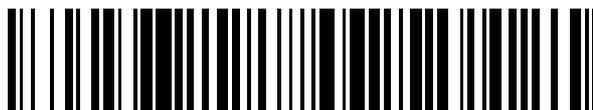


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 711 814**

51 Int. Cl.:

G01N 33/574 (2006.01)

G06F 19/00 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **23.02.2016 PCT/EP2016/053726**

87 Fecha y número de publicación internacional: **01.09.2016 WO16135119**

96 Fecha de presentación y número de la solicitud europea: **23.02.2016 E 16709979 (5)**

97 Fecha y número de publicación de la concesión europea: **26.12.2018 EP 3262416**

54 Título: **Método para el diagnóstico de carcinoma endometrial**

30 Prioridad:

27.02.2015 IT UB20157151

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

07.05.2019

73 Titular/es:

**HOSMOTIC SRL (100.0%)
Via Raffaele Bosco 178
80069 Vico Equense (NA), IT**

72 Inventor/es:

**TROISI, JACOPO;
SCALA, GIOVANNI;
CAMPIGLIA, PIETRO;
ZULLO, FULVIO y
GUIDA, MAURIZIO**

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 711 814 T3

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

DESCRIPCIÓN

Método para el diagnóstico de carcinoma endometrial

5 La presente invención se refiere a un método para el diagnóstico de carcinoma endometrial basado en el análisis metabólico de la sangre y la manipulación bioinformática de los perfiles metabólicos a través de modelos de clasificación.

10 El carcinoma endometrial es el cáncer invasivo más común del tracto genital femenino y es el responsable del 7 % de todos los tumores invasivos en las mujeres (excluyendo los tumores cutáneos).

15 El carcinoma endometrial es raro en mujeres que tienen menos de 40 años. El pico de incidencia es entre 55 y 65 años. Los estudios clínico-patológicos y el análisis molecular apoyan la clasificación del carcinoma endometrial en dos amplias categorías: Tipo I y Tipo II.

20 El tipo I es el más frecuente, con un porcentaje de casos superior al 80 %, este invade las glándulas endometriales proliferativas y se define con el término carcinoma endometriode. En general, surge en un marco de hiperplasia endometrial y, como este, se asocia con obesidad, diabetes, hipertensión, infertilidad y estimulación estrogénica sin oposición. Estudios recientes proporcionan evidencia adicional que apoya la tesis de que la hiperplasia endometrial es un precursor del carcinoma endometrial (Muller GL y otros. Allelotype mapping of unstable microsatellites establishes direct lineage continuity between endometrial precancers and cancers. *Cancers Res* 56:4483, 1996). El carcinoma endometrial tipo II generalmente afecta a las mujeres diez años más tarde que el carcinoma endometrial tipo I (65-75 años) y, a diferencia del tipo I, se desarrolla principalmente en un marco de atrofia endometrial. El tipo II representa menos del 15 % de los casos de carcinoma endometrial y se diferencia escasamente (G3). El subtipo más común es el seroso, que se define así debido a la superposición biológica y morfológica con el carcinoma de ovario. Los subtipos histológicos menos comunes también pertenecen a esta categoría: el carcinoma de células claras y el tumor Mulleriano mixto maligno.

30 Por el momento, no es factible realizar una detección masiva en una población asintomática en edad perimenopáusic y postmenopáusic para el diagnóstico precoz del carcinoma endometrial, tal como se realiza para el carcinoma cervical a través de la prueba de Papanicolaou.

35 Los estudios realizados en una muestra exocervical demuestran una frecuencia de falsos negativos de aproximadamente 40-50 % desde que las células exfoliadas endometriales, que han sufrido la acción del entorno vaginal, presentan alteraciones y, por lo tanto, pierden las características que permiten la diferenciación de la célula tumoral a partir de la célula normal. Además, el pronóstico se relaciona estrictamente con el diagnóstico temprano, de hecho, la supervivencia después de 5 años disminuye drásticamente del 78-98 % en caso de diagnóstico en la etapa I hasta el 3-10 % en caso de diagnóstico en la etapa IV.

40 Hasta la fecha, se han identificado varios miles de metabolitos del suero humano y la aplicación de la metabolómica ha permitido el desarrollo de biomarcadores para muchas enfermedades como la esquizofrenia (Kaddurah-Daouk R., *Metabolic profiling of patients with schizophrenia*, *PLOS Med* 2006; 8:e363), la meningitis (Subramanian A. y otros, *Proton MR/CSF analysis and a new software as predictors for the differentiation of meningitis in children*, *NMR Biomed* 2005; 18:213-25) y el cáncer de colon (Denkert C. y otros, *Metabolite profiling of human colon carcinoma - deregulation of TCA cycle and amino acid turnover*, *Mol. Cancer* 2008; 7:1-15). Sin embargo, el uso de la metabolómica en el campo ginecológico se ha limitado hasta ahora a estudios relacionados al carcinoma de ovario (Fan L. y otros. *Identification of metabolic biomarkers to diagnose epithelial ovarian cancer using a UPLC/QTOF/MS platform* *Acta Oncologica*, 2012; 51:473-479). Hasta la fecha, no hay estudios publicados en la literatura realizados en cromatografía de gases acoplada a espectrometría de masas y con técnicas quimiométricas para el diagnóstico del carcinoma endometrial.

50 Por lo tanto, es muy necesario un sistema de diagnóstico no invasivo que permita realizar una detección en la población con riesgo de edad o de factores de riesgo conocidos, para identificar de manera temprana esta temible neoplasia femenina.

55 Ventajosamente, la presente invención resuelve los problemas mencionados anteriormente mediante un método no invasivo para el diagnóstico del carcinoma endometrial. Hasta hoy, no hay otros métodos de diagnóstico no invasivos que permitan una distinción histológica de este tipo de tumor.

El objeto de la invención se explicará de aquí en adelante en detalle.

60 Breve descripción de las figuras

La Figura 1 muestra el resultado del análisis OPLS-DA basado en datos del perfil metabólico de los pacientes con carcinoma endometrial y de controles sanos.

65

Las gráficas de puntuación discriminan entre las dos clases sin superposiciones. Los triángulos representan a los pacientes afectados por carcinoma endometrial, mientras que los anillos pequeños a los pacientes sanos. Los principales componentes PC1 y PC2 reportados en los ejes respectivamente describen el 16,5 % y el 14,9 % de la varianza global.

5 La Figura 2 muestra, de acuerdo con la invención, la clasificación histológica (carcinoma de tipo I versus carcinoma de tipo II) obtenida con el modelo PLS-DA. Las manchas representan los perfiles metabolómicos de mujeres con carcinoma endometrial de tipo I, mientras que los triángulos a los de los pacientes con carcinoma endometrial de tipo II. El modelo coloca solo una de estas muestras en un área que no es atribuible unívocamente al área correcta.

10 Definiciones

El término "metabolómica", se refiere al análisis de los procesos celulares mediante el estudio del perfil metabolómico de moléculas pequeñas de un organismo.

15 Con el término "análisis metabolómico", los inventores se refieren a la realización de un proceso dirigido a la identificación y la determinación de la concentración del mayor número posible de metabolitos en una muestra biológica.

20 El término "metabolitos" se refiere a las pequeñas moléculas derivadas de los procesos biológicos de tipo anabólico o catabólico de una célula o de un conjunto de células. Con el término "metabolitos", los inventores se refieren a todas las moléculas que tienen un peso molecular inferior a 1000 Dalton, que son potencialmente identificables y medibles dentro de una muestra biológica.

25 El término "perfil metabolómico" se refiere el patrón específico que los metabolitos tienen en la sangre del paciente en dependencia de sus proporciones relativas.

30 El PLS-DA (Análisis Discriminante de Mínimos Cuadrados Parciales) es un método supervisado que usa técnicas de regresión multivariante para extraer a través de combinaciones lineales de las variables originales (X), la información que puede predecir la pertinencia de una determinada clase (Y). Para evaluar la efectividad en la discriminación de las clases, se realiza una prueba de permutación. En cada permutación, se construye un modelo PLS-DA a partir de los datos (X) y las etiquetas de clase conmutadas (Y) mediante el uso de los números óptimos de los componentes determinados por validación cruzada para el modelo basado en la asignación de las clases originales. Se realizan dos tipos de pruebas estadísticas para medir el poder de discriminación entre las clases. La primera se basa en la precisión de la predicción en la fase de capacitación del modelo. La segunda se basa en la distancia de separación de acuerdo con la relación entre la suma de las distancias cuadráticas dentro de las clases y entre las clases (relación B/W).

35 El OPLS-DA (Análisis Discriminante de Mínimos Cuadrados Parciales Ortogonales) es un desarrollo importante de la técnica PLS-DA que se ha propuesto para administrar de forma ortogonal la variación de las clases en la matriz de datos.

40 OPLS-DA aumenta el desempeño de clasificación de los modelos PLS-DA. Los desempeños de clasificación se estiman sobre la base de la "validación cruzada de k-iteraciones" al dividir la matriz de datos en subconjuntos k aleatorios. Para cada ciclo de cálculo, uno de los subconjuntos de F se mantiene aparte como un conjunto de prueba y los subconjuntos k-1 restantes actúan como capacitadores. Cada uno de los subconjuntos K se usan una vez como un conjunto de prueba, generando K valores de precisión. La precisión de la clasificación se calcula como el promedio de las tasas de precisión en subconjuntos k. El modelo se somete a una validación cruzada con el método "validación cruzada dejando uno fuera" (LOOCV) para validarse. La matriz de datos se escala según la media y la varianza unitaria, antes de enviarse a la división en subconjuntos k. En otras palabras, el promedio y la desviación estándar de los datos de capacitación se usan para indicar el centro y para escalar los datos de prueba. Una vez capacitado, el modelo se usa para verificar si los datos han generado un "sobreajuste". Para hacer esto, se genera un conjunto de validación con etiquetas de clase conocidas y, por lo tanto, se verifica si proporciona una tasa de precisión comparable a la de los datos de capacitación. Otro método es una validación de la línea R^2/Q^2 lo que ayuda a evaluar el riesgo de que el modelo actual sea falso, es decir, el modelo se ajusta bien solo a los conjuntos de subconjuntos, pero no predice Y tan bien para las nuevas observaciones. El valor de R^2 es la variación porcentual del conjunto de capacitación que puede explicarse por el modelo.

55 El valor de Q^2 es una medida de la validación cruzada de R^2 . Esta validación compara la bondad del ajuste del modelo original con la bondad del ajuste de diferentes modelos en función de los datos en los que el orden de observaciones Y se permuta aleatoriamente, mientras que la matriz se mantiene intacta. Los criterios para la validez del modelo son los siguientes:

60 1. Todos los valores Q^2 en el conjunto de datos permutados deben ser inferiores al valor Q^2 , estimado en el conjunto de datos actual. Si esto no está verificado, significa que el modelo está sobre ajustado.

2. La línea de regresión (la línea que une el punto real Q^2 al centroide del grupo de Q^2 valores permutados) tiene un valor negativo de la intersección del eje y.

65 Las Máquinas de Vectores de Soporte (SVM) son técnicas supervisadas de aprendizaje automático relativamente nuevas para usos de clasificación. Las SVM fueron propuestas por primera vez en 1982 por Vapnik (Vapnik, V. Estimation of

Dependences Based on Empirical Data; Springer Verlag: New York, 1982). El principio básico de las SVM, que son esencialmente clasificadores binarios, es el siguiente: dado un conjunto de datos con dos clases, se construye un clasificador lineal en forma de un hiperplano, que tiene el margen máximo en la minimización simultánea del error de clasificación empírico y la maximización del margen geométrico. En el caso de conjuntos de datos que no son separables linealmente, los datos originales se asignan a un espacio de características de mayor dimensión y se construye un clasificador lineal en este nuevo espacio (esto se conoce como "kernel"). Considerando un conjunto de datos de capacitación

$$x_i \in \mathbb{R}^n, \quad i = 1, \dots, m$$

10

15

donde cada uno de x_i cae en una de las dos categorías $y_i \in \{1, -1\}$, SVM determina el hiperplano cuyos parámetros se dan por (w, b) según lo obtenido por la solución del siguiente problema de optimización convexa:

$$\min_{w, b, \varepsilon} \frac{1}{2} w^t w + c \sum_{i=1}^m \varepsilon_i$$

20

sujeto a las siguientes condiciones:

$$y_i (w^t x_i + b) \geq 1 - \varepsilon_i$$

25

$$\varepsilon_i \geq 0$$

30

en donde c es el parámetro de regularización, que es un compromiso entre la precisión del aprendizaje y el término predicción, y ε es una medida del número de errores de clasificación. La inclusión del término regularización reduce el problema de sobreajuste.

Árboles de decisión.

35

40

Los árboles de decisión construyen modelos de clasificación basados en la división recursiva de los datos. Típicamente, un algoritmo del árbol de decisión comienza con el conjunto completo de datos, los datos se dividen en dos o más subgrupos en función de los valores de uno o más atributos, y después cada subconjunto se divide repetidamente en subconjuntos más pequeños hasta que el tamaño de cada subconjunto alcanza un nivel apropiado. Todo el proceso de modelado puede representarse en una estructura de árbol, y el modelo generado puede resumirse como un conjunto de reglas "si-entonces". Los árboles de decisión son fáciles de interpretar, poco exigentes desde el punto de vista computacional y capaces de hacer frente a datos ruidosos. La mayoría de los árboles de decisión abordan los problemas de clasificación, como por ejemplo el objeto de esta invención. En este contexto, la técnica también se conoce como árbol de clasificación. En la representación con la estructura de árbol, un nudo representa un conjunto de datos, y todo el conjunto de datos se representa como un nudo en la raíz.

45

El documento JP2007057309 describe un método asistido por computadora para el diagnóstico de carcinoma endometrial basado en un análisis proteómico de la sangre.

50

KARKKAINEN J Y OTROS, EUROPEAN JOURNAL OF CANCER AND CLINICAL ONCOLOGY, vol. 26, núm. 9, 1990, páginas 975-977 describe un análisis de GC-MS de medicamentos administrados (por ejemplo, MPA) en la sangre de pacientes con carcinoma endometrial para el monitoreo del tratamiento.

55

El documento WO 2011/161186 describe el diagnóstico de sepsis y menciona una variedad de tipos específicos de análisis multivariante o métodos de algoritmo de clasificación.

CAO Q JACKIE Y OTROS, International Journal of Gynecological Pathology 23(4), páginas 321-329 describe los perfiles de expresión génica por microarreglo de ADNc en carcinomas endometrioides y serosos del endometrio.

60

El documento JP 2007 147459 describe el ribitol como un estándar interno.

PETER PODOLEC Y OTROS, JOURNAL OF CHROMATOGRAPHY B: BIOMEDICAL SCIENCES & APPLICATIONS., vol. 967, 2014, páginas 134-138, describe el análisis GC-MS/MS de los metabolitos de Trypanosoma brucei.

65

Descripción detallada de la invención

5 La presente invención se refiere a un método asistido por computadora para el diagnóstico del carcinoma endometrial, basado en el análisis metabólico de la sangre y en una integración de los resultados obtenidos a través de un análisis multivariante mediante el uso de modelos de análisis discriminante que se seleccionan en el grupo que consiste en PLS-DA y OPLS-DA, o modelos de aprendizaje por computadora que se seleccionan en el grupo que consiste en SVM y árbol de decisión.

El objeto de la presente invención es un método para el diagnóstico del carcinoma endometrial basado en el análisis metabólico de la sangre, dicho método comprende las siguientes fases:

10 (I) una fase de capacitación que comprende:

- análisis de GCMS o GCxGCMS de muestras de sangre derivadas de pacientes con carcinoma endometrial y controles sanos;

15 - integración de los resultados obtenidos por análisis multivariante mediante el uso de al menos un modelo de análisis discriminante o un modelo de aprendizaje por computadora para capacitar al menos un modelo de clasificación;

(II) una fase de asignación que comprende el análisis de GCMS o GCxGCMS de una muestra de sangre desconocida y su asignación a una clase sobre la base del modelo de clasificación formulado en la fase de capacitación (I).

20 El análisis multivariante, que se realiza en los cromatogramas recolectados mediante el uso de:

- al menos un modelo de análisis discriminante que se selecciona del grupo que consiste en: PLS-DA y OPLS-DA, o

25 - dicho modelo de aprendizaje por computadora que se selecciona del grupo que consiste en: SVM y árbol de decisión;

permite ventajosamente la clasificación dicotómica satisfactoria ("Paciente sano" versus "Paciente afectado por carcinoma endometrial") de muestras desconocidas. El modelo de clasificación obtenido con un análisis multivariante PLS-DA incluso ha permitido la discriminación histológica del carcinoma (carcinoma de tipo I versus carcinoma de tipo II). Hasta la fecha, no hay otros métodos de diagnóstico no invasivos que puedan permitir una discriminación histológica de este tipo de tumor.

30 En dicha fase de capacitación (I) se analizan las muestras derivadas de pacientes afectados por carcinoma endometrial y de mujeres sanas con características físicas (IMC, edad, comorbilidad) y sociales (nivel de educación, condición socio-económica) similares, y de esta manera se capacitan los modelos de clasificación. Esta fase de capacitación tiene como objetivo crear y delimitar las características del perfil metabólico presente en la sangre de los dos grupos. Para tener una buena predicción del modelo de clasificación, es necesario someter a un análisis multivariante un número de muestras de sangre derivadas de pacientes con carcinoma endometrial y de controles sanos igual a al menos el 80 % del número de las variables identificadas de los perfiles metabólicos, tales muestras que pertenezcan al menos a 2 clases diferentes.

35 En tal fase de asignación (II), las muestras desconocidas se someten a análisis de GCMS, y los cromatogramas resultantes se clasifican de acuerdo con los modelos previamente capacitados, estimando la clase de pertinencia más probable.

40 El método de diagnóstico del carcinoma endometrial de la presente invención no se basa en la medición de la concentración de cada metabolito, sino que todo el grupo de metabolitos se considera como biomarcador (perfil metabólico), que, por estar presente de acuerdo con diferentes proporciones en los 2 grupos, permiten la inserción en dos clases diferentes de pertinencia.

50 Preferentemente, dicha fase de capacitación (I) comprende además las siguientes subfases:

- extracción y derivatización de metabolitos de muestras de sangre derivadas de pacientes con carcinoma endometrial y de controles sanos;

55 - análisis de GCMS o GCxGCMS de los metabolitos extraídos y derivatizados para obtener un cromatograma para cada muestra, cada cromatograma es un perfil metabólico;

- creación de la matriz de datos de los perfiles metabólicos de pacientes con carcinoma endometrial y de controles sanos;

60 - estructuración de al menos un modelo de clasificación como un resultado del análisis multivariante de la matriz de datos; en donde dicho análisis multivariante se realiza mediante el uso de al menos un modelo de análisis discriminante o un modelo de aprendizaje por computadora para capacitar al menos un modelo de clasificación.

65 Se pueden usar diferentes modelos de clasificación de acuerdo con la presente invención; preferentemente dichos modelos de clasificación se seleccionan del grupo que consiste en: PLS-DA, OPLS-DA, SVM y Árbol de decisión.

Preferentemente, dicha fase de asignación (II) comprende además las siguientes subfases:

- extracción y derivatización de metabolitos de al menos una muestra de sangre desconocida;
- análisis de GCMS o GCxGCMS de los metabolitos extraídos y derivatizados para obtener al menos un cromatograma

para la muestra de sangre desconocida;

- creación del perfil metabólico a partir de dicho cromatograma de la muestra de sangre desconocida;

- asignación del perfil metabólico a una clase sobre la base del modelo de clasificación capacitado en la fase (I).

Preferentemente, el método de la presente invención contempla un modelo de clasificación capacitado para una clasificación dicotómica "Paciente sano" o "Paciente afectado por carcinoma endometrial". Incluso más preferentemente, dicho modelo de clasificación también se capacita para una clasificación histológica del cáncer "tipo I" o "tipo II".

Preferentemente, dicha extracción se realiza mediante el uso de una mezcla de extracción que consiste en una mezcla acuosa de un alcohol y de un disolvente polar aprótico, preferentemente $\text{CH}_3\text{OH}/\text{H}_2\text{O}/\text{CHCl}_3$, incluso más preferentemente con una relación de volumen 2-3/0,5-0,5/0,5-1.

En una modalidad preferida, dicha subfase de extracción y derivatización comprende:

i) agitación de la muestra obtenida de la adición de una mezcla de extracción;

ii) centrifugación de la muestra obtenida en i);

iii) derivatización del sobrenadante obtenido de ii) por tratamiento con clorhidrato de metoxiamina en piridina;

iv) silanización del sobrenadante de la muestra obtenida en iii) con un agente de silanización que se selecciona del grupo que consiste en: N,O-bis(trimetilsilil) trifluoroacetamida (BSTFA), N-metil-N-(trimetilsilil) trifluoroacetamida (MSTFA), hexametil disilazano (HMDS), 1-(trimetilsilil) imidazol (TMSI), N-terc-butildimetilsilil-N-metiltrifluoroacetamida (MTBSTFA), 1-(terc-butildimetilsilil) imidazol (TBDMSIM) en presencia opcional de trimetildiclorosilano (TMCS).

Preferentemente, dicha extracción de metabolitos se realiza después de añadir a la muestra una alícuota conocida de un compuesto de referencia; preferentemente dicho compuesto de referencia es ribitol.

Para obtener la separación de metabolitos útiles para los fines de la presente invención, es posible trabajar tanto con cromatografía de gases monodimensional como con cromatografía de gases bidimensional; se prefiere la cromatografía de gases bidimensional, ya que el mejor poder de resolución de la técnica ofrece una mejor precisión de clasificación. De todos modos, como se muestra en los EJEMPLOS, también es posible trabajar con la cromatografía de gases monodimensional más común.

Los cromatogramas de gases obtenidos, preferentemente en modo SCAN, se integran para identificar todos los picos que tienen un área mayor que 10 veces el ruido de fondo de la señal del cromatograma.

Mediante el uso del pico del compuesto de referencia (preferentemente ribitol) como una referencia, tanto para el análisis cuantitativo como para centrar los tiempos de retención, cada pico se identifica sobre la base de una señal m/z de cuantificación y al menos 2 señales m/z de calificación. Después de la integración se realiza la cuantificación con el método de áreas de porcentajes normalizados. Los resultados obtenidos de esta cuantificación (áreas de porcentajes normalizados) se transfieren a una matriz en donde cada muestra representa una línea y las columnas se representan por varios metabolitos identificados unívocamente por medio de su tiempo de retención cromatográfico, en comparación con el tiempo de retención del compuesto de referencia. La primera columna de la matriz se usa para definir la clase de pertinencia de la muestra. En el caso más sencillo, solo pueden considerarse dos clases "Paciente sano" y "Paciente afectado por carcinoma endometrial", más adelante se presentan evidencias del funcionamiento de la invención sobre la base de esta clasificación dicotómica.

También es objeto de la presente invención un método como se describe anteriormente que comprende además las siguientes fases:

- integración de cromatogramas, en donde dicha integración proporciona la identificación de todos los picos que tienen un área mayor que 10 veces el ruido de fondo de la señal del cromatograma; mediante el uso del pico del compuesto de referencia como referencia tanto para el análisis cuantitativo como para centrar los tiempos de retención,

donde cada pico se identifica sobre la base de:

- una señal m/z de cuantificación; y

- al menos dos señales m/z de calificación;
- cuantificación con el método de áreas de porcentajes normalizados;
- transferencia de los datos obtenidos de dicha cuantificación a una matriz en la que cada muestra representa una línea y las columnas se representan por varios metabolitos identificados unívocamente por medio de su tiempo de retención cromatográfico.

El análisis estadístico multivariante de los datos (PLS-DA y OPLS-DA) y el aprendizaje automático (SVM y árbol de decisión) se realizan en cromatogramas normalizados y corregidos (basado en el área del pico de ribitol) mediante el uso de SIMPCA-P 13,0 (Umetrics), RapidMiner 5,3 (Rapid-I) y R (Foundation for Statistical Computing, Viena). Los valores se centran en el promedio y se normaliza la varianza. Para el perfil metabólico, el modelo OPLS-DA ha mostrado capacidad satisfactoria de modelado y predicción mediante el uso de un componente predictivo y tres componentes ortogonales ($R^2Y_{cum}=0,995$, $Q^2_{cum}=0,985$). La Figura 1 muestra la separación entre las clases obtenidas con el modelo OPLS-DA.

Además, se construyó una clasificación basada en la histología del carcinoma a través de un modelo PLS-DA. Como se muestra en la Figura 2, solo una muestra se coloca en un área incierta del espacio de definición de las clases.

La presente invención se puede entender mejor a la luz de los siguientes ejemplos no limitantes.

EJEMPLOS

La metodología de diagnóstico, objeto de la presente invención, se desarrolló a partir del análisis metabólico, realizado en muestras de sangre de pacientes con cierto diagnóstico de carcinoma endometrial, antes de la intervención de histerectomía y de un grupo de mujeres de control que tienen características físicas y socio-económicas similares pero con un útero sano. La información sobre el isotipo y la etapa de la neoplasia se recopilaron después de la histerectomía sobre la base de las evidencias anatomopatológicas obtenidas mediante el análisis del órgano explantado.

Recolección de muestras

Las muestras se tomaron de 88 mujeres con carcinoma endometrial y 80 mujeres sanas, que voluntariamente dieron muestras de sangre. El estudio fue aprobado por el comité de ética de la universidad Magna Grecia de Catanzaro y los pacientes y los voluntarios sanos firmaron el consentimiento informado sobre los propósitos del estudio. Las muestras de sangre se tomaron justo antes de la intervención de histerectomía mediante el uso de viales BD Vacutainer®, el suero se congeló a -80 °C hasta el momento del análisis. La sospecha diagnóstica de carcinoma endometrial después de la prueba histerectoscópica con biopsia de la lesión endometrial se confirmó por la prueba anatomopatológica del útero después de la intervención de histerectomía. También se organizó un grupo de control que tomaba muestras de sangre de mujeres que no tenían signos de carcinoma endometrial y con características físicas y socioeconómicas similares (peso, talla, IMC, edad, estado civil, nivel de educación, etc.).

Las características demográficas y clínicas de los casos y de los controles se informan en la Tabla 1, mientras que en la Tabla 2 se enumeran las características anatomopatológicas de los tumores investigados.

Tabla 1: características de la población del estudio

Parámetro	Carcinoma endometrial	Controles	Valor P
Números de casos	88	80	--
Edad (años)	63,3±14,8	63,1±8,3	NS
IMC	27,6±6,7	26,2±4,5	NS

Tabla 2: características anatomopatológicas de los tumores investigados

		Números de casos	Porcentaje de casos
Histotipo	Tipo I	67	76,1 %
	Tipo II	21	23,9 %
Etapa	G1	2	2,3 %
	G2	53	60,2 %
	G3	33	37,5 %

Extracción y derivatización de metabolitos

5 Se transfirieron cincuenta microlitros de suero a viales Eppendorf de 2 ml y 20 µl de una solución de 1 g/l de ribitol y se añadieron 200 µl de una mezcla que consiste en 2,5 partes de metanol, 1 parte de agua y 1 parte de cloroformo (CH₃OH:H₂O:CHCl₃, 2,5:1:1). La solución se mezcló en vórtice durante 30 segundos.

10 Las muestras se centrifugaron a 16 000 rpm durante 10 minutos a 4 °C. Se colectó una alícuota de 200 µl de sobrenadante y se transfirió a nuevos viales Eppendorf de 2 ml y se añadió con 200 µL de H₂O y se mezcló en vórtice durante 30 segundos y se centrifugó nuevamente a 16 000 rpm durante 5 minutos a 4 °C.

Se colectó nuevamente una alícuota de 350 µl del sobrenadante y se transfirió a ampollas de vidrio de 1,5 y se liofilizó.

15 La muestra liofilizada se trató con 50 µl de 20 mg/ml de clorhidrato de metoxiamina en piridina. La reacción se realizó a 37 °C con agitación (350 rpm) durante 90 minutos. Al final, se añadieron 50 µl de di N,O-bis(trimetilsilil)trifluoroacetamida (BSTFA) con 1 % de trimetilclorosilano a cada ampolla y la reacción de silanización se realizó a 37 °C durante 60 minutos bajo agitación (350 rpm).

Análisis de MDGCMS

20 Para la cromatografía de gases bidimensional se usó una columna primaria (colocada en el primer horno), del tipo SLB-5ms 30,0 m x 0,25 mm ID con 1 µm de grosor de película [polímero de silfenileno, prácticamente con polaridad equivalente a poli(5 % difenilo/95 % metilsiloxano)] (J&W Agilent) que se unió a la posición 1 de la interfaz con 7 puertas (SGE).

25 Un BPX-50 5,0 m x 0,50 mm ID con 0,25 µm de grosor de la película se unió a la posición 7 de la interfaz. Se estableció un BPX-50 1,5 m x 0,25 mm ID, 0,25 µm a la posición 6 y se conectó a un detector de ionización de llama (FID) configurado a 320 °C, mientras que la columna analítica de 5,0 m (químicamente idéntica a la conectada a FID) se conectó al sistema qMS.

30 La columna conectada a FID se usó para reducir el flujo en la segunda dimensión y para verificar que el compuesto apenas representativo no se debió a una fluctuación aleatoria de la cromatografía.

Se usó un vaso capilar externo de 40 µl (20 cm x 0,71 mm OD x 0,51 mm ID en acero inoxidable) para conectar las puertas 3 y 4 de la interfaz SGE.

35 El programa térmico igual para los dos hornos fue: 80 °C durante 1 minuto, después se calienta hasta 320 °C a 3 °C/minuto y se mantiene durante 4 minutos.

40 La presión inicial de helio (velocidad lineal constante) se estableció en 129,6 kPa. La presión inicial auxiliar de helio del APC (control avanzado de presión), que también funciona en condiciones de velocidad lineal constante, se estableció en 90,4 kPa.

45 El volumen de inyección de 1 µl con una relación de separación de: 1:5. El período de modulación se estableció en 4,1 s (período de acumulación 4,0 segundos, período de inyección 0,1 segundos). Las condiciones del espectrómetro de masas cuadrupolo fueron: modo de ionización: impacto electrónico (70 eV), intervalo de masa: 40-600 m/z, velocidad de escaneo: 10 000 amu/segundo.

Análisis de GCMS

50 Para la cromatografía de gases monodimensional se usó una columna del tipo CP-Sil 8 CB GC Column, 30 m, 0,25 mm, 1,00 µm, (Agilent J&W).

El programa térmico del GC concibió una temperatura inicial de 100 °C por 1 minuto, después se calentó hasta 320 °C a 4 °C/minuto y 4 minutos de tiempo de espera para un tiempo total de funcionamiento de 60 minutos.

55 La presión inicial de helio (velocidad lineal constante de 39 cm/s) se estableció en 83,7 kPa. El volumen de inyección a 2 µl con una relación de separación: 1:5. Las condiciones del espectrómetro de masas cuadrupolo fueron: modo de ionización: impacto electrónico (70 eV), intervalo de masa: 35-600 m/z, velocidad de escaneo: 3,333 amu/segundo con un tiempo de corte de solvente de 4,5 minutos.

60 Creación de la matriz de datos

En un cromatograma de TIC usualmente se detectan más de 250 señales, algunos de estos picos no se investigaron más, ya que no hubo correspondencias en otras muestras, porque estaban en una concentración demasiado baja o porque tenían una calidad espectral deficiente para ser confirmados como metabolitos.

65

Se detectaron un total de 198 metabolitos endógenos como aminoácidos, ácidos orgánicos, carbohidratos, ácidos grasos y esteroides. Para la identificación del pico, se usó el índice de retención lineal (LRI) como tolerancia máxima, una diferencia entre el índice de Kovats tabulado y el índice experimental de 10, mientras que el mínimo de compatibilidad para la búsqueda en las bibliotecas se estableció en 85 %. Se usaron 2 bibliotecas: el NIST11 y una biblioteca desarrollada intencionadamente por derivatización de más de 500 metabolitos en las mismas condiciones de las muestras analizadas. Las áreas de los picos se normalizaron y se corrigieron con referencia a la señal de ribitol. Los resultados se resumieron en un archivo separado por comas (CSV) y se cargaron en un software adecuado para el procesamiento estadístico. Los cromatogramas de gases obtenidos en modo SCAN se integraron para identificar todos los picos que tienen un área superior a 10 veces el ruido de fondo de la señal del cromatograma de gases. Cada pico se identificó sobre la base de la señal m/z de cuantificación y al menos dos señales m/z de calificación. Después de la integración, se realizó la cuantificación con el método de las áreas de porcentajes normalizados, se usó el pico de ribitol como referencia tanto para el análisis cuantitativo como para centrar los tiempos de retención.

Los resultados obtenidos de esta cuantificación (áreas de porcentajes normalizados) se transfirieron a una matriz en donde cada muestra representa una línea y las columnas se representan por varios metabolitos identificados unívocamente por medio de su tiempo de retención cromatográfica. La primera columna de la matriz se usa para definir la clase de pertinencia de la muestra. En el caso más sencillo, solo pueden considerarse dos clases "Paciente sano" y "Paciente afectado por carcinoma endometrial", más adelante se presentan evidencias del funcionamiento de la invención sobre la base de esta clasificación dicotómica. Se obtuvieron evidencias adicionales sobre la posibilidad de probar diferentes modelos de clasificación para predecir el histotipo de la neoplasia y la clasificación.

Análisis estadístico

El análisis estadístico multivariante de los datos (PLS-DA y OPLS-DA) y el aprendizaje automático (SVM y el árbol de decisión) se llevaron a cabo en los cromatogramas normalizados y corregidos (basado en el área del pico de ribitol) mediante el uso de SIMPCA-P 13,0 (Umetrics), RapidMiner 5,3 (Rapid-I) y R (Foundation for Statistical Computing, Vienna).

Los valores se centraron en el promedio y se normalizó la varianza.

Resultados

Para un perfil metabólico, el modelo OPLS-DA ha mostrado capacidad satisfactoria de modelado y predicción mediante el uso de un componente predictivo y tres componentes ortogonales ($R^2Y_{cum} = 0,995$, $Q^2_{cum} = 0,985$). Los otros modelos de clasificación han mostrado buenas habilidades de clasificación (incluso si son más bajas que OPLS-DA). Son posibles diferentes enfoques para la asignación final de la clase de pertinencia de la muestra desconocida. Puede usarse la respuesta de un modelo único o pueden integrarse las respuestas de los distintos modelos en un algoritmo de decisión más complejo.

La Tabla 3 informa algunos índices de la evaluación del desempeño diagnóstico que se usó para evaluar los modelos investigados. La sensibilidad se calculó como $TP/(TP+FN)$, en donde TP representa el número de verdaderos positivos, es decir, específicamente las muestras correctamente diagnosticadas como afectadas por el carcinoma endometrial por el modelo propuesto, y FN es el número de falsos negativos, es decir, específicamente las muestras erróneamente identificadas como negativos. La especificidad se calculó como $TN/(TN+FP)$, en donde TN representa el número de verdaderos negativos, es decir, específicamente muestras correctamente diagnosticadas como sanas y FP representa los falsos positivos, es decir, específicamente el número de personas erróneamente diagnosticadas como sanas. La relación de probabilidad positiva (PLR) se calculó como $Sensibilidad/(1-Especificidad)$, mientras que la negativa (NLR) como $(1-Sensibilidad)/Especificidad$.

El valor predictivo (NPV) se calculó como $TN/(TN+FN)$, mientras que el valor positivo (VPP) como $TP/(TP+FP)$. La precisión representa el porcentaje de todas las asignaciones correctas y se calculó como $(TP+TN)/(TP+FP+TN+FN)$ mientras que la repetibilidad como los números de reasignaciones correctas en 10 repeticiones del análisis de una muestra.

55

Tabla 3 - Desempeño diagnóstico de los modelos investigados

Parámetro	OPLS-DA	PLS-DA	SVM	Árbol de decisión
5 <i>Sensibilidad</i>	No hay error de clasificación	0,989	0,966	0,977
<i>Especificidad</i>		0,988	0,974	0,963
10 <i>PLR</i>		79,1	37,7	26,1
<i>NLR</i>		0,012	0,035	0,024
<i>NPV</i>		0,988	0,962	0,975
<i>PPV</i>		0,989	0,977	0,966
15 <i>Exactitud</i>	No hay error de clasificación	0,988	0,970	0,970
<i>Repetibilidad</i>		>99 %	>99 %	>99 %

20 Para identificar los metabolitos que contribuyeron mucho más a la separación de las clases, se calculó la puntuación de las variables importantes en la proyección (VIP) para cada componente. Las puntuaciones VIP representan la suma ponderada de los cuadrados de carga de los pls, considerando la cantidad de y-varianza en cualquier dimensión. Dos picos muestran una puntuación VIP mayor que 2 tanto en los modelos PLS-DA como en OPLS-DA (tanto en la clasificación de carcinoma endometrial versus control y en la clasificación de tipo I versus tipo II). Estos se identificaron como nudos importantes también en el árbol de decisión; estas observaciones sugieren una gran importancia de estas variables en los procesos de clasificación (datos no informados). El primer metabolito (puntuación VIP=2,3; similitud espectrométrica =91 %; δ LRI = 11) resultó ser una señal atribuible al aminoácido glutamina, mientras que el segundo (puntuación VIP=2,1; similitud espectrométrica=89 % δ LRI = 16) resultó ser atribuible a glucono δ -lactona.

30

Reivindicaciones

- 5 1. Un método asistido por computadora para el diagnóstico de carcinoma endometrial basado en el análisis metabolómico de la sangre, dicho método que comprende las siguientes fases:
 - (I) una fase de capacitación que comprende:
 - análisis de GCMS o GCxGCMS de muestras de sangre derivadas de pacientes con carcinoma endometrial y controles sanos;
 - 10 - integración de los resultados obtenidos por un análisis multivariante mediante el uso de al menos un modelo de análisis discriminante o un modelo de aprendizaje por computadora para capacitar al menos un modelo de clasificación;
 - (II) una fase de asignación que comprende el análisis de GCMS o GCxGCMS de una muestra de sangre desconocida y su asignación a una clase de pertinencia sobre la base del modelo de clasificación formulado en la fase de capacitación (I).
- 20 2. El método de acuerdo con la reivindicación 1, en donde
 - al menos un modelo de análisis discriminante se selecciona del grupo que consiste en: PLS-DA y OPLS-DA, o
 - dicho modelo de aprendizaje por computadora se selecciona del grupo que consiste en: SVM y árbol de decisión.
- 25 3. El método de acuerdo con una o más de las reivindicaciones anteriores, en donde la fase de capacitación (I) comprende las siguientes subfases:
 - extracción y derivatización de metabolitos de muestras de sangre derivadas de pacientes con carcinoma y de controles sanos;
 - análisis de GCMS o GCxGCMS de metabolitos extraídos y derivatizados para obtener un cromatograma para cada muestra;
 - 30 - creación de la matriz de datos de los perfiles metabólicos de los pacientes con carcinoma endometrial y de controles sanos;
 - estructuración de al menos un modelo de clasificación como un resultado del análisis multivariante de la matriz de datos; en donde dicho análisis multivariante se lleva a cabo mediante el uso de al menos un modelo de análisis discriminante o un modelo de aprendizaje por computadora para capacitar al menos un modelo de clasificación.
- 35 4. El método de acuerdo con una o más de las reivindicaciones precedentes, en donde dicha fase (II) comprende además las siguientes subfases:
 - extracción y derivatización de metabolitos de al menos una muestra de sangre desconocida;
 - 40 - análisis de GCMS o GCxGCMS de los metabolitos extraídos y derivatizados para obtener al menos un cromatograma para la muestra de sangre desconocida;
 - creación del perfil metabólico a partir de dicho cromatograma de la muestra de sangre desconocida;
 - asignación del perfil metabólico a una clase sobre la base del modelo de clasificación capacitado en la fase (I).
- 45 5. El método de acuerdo con una o más de las reivindicaciones anteriores, en donde el número de muestras de sangre derivadas de pacientes con carcinoma endometrial y de controles sanos es igual a al menos el 80 % del número de variables identificadas de perfiles metabólicos.
- 50 6. El método de acuerdo con una o más de las reivindicaciones anteriores, en donde dicho modelo de clasificación se capacita para una clasificación dicotómica "Paciente sano" o "Paciente afectado por carcinoma endometrial".
7. El método de acuerdo con una o más de las reivindicaciones anteriores, en donde dicho modelo de clasificación se capacita además para una clasificación histológica de cáncer "tipo I" o "tipo II".
- 55 8. El método de acuerdo con una o más de las reivindicaciones anteriores, en donde dicha extracción y derivatización comprenden:
 - i) agitación de la muestra obtenida de la adición de una mezcla de extracción;
 - ii) centrifugación de la muestra obtenida en i);
 - 60 iii) derivatización del sobrenadante obtenido en ii) por tratamiento con clorhidrato de metoxiamina en piridina;
 - iv) silanización del sobrenadante de la muestra obtenida en iii) con un agente de silanización que se selecciona del grupo que consiste en: N,O-bis(trimetilsilil) trifluoroacetamida (BSTFA), N-metil-N-(trimetilsilil) trifluoroacetamida (MSTFA), hexametil disilazano (HMDS), 1-(trimetilsilil) imidazol (TMSI), N-terc-butildimetilsilil-N-metiltrifluoroacetamida (MTBSTFA), 1-(terc-butildimetilsilil) imidazol (TBDMSIM);

en donde dicha mezcla de extracción consiste en una mezcla acuosa de un alcohol y un disolvente polar aprótico.

5 9. El método de acuerdo con una o más de las reivindicaciones anteriores, en donde dicha extracción de metabolitos se realiza añadiendo una alícuota de un compuesto de referencia, preferentemente ribitol.

10. El método de acuerdo con una o más de las reivindicaciones anteriores, que comprende además las siguientes fases:

10 - integración de los cromatogramas obtenidos de acuerdo con una o más de las reivindicaciones anteriores, en donde dicha integración proporciona la identificación de todos los picos que tienen un área mayor que 10 veces el ruido de fondo de la señal del cromatograma; mediante el uso del pico del compuesto de referencia como referencia tanto para el análisis cuantitativo como para centrar los tiempos de retención,

15 donde cada pico se identifica sobre la base de:

- una señal m/z de cuantificación; y
 - al menos dos señales m/z de calificación;
 - cuantificación con el método de áreas de porcentajes normalizados;
 - transferencia de los datos obtenidos de dicha cuantificación a una matriz en la que cada muestra representa una línea y las columnas se representan por varios metabolitos identificados unívocamente por medio de su tiempo de retención cromatográfica.
- 20

25

