



19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 360 365**

51 Int. Cl.:  
**G01N 15/14** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **06121243 .7**  
96 Fecha de presentación : **26.09.2006**  
97 Número de publicación de la solicitud: **1770387**  
97 Fecha de publicación de la solicitud: **04.04.2007**

54 Título: **Generación de archivos de datos de citometría de flujo con un número potencialmente infinito de dimensiones derivadas de la fusión de un grupo de archivos de datos de citometría de flujo separados y su reconstrucción multidimensional tanto con datos de citometría de flujo medidos realmente como estimados.**

30 Prioridad: **30.09.2005 US 240167**

45 Fecha de publicación de la mención BOPI:  
**03.06.2011**

45 Fecha de la publicación del folleto de la patente:  
**03.06.2011**

73 Titular/es: **Universidad de Salamanca  
Patio de Escuelas Menores, nº 1  
37008 Salamanca, ES**

72 Inventor/es:  
**Orfao de Matos Correia e Vale, José Alberto;  
Pedreira, Carlos Eduardo y  
Sobral da Costa, Elaine**

74 Agente: **Pons Ariño, Ángel**

ES 2 360 365 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

5 Generación de archivos de datos de citometría de flujo con un número potencialmente infinito de dimensiones derivadas de la fusión de un grupo de archivos de datos de citometría de flujo separados y su reconstrucción multidimensional tanto con datos de citometría de flujo medidos realmente como estimados

### CAMPO DE LA INVENCION

10 Esta invención se refiere al campo de la citometría de flujo y más particularmente a la generación de nuevos archivos de datos de citometría de flujo con un número potencialmente infinito de dimensiones, teniendo cada evento individual incluido en estos nuevos archivos de datos información asociada que ha sido medida realmente o estimada, para todos los parámetros individuales evaluados. Estos nuevos archivos de datos de citometría de flujo se crean: 1) fusionando un grupo de archivos de datos de citometría de flujo separados que contienen información acerca de eventos medidos en diferentes partes alícuotas procedentes de la misma muestra; estos archivos de datos originales contienen, en común, 15 datos acerca de uno o más parámetros, al mismo tiempo que contienen información acerca de uno o más parámetros diferentes; 2) estimando la magnitud de los valores y de una correspondiente medida de incertidumbre de las variables que difieren entre uno o más de los archivos de datos originales que han sido fusionados para cada evento individual contenido en el archivo de datos fusionado para el que esa variable no fue medida directamente en la citometría de flujo y; 3) reconstruyendo multidimensionalmente el nuevo archivo de datos que contiene tanto los datos reales medidos en la citometría de flujo para cada 20 evento de los archivos de datos originales individuales fusionados como los datos estimados para aquellos parámetros no medidos en un grupo de eventos procedentes de cada uno de los archivos de datos originales. Esta invención permite la generación de archivos de datos que contienen información para eventos celulares individuales procedentes de una muestra acerca de un número más alto de parámetros que aquellos medidos realmente en cada uno de los archivos de datos originales fusionados; 25 el número global de parámetros para los cuales se asignan valores a cada evento celular individual incluido en el nuevo archivo de datos solo está limitado por el número total de parámetros medidos en todo el grupo de archivos de datos fusionados. Esto permite una más potente identificación, enumeración y caracterización de diferentes poblaciones de eventos contenidos en una muestra.

### ANTECEDENTES DE LA INVENCION

30 Loken y Terstappen han descrito previamente en la patente de EE.UU. Nº 5.047.321, el análisis multiparamétrico de componentes celulares en un cuerpo fluido compuesto de sangre y médula ósea. Estos autores pudieron discriminar entre diversos componentes celulares de la sangre y la médula ósea, contar el número de células dentro de cada componente y proporcionar un análisis diferencial de cada uno de ellos usando una combinación de dos colorantes de ácido nucleico —el colorante de ADN LDS-751 35 (Exciton), el colorante de ARN naranja de tiazol (TO, Molecular Probes, Inc)— un anticuerpo monoclonal anti-CD45 marcado fluorescentemente y dos parámetros de dispersión de luz (dispersión de luz frontal y lateral). Este procedimiento permitía la identificación y diferenciación entre glóbulos rojos nucleados, eritrocitos, reticulocitos, plaquetas, linfocitos, monocitos, granulocitos neutrófilos, granulocitos basófilos, granulocitos eosinófilos, y precursores de todas las células nucleadas. A pesar de esto, no pudieron 40 demostrar que con este procedimiento fueran capaces de diferenciar específicamente entre células normales y neoplásicas que coexisten en la misma muestra o de caracterizar además estos subconjuntos de células.

45 En la patente de EE.UU. Nº 6.287.791, Terstappen y Chen describían un nuevo refinamiento de la patente de EE.UU. Nº 5.047.321, pero no demostraron ninguna caracterización mejor de las diferentes poblaciones de leucocitos.

50 En la patente de EE.UU. Nº 02380098.0, Orfao describía un procedimiento para el análisis diferencial multidimensional de leucocitos de la sangre, la médula ósea y otros fluidos corporales, que permitía específicamente una identificación adicional de células dendríticas y sus subconjuntos, además de glóbulos rojos nucleados, linfocitos, monocitos, granulocitos neutrófilos, granulocitos basófilos, granulocitos eosinófilos y precursores de todas las células nucleadas.

55 A su vez, en la patente de EE.UU. Nº 5538855, Orfao describía un procedimiento que permitía un análisis más detallado de los compartimentos linfoides a través de la identificación simultánea de hasta 12 subconjuntos diferentes de células T, B y Nk en la sangre y la médula ósea entre otros tipos de muestras biológicas. El autor usó una tinción combinada para los antígenos CD3, CD19, CD56 (y/o CD16), CD4 y CD8 en una única tinción de 3 colores. Sin embargo, usando este procedimiento no pudo caracterizar más los subconjuntos de células identificados, al mismo tiempo, ni todos los subconjuntos de células no linfoides presentes en una sangre periférica normal pudieron identificarse específicamente (por ejemplo: las subpoblaciones de células dendríticas).

En ninguno de los procedimientos referidos, los procedimientos descritos permitieron directamente una caracterización más detallada de las células identificadas. En tal caso, se requiere el uso de un mayor número de tinciones asociadas con emisiones de fluorescencia distinguibles y de instrumentos de citometría de flujo capaces de medir un mayor número de emisiones de fluorescencia diferentes. En las últimas dos décadas, el número de emisiones de fluorescencia diferentes que pueden medirse simultáneamente por un citómetro de flujo ha aumentado notablemente pasando de 3 hasta 17 colores. Sin embargo, este número aún es limitado ya que frecuentemente se requieren más de 20 marcadores para una caracterización detallada de los componentes celulares presentes en diferentes muestras normales y patológicas. Como ejemplo, la evaluación del repertorio TCRVbeta de células TCRalfa-beta+ T de sangre periférica, requiere típicamente tinción para más de 20 marcadores diferentes dirigidos contra diferentes miembros de las familias de proteínas TCRVbeta; igualmente, el análisis inmunofenotípico de diagnóstico rutinaria de muestras leucémicas está basado en la evaluación de los patrones de tinción de médula ósea y/o células sanguíneas para hasta varias decenas de marcadores diferentes (típicamente entre 20 y 40 antígenos). De este modo, incluso con la capacidad de medir simultáneamente 17 emisiones de fluorescencia diferentes, los instrumentos más avanzados de citometría de flujo aún tienen capacidades multicolores limitadas. Como se mencionó anteriormente, el número máximo de emisiones de fluorescencia medidas simultáneamente para un evento individual o para un grupo de eventos contenidos en un archivo de datos de citometría de flujo, depende de la disponibilidad de fluorocromos con emisiones de fluorescencia compatibles, distinguibles y del número máximo de emisiones de fluorescencia que pueden detectarse en el citómetro de flujo. Debido a tal limitación, la tinción de una muestra para un número de marcadores superior al número de emisiones de fluorescencia que el instrumento citómetro de flujo disponible es capaz de medir se hace normalmente en dos o más partes alícuotas separadas medidas una detrás de otra. Como ejemplo de tal estrategia, en la patente de EE.UU. N° 5.137.809, Loken y Sha describían un procedimiento para un análisis multiparamétrico de componentes celulares en médula ósea. Los autores describían el uso, en una primera etapa, de una combinación de anticuerpos monoclonales marcados cada uno con un fluorocromo diferente, para teñir todos los leucocitos y de más combinaciones para teñir poblaciones seleccionadas de leucocitos, en una segunda etapa. Sin embargo, tales procedimientos no permiten vincular automáticamente y comparar directamente la información sobre la cantidad de dispersión de luz y emisiones de fluorescencia medidas para las células individuales contenidas 1) en partes alícuotas diferentes procedentes de la misma muestra, 2) en muestras diferentes derivadas de tejidos idénticos o diferentes procedentes del mismo individuo o de individuos diferentes, o; 3) en partes alícuotas de muestras diferentes que han sido medidas bajo condiciones diferentes.

Todos los procedimientos descritos anteriormente permitían la identificación de un número variable de poblaciones diferentes de leucocitos normales presentes en la sangre, la médula ósea y otras muestras y solo permiten la identificación de subpoblaciones seleccionadas de células dependiendo de la combinación específica de anticuerpos monoclonales y los colorantes de ácidos nucleicos usados; no obstante, no pudieron proporcionar un procedimiento para la identificación específica y reproducible de células neoplásicas mezcladas natural o artificialmente con células normales en una muestra.

En la patente provisional de EE.UU. N° 10/791.994, Orfao, Pedreira y Sobral da Costa describían un procedimiento para la detección multidimensional de fenotipos aberrantes en células neoplásicas que ha de usarse para distinguir las de las células normales, para monitorizar números mínimos de células neoplásicas en muestras de sangre, médula ósea, líquido espinal y ganglio linfático, usando mediciones de citometría de flujo. Este procedimiento también permite una comparación objetiva entre datos de flujo citométrico adquiridos en diferentes mediciones que correspondían a: 1) diferentes partes alícuotas procedentes de la misma muestra, 2) diferentes partes alícuotas procedentes de diferentes muestras procedentes del mismo sujeto y 3) diferentes partes alícuotas procedentes de distintas muestras procedentes de diferentes individuos, incluso en casos donde fueron medidas bajo diferentes condiciones. Con este procedimiento, se generan grandes archivos de datos de citometría de flujo combinando datos procedentes de dos o más archivos de datos diferentes que contienen datos procedentes de una o más muestras. En estos archivos de datos, la información acerca de un número infinito de parámetros puede almacenarse junta; sin embargo, cada evento del archivo de datos de modos de listas solo está asociado a datos derivados de los parámetros medidos realmente en el citómetro de flujo para ese evento particular y no puede hacerse ningún vínculo para eventos individuales entre los parámetros que han sido medidos y aquellos que no han sido medidos para ese evento particular, aunque esos parámetros fueran evaluados para otros eventos contenidos en el archivo de datos fusionados.

#### **DESCRIPCIÓN DE LA INVENCIÓN**

La presente invención se refiere a un procedimiento que genera nuevos archivos de datos de citometría de flujo con un número potencialmente infinito de dimensiones, teniendo los eventos contenidos en estos archivos de datos información asociada para cada uno de todo el conjunto de parámetros evaluados, correspondiendo tal información a datos que han sido medidos directamente en el citómetro

de flujo, o estimados después. La invención comprende las etapas de: 1) fusionar dos o más archivos de datos de citometría de flujo separados que contienen información acerca de eventos medidos en diferentes partes alícuotas procedentes de la misma muestra; estos archivos de datos originales contienen, en común, datos acerca de uno o más parámetros al mismo tiempo que contienen información acerca de uno o más parámetros diferentes, 2) estimar la magnitud de los valores y de una correspondiente medida de incertidumbre de las variables que difieren entre uno o más de los archivos de datos originales que han sido fusionados, para cada evento individual contenido en el archivo de datos fusionado, para el que esa variable no fue medida directamente en la citometría de flujo y; 3) reconstruir multidimensionalmente el nuevo archivo de datos que contiene tanto los datos reales medidos en la citometría de flujo para cada evento de los archivos de datos originales individuales fusionados, como los datos estimados para aquellos parámetros no medidos en un grupo de eventos procedentes de uno o más de los archivos de datos originales. Esta invención permite la generación de archivos de datos que contienen información para eventos celulares individuales procedentes de una muestra acerca de un número más alto de parámetros que aquellos medidos realmente en el citómetro de flujo para los eventos individuales contenidos en cada uno de los archivos de datos originales fusionados; el número global de parámetros para los que se asignan valores a cada evento celular individual incluido en el nuevo archivo de datos puede ser tan grande como el número total de parámetros medidos en todo el grupo de archivos de datos fusionados. Esto permite una más potente identificación, enumeración y caracterización de diferentes poblaciones de eventos contenidos en una muestra.

Según esta invención, antes de fusionar los archivos de datos de citometría de flujo originales, dos o más partes alícuotas de una muestra son teñidas por separado con un panel de anticuerpos monoclonales. La muestra puede contener células normales, células neoplásicas o una mezcla tanto de células normales como neoplásicas. La muestra puede comprender suspensiones de sangre periférica, médula ósea, ganglio linfático, bazo, líquido ascítico, líquido espinal, líquido sinovial, efusiones pleurales y de células individuales preparadas a partir de tejido sólido. El panel de reactivos usados para teñir la muestra está formado por dos o más combinaciones de anticuerpos monoclonales directamente conjugados a fluorocromos. Estas combinaciones de anticuerpos monoclonales conjugados a fluorocromos consisten en una combinación de reactivos de anticuerpos monoclonales, que son idénticos en todas las combinaciones, y reactivos de anticuerpos monoclonales que solo se usan en una cualquiera o parte de todas las combinaciones de reactivos de anticuerpos monoclonales. Cada combinación de reactivos de anticuerpos monoclonales puede contener, además de los anticuerpos monoclonales, otras sondas para teñir la muestra. Dichas sondas adicionales consisten en fluorocromos específicos para colorantes de ácidos nucleicos, mitocondrias y otros componentes celulares (por ejemplo, gránulos de zimógeno, cromosomas, iones, proteínas, lípidos, carbohidratos) o funciones celulares (por ejemplo, fagocitosis, secreción de proteínas, activación, proliferación, diferenciación). Cada anticuerpo monoclonal en cada combinación está conjugado a un fluorocromo diferente y cada combinación de anticuerpos monoclonales tiene, en común, uno o más reactivos de anticuerpos monoclonales conjugados a fluorocromos. El número de emisiones de fluorescencia en cada combinación de múltiples anticuerpos monoclonales comprende dos o más fluorocromos diferentes, cada uno vinculado a un anticuerpo monoclonal diferente, cuya emisión de fluorescencia es distinguible de la de los otros reactivos de anticuerpos monoclonales conjugados a monocromos en esa combinación. Los fluorocromos que pueden usarse en esta invención incluyen isotiocianato de fluoresceína (FITC), ficoeritrina (PE), proteína peridina clorofila (PerCP), aloficocianina, alexa flúor 488, alexa 647, alexa flúor 610, alexa 710, alexa flúor 405, cianina 5 (Cy5), cianina 5.5 (Cy5.5), rojo de Texas (TR), azul Pacífico (PB), amarillo cascada, azul cascada y conjugados de los mismos unidos a PE, a APC, o a PerCP (por ejemplo, PE/Cy5, PE/Cy5.5, PE/Cy7, PE/TR, PerCP/Cy5.5, APC/Cy7), puntos cuánticos o cualquier fluorocromo compatible adicional o tándem de fluorocromos.

Los procedimientos exactos de preparación de muestras usados para teñir las células antes de medirlas en un citómetro de flujo se han descrito previamente en detalle y se han usado extensamente, y pueden implementarse fácilmente por un experto en el campo.

Después de prepararse y teñirse la muestra, las emisiones de fluorescencia asociadas a grandes números de células a partir de partes alícuotas de la muestra teñidas con cada una de las combinaciones de anticuerpos monoclonales se miden en un citómetro de flujo según procedimientos convencionales y se almacenan dos o más archivos de datos de modos de listas; cada uno de los archivos de datos almacenados contiene información sobre la dispersión de luz específica y características de emisiones de fluorescencia de las células individuales analizadas en cada parte alícuota de la muestra medida. Para cada parte alícuota de la muestra puede medirse un número idéntico o diferente de eventos en el citómetro de flujo. Alternativamente, la información acerca de dos o más partes alícuotas de muestras diferentes puede almacenarse en un único archivo de datos, directamente después de medidas en la citometría de flujo.

Según esta invención, todos aquellos parámetros comunes a cada par de partes alícuotas de la

5 muestra también son comunes a todas las demás partes alícuotas de la muestra medidas cuya información estará contenida en el archivo de datos fusionado; alternativamente, solo parte de todos aquellos parámetros comunes a cada par de partes alícuotas de la muestra también son comunes a todas las demás partes alícuotas de la muestra medidas, cuya información estará contenida en el archivo de datos fusionado.

10 Desde un punto de vista estructural, cada uno de los archivos de datos almacenados consiste en una matriz de datos en la que la información para cada evento celular diferente medido está alineada en una sola línea, estando los eventos que han sido medidos consecutivamente colocados en líneas consecutivas en la matriz de datos; para cada evento, la información sobre su dispersión de luz - dispersión de luz frontal (FSC) y lateral (SSC)- y sus parámetros asociados a la emisión de fluorescencia se coloca en una columna diferente de la matriz de datos.

15 Para fusionar dos o más archivos de datos de modos de listas, los datos procedentes de los archivos que han de fusionarse se pegan en una base de datos en una secuencia conocida. En el archivo de datos fusionado, cada parámetro diferente medido en todo el conjunto de archivos de datos, relacionado con las características de dispersión de la luz (FSC y SSC) y las emisiones de fluorescencia de una célula se coloca en una columna diferente, mientras que los datos sobre cada evento celular medido se colocan en una línea diferente. Los eventos que proceden del mismo archivo de datos inicial se colocan secuencialmente uno detrás de otro, ordenados en una secuencia que está relacionada directa o inversamente con la secuencia en la que fueron medidos en el citómetro de flujo. Alternativamente, los 20 archivos de datos originales y los eventos contenidos en cada uno de dichos archivos de datos se pegan en una secuencia que no está relacionada con la secuencia en la que fueron medidos en el citómetro de flujo. Como solo se midió una parte de todos los parámetros para eventos que corresponden a cada una de las partes alícuotas de la muestra, en el archivo de datos fusionado, para cada línea (evento celular), habrá columnas rellenas y sin rellenar que corresponden a aquellos parámetros medidos directamente y a aquellos parámetros no evaluados en el citómetro de flujo para ese evento particular, respectivamente. Las columnas de la base de datos pueden ordenarse en una secuencia, que está relacionada directa o inversamente con el orden en que se midió cada parámetro o grupo de parámetros. Alternativamente, las columnas de la matriz de base de datos fusionados pueden seguir un orden diferente. Las columnas vacías de la base de datos para cada evento celular (línea) se rellenarán entonces usando procedimientos de estimación estadística basados en la similitud para calcular valores y una medida de incertidumbre correspondiente para aquellos parámetros no medidos para un grupo de eventos del 30 archivo de datos fusionado.

35 Para rellenar todas las columnas vacías de la base de datos para cada evento individual, todas las emisiones de fluorescencia comunes medidas más las medidas de dispersión de luz se consideran como 'parámetros comunes' a todos los archivos de datos originales y pueden ser de tipo y número variable ( $k$ =número total de 'parámetros comunes' medidos). A su vez, las medidas de todas las demás emisiones de fluorescencia registradas en el citómetro de flujo solo para los eventos contenidos en parte de todos los archivos de datos originales fusionados se consideran como 'parámetros no comunes'. Todos los datos incluidos en un archivo de datos fusionado que corresponden tanto a los 'parámetros comunes' como a los 'parámetros no comunes' se miden en un número variable de partes alícuotas de la muestra ( $n$ =número de partes alícuotas de la muestra medidas). Por consiguiente,  $C_{ih}$  representa el  $h^{\text{ésimo}}$  parámetro no común de la  $i^{\text{ésima}}$  parte alícuota de la muestra, donde  $i=1, \dots, n$  y  $h=1, \dots, k$ . Igualmente,  $nc_{ih}$  representa el  $h^{\text{ésimo}}$  parámetro no común de la  $i^{\text{ésima}}$  parte alícuota de la muestra, donde  $i=1, \dots, n$  y  $h=1, \dots, m$ . Si el número de eventos en cada archivo de datos iniciales que corresponde a una sola parte 45 alícuota de la muestra se denota como  $j$ ,  $C_{ih}$  y  $nc_{ih}$  pueden representarse como vectores en  $\mathfrak{R}^j$  para cada  $i=1, \dots, n$ , y  $h=1, \dots, k$  y para cada  $i=1, \dots, n$ , y  $h=1, \dots, m$ , respectivamente. Además, si  $C_{ih}(g)$ , y  $nc_{ih}(g)$  (donde  $g$  representa los eventos medidos en una parte alícuota de la muestra) se consideran como los  $g^{\text{ésimos}}$  componentes de los vectores  $C_{ih}$  y  $nc_{ih}$ , respectivamente, para un evento dado 's' (para cada parte alícuota

de la muestra  $i=1, \dots, n$ )  $C_i(s) \equiv (C_{i1}(s), C_{i2}(s), \dots, C_{ik}(s)) \in \mathfrak{R}^k$  y para un evento dado 'p' (para cada tubo  $i=1, \dots, n$ )  $nc_i(p) \equiv (nc_{i1}(p), nc_{i2}(p), \dots, nc_{im}(p)) \in \mathfrak{R}^m$ .

Basándose en las notaciones definidas anteriormente, si se consideran dos de los archivos de datos fusionados que corresponden a dos  $n_\alpha$  y  $n_\beta$  partes alícuotas de la muestra, donde  $n_\alpha \neq n_\beta$  y  $n_\alpha, n_\beta = 1, 2, \dots, n$ , cada una conteniendo datos medidos para  $j$  eventos, se ejecuta la siguiente rutina:

55 a) Considerar un evento  $s$ , o un conjunto de eventos  $\Omega$ . Tomar el valor de  $c_\alpha(s)$  (en la parte alícuota  $\alpha$  de la muestra). Alternativamente, calcular cualquier función  $F_C(\Omega) \in \mathfrak{R}^k$  (por ejemplo

$F_c(\Omega) = \text{media}(c_\alpha(\xi))$  para  $\xi \in \Omega$ ). Obsérvese que uno puede tener  $\Omega = s$ , y  $F_c(\Omega) \equiv c_\alpha(\Omega) = c_\alpha(s)$ , y de ese modo el último es, de hecho, una generalización del primero.

5 b) Buscar en  $n_\beta$  aquellos eventos "q" con cualquier similitud mensurable (por ejemplo, la distancia mínima, en cuyo caso se podría usar el principio de los vecinos más cercanos), con  $c_\alpha(s)$ . Alternativamente, el mismo procedimiento con  $F_c(\Omega)$  (en lugar de  $c_\alpha(s)$ ). Supongamos que estos "q" eventos están marcados como L ( $L \in \mathcal{R}^q$  es un conjunto de marcas).

Entonces, desarrollar un conjunto  $NC_\beta(L)$ .

c) Evaluar la distribución de  $nc_\beta(L)$  y tratar la situación de acuerdo con la dispersión de  $nc_\beta(L)$ .

10 Al final de este proceso y después de ejecutar la rutina para todos los eventos que corresponden a la parte alícuota de la muestra  $n_\alpha$ , estarán rellenas todas las columnas vacías para los 'j' eventos que corresponden a la parte alícuota de la muestra  $n_\alpha$ . A su vez, después de ejecutar la rutina para todos los eventos y todos los pares de partes alícuotas de las muestras, se habrá construido una matriz con  $(n \times j)$  líneas y  $k + m$  columnas que contienen la información de  $k + m$  parámetros para todos los  $(n \times j)$  eventos.

15 Según lo que se describe anteriormente, el cálculo de valores y una medida de incertidumbre correspondiente para aquellos parámetros no medidos para un grupo de eventos en el archivo de datos fusionado, se realiza para cada evento 's' individual; alternativamente, el cálculo de valores y una medida de incertidumbre correspondiente para aquellos parámetros no medidos para un grupo de eventos en el archivo de datos fusionado, se realiza para un grupo 'Ω' de dos o más eventos.

20 Además, también puede estimarse uno o más de los parámetros medidos en una o más partes alícuotas de la muestra. Tal estimación puede usarse como un control de calidad interno del procedimiento de estimación.

25 Según esta invención, para calcular valores y una medida de incertidumbre correspondiente para aquellos parámetros no medidos para un grupo de eventos en el archivo de datos fusionado, se usan todos los parámetros comunes a dos o más partes alícuotas de la muestra o solo parte de ellos. Además, en esta invención, todos los eventos del archivo de datos fusionado o solo parte de ellos, se usan para calcular valores y una medida de incertidumbre correspondiente para aquellos parámetros no medidos directamente en la citometría de flujo para un grupo de eventos del archivo de datos fusionado.

30 Los datos contenidos en esta nueva matriz rellena pueden analizarse usando cualquier programa de software convencional desarrollado para el análisis específico de archivos de datos de citometría de flujo.

La invención se ilustrará mediante un ejemplo que no limita sus áreas de aplicación de la siguiente manera:

**Ejemplo 1.**

**1.- Muestra:**

35 Se obtuvieron mediante punción venosa cinco ml de sangre periférica (PB) de un voluntario sano y se colocaron en un tubo VACUTAINER™ (Becton Dickinson, New Jersey, NJ) que contiene EDTA como anticoagulante.

**2.- Preparación de la muestra:**

40 Después de mezclar cuidadosamente la muestra, 200 µl de la muestra de PB que contienen  $10^6$  células nucleadas fueron colocados en cinco tubos de replicación diferentes. Después, a cada tubo se añadió una de las siguientes cinco combinaciones de 6 colores de anticuerpos monoclonales, cada anticuerpo monoclonal conjugado con un fluorocromo diferente (FITC/PE/PE-TR/PerCP-Cy5.5/APC/APC-Cy7) añadiéndose en cantidades de saturación en un volumen de 5 µl:

- 45
- 1)CD22/CD23/CD19/CD45/CD5/CD20,
  - 2)CD43/CD79b/CD19/CD45/CD5/CD20,
  - 3)CD9/CD27/CD19/CD45/CD5/CD20,
  - 4)CD11c/CD10/CD19/CD45/CD5/CD20 y,
  - 5)CD103/CD25/CD19/CD45/CD5/CD20. Después de mezclar cuidadosamente, las partes

alícuotas de la muestra fueron incubadas durante 15 minutos a temperatura ambiente en la oscuridad. Después de esta incubación, se añadieron 2 ml de solución FACSlising (Becton Dickinson Biosciences, San José, CA), diluidos al 1/10 (vol/vol) en agua destilada. Después de mezclar cuidadosamente, las partes alícuotas de la muestra teñidas y lisadas fueron incubadas durante otros 10 minutos a temperatura ambiente en la oscuridad. Después, las células fueron centrifugadas (5 minutos a 540 g), lavadas en 2 ml de tampón fosfato salino (PBS, pH=7,6) y vueltas a suspender en 500  $\mu$ l de PBS/tubo.

### 3.- Adquisición de datos:

Después de mezclar cuidadosamente cada parte alícuota de la muestra teñida, la dispersión de luz y las emisiones de fluorescencia de las células teñidas fueron medidas en un citómetro de flujo FACSARIA (Beckton/Dickinson Biosciences –BDB-, San José, CA) equipado con dos luces láser que emiten a 488 y 635 nm, usando el programa de software FACSDIVA (BDB). Para cada parte alícuota de la muestra analizada, se recopiló un único archivo de datos que contenía información sobre la dispersión de luz frontal (FSC) y lateral (SSC), y las emisiones de fluorescencia de FITC, PE, PE-TR, PE-Cy5.5, APC y APC-Cy7 para hasta  $5 \times 10^5$  células/parte alícuota de la muestra. Además, también se registró la información que corresponde a la secuencia de adquisición (tiempo de adquisición) de cada evento. Se realizó adquisición de datos para los diferentes tubos en la secuencia definida en la sección anterior acerca de la “preparación de la muestra”, usando un umbral de FSC establecido en 200.

### 4.- Manipulación de datos y análisis de datos:

Los datos contenidos en cada uno de los cinco archivos de datos generados fueron exportados luego a una base de datos Excel (Microsoft, Madrid, España) usando el programa de software asistente FCS 1a.a (Universidad de Stanford, Stanford, CA). Una vez colocados en la base de datos Excel, las cinco bases de datos individuales, que corresponden cada una a una parte alícuota diferente de la muestra, fueron fusionadas en una sola base de datos donde el primer evento medido de la segunda parte alícuota de la muestra fue colocado en una línea inmediatamente posterior a la que corresponde al último evento medido para la primera parte alícuota de la muestra, el primer evento medido de la tercera parte alícuota de la muestra fue colocado en una línea inmediatamente posterior a la que corresponde al último evento medido para la segunda parte alícuota de la muestra, el primer evento medido de la cuarta parte alícuota de la muestra fue colocado en una línea inmediatamente posterior a la que corresponde al último evento medido para la tercera parte alícuota de la muestra y el primer evento medido de la quinta parte alícuota de la muestra fue colocado en una línea inmediatamente posterior a la que corresponde al último evento medido para la cuarta parte alícuota de la muestra. Además, fueron creadas las siguientes columnas: columna 1, FSC; columna 2, SSC; columna 3, CD22; columna 4, CD23; columna 5, CD19; columna 6, CD45; columna 7, CD5; columna 8, CD20; columna 9, CD43; columna 10, CD79b; columna 11, CD9; columna 12, CD27; columna 13, CD11c; columna 14, CD10; columna 15, CD103; y columna 16: CD25. Los datos que corresponden a la parte alícuota de la muestra teñida con la combinación multicolor CD22/CD23/CD19/CD45/CD5/CD20 de reactivos de anticuerpos monoclonales fueron registrados en las columnas 1 a 8, ambas incluidas, desde la línea 1 hasta la línea 500.000; los datos medidos para la segunda parte alícuota de la muestra (tinción multicolor CD43/CD79b/CD19/CD45/CD5/CD20) fueron almacenados en las columnas 1, 2 y 5 a 10, desde la línea 500.001 hasta la línea 1.000.000; para la tercera parte alícuota de la muestra (tinción CD9/CD27/CD19/CD45/CD5/CD20) la información fue almacenada en las columnas 1, 2, 5 a 8, 11 y 12, desde la línea 1.000.001 hasta la línea 1.500.000; los datos medidos para la cuarta parte alícuota de la muestra (tinción CD11c/CD10/CD19/CD45/CD5/CD20) fueron almacenados en las columnas 1, 2, 5 a 8, 13 y 14, desde la línea 1.500.001 hasta la línea 2.000.000; para la quinta parte alícuota de la muestra (tinción multicolor CD103/CD25/CD19/CD45/CD5/CD20), la información fue registrada en las columnas 1, 2, 5 a 8, 15 y 16, desde las líneas 2.000.001 hasta la 2.500.000. Las celdas de la base de datos que permanecían vacías fueron rellenas después con valores de datos estimados, calculados como se describe más adelante.

Las emisiones de fluorescencia asociadas a anticuerpos comunes medidas (CD19-PE-TR/CD45-PE-Cy5.5/CD5-APC/CD20-APC-Cy7) más las medidas de dispersión de luz tanto frontal (FSC) como lateral (SSC) fueron consideradas como los ‘parámetros comunes’ “k” a todos los archivos de datos originales (k=6, ya que el número global de ‘parámetros comunes’ fue de 6). Las medidas de emisión de fluorescencia para todos los demás anticuerpos monoclonales (CD22-FITC/CD23-PE/CD43-FITC/CD79b-PE/CD9-FITC/CD27-PE/CD11c-FITC/CD10-PE/CD103-FITC/CD25-PE) fueron consideradas como los ‘parámetros no comunes’ “m” (m = 10, ya que el número global de ‘parámetros no comunes’ fue de 10). Estos parámetros comunes y no comunes fueron medidos en las cinco partes alícuotas de la muestra “n” (n = 5, que corresponde a las cinco partes alícuotas de la muestra teñidas, preparadas y medidas en el citómetro de flujo). Por consiguiente, cada uno de los cinco archivos de datos originales almacenados que corresponden a cada una de las cinco partes alícuotas de la muestra medidas, contenían seis parámetros comunes y dos parámetros diferentes donde  $c_{ih}$  representa el  $h^{\text{ésimo}}$  ( $h=1, \dots, 6$ ) parámetro común del tubo  $i^{\text{ésimo}}$  ( $i=1, 2, \dots, 5$ ) y  $nc_{ih}$  representa el  $h^{\text{ésimo}}$  ( $h=1, \dots, 10$ ) parámetro no común del tubo  $i^{\text{ésimo}}$  ( $i=1, 2, \dots, 5$ ). El

número total de eventos medidos para cada parte alícuota de la muestra ("j") fue de 500.000 eventos

(j=500.000); por lo tanto  $c_{ih}$  y  $nc_{ih}$  fueron representados como vectores en  $\mathcal{R}^{500000}$  para cada  $i=1, \dots, 5$  y  $h=1, \dots, 6$  y para cada  $i=1, \dots, 5$  y  $h=1, \dots, 10$ , respectivamente. Además, al considerar que  $c_{ih}(g)$  y  $nc_{ih}(g)$  (para  $g=1, 2, \dots, 500.000$ ) son las  $g^{\text{ésimas}}$  componentes de los vectores  $c_{ih}$  y  $nc_{ih}$ , respectivamente, para un evento dado 's' y un evento dado 'p' (para cada parte alícuota de la muestra  $i=1, \dots, 5$ ) consideramos  $c_i(s) \equiv (c_{i1}(s),$

$c_{i2}(s), \dots, c_{i6}(s)) \in \mathcal{R}^6$  y  $nc_i(p) \equiv (nc_{i1}(p), nc_{i2}(p), \dots, nc_{i10}(p)) \in \mathcal{R}^{10}$ , respectivamente. Considerando la notación definida previamente, se ejecutó una rutina para las partes alícuotas de la muestra  $n_1$  y  $n_2$ , para cada uno de los j eventos de  $n_1$ , de la siguiente manera:

10 Para un evento dado 's', se calculó el valor de  $c_1(s)$  en la parte alícuota de la muestra  $n_1$ . Luego se buscaron los q vecinos más cercanos a  $c_1(s)$  en  $n_2$ . Después, se construyó un conjunto de

$nc_2(L)$  puntos de datos, suponiendo que estos q puntos están marcados como L ( $L \in \mathcal{R}^q$  es un conjunto de marcas).

15 Si el  $nc_2(L)$  construido era compacto y estaba formado por una sola nube de puntos de datos, entonces estimamos  $nc_{1\_EST} = \text{mediana}(nc_2(L_{SUB}))$ , donde  $L_{SUB}$  es cualquier subconjunto de L. A su vez, si  $nc_2(L)$  no era compacto y había dos o más nubes de puntos de datos, calculamos la probabilidad de aparición de cada nube, hicimos un sorteo teniendo en cuenta la probabilidad de aparición de cada una de las nubes, escogimos una de estas nubes y estimamos  $nc_{1\_EST} = \text{mediana}(nc_2(L_{SUB}))$ , donde  $L_{SUB}$  es cualquier subconjunto de L, solamente para los puntos de esta nube.

20 Al final de este procedimiento y después de ejecutar la rutina para todos los eventos que corresponden a la parte alícuota de la muestra  $n_1$ , fueron rellenas todas las columnas vacías para los 'j' eventos que corresponden a la parte alícuota de la muestra  $n_1$ . A su vez, después de ejecutar la rutina para todos los eventos y todos los pares de partes alícuotas de la muestra, se construyó una matriz con 2.500.000 (500.000 x 5) líneas y 16 (6 + 10) columnas que contenía la información de 16 parámetros para los 2.500.000 eventos. Los datos contenidos en esta nueva matriz rellena fueron analizados luego usando un programa de software de citometría de flujo convencional (software FACSDIVA).

## REIVINDICACIONES

1. Un procedimiento para la generación de nuevos archivos de datos de citometría de flujo a partir de archivos de datos originales que contienen datos de parámetros acerca de eventos medidos en diferentes partes alícuotas de la misma muestra, en el que cada archivo de datos originales comprende datos de parámetros comunes que corresponden a parámetros medidos para todas las partes alícuotas y datos de parámetros no comunes que corresponden a parámetros medidos solo para algunas partes alícuotas, comprendiendo el procedimiento la etapa de:
- 5 a) fusionar dos o más archivos de datos originales en un archivo;
- caracterizado por** comprender además las etapas de,
- 10 b) para cada evento del archivo fusionado, estimar los datos de parámetros no comunes que no fueron medidos directamente en el citómetro de flujo basándose en eventos que pertenecen a partes alícuotas para las cuales dicho parámetro no común ha sido medido por el citómetro de flujo, así como una medida de incertidumbre correspondiente, y;
- 15 c) reconstruir un nuevo archivo de datos que contiene datos de parámetros medidos y datos de parámetros estimados.
2. Procedimiento según la reivindicación 1, en el que antes de fusionar los archivos de datos originales, dos o más partes alícuotas de una muestra son teñidas por separado con un panel de anticuerpos monoclonales.
3. Procedimiento según una cualquiera de las reivindicaciones 1 y 2, en el que la muestra contiene células normales, células neoplásicas o una mezcla tanto de células normales como neoplásicas.
- 20 4. Procedimiento según una cualquiera de las reivindicaciones 1 a 3, en el que la muestra contiene sangre periférica.
5. Procedimiento según una cualquiera de las reivindicaciones 1 a 3, en el que la muestra comprende médula ósea.
- 25 6. Procedimiento según una cualquiera de las reivindicaciones 1 a 3, en el que la muestra contiene líquido espinal.
7. Procedimiento según una cualquiera de las reivindicaciones 1 a 3, en el que la muestra comprende ganglio linfático.
8. Procedimiento según una cualquiera de las reivindicaciones 1 a 3, en el que la muestra comprende líquido ascítico.
- 30 9. Procedimiento según una cualquiera de las reivindicaciones 1 a 3, en el que la muestra comprende efusión pleural.
10. Procedimiento según una cualquiera de las reivindicaciones 1 a 3, en el que la muestra comprende líquido sinovial.
- 35 11. Procedimiento según una cualquiera de las reivindicaciones 1 a 3, en el que la muestra comprende una suspensión de células individuales preparada a partir de tejido sólido.
12. Procedimiento según una cualquiera de las reivindicaciones 1 a 11, en el que el panel de reactivos está formado por dos o más combinaciones de anticuerpos monoclonales directamente conjugados a fluorocromos.
- 40 13. Procedimiento según una cualquiera de las reivindicaciones 1 a 12, en el que las diferentes combinaciones de anticuerpos monoclonales conjugados a fluorocromos usadas para teñir distintas partes alícuotas de la muestra consisten en reactivos de anticuerpos monoclonales que son idénticos en todas las combinaciones y reactivos de anticuerpos monoclonales que solo se usan en una cualquiera o parte de todas las combinaciones de reactivos de anticuerpos monoclonales usadas para teñir la muestra.
- 45 14. Procedimiento según una cualquiera de las reivindicaciones 1 a 13, en el que cada combinación de reactivos de anticuerpos monoclonales contiene, además de los anticuerpos monoclonales, otras sondas que consisten en fluorocromos específicos para la medición de colorantes de colorantes de ácidos nucleicos, mitocondrias y otros componentes celulares y funciones celulares.
15. Procedimiento según una cualquiera de las reivindicaciones 1 a 14, en el que cada anticuerpo

monoclonal en cada combinación está conjugado a un fluorocromo diferente y cada combinación de anticuerpos monoclonales tiene, en común, uno o más reactivos de anticuerpos monoclonales conjugados a fluorocromos.

- 5 16. Procedimiento según una cualquiera de las reivindicaciones 1 a 15, en el que el número de emisiones de fluorescencia en cada combinación de múltiples anticuerpos monoclonales comprende dos o más fluorocromos diferentes, cada uno vinculado a un anticuerpo monoclonal diferente, cuya emisión de fluorescencia es distinguible de la de los otros reactivos de anticuerpos monoclonales conjugados a fluorocromos en la combinación.
- 10 17. Procedimiento según una cualquiera de las reivindicaciones 1 a 16, en el que una combinación de fluorocromos compatibles se selecciona de isotiocianato de fluoresceína (FITC), ficoeritrina (PE), proteína peridina clorofila (PerCP), alofocianina, alexa flúor 488, alexa 647, alexa flúor 610, alexa 710, alexa flúor 405, cianina 5 (Cy5), cianina 5.5 (Cy5.5), azul Pacífico (PB), amarillo cascada, azul cascada y conjugados de los mismos unidos a PE, a APC, o a PerCP, puntos cuánticos o cualquier fluorocromo compatible adicional o tándem de fluorocromos.
- 15 18. Procedimiento según una cualquiera de las reivindicaciones 1 a 17, en el que se almacena un archivo de datos para cada parte alícuota de la muestra medida en el citómetro de flujo.
19. Procedimiento según una cualquiera de las reivindicaciones 1 a 17, en el que se almacena un archivo de datos que contiene datos para dos o más partes alícuotas de la muestra medidas en el citómetro de flujo.
- 20 20. Procedimiento según una cualquiera de las reivindicaciones 1 a 19, en el que se mide un número idéntico de eventos para cada parte alícuota de la muestra.
21. Procedimiento según una cualquiera de las reivindicaciones 1 a 19, en el que se miden diferentes números de eventos para cada parte alícuota de la muestra.
- 25 22. Procedimiento según una cualquiera de las reivindicaciones 1 a 21, en el que tanto los archivos de datos originales como el archivo de datos fusionado consisten en matrices de datos en las que cada dato de parámetro diferente medido relacionado con las características de dispersión de la luz y las emisiones de fluorescencia de una célula se coloca en una columna diferente y los datos sobre cada evento celular medido se colocan en una línea diferente.
- 30 23. Procedimiento según una cualquiera de las reivindicaciones 1 a 22, en el que los archivos de datos originales y los eventos contenidos en cada uno de dichos archivos de datos se pegan en una secuencia relacionada directamente con la secuencia en la que fueron medidos en el citómetro de flujo.
24. Procedimiento según una cualquiera de las reivindicaciones 1 a 22, en el que los archivos de datos originales y los eventos contenidos en cada uno de dichos archivos de datos se pegan en una secuencia relacionada inversamente con la secuencia en la que fueron medidos en el citómetro de flujo.
- 35 25. Procedimiento según una cualquiera de las reivindicaciones 1 a 22, en el que los archivos de datos originales y los eventos contenidos en cada uno de dichos archivos de datos se pegan en una secuencia que no está relacionada con la secuencia en la que fueron medidos en el citómetro de flujo.
- 40 26. Procedimiento según una cualquiera de las reivindicaciones 1 a 25, en el que la secuencia de parámetros en las columnas del archivo de datos fusionado está relacionada directamente con la secuencia en la que fueron medidos en el citómetro de flujo.
27. Procedimiento según una cualquiera de las reivindicaciones 1 a 25, en el que la secuencia de parámetros en las columnas del archivo de datos fusionado está relacionada inversamente con la secuencia en la que fueron medidos en el citómetro de flujo.
- 45 28. Procedimiento según una cualquiera de las reivindicaciones 1 a 25, en el que la secuencia de parámetros en las columnas del archivo de datos fusionado no está relacionada con la secuencia en la que fueron medidos en el citómetro de flujo.
29. Procedimiento según una cualquiera de las reivindicaciones 1 a 28, en el que el número global de parámetros estimados para los cuales se asignan valores a cada evento celular individual incluido en el nuevo archivo de datos fusionado es igual o inferior al número total de parámetros que difieren entre dos o más de las partes alícuotas de la muestra medidas en el citómetro de flujo.
- 50 30. Procedimiento según las reivindicaciones 1 a 29 en el que se usan procedimientos de estimación estadística basados en la similitud para calcular los valores y una medida de incertidumbre

correspondiente para parámetros no comunes.

31. Procedimiento según una cualquiera de las reivindicaciones 1 a 30, en el que para calcular los valores y una medida de incertidumbre correspondiente para parámetros no comunes, se usan todos los parámetros comunes.
- 5 32. Procedimiento según una cualquiera de las reivindicaciones 1 a 30, en el que para calcular los valores y una medida de incertidumbre correspondiente para parámetros no comunes, solo se usa una parte de los parámetros comunes.
- 10 33. Procedimiento según una cualquiera de las reivindicaciones 1 a 32, en el que para calcular los valores y una medida de incertidumbre correspondiente para parámetros no comunes, se usan todos los eventos del archivo de datos fusionado.
34. Procedimiento según una cualquiera de las reivindicaciones 1 a 32, en el que para calcular los valores y una medida de incertidumbre correspondiente para parámetros no comunes, solo se usa parte de todos los eventos del archivo de datos fusionado.
- 15 35. Procedimiento según una cualquiera de las reivindicaciones 1 a 34, en el que el cálculo de los valores y una medida de incertidumbre correspondiente para parámetros no comunes, se realiza para eventos individuales.
36. Procedimiento según una cualquiera de las reivindicaciones 1 a 34, en el que el cálculo de los valores y una medida de incertidumbre correspondiente para parámetros no comunes, se realiza para un grupo de dos o más eventos.
- 20 37. Procedimiento según una cualquiera de las reivindicaciones 1 a 36, en el que uno o más de los parámetros medidos en una o más partes alícuotas de la muestra se estiman y usan como un control de calidad interno del procedimiento de estimación.
- 25 38. Procedimiento según una cualquiera de las reivindicaciones 1 a 37, en el que el nuevo archivo de datos que contiene los datos medidos y los datos estimados se analiza usando programas de software de citometría de flujo convencionales.