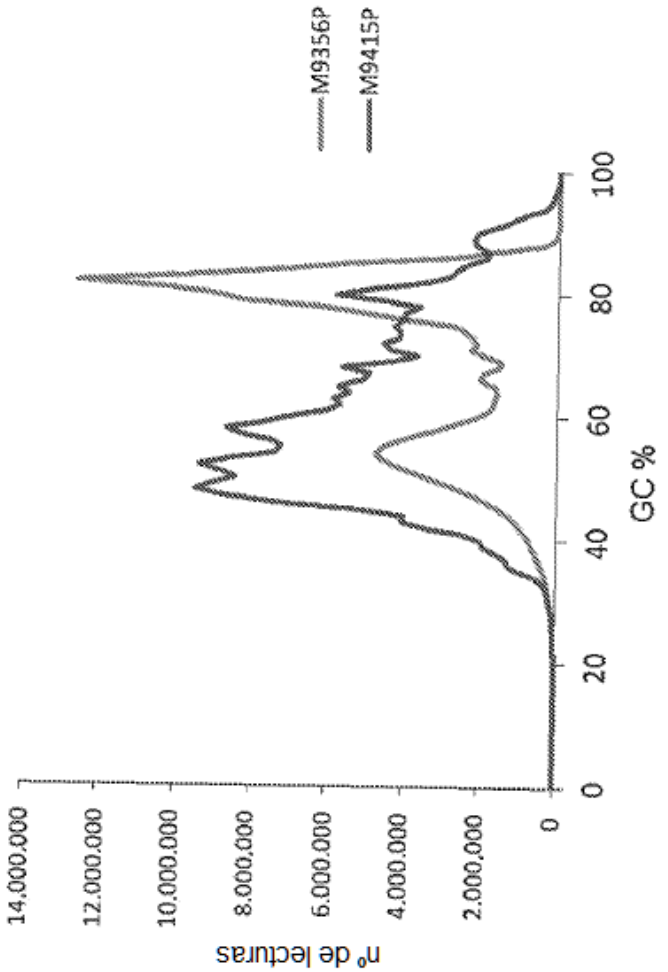
M9356P
M9415P
nº de lecturas
14.000.000
12.000.000
10.000.000
8.000.000
6.000.000
4.000.000
2.000.000
0
GC %
0
20
40
60
80
100

FIG. 6A

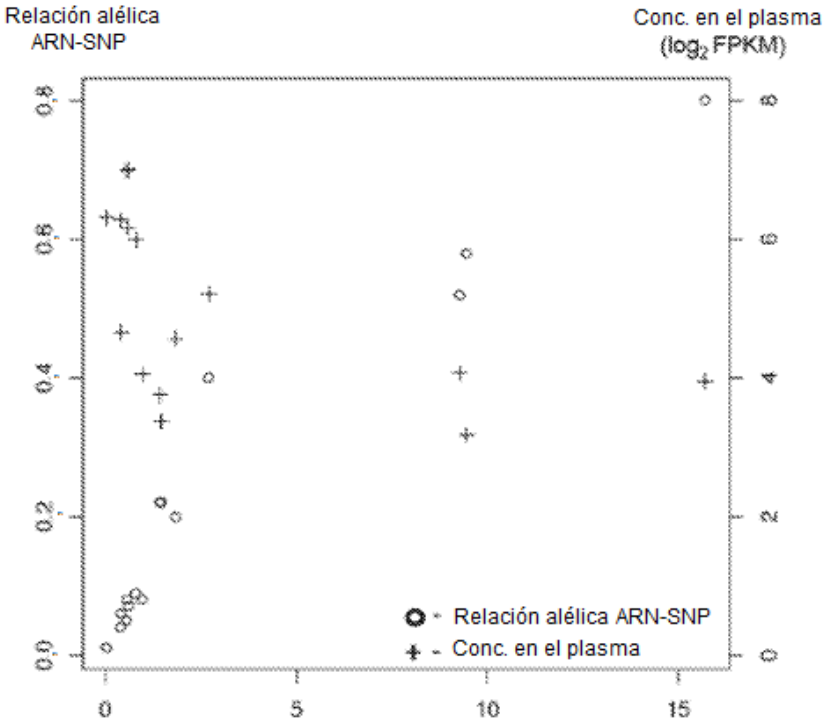
Relación alélica
ARN-SNP
Conc. en el plasma
(log2 FPKM)
0,0
0,2
0,4
0,6
0,8
0
2
4
6
8
0
5
10
15
Relación alélica ARN-SNP
Conc. en el plasma
Nivel de expresión de placenta/células sanguíneas

FIG. 7

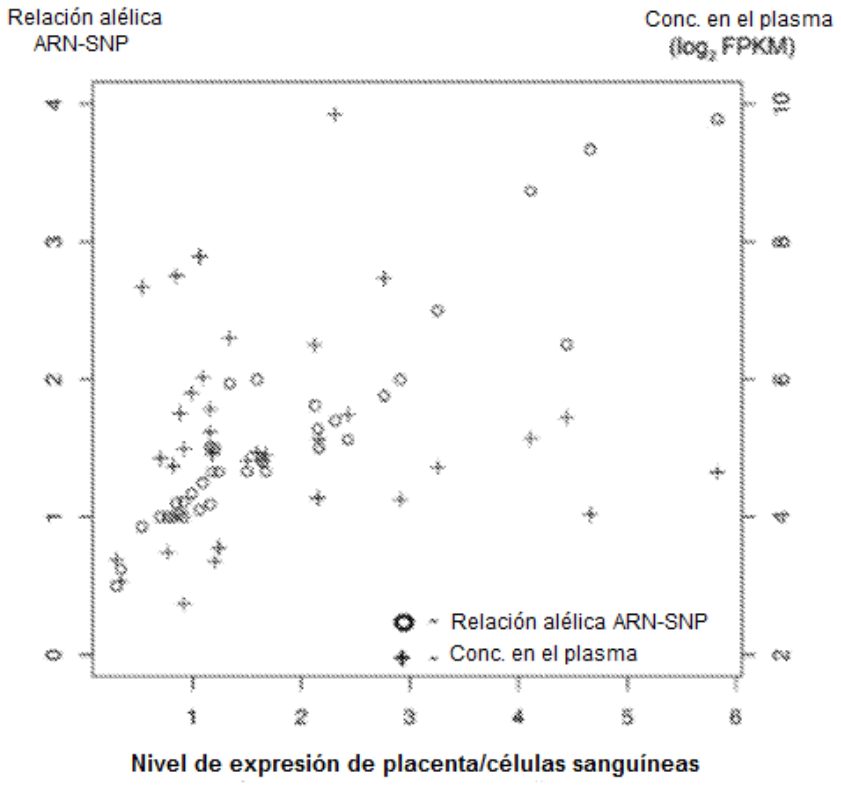
Relación alélica
ARN-SNP
Conc. en el plasma
(log2 FPKM)
0
1
2
3
4
0
2
4
6
8
10
1
2
3
4
5
6
Relación alélica ARN-SNP
Conc. en el plasma
Nivel de expresión de placenta/células sanguíneas

FIG. 8

Alelos de SNP específicos del feto (Materno: AA, Fetal: AB)	SNP	Relación alelo fetal/alelo compartido	Relación ARN-SNP media por gen	FPKM de placenta	FPKM de células sanguíneas	Veces de diferencia (placenta/células sanguíneas)	FPKM del plasma
AFF1	SNP 7:rs342462	0,28	0,52	149,731	16,1598	9,27	16,7037
	SNP 8:rs342459	1					
	SNP 9:rs10029915	0,29					
SH3BGRL2	SNP 13:rs1507	0,07	0,07	24,7477	40,3078	0,61	128,2908
SLC38A1	SNP 3:rs3742058	0,17	0,58	142,8723	15,1123	9,45	9,0209
	SNP 4:rs3742059	1					
PTTG1P	SNP 6:rs873317	0,4	0,4	152,8104	56,4153	2,71	37,1191
GIMAP1-GIMAP5	SNP 14:rs1046355	0,04	0,04	15,9771	39,9271	0,4	77,4282
CTSS	SNP 1:rs3754212	0,01	0,01	11,0084	382,8589	0,03	79,3012
RHOBTB3	SNP 10:rs34896	1	0,8	338,5763	21,5841	15,69	15,4046
	SNP 11:rs34898	0,6					
	SNP 12:rs34899	0,8					
ARHGEF12	SNP 2:rs192719576	0,09	0,09	88,5256	110,1833	0,8	63,7037
TXNRD1	SNP 5:rs11111979	0,2	0,2	65,9309	35,795	1,84	23,6787
HPCAL1	SNP 11:rs2270302	0,06	0,06	17,2498	42,5469	041	25,2768
RAB4A	SNP 2:rs15452	0,08	0,08	12,7741	22,2066	0,58	72,2382
PGM2	SNP 13:rs3752683	0,22	0,22	33,0576	23,045	1,43	13,4293
PSMB3	SNP 10:rs228274	0,05	0,05	65,0683	117,5188	0,55	127,3262
ADIPOR2	SNP 4:rs12342	0,08	0,08	18,1942	18,7699	0,97	16,5717
	SNP 5:rs1044471	0,08					
CHFR	SNP 9:rs35011845	0,22	0,22	21,9501	15,0583	1,46	10,3748

FIG. 9

Alelos de SNP específicos del feto (Materno: AA, Fetal: AB)	SNP	Relación alelo fetal/alelo compartido	Relación ARN-SNP media por gen	FPKM de placenta	FPKM de células sanguíneas	Veces de diferencia (células sanguíneas/ placenta)	FPKM del plasma
AKAP10	SNP 18:rs2108978	1,33	1,33	22,414	18,1324	1,24	11,7168
EIF252	SNP 37:rs4911406	1,25	1,25	39,3008	35,9748	1,09	65,1988
JHDM1D	SNP 51:rs76676651	1,4	1,4	43,0656	26,1836	1,64	28,1734
	SNP 5:rs11022079	0,5	0,5	10,735	36,1867	0,3	10,3945
USP47	SNP 6:rs148238084	0,5					
PPP4R2	SNP 40:rs61188513	1,09	1,09	80,3628	69,563	1,16	47,4149
	SNP 15:rs4779520	0,78	1,64	22,9283	10,6869	2,15	19,3486
KLF13	SNP 16:rs6493629	2,5					
FAM107B	SNP 4:rs11259151	2	2	24,1097	15,1776	1,59	30,5864
APOL2	SNP 39:rs132759	1,33	1,33	17,9996	12,0226	1,5	28,1601
	SNP 10:rs238991D	1,56	1,11	27,2282	29,7396	0,92	31,7672
RAP2A	SNP 11:rs12873919	0,67					
	SNP 1:rs117963129	4	3,37	48,3896	11,7589	4,11	35,3443
	SNP 2:rs1055184	5					
RIT1	SNP 3:rs493446	1,12					
STK17A	SNP 48:rs56286238	1,33	1,33	61,0052	36,447	1,67	30,0074
	SNP 23:rs14031	0,5	0,62	11,0432	32,3155	0,34	8,3988
C19orf63	SNP 24:rs1045246	0,75					
SMEK1	SNP 14:rs10135595	1	1	27,1212	33,0441	0,82	26,8366
	SNP 42:rs3797341	1,5	1,02	26,837	30,3427	0,88	45,4491
	SNP 43:rs3203922	0,78					
TNFAIP8	SNP 44:rs1045242	0,79					
CPEB4	SNP 46:rs12659398	1	1	28,4523	40,5723	0,7	28,802
PTPRJ	SNP 7:rs117902837	2	2	59,7568	20,5437	2,91	19,0642
RPL36	SNP 22:rs3810220	1,7	1,7	152,2501	65,9528	2,31	921,8925
	SNP 25:rs1137288	1,62	1,88	114,7107	41,6213	2,76	176,8333
	SNP 26:rs3087895	2,17					
	SNP 27:rs3768670	1,86					
YPEL5	SNP 28:rs3768671	1,86					
	SNP 53:rs114261839	1	1,33	51,7974	43,9599	1,18	30,1875
	SNP 54:rs208753	1,4					
	SNP 55:rs17635381	2,27					
PCM1	SNP 56:rs6991775	0,67					
UBR2	SNP 47:rs6916713	2,25	2,25	102,9892	23,2007	4,44	43,5569
	SNP 32:rs3747	1,22	1,97	50,467	37,5349	1,34	96,8859
	SNP 33:rs9579	2,17					
	SNP 34:rs3088214	2,89					
TNS1	SNP 35:rs1063281	1,62					
EIF4A1	SNP 17:rs1057086	0,93	0,93	70,1974	132,2932	0,53	161,6599
BACH1	SNP 38:rs15092	2,5	2,5	112,4409	34,6219	3,25	26,4284
	SNP 49:rs6850	1,11	1,05	66,7742	62,9202	1,06	220,7639
PPIA	SNP 50:rs6904	1					
TMX1	SNP 12:rs7160810	1	1	11,8758	12,8978	0,92	6,7127
ANKRD44	SNP 29:rs3731569	3,67	3,67	65,5013	14,0504	4,66	16,412
	SNP 57:rs3398	0,33	1,51	43,1289	37,048	1,16	37,6685
ZFAND5	SNP 58:rs969	2,69					
	SNP 19:rs12327477	0,67	3,89	171,6269	29,4281	5,83	25,153
	SNP 20:rs2304859	7					
SMCHD1	SNP 21:rs7238459	4					
NR3C1	SNP 45:rs6191	1	1	28,0561	36,5516	0,77	11,2138
CRNLK1	SNP 36:rs2255255	1,5	1,5	15,8244	13,148	1,2	10,1897
CALM1	SNP 13:rs5871	1,1	1,1	76,0647	89,0949	0,85	181,3474
FDFT1	SNP 52:rs79708434	1,56	1,56	83,9629	34,6215	2,43	44,9826
	SNP 30:rs1051677	1	1,17	74,6452	75,2029	0,99	55,8144
XRCC5	SNP 31:rs2440	1,33					
	SNP 8:rs3741380	2,75	1,81	29,7293	14,0146	2,12	90,3776
EHBP1L1	SNP 9:rs6591182	0,88					
OPA1	SNP 41:rs7643844	1,5	1,5	49,1647	22,7367	2,16	34,8335

FIG. 10

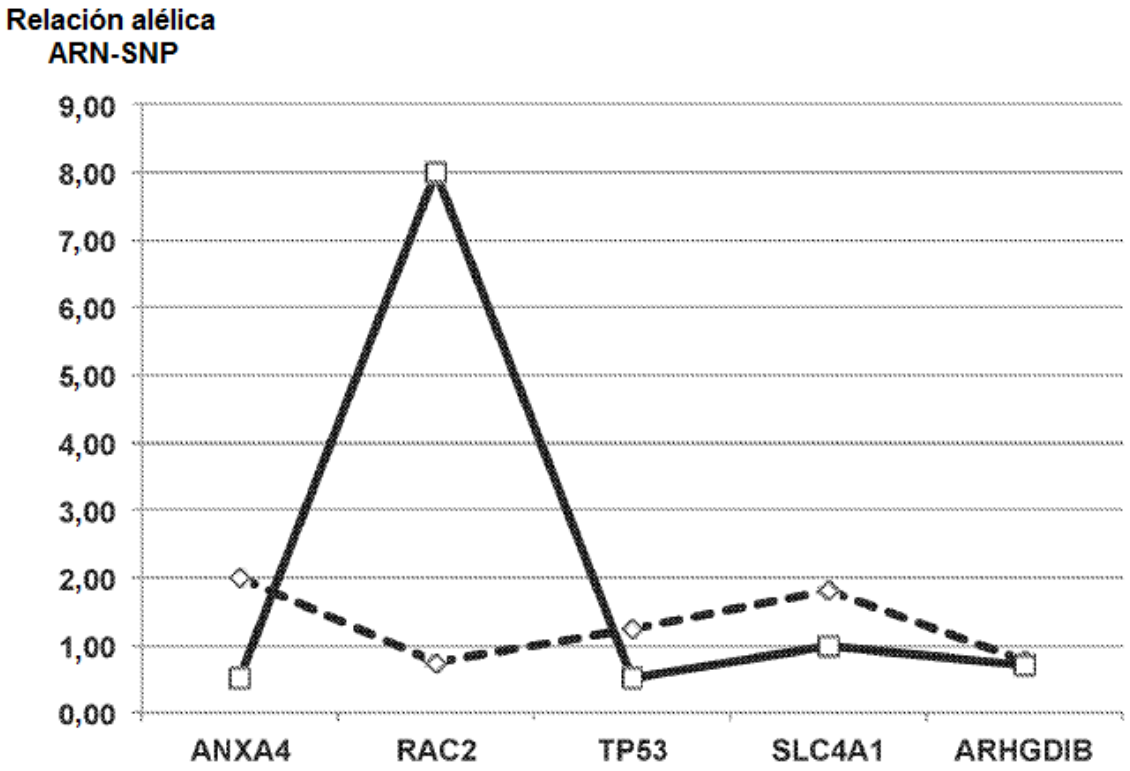

FIG. 11A

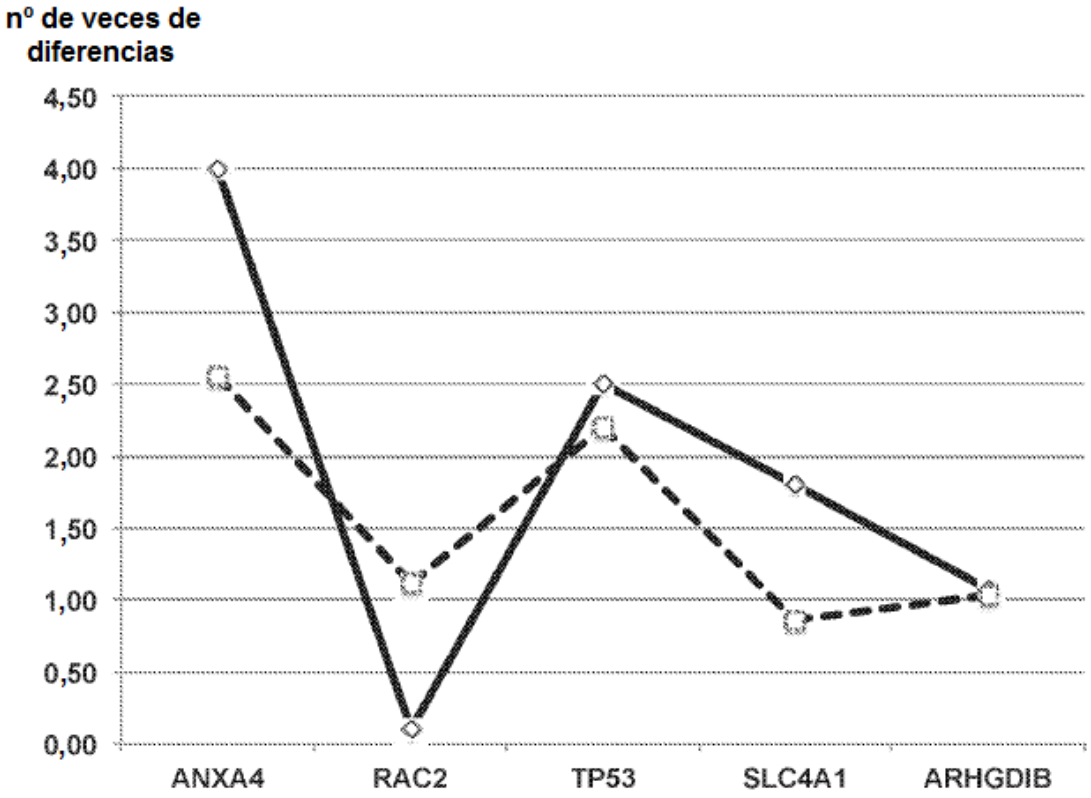

FIG. 11B

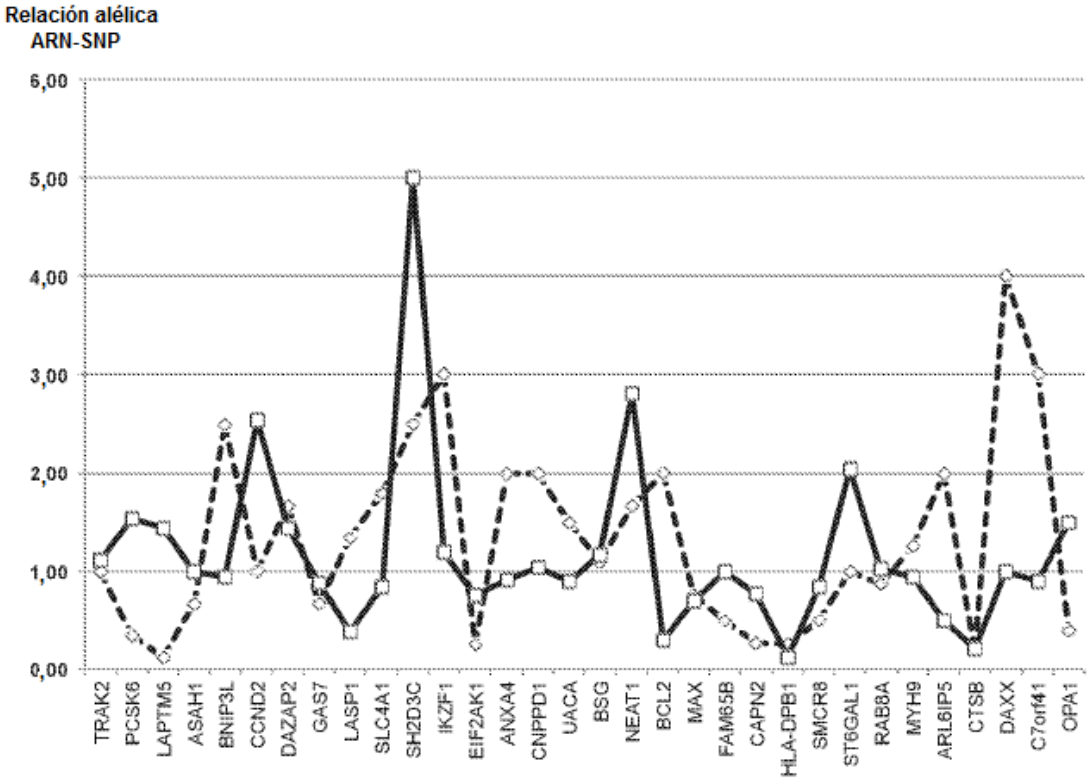

FIG. 12A

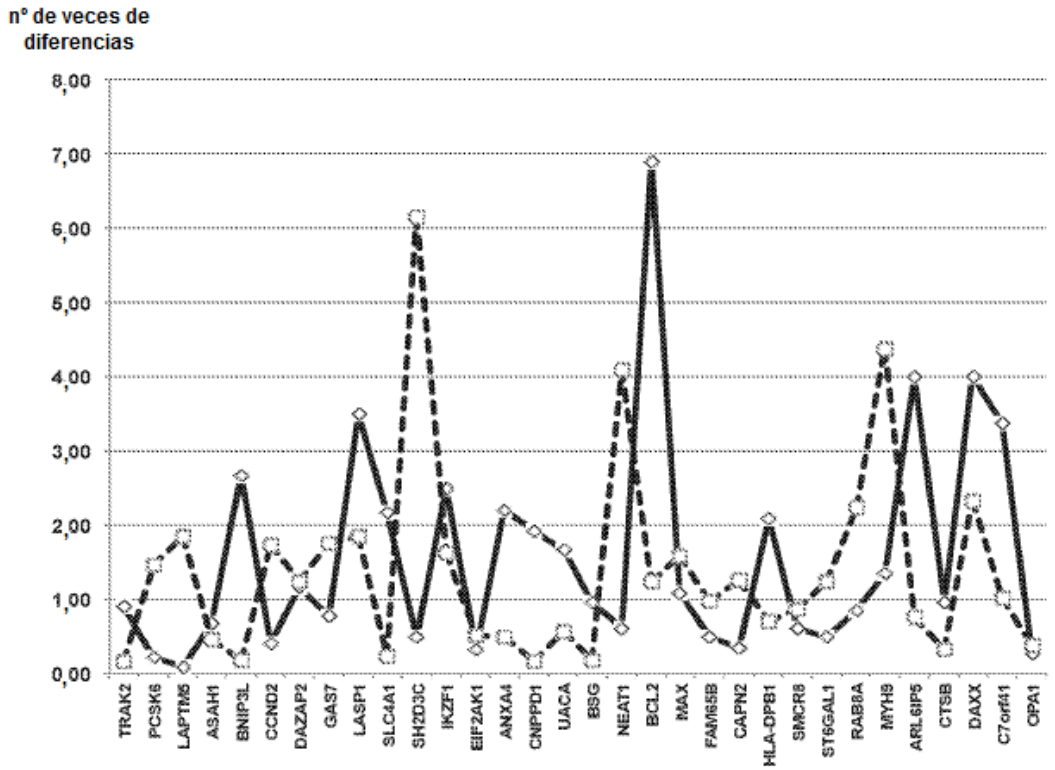

FIG. 12B

	5641 (PET) (GA: 36 6/7)			7171 (Normal) (GA: 36 6/7)			Nº de veces de cambios de relación ARN-SNP para cada gen (5641 frente a 7171)	Nivel de expresión en el plasma		
Alelos de SNP específicos de la madre (Materno: AB, Fetal: AA)	**SNP**	**Relación alelo fetal/alelo compartido**	**Relación ARN-SNP media por gen**	**SNP**	**Relación alelo materno/alelo compartido**	**Relación ARN-SNP media por gen**		**FPKM de 5641**	**FPKM de 7171**	**Veces de cambios de FPKM (5641 frente a 7171)**
ANXA4	SNP 1:kgp9682337	2	2	SNP 9:rs6546548	0,5	0,5	4	12,1379	4,7565	2,551855356
RAC2	SNP 3:rs6572	0,75	0,75	SNP 9:rs2239774	8	8	0,09375	319,973	288,0231	1,110928255
TP53	SNP 2:rs17879353	0,5	1,25	SNP 4:rs1042522	0,5	0,5	2,5	28,9503	13,1548	2,200740414
	SNP 4:rs1042522	0,2								
SLC4A1	SNP 5:rs5035	1,8	1,8	SNP 7:rs2072081	1	1	1,8	165,1352	192,3307	0,858600317
ARHGDIB	SNP 6:kgp6869929	0,76	0,76	SNP 6:kgp6869929	0,71	0,71	1,070422535	665,6615	640,9633	1,038532939

FIG. 13

ES 2 670 544 T3

	5641 (PET) (GA: 36 6/7)			9356 (Normal) (GA: 37 2/7)			Nº de veces de cambios de relación ARN-SNP para cada gen (5641 frente a 7171)	Nivel de expresión en el plasma		
Alelos de SNP específicos de la madre (Materno: AB, Fetal: AA)	SNP	Relación alelo materno/ alelo compartido	Relación ARN-SNP media por gen	SNP	Relación alelo materno/ alelo compartido	Relación ARN-SNP media por gen		FPKM de 5641	FPKM de 9356	Veces de cambios de FPKM (5641 frente a 7171)
TRAK2	SNP 2:kgp14544935	1	1	SNP 62:rs11898629	0,8	1,11	0,800900501	95,0691	618,1783	0,153789125
				SNP 63:rs3795966	1,47					
				SNP 64:rs13022344	1,04					
PCSK8	**SNP 26:kgp8258505**	0,33	0,33	SNP 48:rs1058260	1,54	1,54	8,214285714	22,3436	15,3949	1,451376755
LAPTM5	SNP 7:kgp15703090	0,12	0,12	SNP 38:rs8048	1,43	1,43	0,083916084	215,8869	117,3578	1,83956158
ASAH1	SNP 3:kgp3689333	0,67	0,67	SNP 85:rs71526182	1	1	0,67	35,6956	78,8371	8,452776675
BNIP3L	SNP 5:rs17310286	2,5	2,5	SNP 5:rs17310286	1,09	0,94	2,659574468	139,6538	783,6419	0,178210609
				SNP 86:rs1042992	0,78					
CCND2	SNP 11:rs1049606	1	1	SNP 11:rs1049606	1,07	2,54	0,393700787	39,864	22,9748	1,735118478
				SNP 42:rs3217925	4					
DAZAP2	**SNP 25:kgp1918624**	1,67	1,67	SNP 43:rs1049467	1,43	1,43	1,167832168	104,3778	84,1193	1,240830582
GAS7	SNP 31:rs2240739	0,67	0,67	SNP 48:rs14383	0,25	0,88	0,761363636	12,6931	7,2272	1,756295661
				SNP 50:rs16958968	1,5					
LASP1	SNP 29:kgp8909974	1,33	1,33	SNP 53:rs11539845	0,38	0,38	3,5	72,976	39,5819	1,843670971
SLC4A1	SNP 4:rs5035	1,8	1,8	SNP 54:rs13306788	0,83	0,83	2,168674685	165,1352	726,2616	0,227377022
SH2D3C	SNP 17:rs514024	2,5	2,5	SNP 87:rs35019975	5	5	0,5	31,7735	5,1573	6,160878754
KZF1	SNP 21:kgp5028738	3	3	SNP 82:rs72645704	1,6	1,2	2,5	45,9371	28,2684	1,625033609
				SNP 83:rs72645705	0,8					
EF2AK1	SNP 22:rs2639	0,25	0,25	SNP 79:rs2640	0,8	0,75	0,333333333	69,0985	133,6252	0,517106803
				SNP 22:rs2639	0,71					
ANXA4	SNP 1:kgp9682337	2	2	SNP 61:rs2228203	0,91	0,91	2,197802198	12,1379	24,6486	0,4884742
CNPPD1	SNP 8:rs1043160	2	2	SNP 65:rs11236	1,1	1,04	1,920076923	26,7431	163,7143	8,163352255
				SNP 66:rs1127102	1,08					
				SNP 8:rs1043160	1,36					
				SNP 67:rs1127101	0,61					
UACA	SNP 16:rs934005	1,5	1,5	SNP 16:rs934005	1,6	0,9	1,666666667	12,9841	23,1721	0,580333332
				SNP 47:rs4776525	0,2					
				SNP 47:rs4776525	0,2					
BSG	SNP 10:rs2229664	1,1	1,1	SNP 56:rs4682	1,2	1,16	8,948275862	151,0878	867,8964	0,174084468
				SNP 57:rs8637	1,32					
				SNP 58:rs8358	0,96					
NEAT1	SNP 32:rs680413	1,67	1,67	SNP 32:rs8801413	1,12	2,81	0,59430605	83,9749	20,5134	4,093660729
				SNP 41:rs2238895	4,5					
BCL2	SNP 36:rs1564483	2	2	SNP 55:rs4987857	0,29	0,29	6,858551724	13,6149	11,001	1,237605672
MAX	SNP 18:rs4902359	0,75	0,75	SNP 44:rs4902357	0,62	0,69	1,006956522	186,4324	117,3962	1,588061624
				SNP 45:rs1957949	0,75					
				SNP 46:rs1957848	0,74					
				SNP 18:rs4902359	0,64					
FAM65B	SNP 12:rs12183109	0,5	0,5	SNP 77:rs11285	1	1	0,5	38,6914	39,9277	8,869036534
CAPN2	SNP 22:rs17598	0,33	0,27	SNP 38:rs3736383	0,92	0,77	0,350649351	59,5279	47,4986	1,250255885
	SNP 13:kgp4535923	0,2		SNP 6:rs17598	0,48					
				SNP 40:rs10961	0,93					
HLA-DPB1	SNP 27:kgp4559776	0,25	0,25	SNP 78:rs41558014	0,12	0,12	2,003333333	62,8314	90,6768	0,692915939
SMCR8	SNP 35:rs3829956	0,5	0,5	SNP 51:rs1563632	1	0,83	0,602409639	4,3801	5,1181	0,855805865
				SNP 52:rs921985	0,67					
ST6GAL1	SNP 19:rs2284749	1	1	SNP 18:rs2284749	1,27	2,04	0,490196070	15,0179	12,2028	1,230692956
				SNP 73:rs4886848	2					
				SNP 74:rs1801380	4,5					
				SNP 75:rs7559	0,4					
RAB8A	SNP 9:rs3745314	0,71	0,87	SNP 59:rs3745313	1,86	1,02	0,052341176	150,8067	67,5592	2,232215590
	SNP 20:rs17722795	1,33		SNP 9:rs3745314	0,92					
	SNP 30:rs10409364	0,44		SNP 20:rs17722795	1,17					
	SNP 37:rs1043346	1		SNP 37:rs1043346	0,38					
				SNP 60:rs1043407	0,78					
MYH9	SNP 28:rskgp7015684	1,25	1,25	SNP 68:rs2481	1,39	0,93	1,344086022	547,5822	125,4471	4,365044708
				SNP 68:rs7078	0,48					
				SNP 70:rs12107	0,47					
				SNP 71:rs11089787	1,38					
ARL6P5	SNP 33:rs7038	2	2	SNP 72:rs60050966	0,5	0,5	4	57,8305	77,8226	0,743104848
CTSB	SNP 34:kgp8442325	0,2	0,2	SNP 84:rs138489258	0,21	0,21	0,952380952	40,7539	124,8545	0,326411143
DAXX	SNP 23:rs1059231	4	4	SNP 23:rs1059231	1	1	4	33,1412	14,2532	2,328176101
C7orf41	SNP 15:rs2110394	3	3	SNP 80:rs702829	0,67	0,89	3,370786517	87,1367	86,8271	1,000565707
				SNP 81:rs3750089	1,11					
OPA1	SNP 14:kgp7301324	0,4	0,4	SNP 78:rs7643844	1,5	1,5	0,266666667	13,0804	34,8335	0,375512078
	SNP 24:kgp8804495	0,4								

FIG. 14

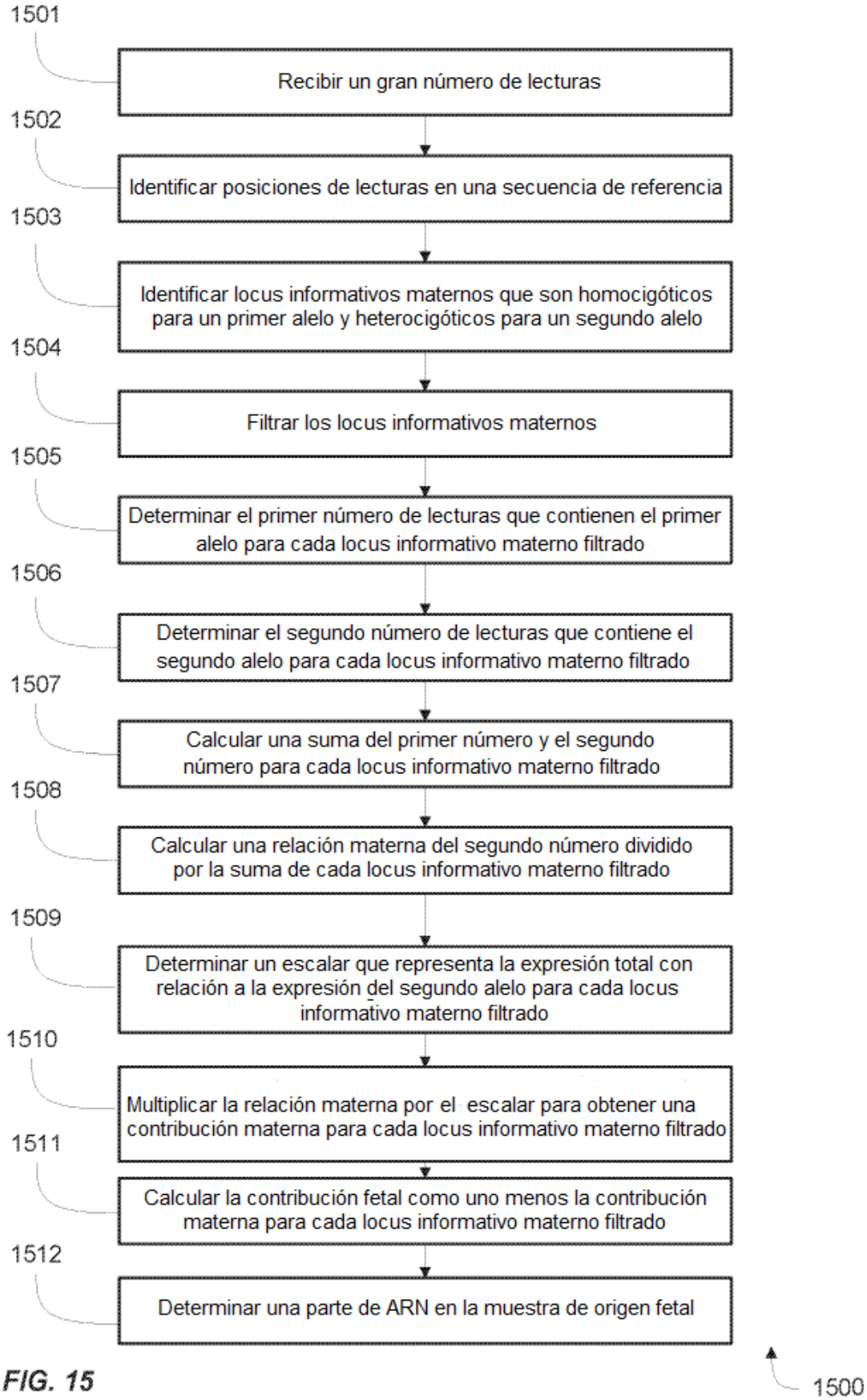
1501
Recibir un gran número de lecturas
1502
Identificar posiciones de lecturas en una secuencia de referencia
1503
Identificar locus informativos maternos que son homocigóticos para un primer alelo y heterocigóticos para un segundo alelo
1504
Filtrar los locus informativos maternos
1505
Determinar el primer número de lecturas que contienen el primer alelo para cada locus informativo materno filtrado
1506
Determinar el segundo número de lecturas que contiene el segundo alelo para cada locus informativo materno filtrado
1507
Calcular una suma del primer número y el segundo número para cada locus informativo materno filtrado
1508
Calcular una relación materna del segundo número dividido por la suma de cada locus informativo materno filtrado
1509
Determinar un escalar que representa la expresión total con relación a la expresión del segundo alelo para cada locus informativo materno filtrado
1510
Multiplicar la relación materna por el escalar para obtener una contribución materna para cada locus informativo materno filtrado
1511
Calcular la contribución fetal como uno menos la contribución materna para cada locus informativo materno filtrado
1512
Determinar una parte de ARN en la muestra de origen fetal
1500

FIG. 15

	Fase inicial del embarazo		Fase final del embarazo	
	nº de SNP	**nº de genes**	**nº de SNP**	**nº de genes**
Alelos de SNP específicos del feto (feto=AB, madre=AA)				
SNP/genes informativos totales	16.051	6.714	20.643	7.788
SNP/genes después del filtrado ASE	NA[b]		19.463	7.581
Transcritos en plasma materno	4.470	2.647 (39,43%)[c]	7.114	3.813 (48,96%)
Contribución fetal (total)	109	98 (3,70%)[d]	504	430 (11,28%)
Alta contribución fetal	25	24 (0,91%)[e]	112	96 (2,52%)
Alelos de SNP específicos de la madre (feto=AA, madre=AB)				
SNP/genes informativos totales	15.962	6.753	19.833	7.761
SNP/genes después del filtrado ASE	NA		19.343	7.656
Transcritos en plasma materno	4.414	2.654 (39,30%)	6.995	3.717 (47,89%)
Contribución materna (total)	3.156	2.041 (76,90%)	5.023	2.911 (78,32%)
Alta contribución materna	1.634	1.130 (42,58%)	3.002	1.895 (50,98%)

a. Número de genes que contenían SNP informativos.

b. La filtración de la expresión específica del alelo (ASE) no se realizó para muestras del inicio del embarazo.

c. % de genes informativos totales que estaban presentes en el plasma materno.

d. % de transcritos de plasma materno aportado por el feto.

e. % de transcritos de plasma materno muy aportado por el feto.

FIG. 16

	Fase inicial del embarazo	
	nº de SNP	**nº de genes[a]**
Alelos de SNP específicos del feto (feto=AB, madre=AA):		
SNP/genes informativos totales	20.643	7.788
Transcritos en plasma materno	7.303	3.839 (49,29%)[b]
Contribución fetal (total)	581	485 (12,63%)[c]
Alta contribución fetal	132	111 (02,89%)[d]
Alelos de SNP específicos de la madre (feto=AA, madre=AB):		
SNP/genes informativos totales	19.833	7.761
Transcritos en plasma materno	7.365	3.829 (49,34%)
Contribución materna (total)	5.326	3.014 (78,72%)
Alta contribución materna	3.321	1.986 (51´87%)
		1.130 (42,58%)

a. Número de genes que contenían SNP informativos.

b. % de genes informativos totales que estaba presente en el plasma materno.

c. % de transcritos de plasma materno aportado por el feto.

d. % de transcritos de plasma materno muy aportado por el feto.

FIG. 17

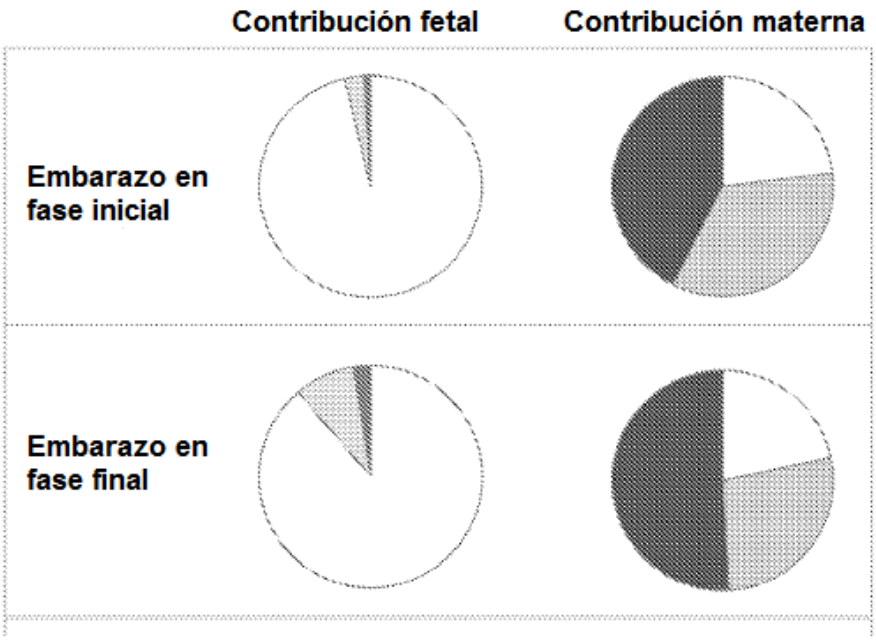

FIG. 18A

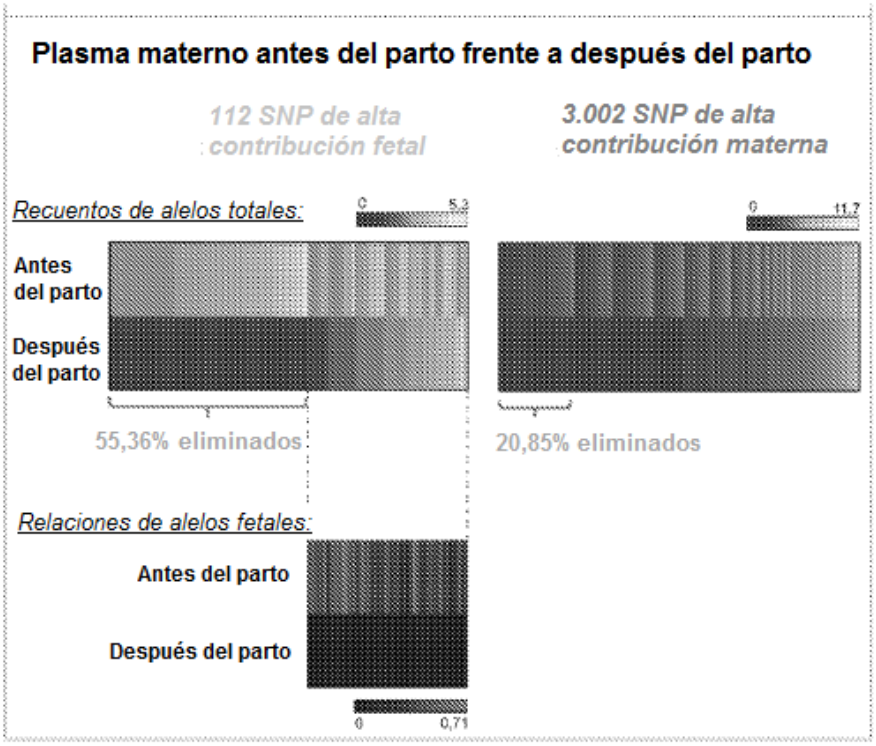

FIG. 18B

GNAS (locus de SNP: rs7121)	Genotipado	Recuento de alelos A (alelos compartidos)	Recuento de alelos B (específicos de la madre)	Contribución de fracción materna, M = 2*alelo materno/alelo materno+fetal	Contribución de fracción fetal, F = 1-M	FPKM
7171 (Referencia)	Madre: AB, Feto: AA	7	6	0,92	0,08	310,3

GNAS (locus de SNP: rs7121)	Genotipado	Recuento de alelos A (alelos compartidos)	Recuento de alelos B (específicos del feto)	Contribución de fracción fetal, M = 2*alelo fetal/alelo materno+fetal	Contribución de fracción materna, F = 1-M	FPKM	% cambio en contribución de fracción materna frente a referencia	% cambio en contribución de fracción fetal frente a referencia
5641 (PET)	Madre: AA, Feto: AB	81	4	0,09	0,91	376,2	(0,91-0,92)/(0,92)*100 = -1,09%	),91-,92)/(0,92)*100 = 12,5%

FIG. 19

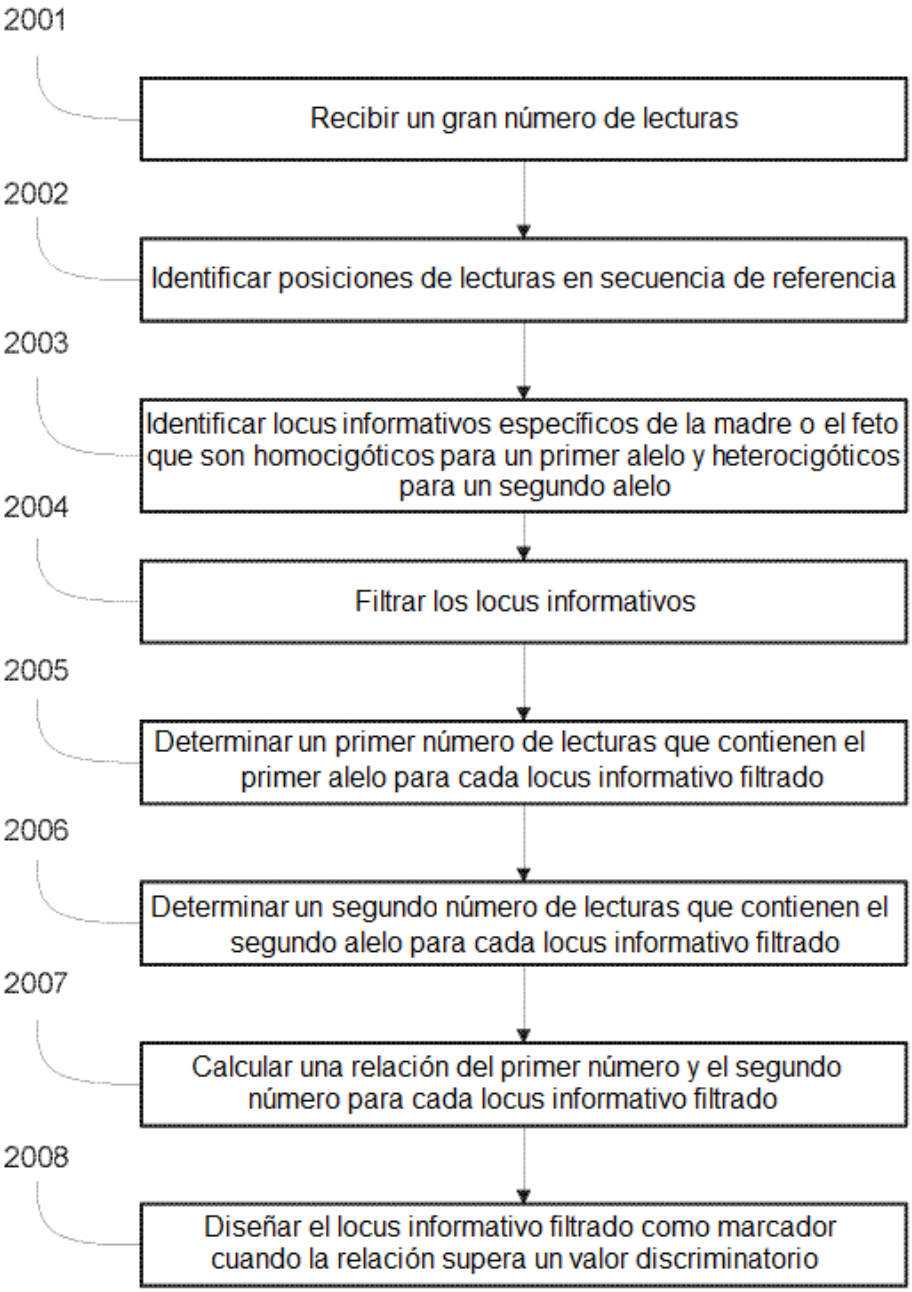
2001
Recibir un gran número de lecturas
2002
Identificar posiciones de lecturas en secuencia de referencia
2003
Identificar locus informativos específicos de la madre o el feto que son homocigóticos para un primer alelo y heterocigóticos para un segundo alelo
2004
Filtrar los locus informativos
2005
Determinar un primer número de lecturas que contienen el primer alelo para cada locus informativo filtrado
2006
Determinar un segundo número de lecturas que contienen el segundo alelo para cada locus informativo filtrado
2007
Calcular una relación del primer número y el segundo número para cada locus informativo filtrado
2008
Diseñar el locus informativo filtrado como marcador cuando la relación supera un valor discriminatorio
2000

FIG. 20

Gen	Caso 9356			Caso 9415			Notas (referencias)
	Plasma antes del parto (nivel de transcripción)	Plasma después del parto (nivel de transcripción)	Veces de cambio	Plasma antes del parto (nivel de transcripción)	Plasma después del parto (nivel de transcripción)	Veces de cambio	
CSH1	1391,00	17,66	78,78	796,21	1,13	704,89	*a* (1)
KISS1	433,89	8,25	52,56	17,36	7,26	20,29	*a* (2)
STAT1	227,34	104,49	2,18	275,22	85,30	3,23	*b*
CGA	270,02	6,98	38,67	134,19	1,00	134,19	
CSH2	136,04	3,89	34,96	185,78	2,74	67,84	
TFPI2	216,09	15,11	14,30	85,16	3,55	24,02	*a* (2)
GBP1	99,47	45,89	2,17	153,00	63,58	2,41	*b*
PLAC4	177,87	1,03	172,41	69,32	1,00	69,32	*a* (3)
HSD17B1	140,14	5,94	23,58	93,76	25,56	3,67	*b*
CSHL1	140,57	1,00	140,57	66,97	1,00	66,97	*a* (4)
KRT8	86,47	23,42	3,69	76,78	13,28	5,78	*c*
KRT18	86,76	25,92	3,35	64,11	20,27	3,16	*b*
HPGD	87,76	38,45	2,28	39,05	13,36	2,92	*d*
GADD45G	48,82	20,75	2,35	73,22	23,37	3,13	*b*
LGALS14	56,55	1,00	56,55	62,86	1,00	62,86	
HSD3B1	80,62	3,51	17,28	48,26	1,00	48,26	
KRT19	77,95	24,04	3,24	30,71	3,52	8,71	
SERPINE1	37,39	15,29	2,45	60,90	12,71	4,79	*a* (5)
GDF15	62,32	3,68	16,92	33,52	12,10	2,77	
CYP19A1	70,84	4,90	14,47	18,65	1,00	18,65	
PSG4	36,10	1,45	24,89	47,82	1,00	47,82	
PSG3	53,00	1,00	53,00	23,03	1,00	23,03	
CRYAB	15,46	1,00	15,46	58,22	4,50	12,93	
HSPB8	35,59	6,62	5,37	36,79	13,47	2,73	
PKIB	45,08	8,79	5,13	26,12	8,09	3,23	
PGF	44,72	5,40	8,29	24,94	3,80	6,57	
CYP11A1	40,77	3,64	12,20	28,04	1,71	16,41	
PSG5	42,56	1,00	42,56	25,66	1,00	25,66	
PSG1	37,70	1,49	25,29	28,17	1,00	28,17	
PAPPA	40,90	5,91	8,92	20,03	1,00	20,03	*a* (6)
ADAM12	44,48	4,01	11,09	15,16	1,00	15,16	*a* (7)
PSG2	29,16	1,00	29,16	29,95	1,00	29,95	
VGLL3	34,80	1,98	17,54	20,36	1,80	11,34	
KRT7	27,20	10,39	2,62	26,54	9,75	2,72	
TMEM54	32,99	5,94	5,56	18,70	5,22	3,58	
PRKCZ	28,22	10,05	2,81	23,12	8,76	2,64	
SDC1	30,62	2,62	11,68	17,43	1,84	9,45	
LGALS13	32,32	1,00	32,32	15,60	1,00	15,60	
EFHD1	15,55	2,80	5,56	30,55	3,94	7,76	
CAPN6	23,18	1,64	14,14	22,90	1,00	22,90	
XAGE3	25,08	2,59	9,67	13,08	1,00	13,08	
EBI3	18,79	6,48	2,90	17,96	7,30	2,46	
GH2	19,16	1,00	19,16	17,22	1,00	17,22	*a* (7)
PAGE4	28,89	2,99	9,67	6,28	1,00	6,28	

FIG. 21

Gen	Caso 9356			Caso 9415			Notas (referencias)
	Plasma antes del parto (nivel de transcripción)	Plasma después del parto (nivel de transcripción)	Veces de cambio	Plasma antes del parto (nivel de transcripción)	Plasma después del parto (nivel de transcripción)	Veces de cambio	
ALPP	19,62	1,54	12,73	14,03	1,00	14,03	
INHBA	18,99	2,71	7,01	14,51	4,29	3,38	*a* (6)
LOC100505659	17,53	1,00	17,53	15,65	1,00	15,65	
FBLN1	10,12	2,12	4,76	20,01	1,28	15,62	
SEMA3B	16,06	2,52	6,38	12,90	1,00	12,90	
GPC3	18,22	1,00	18,22	10,12	1,00	10,12	
PLAC2	19,86	1,99	9,96	8,04	1,97	4,09	
PSG9	7,10	1,00	7,10	20,32	1,00	20,32	*a* (8)
FN1	17,30	6,57	2,63	9,72	3,81	2,55	
NOS3	10,01	3,45	2,90	16,07	7,08	2,27	
LOC100506655	12,29	1,00	12,29	13,53	1,00	13,53	
PSG11	19,48	1,83	10,63	5,77	1,00	5,77	
SPTLC3	19,82	1,00	19,82	4,99	1,00	4,99	
EXPH5	12,58	4,48	2,81	12,09	2,24	5,40	
HSPA2	12,86	5,85	2,20	11,40	4,87	2,34	
PSG6	10,66	1,00	10,66	13,38	1,00	13,38	
PLAC1	17,08	1,00	17,08	6,60	1,00	6,60	*a* (3)
TACC2	5,93	1,61	3,69	16,24	1,64	9,89	
PRPF40B	8,27	3,98	2,08	12,53	5,32	2,36	
LOC388948	12,97	1,00	12,97	7,71	1,00	7,71	
SERPINB2	13,48	1,36	9,91	6,86	1,44	4,77	*a* (9)
CRH	15,40	1,00	15,40	4,75	1,00	4,75	*a* (10)
GBP1P1	6,35	1,19	5,32	13,04	5,04	2,59	
CLDN4	12,50	1,00	12,50	6,45	2,84	2,27	
C2orf72	11,14	3,29	3,38	7,26	1,74	4,18	
PAPPA2	10,92	1,53	7,13	7,48	1,00	7,48	
HIST2H3A	14,75	7,27	2,03	2,44	1,02	2,39	
HIST2H3C	14,75	7,27	2,03	2,44	1,02	2,39	
TCL6	12,92	1,00	12,92	3,93	1,00	3,93	
MFSD2A	8,09	3,35	2,42	8,72	1,00	8,72	
ZFAT-AS1	9,10	3,14	2,90	7,25	3,31	2,19	
INSL4	13,90	2,40	5,80	2,01	1,00	2,01	
DHRS2	12,24	6,11	2,00	3,30	1,00	3,30	
HES2	11,03	3,11	3,54	4,50	1,46	3,08	
WLS	11,06	4,77	2,32	4,40	1,68	2,63	
PLCXD3	10,49	3,62	2,90	4,88	2,01	2,43	
LOC100505483	5,40	1,00	5,40	9,39	2,62	3,58	
C1orf130	11,60	1,85	6,28	2,48	1,00	2,48	
FOSB	3,02	1,00	3,02	10,65	4,94	2,16	
GCM1	5,93	1,60	3,71	6,94	1,00	6,94	*a* (8)
TRPV6	9,31	1,00	9,31	3,41	1,00	3,41	
TFAP2A	4,02	1,00	4,02	8,59	1,46	5,87	
MMP11	10,08	1,96	5,16	2,46	1,00	2,46	
TUSC3	9,90	2,66	3,73	2,55	1,00	2,55	
HMGCR	7,15	2,25	3,18	5,13	2,26	2,27	

FIG. 21 cont.

Gen	Caso 9356 Plasma antes del parto (nivel de transcripción)	Caso 9356 Plasma después del parto (nivel de transcripción)	Caso 9356 Veces de cambio	Caso 9415 Plasma antes del parto (nivel de transcripción)	Caso 9415 Plasma después del parto (nivel de transcripción)	Caso 9415 Veces de cambio	Notas (referencias)
CCK	7,06	1,00	7,06	4,98	1,00	4,98	
LOC100129935	8,56	1,61	5,32	3,38	1,00	3,38	
C8orf39	4,02	1,00	4,02	7,86	2,93	2,69	
GLDN	9,39	1,00	9,39	2,44	1,00	2,44	
PGAP3	6,10	2,19	2,78	4,84	2,32	2,09	
MSX2P1	3,63	1,00	3,63	6,87	3,36	2,05	
GH1	4,33	1,00	4,33	6,02	1,00	6,02	
SVEP1	7,79	2,98	2,61	2,26	1,00	2,26	
PPP1R32	2,78	1,00	2,78	7,24	1,35	5,37	
PSG8	5,01	1,00	5,01	4,94	1,00	4,94	
ENDOU	7,24	2,40	3,02	2,27	1,00	2,27	
EGFR	6,83	1,00	6,83	2,54	1,00	2,54	
DUSP4	5,24	1,61	3,26	4,13	1,41	2,92	
PHYHIPL	5,53	1,00	5,53	3,48	1,39	2,51	
CTSF	6,01	1,00	6,01	2,77	1,03	2,69	
TRIM29	3,54	1,00	3,54	5,13	1,00	5,13	
RCN3	3,07	1,00	3,07	5,34	1,12	4,77	
SPIRE2	5,93	2,27	2,61	2,48	1,00	2,48	
LOC100216001	4,60	1,06	4,35	3,55	1,49	2,39	
FAM176A	5,76	2,10	2,74	2,36	1,00	2,36	
SCIN	5,18	1,00	5,18	2,82	1,00	2,82	
ZNF500	2,41	1,00	2,41	5,52	2,23	2,47	
PRR16	2,83	1,00	2,83	4,91	1,18	4,18	
LOC100128054	4,55	1,05	4,35	3,08	1,00	3,08	
OLR1	4,53	1,17	3,87	2,71	1,00	2,71	
SLC30A2	5,18	1,00	5,18	2,00	1,00	2,00	
LOC285972	3,49	1,00	3,49	3,64	1,02	3,58	
HESX1	4,25	1,26	3,38	2,64	1,00	2,64	
TMEM139	4,61	1,00	4,61	2,25	1,00	2,25	
ZNF727	3,82	1,00	3,82	2,58	1,24	2,09	
TM4SF19	4,04	1,39	2,90	2,34	1,00	2,34	
EFS	3,88	1,00	3,88	2,38	1,00	2,38	
TIMD4	2,08	1,00	2,08	3,62	1,51	2,39	
ALDH3B2	2,91	1,00	2,91	2,76	1,00	2,76	
KRT81	2,61	1,00	2,61	2,92	1,00	2,92	
MUC15	2,94	1,00	2,94	2,19	1,00	2,19	
PRSS8	2,20	1,00	2,20	2,87	1,00	2,87	
SH2D5	2,41	1,00	2,41	2,42	1,00	2,42	
LOC728175	2,28	1,00	2,28	2,31	1,10	2,09	
GRHL2	2,45	1,00	2,45	2,13	1,00	2,13	
PABPNIL	2,18	1,00	2,18	2,37	1,00	2,37	
CORO6	2,08	1,00	2,08	2,07	1,00	2,07	

Notas:

a. Transcritos anteriormente descritos por estar relacionados con el embarazo en el plasma materno

b. Transcritos recién identificados en este estudio por RNA-seq del plasma materno y validados por RT-PCR en tiempo real (Datos complementarios Fig. 6)

c. ARNm de *KRT8* no se estudió por RT-PCR ya que no pudo diseñarse el ensayo de detección

d. El resultado de RT-PCR en tiempo real demuestra que el ARNm de *HPGD* no presentaba patrón de eliminación tras el parto (Datos complementarios Fig. 8)

Referencias:

1. Ng EKO, Tsui NBY, Lau TK, Leung TN, Chiu RWK, Panesar NS, *et al. Proc. Natl. Acad. Sci. U S A* 2003; 100: 4748-53.
2. Tsui NBY, Chim SSC, Chiu RWK, Lau TK, Ng EKO, Leung TN, *et al*. *J. Med. Genet*. 2004;41: 461-7.
3. Lo YMD, Tsui NBY, Chiu RWK, Lau TK, Leung TN, Heung MMS, *et al*. *Nat. Med*. 2007; 13: 218-23.
4. Heung MMS, Jin S, Tsui NBY, Ding C, Leung TY, Lau TK, *et al. PLoS One* 2009; 4:e5858.
5. Purwosunu Y, Sekizawa A, Koide K, Farina A, Wibowo N, Wiknjosastro GH, *et al*. *Clin. Chem.* 2007;53 :399-404.
6. Wong BCK, Chiu RWK, Tsui NBY, Chan KCA, Chan LW, Lau TK, *et al*. *Clin. Chem*. 2005;51 : 1786-95.
7. Pang WW, Tsui MH, Sahota D, Leung TY, Lau TK, Lo YM, Chiu RW. *Prenat. Diagn*. 2009:29: 495-504.
8. Go AT, Visser A, Mulders MA, Blankenstein MA, Van Vugt JM, Oudejans CB. *Clin. Chem*. 2004;50:1413-4.
9. Tsui NBY, Wong BCK, Leung TY, Lau TK, Chiu RWK, Lo YMD. *Prenat. Diagn*. 2009; 29: 1031-7.
10. Ng EKO, Leung TN, Tsui NBY, Lau TK, Panesar NS, Chiu RWK, Lo YMD. *Clin. Chem*. 2003;49: 727-31.

FIG. 21 cont.

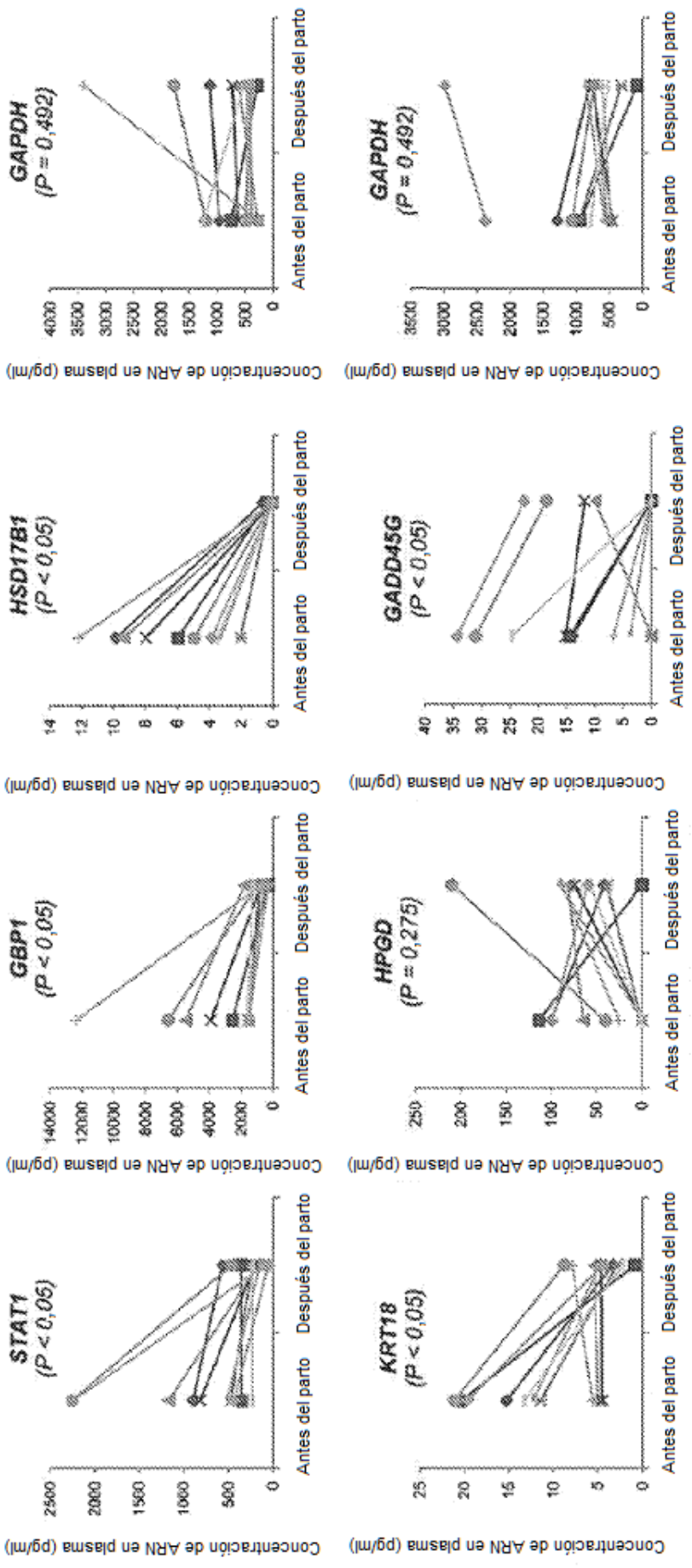
GAPDH
(P = 0,492)
Concentración de ARN en plasma (pg/ml)
Antes del parto
Después del parto
HSD17B1
(P < 0,05)
GADD45G
(P < 0,05)
GBP1
(P < 0,05)
HPGD
(P = 0,275)
STAT1
(P < 0,05)
KRT18
(P < 0,05)

FIG. 22

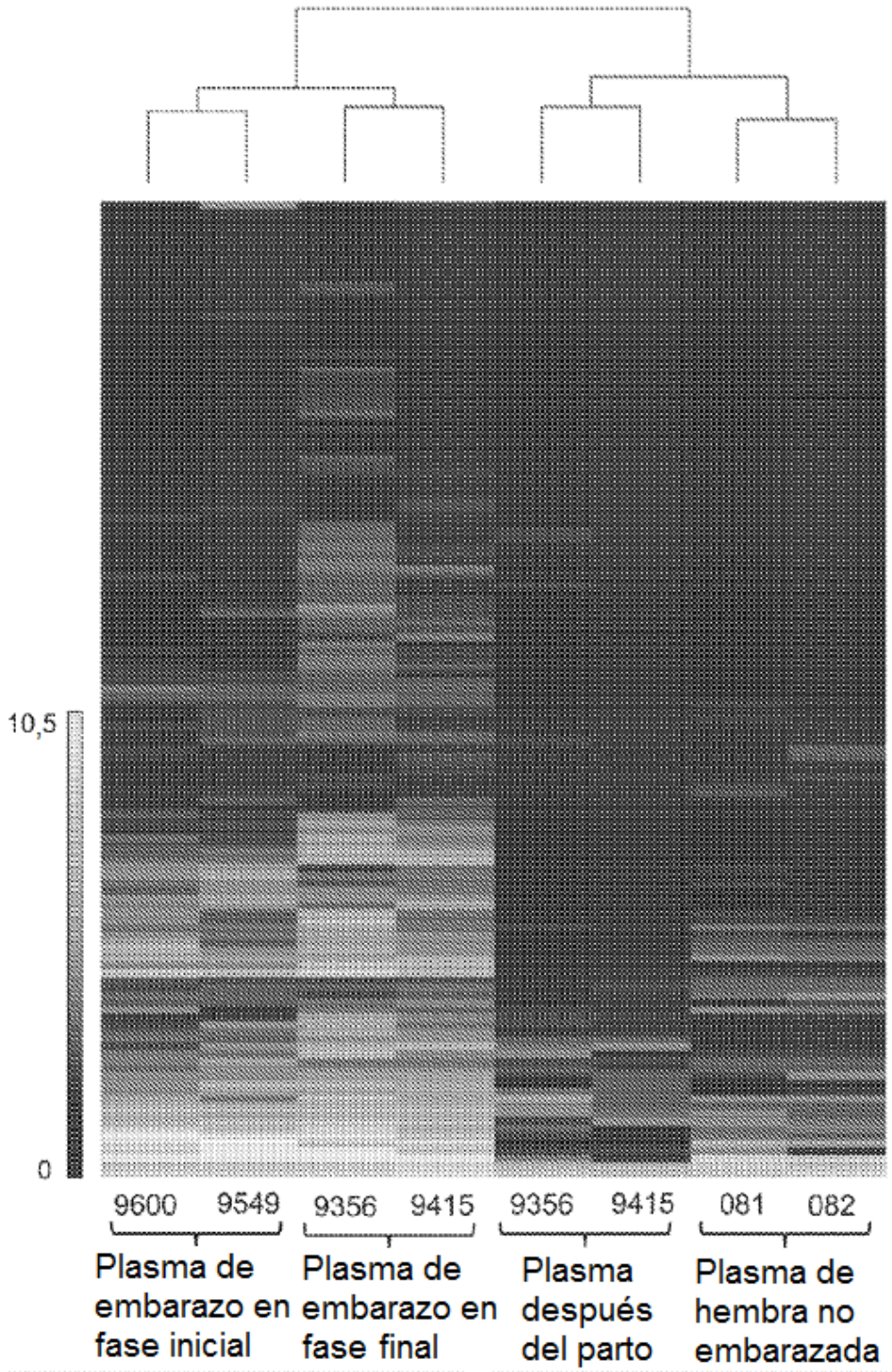
10,5
0
9600
9549
9356
9415
9356
9415
081
082
Plasma de embarazo en fase inicial
Plasma de embarazo en fase final
Plasma después del parto
Plasma de hembra no embarazada

FIG. 23

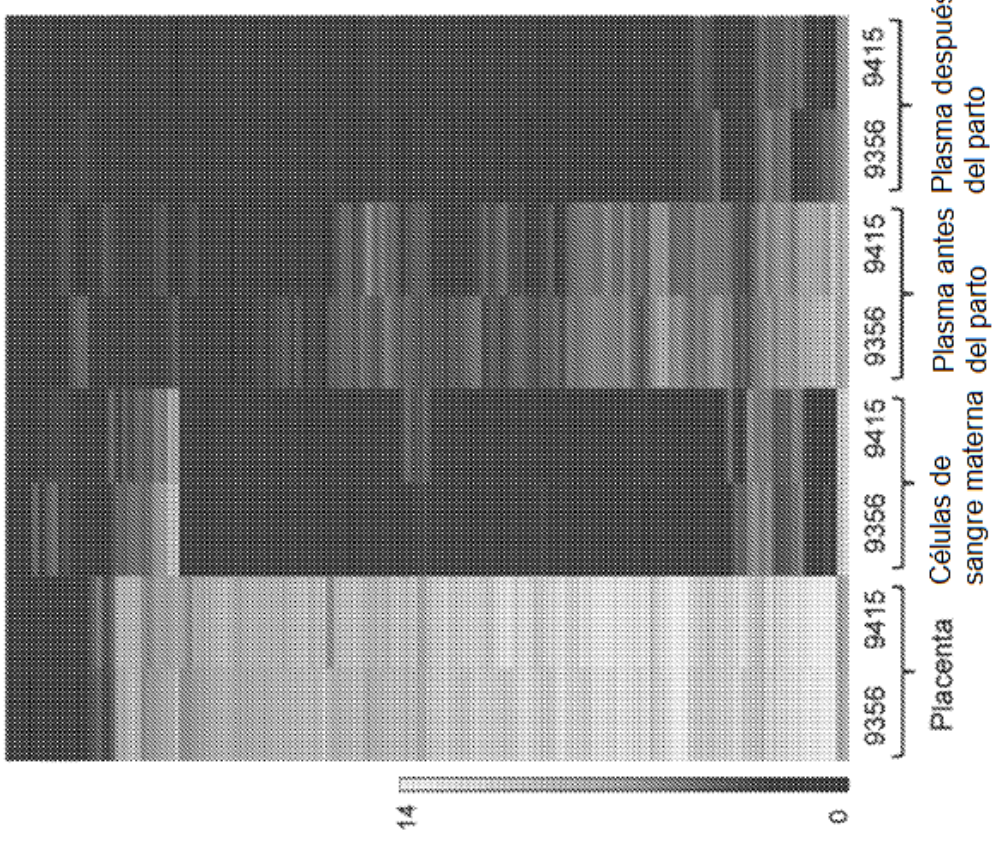

FIG. 24A

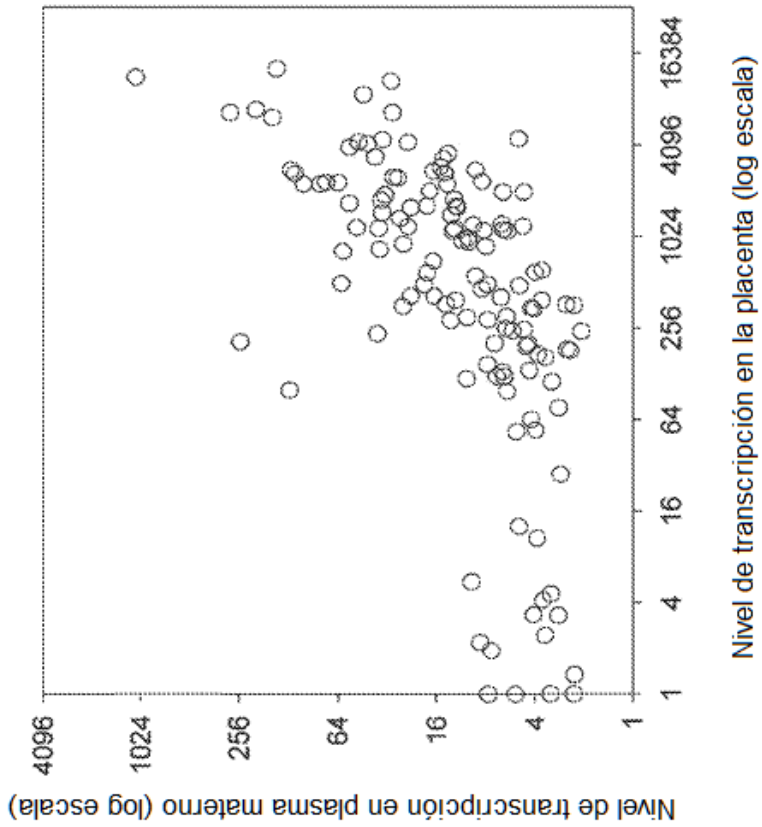

FIG. 24B

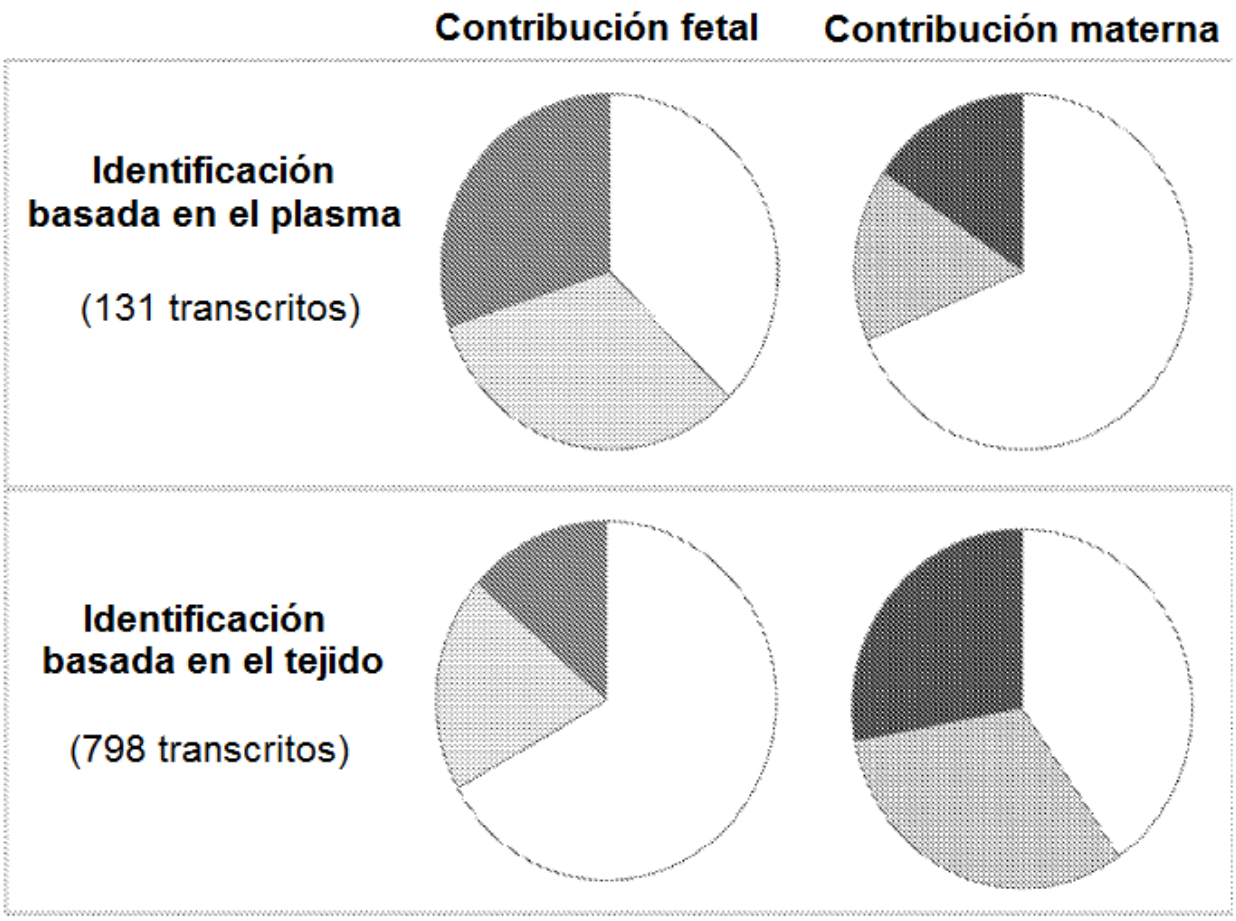
Contribución fetal
Contribución materna
Identificación basada en el plasma
(131 transcritos)
Identificación basada en el tejido
(798 transcritos)

Fig. 25

ARN	Preeclampsia	Normal	ARN	Preeclampsia	Normal
ADAM9	26,78	12,26	LY6E	84,52	31,38
AGBL5	144,47	65,14	MEIS1	79,49	18,77
APOBR	64,12	26,31	MLH3	150,61	36,10
APOL2	59,24	29,05	MYADM	51,11	16,60
APP	212,58	84,93	NCF1	107,35	41,90
ASAH1	167,14	47,38	NCF1B	34,80	9,41
ATP5I	23,85	9,20	NDUFA1	20,73	7,36
ATP6V0E1	60,73	20,56	NES	176,78	58,83
B2M	1765,75	654,19	NFAM1	49,17	17,91
BCAP31	42,85	14,86	NGFRAP1	106,72	19,52
BRK1	91,46	41,63	NPTN	80,63	29,97
C12orf76	16,97	4,56	NT5C3	173,39	65,34
C19orf59	22,94	6,42	OST4	404,86	100,63
C19orf79	13,22	3,00	PARK7	97,39	40,79
Corf151-NBL1	25,89	11,19	PARP10	52,41	19,75
C21orf7	308,80	83,41	PDGFA	55,19	24,29
C7orf53	8,04	0,73	PF4	680,43	208,45
CARD16	85,90	29,13	PGRMC1	332,76	78,90
CD97	59,09	28,74	PRIC285	72,34	20,44
CEBPD	147,37	44,30	PSMB9	73,17	34,36
CTSA	341,67	138,37	RAB32	51,90	25,20
CYB5R1	25,14	10,12	RABAC1	25,10	7,80
DDX11L10	38,97	7,34	RBX1	36,44	15,06
DUSP1	168,06	30,57	RNF213	316,69	147,00
DYNLRB1	166,30	59,63	RPPH1	242,37	63,73
EMP3	99,80	39,75	S100P	48,92	14,31
ENKUR	64,87	15,45	S1PR3	9,45	1,41
GALM	31,68	5,09	SAT1	213,55	100,86
GDI1	91,22	41,05	SEPT3	14,54	6,35
HIST1H2AC	271,60	72,77	SERF2	310,66	144,79
HIST1H3H	131,91	40,96	SHISA5	141,53	59,85
HIST1H4A	20,40	5,53	SIGLEC1	18,55	3,32
HIST1H4B	25,44	5,58	SIGLEC14	10,79	0,95
HIST1H4E	74,16	26,59	SNN	348,09	171,73
HIST1H4H	69,35	18,31	SOD2	569,41	229,85
HIST1H4L	20,99	6,89	SPARC	914,92	286,17
HLA-C	417,81	203,10	TIMP1	76,27	27,68
HLA-L	11,78	3,90	TMEM140	121,93	29,08
HRC	14,87	0,16	TMEM185A	36,93	12,09
HSPA1A	37,29	14,90	TMEM50A	139,68	32,72
IFI6	171,27	18,29	TRIM22	135,78	58,79
ISG20	86,19	23,09	TSC22D3	330,23	108,27
JAM3	121,57	52,31	UBE2LR	171,88	63,40
KLF6	127,12	62,64	VKORC1L1	63,06	25,17
LEPR	43,19	17,71	VTRNA1-1	7,23	0,08
LEPROT	70,13	34,38	YIF1B	36,33	16,96
LILRA5	31,45	11,05	YPEL3	150,75	73,23
LOC146336	13,75	4,89	YWHAH	236,85	116,18
LRRC32	102,05	34,23	ZNF485	15,35	6,38

FIG. 26

PAPPA SNP rsID	Genotipo de ADN		Placenta		Células sanguíneas maternas		Plasma materno antes del parto		Plasma materno antes del parto	
	Feto	Madre	A	G	A	G	A	G	A	G
rs386088	GA	AA	2421	2164	0	0	16	21	0	0

FIG. 27

H19 SNP rsID	Genotipo de ADN		Placenta		Celulas sanguíneas maternas		Plasma materno antes del parto		Plasma materno después del parto	
	Feto	Madre	Compartido	Específico de la madre	Compartidos	Específicos de la madre	Compartido	Específico de la madre	Compartido	Específico de la madre
rs2839698[a]	AA	AG	3418 (A)	10 (G)	0 (A)	0 (G)	246 (A)	1679 (G)	0 (A)	808 (G)
rs2839701	GG	CG	4959 (G)	8 (C)	0 (G)	0 (C)	33 (G)	219 (C)	1 (G)	78 (C)
rs2839702	CC	AC	4996 (C)	8 (A)	0 (C)	0 (A)	38 (C)	233 (A)	2 (C)	92 (A)
rs3741219	GG	AG	2255 (G)	15 (A)	0 (G)	0 (A)	17 (G)	158 (A)	0 (G)	96 (A)

a: El alelo A compartido se imprimió mientras que el alelo G específico de la madre se transcribió. Detalles del estado de metilación se muestran en los Datos complementarios de la Fig. 5.

FIG. 28

ES 2 670 544 T3

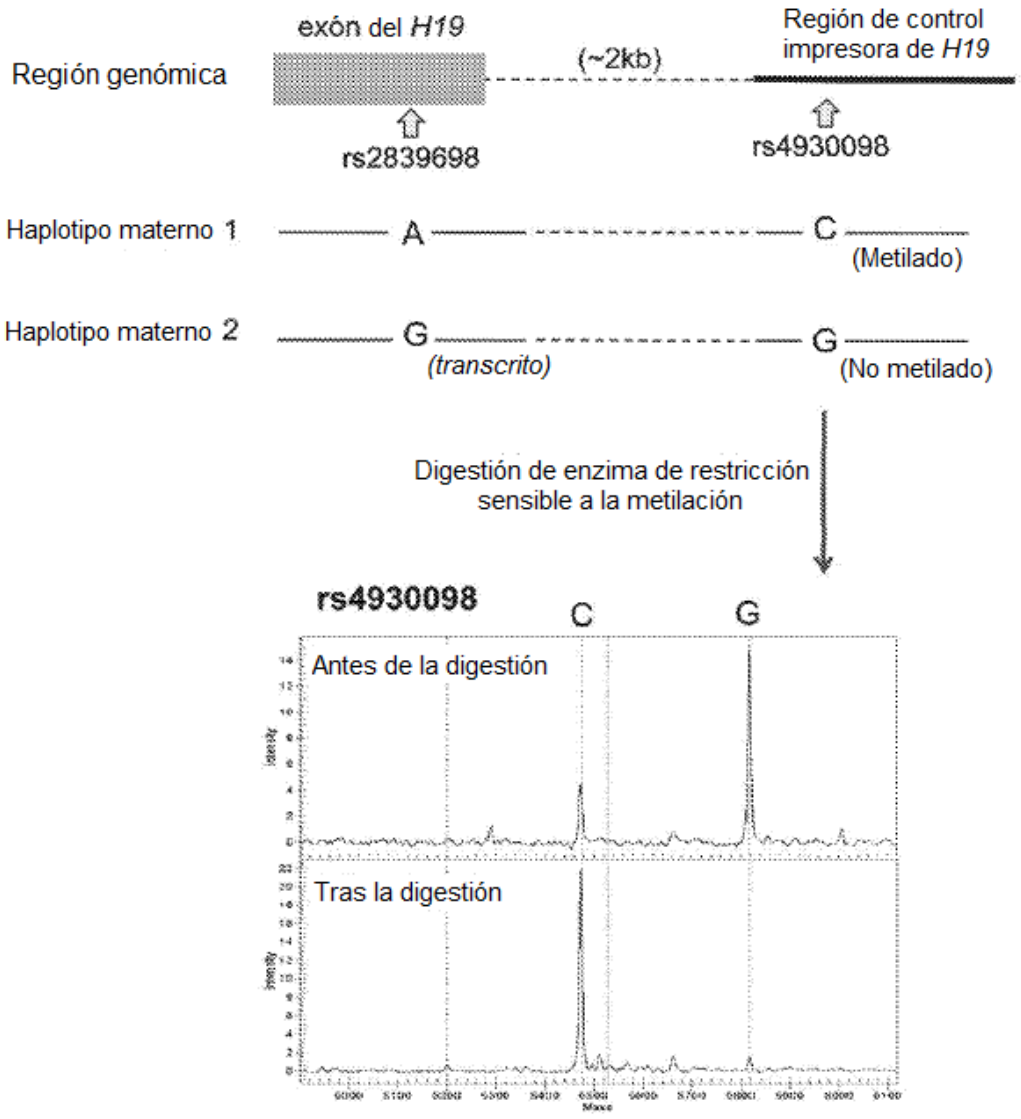
Región genómica
exón del H19
(~2kb)
Región de control impresora de H19
rs2839698
rs4930098
Haplotipo materno 1
A
C
(Metilado)
Haplotipo materno 2
G
(transcrito)
G
(No metilado)
Digestión de enzima de restricción sensible a la metilación
rs4930098
C
G
Antes de la digestión
Tras la digestión

FIG. 29

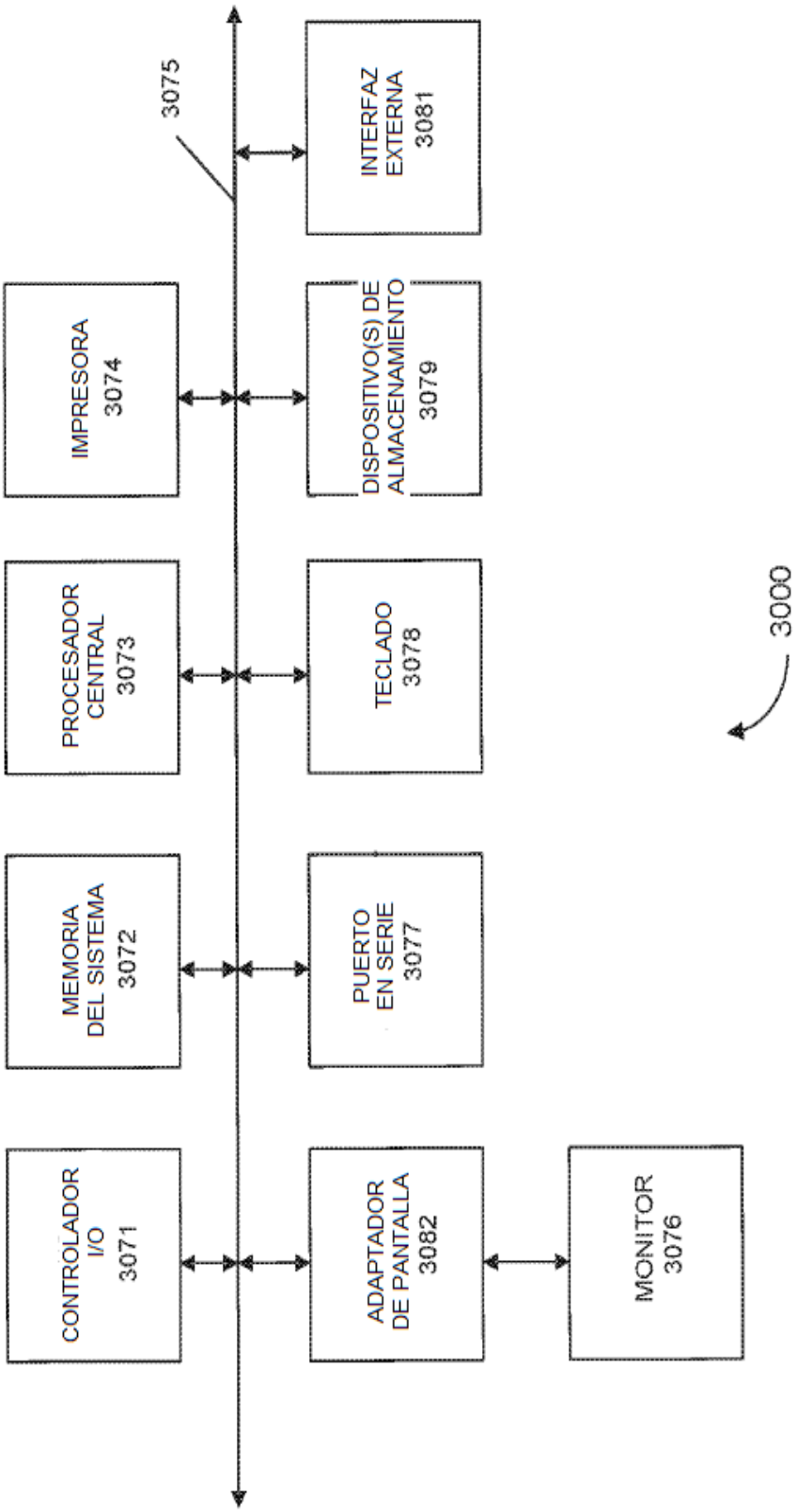
3000
3075
CONTROLADOR I/O 3071
MEMORIA DEL SISTEMA 3072
PROCESADOR CENTRAL 3073
IMPRESORA 3074
ADAPTADOR DE PANTALLA 3082
PUERTO EN SERIE 3077
TECLADO 3078
DISPOSITIVO(S) DE ALMACENAMIENTO 3079
MONITOR 3076
INTERFAZ EXTERNA 3081

FIG. 30