

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 609 412**

51 Int. Cl.:

G06F 19/24 (2011.01)
G06K 9/62 (2006.01)
G01N 33/569 (2006.01)
G01N 33/574 (2006.01)
G01N 33/50 (2006.01)
G06K 9/00 (2006.01)
G06F 19/20 (2011.01)
G06F 19/26 (2011.01)
G01N 15/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **17.02.2006 PCT/US2006/005778**
- 87 Fecha y número de publicación internacional: **24.08.2006 WO06089190**
- 96 Fecha de presentación y número de la solicitud europea: **17.02.2006 E 06735440 (7)**
- 97 Fecha y número de publicación de la concesión europea: **28.09.2016 EP 1859377**

54 Título: **Sistema para detectar células anómalas empleando un análisis multidimensional**

30 Prioridad:

18.02.2005 US 654265 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
20.04.2017

73 Titular/es:

**HEMATOLOGICS, INC. (100.0%)
3161 Elliott Avenue, Suite 200
Seattle, WA 98121, US**

72 Inventor/es:

**LOKEN, MICHAEL, R. y
JOSHI, SANJAYA, N.**

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 609 412 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

Sistema para detectar células anómalas empleando un análisis multidimensional

Antecedentes de la invención

Campo de la invención

- 5 Esta invención se dirige a un análisis multidimensional de características celulares medidas y, en particular, a un sistema, un método y un artículo para detectar células anómalas en un conjunto de ensayo de células empleando un análisis multidimensional de características celulares medidas empleando la citometría de flujo.

Descripción de la técnica relacionada

10 Un método para caracterizar poblaciones de células heterogéneas es mediante la citometría de flujo, originariamente desarrollada por Herzenberg y colaboradores (Science, 1969, 166(906):747-749; J. Histochem. Cytochem., 1976, 24(1):284-291; Clin. Chem., 1973, 19(8):813-816; Ann. N.Y. Acad. of Sci., 1975, 254:163-171). Empleando esta tecnología, las células se marcan con anticuerpos conjugados a tintes. La citometría de flujo puede detectar habitualmente 3, 4 o más marcadores inmunofluorescentes simultáneamente de una manera cuantitativa. Mediante la combinación de múltiples marcadores inmunofluorescentes con las propiedades de dispersión de luz de las células es posible distinguir no solo entre células de diferentes linajes, sino también entre células en diversos estadios de maduración dentro de esos linajes. Esto se determina basándose en los patrones de expresión de antígenos exclusivo de la superficie celular (véase, por ejemplo, Loken M.R., *et al.*, en Flow Cytometry in Hematology, Laerum O.D., Bjerksnes R. eds., Academic Press, Nueva York, pp. 31-42, 1992; Civin C.I., *et al.*, en "Concise Reviews in Clinical and Experimental Hematology", Martin J. Murphy ed., AlphaMed Press, Dayton, OH, 1992, pp. 149-159). Las poblaciones identificadas mediante el citómetro de flujo después pueden aislarse empleando los componentes electrónicos disponibles en el instrumento.

15 La citometría de flujo de múltiples parámetros se emplea en la actualidad para detectar una diversidad de leucemias. Sin embargo, las técnicas actuales requieren que un profesional, concretamente alguien versado en la citometría de flujo y la hematopatología, tal como un médico, realice el análisis de los datos, que es un proceso muy largo. Es necesario un largo proceso de aprendizaje para educar al profesional para que distinga entre poblaciones celulares normales y anómalas. Además, cuando se emplea la citometría de flujo para controlar la respuesta de un paciente a una terapia, las técnicas convencionales requieren el uso de paneles específicos de paciente para detectar la enfermedad residual.

20 Por consiguiente, en la técnica es necesaria una tecnología para mejorar la precisión de la detección y para simplificar el análisis de los datos. La presente invención puede satisfacer estas y otras necesidades.

Breve resumen de la invención

Las realizaciones de la presente invención son:

1. Un sistema de diagnóstico que comprende un medio para llevar a cabo un método de caracterización de un conjunto de ensayo de células biológicas que comprende:
- 35 a) cartografiar cada célula en el conjunto de ensayo de células biológicas a un punto correspondiente en un espacio n-dimensional, en el que $n \geq 3$ empleando un protocolo definido en el que al menos algunas de las n-dimensiones son datos de intensidad fluorescente de las células, y los puntos correspondientes forman un conjunto de ensayo de puntos;
- 40 b) comparar el conjunto de ensayo de puntos con un conjunto definido de agrupaciones de células biológicas normales, es decir, un conjunto definido de agrupaciones normales, en el espacio n-dimensional, en el que cada agrupación en el conjunto definido de agrupaciones normales se corresponde con un nivel de maduración dentro de un linaje celular, y en el que, para el conjunto de agrupaciones normales, se define una línea centroide de referencia y un radio, y la línea centroide de referencia y el radio forman diversas conformaciones de agrupaciones, en el que la línea centroide de referencia para el conjunto de agrupaciones se basa en un conjunto de puntos de referencia, y en el que el radio es una función de una posición en la línea centroide de referencia, en el que la línea centroide de referencia y el radio se ajustan de una manera iterativa empleando al menos un análisis estadístico y la entrada de un usuario, en el que el ajuste incluye determinar una distancia desde un punto centroide identificado de cada agrupación y un punto más cercano en la línea centroide de referencia;
- 45 c) determinar, para cada punto del conjunto de ensayo de puntos, basándose en la línea centroide de referencia y el radio determinados, si se encuentra fuera del conjunto definido de agrupaciones normales basándose en la comparación; y
- 50 d) caracterizar el conjunto de ensayo de células biológicas basándose al menos en parte en una determinación de una cantidad de células anómalas en el conjunto de ensayo de células biológicas basándose en la determinación de la etapa c).

2. El sistema de diagnóstico de la realización 1, en el que la línea centroide de referencia para el conjunto de agrupaciones se basa en un conjunto de puntos de referencia y en los puntos centroides de las agrupaciones en el conjunto de agrupaciones normales.
- 5 3. El sistema de diagnóstico de la realización 1, que es para diagnosticar el cáncer y comprende, además de las etapas de la realización 1, las siguientes etapas:
- exponer cada célula en un conjunto normal de células biológicas a una pluralidad de cuatro o más reactivos empleado un primer protocolo;
- medir una correspondiente pluralidad de intensidades de fluorescencia de cada célula en el conjunto normal de células biológicas empleando un segundo protocolo;
- 10 cartografiar cada célula en el conjunto normal de células biológicas a un punto correspondiente en un espacio n-dimensional, basándose al menos en parte en la pluralidad de intensidades de fluorescencia medidas en la célula en el conjunto normal de células biológicas, en el que los puntos correspondientes forman un conjunto normal de puntos, en el que la definición de la línea centroide de referencia y el radio para el conjunto de agrupaciones normales se basa en el cartografiado del conjunto normal de puntos en el espacio n-dimensional;
- 15 exponer cada célula en el conjunto de ensayo de células biológicas a la pluralidad de reactivos empleando el primer protocolo;
- medir una correspondiente pluralidad de intensidades de fluorescencia de cada célula en el conjunto de ensayo de células biológicas empleando el segundo protocolo;
- 20 cartografiar cada célula en el conjunto de ensayo de células biológicas a un punto correspondiente en un espacio n-dimensional, basándose al menos en parte en la pluralidad de intensidades de fluorescencia medidas en la célula en el conjunto de ensayo de células biológicas, en el que los puntos correspondientes forman un conjunto de ensayo de puntos;
- comparar el conjunto de ensayo de puntos con el conjunto de agrupaciones normales; y
- caracterizar y diagnosticar el conjunto de ensayo de células biológicas basándose en la comparación.
- 25 4. El sistema de diagnóstico de la realización 1, que comprende además la etapa de:
- cartografiar cada célula en el conjunto normal de células biológicas a un punto correspondiente en un espacio n-dimensional utilizando el protocolo definido, en el que los puntos correspondientes forman un conjunto normal de puntos, en el que la definición de la línea centroide de referencia y el radio para el conjunto de agrupaciones normales en el espacio n-dimensional se basa en el cartografiado del conjunto normal de puntos en el espacio n-
- 30 dimensional.
5. El sistema de diagnóstico de la realización 1, que comprende además las etapas de:
- representar el conjunto de ensayo en un diagrama de coordenadas cartesianas que comprende un primer eje que se corresponde con una maduración de la célula dentro de un linaje celular, y un segundo eje que se corresponde con una frecuencia de aparición; y
- 35 representar, en el diagrama de coordenadas cartesianas, el conjunto de agrupaciones normales en el espacio n-dimensional.
6. El sistema de diagnóstico de cualquiera de las realizaciones anteriores para diagnosticar el cáncer, en el que al menos una agrupación en el conjunto de agrupaciones es un hiperelipsoide.
- 40 En una realización, un conjunto normal de células se caracteriza empleando la citometría de flujo. Se define un centroide y un radio para un conjunto de agrupaciones en un espacio n-dimensional que se corresponden con una maduración normal para un linaje celular en el conjunto normal de células. Un conjunto de ensayo de células se caracteriza empleando la citometría de flujo, y la caracterización se compara con el conjunto de agrupaciones. Esta estrategia facilita la detección de niveles bajos de células tumorales basándose en sus diferencias fenotípicas con respecto a sus homólogos normales, según se evalúa mediante un análisis de los datos complejos procedentes de
- 45 poblaciones normales y anómalas.
- En un aspecto, el sistema de diagnóstico comprende un medio para llevar a cabo un método que comprende exponer una célula a una pluralidad de cualquier número de reactivos. Algunos instrumentos son capaces de producir nueve o más colores. El uso de más reactivos y colores facilita la caracterización de las células.
- 50 En otro aspecto, una realización de un sistema de diagnóstico comprende: un controlador; una memoria; una interfase de datos; una interfase de control; y un motor gráfico, en el que el sistema de diagnóstico está configurado para comparar un conjunto de ensayo de datos con un conjunto de agrupaciones normales en un espacio n-

dimensional definido por un centroide y un radio, y en el que una agrupación en el conjunto de agrupaciones normales se corresponde con un nivel de maduración normal dentro de un linaje celular.

5 En otro aspecto, una realización de un sistema para diagnosticar un conjunto de ensayo de células comprende: un medio para definir un conjunto de agrupaciones normales que se corresponden con un linaje celular normal; y un medio para comparar el conjunto de ensayo de células con el conjunto de agrupaciones normales.

Estos y otros aspectos de la presente invención serán evidentes tras remitirse a la siguiente descripción detallada y dibujos adjuntos.

Breve descripción de los dibujos

10 La figura 1 es un diagrama de bloques funcional de un sistema que ejecuta una realización de un método para diagnosticar un conjunto de ensayo de células.

La figura 2 es un diagrama esquemático de una estructura de datos adecuada para almacenar datos relacionados con un conjunto de células biológicas.

La figura 3 es un diagrama esquemático de una estructura de datos adecuada para almacenar información relacionada con el procesamiento de los datos contenidos en la estructura de datos ilustrada en la figura 2.

15 Las figuras 4A a 9A y 4B a 9B son ilustraciones de los datos multidimensionales proyectados en presentaciones bidimensionales generadas por un sistema, tal como el sistema ilustrado en la figura 1.

Las figuras 10A y 10B son ilustraciones de los datos multidimensionales proyectados en presentaciones tridimensionales generadas por un sistema, tal como el sistema ilustrado en la figura 1.

20 Las figuras 11A y 11B ilustran un menú de una interfase gráfica del usuario generada por un sistema, tal como el sistema ilustrado en la figura 1.

Las figuras 12A a 17A y 12B a 17B son ilustraciones de los datos multidimensionales proyectados en presentaciones pseudo-bidimensionales generadas por un sistema, tal como el sistema ilustrado en la figura 1.

Las figuras 18A a 18C son diagramas de flujo que ilustran la operación de un sistema para definir un centroide y un radio normal para un conjunto de agrupaciones normales que se corresponden con un linaje celular normal.

25 La figura 19 es un diagrama de flujo que ilustra la operación de un sistema para definir una línea centroide para un conjunto de agrupaciones normales que se corresponden con un linaje celular normal.

Las figuras 20A y 20B son ilustraciones de los datos multidimensionales proyectados en presentaciones pseudo-tridimensionales generadas por un sistema, tal como el sistema ilustrado en la figura 1.

30 La figura 21 es un diagrama de flujo que ilustra la operación de un sistema para definir una línea centroide y un radio normal para un conjunto de agrupaciones normales que se corresponden con un linaje celular normal.

La figura 22 es un diagrama de flujo que ilustra la operación de un sistema para determinar qué puntos en el conjunto de ensayo de datos están contenidos dentro de un conjunto de agrupaciones normales en un espacio n-dimensional.

35 Las figuras 23A y 23B son ilustraciones de los datos multidimensionales proyectados en presentaciones tridimensionales generadas por un sistema, tal como el sistema ilustrado en la figura 1.

La figura 24 es un diagrama de flujo que ilustra la operación de un sistema para comparar un conjunto de ensayo de datos con puntos centroides definidos para un conjunto de agrupaciones normales.

La figura 25 es un diagrama esquemático de una estructura de datos adecuada para almacenar información para definir un centroide y un radio para un conjunto de agrupaciones normales en un espacio n-dimensional.

40 La figura 26 es una ilustración de los datos multidimensionales proyectados en una presentación bidimensional generada por un sistema, tal como el sistema ilustrado en la figura 1.

Descripción detallada de la invención

45 Los productos génicos pueden identificarse sobre la superficie celular o en el citoplasma de células empleando anticuerpos monoclonales específicos. Puede emplearse la citometría de flujo para detectar múltiples marcadores inmunofluorescentes simultáneamente de una manera cuantitativa. La técnica de la tinción inmunofluorescente es muy conocida y puede realizarse según cualquiera de una diversidad de protocolos, tales como los descritos en Current Protocols in Cytometry (John Wiley & Sons, NY, NY, eds. J. Paul Robinson, *et al.*). En general, una muestra biológica, tal como sangre periférica, médula ósea, tejido de nódulo linfático, sangre de cordón umbilical, tejido de timo, tejido de un sitio de infección, tejido de bazo, tejido tumoral y similares, se recoge de un sujeto y las células se

aíslan de ella empleando técnicas conocidas en la técnica. En una realización, la muestra se recoge de un sujeto y los eritrocitos maduros se lisan empleando un tampón, tal como NH₄Cl tamponado. El resto de los leucocitos se lavan y después se incuban con anticuerpos (por ejemplo, anticuerpos monoclonales) conjugados con cualquiera de una diversidad de tintes (fluoróforos) conocidos en la técnica (véase, por ejemplo, <http://colon.barradoble.com/glen/filters/ffflluorpn.htm> o <http://colon.barradoble.com/cellscience/bio-rad/fluorescence/fluorophoradata.htm>). Los tintes representativos en este contexto incluyen, pero no se limitan a FITC (isotiocianato de fluoresceína), R-ficoeritrina (PE), alofocianina (APC), Cy7®, y rojo de Texas.

Una amplia diversidad de anticuerpos conocidos en la técnica y anticuerpos específicos generados empleando técnicas muy conocidas en la técnica son útiles en el contexto de las realizaciones descritas en la presente. En general, los anticuerpos para su uso en los métodos descritos en la presente son específicos para un marcador celular de interés, tal como cualquiera de los marcadores de la superficie celular CD (véase, por ejemplo, el índice de CD en <http://colon.barradoble.com/ncbi/nlm/nih/gov/PROW/guide/45277084.html>; o Current Protocols in Immunology, John Wiley & Sons, NY, NY), citoquinas, proteínas de adhesión, marcadores de la superficie celular del desarrollo, antígenos tumorales u otras proteínas expresada por una población celular de interés. Un anticuerpo específico para casi cualquier proteína expresada por una célula es útil en el contexto de la presente descripción. Los anticuerpos ilustrativos incluyen, pero no se limitan a anticuerpos que se unen específicamente a CD3, CD33, CD34, CD8, CD4, CD56, CD19, CD14, CD15, CD16, CD13, CD38, CD71, CD11b, HLA-DR, glicoforina, CD45, CD20, CD5, CD7, CD2, CD10 y TdT.

Después de un periodo de incubación con un anticuerpo conjugado con un tinte, generalmente durante aproximadamente 20 minutos en la oscuridad (los tiempos de incubación y las condiciones pueden variar según los protocolos concretos), los leucocitos se lavan con disolución salina tamponada y se resuspenden en disolución salina tamponada que contiene una proteína para su introducción en un citómetro de flujo.

El citómetro de flujo analiza la población celular heterogénea célula a célula y puede clasificar las células basándose en la unión del anticuerpo monoclonal inmunofluorescente y las propiedades de dispersión de luz de cada célula (véase, por ejemplo, Immunol. Today, 2000, 21(8):383-390). La detección de la fluorescencia se realiza empleando tubos fotomultiplicadores; el número de detectores (canales) determina el número de parámetros ópticos que el instrumento puede examinar simultáneamente, mientras que los filtros de paso en banda aseguran que solo se recogen las longitudes de onda previstas. Así, la citometría de flujo puede detectar habitualmente múltiples marcadores inmunofluorescentes simultáneamente de una manera cuantitativa y puede medir otros parámetros, tales como la dispersión de luz directa (que es una indicación del tamaño de la célula) y la dispersión de luz de ángulo derecho (que es una indicación de la granularidad de la célula). Por consiguiente, puede diferenciarse una amplia diversidad de poblaciones celulares y clasificarse empleando la inmunofluorescencia y la citometría de flujo.

Por ejemplo, combinando 4 colores de inmunofluorescencia con los parámetros físicos de dispersión de luz directa (medición del tamaño de la célula) y dispersión de luz de ángulo derecho (medición de la granularidad de la célula), puede generarse un espacio de datos de seis dimensiones, en el que las poblaciones celulares específicas que se encuentran en la sangre o la médula ósea normales están restringidas a pequeñas porciones del espacio de datos. Tal como reconocerán los expertos en la técnica después de estudiar la memoria descriptiva, también pueden utilizarse más o menos de 4 colores de marcadores inmunofluorescentes. La excitación de los fluoróforos no se limita a la luz en el espectro visible; varios tintes, tales como la serie Indo (para medir el calcio intracelular) y la serie Hoechst (para análisis del ciclo celular) se excitan en la gama del ultravioleta. Así, algunos de los instrumentos disponibles en la técnica en la actualidad están configurados con fuentes emisoras de ultravioleta, tal como Becton Dickinson LSR II de 10 colores y 4 láseres. Además, mediante el empleo de un clasificador celular activado por fluorescencia disponible en el mercado, tal como FACSVANTAGE™ (Becton Dickinson, San José, CA), el clasificador EPICS® ALTRA™ (Beckman Coulter, Fullerton, CA) o el clasificador MOFLO® (DakoCytomation, Inc., Carpintería, CA), las poblaciones celulares también pueden clasificarse en fracciones purificadas.

La expresión génica observada durante el desarrollo de las células sanguíneas desde las células pluripotenciales hematopoyéticas hasta las células maduras que se encuentran en la sangre es un proceso muy regulado. Véase Civin C.I., Loken M.R., Cell Surface Antigens on Human Marrow Cells: Dissection of Hematopoietic Development Using Monoclonal Antibodies and Multiparameter Flow Cytometry, Int'l J. Cell Cloning, 5:1-16 (1987). Así, de modo específico, esta expresión muy controlada de los genes se produce no solo dentro de diferentes linajes de células sanguíneas, sino también durante diferentes estadios de maduración dentro de esos linajes. Véase Loken, M.R., Terstappen L.W.M.M., Civin C.I., Fackler, M.J., Flow Cytometry Characterization of Erythroid, Lymphoid and Monomyeloid Lineages in Normal Human Bone Marrow, Flow Cytometry in Hematology, Laerum O.D., Bjerksnes R., eds., Academic Press, Nueva York, pp. 31-42 (1992). No solo estos productos génicos aparecen y/o desaparecen en estadios concretos de la maduración, sino que la cantidad de estas glicoproteína se regula dentro de unos límites muy estrechos en células normales. Se ha demostrado que estas relaciones antigénicas se establecen en un estadio temprano del desarrollo fetal y que son constantes a lo largo de la vida adulta en células sanguíneas que están sufriendo un constante recambio y reposición. Véase LeBein T.W., Wormann B., Villablanca J.G., Law C.L., Shah V.O., Loken M.R., Multiparameter Flow Cytometric Analysis of Human Fetal Bone Marrow B Cells, Leukemia, 4:354-358 (1990). Estos patrones y relaciones de expresión génica durante la maduración de las células normales se mantienen después de una quimioterapia o incluso tras un trasplante de médula ósea. Véase Wells D.A., Sale G.E., Shulman H.E., Myerson D., Bryant E., Gooley T., Loken M.R., Multidimensional Flow Cytometry of Marrow Can

Differentiate Leukemic Lymphoblasts From Normal Lymphoblasts and Myeloblasts Following Chemotherapy and/or Bone Marrow Transplant, *Am. J. Clin. Path.*, 110:84-94 (1998). Por tanto, existe una regulación fuertemente coordinada de múltiples genes durante el desarrollo normal de las células sanguíneas, tanto en términos de temporización de la expresión, como en la regulación de las cantidades de productos génicos expresados sobre las superficies celulares.

Una comparación de la expresión normal de antígenos con los procesos neoplásicos indica que la regulación de la expresión génica está alterada en células neoplásicas. Esta alteración produce una relaciones antigénicas diferentes de las observadas durante la maduración normal de las células. Véase Hurwitz, C.A., Loken M.R., Graham M.L., Karp J.E., Borowitz M.J., Pullen D.J., Civin C.I., *Asynchronous Antigen Expression in B Lineage Acute Lymphoblastic Leukemia, Blood*, 72:299-307 (1998). Estos no son antígenos nuevos, sino que son los productos génicos normalmente expresados que han perdido la regulación coordinada que se encuentra en las células normales. Tanto la leucemia linfoblástica aguda ("ALL") como la leucemia mieloblástica aguda ("AML") expresan antígenos de forma anómala. Véase Terstappen L.W.M.M., Loken M.R., *Myeloid Cell Differentiation in Normal Bone Marrow and Acute Myeloid Leukemia Assessed by Multi-Dimensional Flow Cytometry, Anal. Cell Path.*, 2:229-240 (1990).

Los tipos de anomalías incluyen:

(1) Infidelidad de linaje, definida como la expresión de antígenos que no pertenecen al linaje;

(2) Asincronía antigénica, por ejemplo, la expresión sobre células maduras de antígenos que normalmente aparecen en células inmaduras;

(3) Ausencia antigénica; y

(4) Anomalías cuantitativas.

Véase Terstappen L.W.M.M., Konemann S., Safford M., Loken M.R., Zurlutter K., Buchner Th., Hiddemann W., Wormann B., *Flow Cytometric Characterization of Acute Myeloid Leukemia, Part II, Phenotypic Heterogeneity at Diagnosis, Leukemia*, 6:70-80 (1991).

No solo son los fenotipos de las células leucémicas diferentes de las normales, sino también las relaciones entre los antígenos son diferentes de un caso a otro, lo cual sugiere que cada transformación leucémica provoca una pérdida de la regulación génica coordinada, que resulta en un patrón fenotípico exclusivo para cada leucemia. En 120 casos pediátricos de ALL y 86 casos de AML en adultos, cada fenotipo detallado fue diferente del normal y diferente del resto de casos. Véase *Id.*; Hurwitz, *supra*. Así, la transformación neoplásica afecta a la secuencia de ADN primaria (genotipo) y la regulación de los genes normales, de modo que son expresados de modo inapropiado en el momento equivocado durante el desarrollo, son expresados en las cantidades equivocados y/o son expresados en un contexto con otros genes que no se observa en células normales (fenotipo). La pérdida de la regulación génica coordinada parece ser una característica de la transformación neoplásica que produce fenotipos anómalos, en los que cada clon leucémico es diferente del normal y es diferente de otras leucemias del mismo tipo.

Debe advertirse que las realizaciones no se limitan al análisis de las células leucémicas (por ejemplo, leucemias linfocíticas agudas y crónicas (ALL, CLL) y leucemia mielógena aguda y crónica (AML, CML)) y otras células neoplásicas hematopoyéticas y linfoides. Las realizaciones pueden aplicarse al análisis de cualquiera de una diversidad de malignidades, por ejemplo, linfoma, mieloma o premalignidades, tales como mielodisplasia y otros trastornos, que incluyen cualquiera de una diversidad de trastornos hematopoyéticos.

La citometría de flujo puede adoptarse para emplear esta diferencia fenotípica desde la situación normal para ayudar en el diagnóstico de la leucemia, así como para controlar la respuesta a la terapia. La citometría de flujo se ha empleado en hematopatología para fenotipificar los tumores, por ejemplo, para diferenciar la AML de la ALL. Sin embargo, las estrategias convencionales requieren que las células de interés formen una porción predominante de las células totales examinadas y que el proceso de enfermedad esperado sea conocido antes de realizar el análisis, tal como cuando un examen morfológico identifica una población de células leucémicas de un subtipo incierto. El enfoque en las células neoplásicas puede extenderse a la detección de una enfermedad residual. Sin embargo, las técnicas de detección de enfermedades residuales convencionales que emplean la citometría de flujo requieren un panel de reactivos específicos de paciente para identificar el fenotipo específico observado en el diagnóstico. Véase Reading C.I., Estey E.H., Huh Y.O., Claxton D.F., Sánchez G., Terstappen L.W., O'Brien M.C., Baron S., Deisseroth A.B., *Expression of Unusual Immunophenotype Combinations in Acute Myelogenous Leukemia, Blood*, 81:3083-3090 (1993). Estos paneles específicos de paciente se han empleado para detectar ALL y AML residual hasta unos niveles de 0,03-0,05%. Véase Coustan-Smith E., Sancho J., Hancock M.L., Boyett J.M., Behm F.G., Raimondi S.C., Sandlund J.T., Rivera G.K., Rubnitz J.E., Ribeiro R.C., Pui C.H., Campana D., *Clinical Importance of Minimal Residual Disease in Childhood Acute Lymphoblastic Leukemia, Blood*, 96:2691-2696 (2001); San Miguel J.F., Vidriales M.B., López-Berges C., Díaz-Mediavilla J., Gutiérrez N., Canizo C., Ramos F., Calmunitia M.J., Pérez J., González M., Orfao A., *Early Immunophenotypical Evaluation of Minimal Residual Disease in Acute Myeloid Leukemia Identifies Different Patient Risk Groups and may Contribute to Postinduction Treatment Stratification, Blood*, 98:1746-1751 (2002).

Sin embargo, la detección convencional de enfermedades residuales que emplean paneles de reactivos específicos de paciente tiene las siguientes limitaciones:

1. Es necesario un espécimen de diagnóstico con un fenotipo aberrante para construir un panel. En 25% de los casos no puede identificarse un fenotipo aberrante. Véase Vidriales, *supra*.
- 5 2. El tiempo de procesamiento es sustancial, porque un técnico debe examinar los análisis previos para el paciente concreto para determinar la combinación de reactivos que se vaya a utilizar en cada caso.
3. El fenotipo de una población de células leucémicas que sea diferente del fenotipo originariamente diagnosticado puede no ser detectado. Por ejemplo, el fenotipo puede cambiar desde el momento del diagnóstico al de la recaída como resultado de una evolución clonal o un brote de un subclón resistente a la quimioterapia minoritario. Véase San Miguel, *supra*.
- 10 4. Pueden pasarse por algo anomalías inesperadas o no previstas, tales como una mielodisplasia secundaria o anomalías en otros linajes.

La evaluación de la enfermedad residual empleando paneles específicos de paciente pueden tener una actuación satisfactoria en un entorno controlado, tal como un estudio de investigación en el que se puede acceder a todos los especímenes secuenciales y exista una alta conformidad para obtener especímenes en momentos específicos en la terapia. Sin embargo, en la práctica clínica, un laboratorio de citometría de flujo puede recibir el encargo de realizar un análisis de enfermedad residual, pero no fue en ese laboratorio en el que realizó el diagnóstico inicial. A menudo no está disponible un inmunofenotipo detallado o este está incompleto.

La detección de la enfermedad residual también puede realizarse empleando paneles estandarizados y emplear las diferencias con respecto a la normalidad como el marcador específico de tumor. La expresión génica coordinada es tan precisa que una divergencia de media década en la expresión del antígeno es suficiente para la discriminación entre células normales y neoplásicas aberrantes. En esta estrategia, se emplean paneles de reactivos específicos para cada linaje sospechoso, por ejemplo, linaje B de ALL; linaje T de ALL; AML; linaje B de linfoma no hodgkiniano ("B-NHL"); y linaje T de NHL ("T-NHL"), así como MDS y mieloma. Las poblaciones tumorales pueden ser identificadas identificando, en primer lugar, patrones esperados de células normales, y después centrándose en las células que no se corresponden con los patrones esperados de células normales. Esta estrategia para detectar la enfermedad residual ha sido empleada por the Fred Hutchinson Cancer Research Center durante varios años y ha tenido éxito en la predicción de resultados en neoplasmas hematopoyéticos. Por ejemplo:

1. En trasplantes de células pluripotenciales hematopoyéticas para ALL, se ha demostrado que la citometría de flujo es más sensible y más específica que la morfología, la citogenética o las dos tecnologías combinadas para predecir la recaída para 120 pacientes. Véase Wells, D.A., *supra*.

2. En la AML pediátrica, la detección mediante citometría de flujo de la enfermedad residual fue el mejor predictor del resultado en 252 pacientes estudiados. Sievers, E. L., Lange, B. J., Alonzo, T. A., Gerbing, R. B., Bernstein, I. D., Smith, F. O., Arceci, R. J., Woods, W.G., Loken, M. R., Immunophenotypic evidence of leukemia after induction therapy predicts relapse: results from a prospective Children's Cancer Group study of 252 patients with acute myeloid leukemia, *Blood*, 101: 3398-3406 (2003). Los pacientes con un tumor detectable en cualquier momento durante la terapia presentaban una probabilidad 4 veces mayor de recaída y una probabilidad 3 veces mayor de morir que los pacientes en los que no se detectó ningún tumor.

3. En trasplantes de células pluripotenciales hematopoyéticas, la citometría de flujo fue capaz de distinguir entre blastos en regeneración y el tumor recurrente basándose en la expresión de antígenos aberrantes. Véase Shulman H., Wells D., Gooley T., Myerson D., Bryant E., Loken M., The biologic significance of rare peripheral blasts after hematopoietic cell transplant is predicted by multidimensional flow cytometry, *Am. J. Clin. Path.*, 112:513-523 (1999). Los pacientes pueden mostrar 20% de blastos normales en la sangre o puede presentar hasta 50% de blastos en regeneración en la médula ósea sin detección de células neoplásicas.

La detección de fenotipos anómalos de pequeñas poblaciones de células en la sangre o la médula ósea extiende la utilidad de la citometría de flujo a otras aplicaciones más allá de simplemente la fenotipificación de leucemias. La citometría de flujo se ha empleado para demostrar que una proporción significativa (10%) de pacientes con un diagnóstico de mielodisplasia han recibido un diagnóstico equivocado y tienen anomalías linfoides y no mieloides. Véase Wells D.A., Hall M.C., Shulman H.E., Loken M.R., Occult B cell malignancies can be detected by three-color flow cytometry in patients with cytopenias, *Leukemia*, 12:2015-2023 (1998). La citometría de flujo también ha permitido el desarrollo de un sistema de puntuación para estratificar pacientes con mielodisplasia basándose en el grado de anomalías detectado entre las células mieloides en maduración. Véase Wells, D., Benesch, M., Loken, M., Vallejo, C., Myerson, D., Leisenring, W., Deeg, H., Myeloid and monocytic dyspoiesis as determined by flow cytometric scoring in myelodysplastic syndrome correlates with the IPSS and with outcome after hematopoietic stem cell transplantation, *Blood*, 102: 394-403 (2003). Los pacientes con células mieloides que muestran más aberraciones en la expresión génica, según se evidencia por medio de un inmunofenotipo anómalo, tienen una mayor tasa de recaída y muerte tras un trasplante con células pluripotenciales cuando se comparan con pacientes con menos anomalías detectables. También existe una alta correlación con el sistema de puntuación de pronóstico

internacional (International Prognostic Scoring System, IPSS). Además, una puntuación alta de citometría de flujo divide el grupo intermedio I de pacientes en el sistema IPSS en grupos estadísticamente significativos basándose en la recaída tras un trasplante de células pluripotenciales.

Existen varias ventajas en la detección de tumores basadas en la diferencia con respecto a la situación normal.

- 5 1. La técnica no requiere un espécimen de diagnóstico para la creación de un panel específico.
2. La estrategia permite el procesamiento rápido de especímenes en un laboratorio de alto volumen de trabajo, ya que se emplean paneles idénticos para pacientes diferentes.
3. Los resultados no se ven afectados por un cambio en el fenotipo después de la terapia.
- 10 4. La correcta selección del panel estandarizado permite la detección de descubrimientos inesperados o no previstos que son el resultado de anomalías hematológicas.

15 La distinción convencional entre poblaciones de células normales y aberrantes tiene limitaciones significativas. El análisis de los datos debe ser realizado, de modo convencional, por un profesional (un médico versado en la citometría de flujo y en hematopatología) y no por un técnico, puesto que diversas situaciones clínicas pueden indicar si las anomalías observadas son normales o anómalas. Es necesario un largo proceso de aprendizaje para formar al profesional para que distinga entre poblaciones celulares normales y anómalas. Un hematopatólogo bien formado puede tardar de 6 meses a un año en aprender las técnicas. En la actualidad, la evaluación de una situación normal frente a otra anómala por parte de un profesional se basa en la experiencia, con todas las dificultades inherentes de un análisis subjetivo, similares a la formación en la microscopía de diagnóstico. Es difícil extender el análisis a otros sitios y mantener la misma sensibilidad y especificidad. En los casos difíciles, dos o más profesionales deben llegar a un consenso para el diagnóstico final.

20 Por ejemplo, Weir, *et al.* describen un "molde" normal que resulta de un análisis de citometría de flujo de cuatro colores de precursores de células B normales con el que pueden compararse muestras de tumor. Véase Weir, E. G., *et al.*, *Leukemia* (1999), 13:558-567. Sin embargo, a diferencia de la presente invención, este molde es un conjunto específico y fijado de regiones geométricas trazadas alrededor de los acontecimientos de la gráfica de puntos presentada, que después se emplean como los límites de la normalidad. Tal como indican Weir, *et al.*, los acontecimientos aislados de naturaleza incierta presentes en muestras normales que se encuentran fuera de los límites de la normalidad definidos por el molde son un problema grave que aún ha de ser resuelto con su método, en particular en el entorno de la detección de enfermedad residual mínima. Además, al igual que otros métodos de la técnica anterior, es necesario un análisis por un individuo muy formado para comparar las muestras de los pacientes con el molde.

25 Además, las poblaciones identificadas mediante múltiples anticuerpos monoclonales en la médula ósea normal no son nubes esféricas diferenciadas en el espacio multidimensional. Por el contrario, los datos pueden describirse como una serie de tubos o serpientes que cambian de tamaño y posición a medida que los linajes de células se trasladan de las formas inmaduras a las formas maduras, desplazándose desde la cabeza a la cola en el espacio de datos multidimensional. Así, los programas de análisis de agrupaciones que tratan a los datos como nubes esféricas producen resultados con las limitaciones indicadas anteriormente.

30 Por contraste, las realizaciones descritas a continuación en la presente implican un método para determinar, entre otras cosas, una línea centroide y un radio de una o más agrupaciones de acontecimientos que se corresponden con un linaje de maduración de células normal. De esta manera, puede emplearse un análisis estadístico para determinar si un acontecimiento representa un acontecimiento anómalo (es decir, un cáncer).

35 La médula ósea normal está compuesta de múltiples linajes, cada uno de ellos sometido a una maduración continua y constante. Mediante la evaluación en primer lugar de las células normales, puede definirse una medición estadística de lo que constituye lo normal y de lo que constituye una anomalía. Esta definición entonces se convierte en el patrón para el análisis. La automatización de la identificación de las células que están dentro de las posiciones esperadas y definidas de la normalidad facilitaría la enseñanza a nuevos profesionales y técnicos de lo que se considera fenotípicamente anómalo. También permitiría la estandarización del análisis en múltiples sitios que proporcionaría una coherencia entre los analistas para identificar a las poblaciones anómalas.

40 La automatización de la identificación de las células anómalas también permite una mayor sensibilidad. La evaluación manual actual se realiza empleando tres anticuerpos en combinación con la dispersión de luz directa y de ángulo derecho que recoge 10.000 acontecimientos para cada tubo. Un panel consiste en entre siete y catorce tubos diferentes, cada uno con una combinación diferente de anticuerpos. Empleando este sistema actual, los tumores pueden detectarse con una especificidad que se acerca al 100%. Véase *Am. J. Clin. Path.*, 110:84-94, *supra*; *Blood*, 98:1746-1751, *supra*; *Blood*, 101:3398-3406, *supra*. Es posible que un único profesional analice e indique entre 20-30 de estos casos en un único día. El aumento de la sensibilidad es una limitación de las estrategias convencionales, porque el profesional debe dedicar más tiempo a analizar cada caso. La automatización de la identificación de las células anómalas permitiría unos conjuntos de datos más grandes (para contar más células) y la aplicación de más anticuerpos sin aumentar el tiempo que debe dedicar el analista a cada espécimen.

El análisis estadístico puede emplearse para identificar cambios más sutiles en las anomalías hematopoyéticas. Esto es especialmente importante para el análisis del síndrome mielodisplásico ("MDS"), en el que las anomalías se observan en las células más maduras y no únicamente en los blastos inmaduros. El análisis estadístico identificará bultos en los tubos o desplazamientos en la línea centroide que puedan indicar la regulación anómala de las células. También puede definir puntos reguladores y tasas de avance a través del proceso del desarrollo, permitiendo una mejor comprensión de la pérdida de la regulación de genes coordinada observada durante la transformación neoplásica.

La figura 1 es un diagrama de bloques funcional de un sistema 100 que ejecuta una realización de un sistema para detectar células anómalas empleando un análisis multidimensional. El sistema 100 comprende un sistema de medición 102 y un sistema de diagnóstico 104.

El sistema de medición 102 mide las características de las células en una muestra de células y, tal como se ilustra, comprende un citómetro de flujo 106 y un programa de formateado de datos 108. Puede emplearse más de un citómetro de flujo 106, aunque normalmente las mediciones para una muestra concreta se realizarán con un instrumento. Por ejemplo, tal como se analiza con más detalle a continuación, las mediciones para un conjunto normal de células pueden tomarse con un citómetro de flujo, mientras que las mediciones procedentes de un conjunto de ensayo de células pueden tomarse con otro citómetro de flujo. Pueden emplearse otros dispositivos de medición en el sistema de medición 102, tales como un microscopio (es decir, un microscopio de alta capacidad de procesamiento).

El sistema de medición 102 puede contener un programa de formateado de datos separado 108 para dar formato a los datos recogidos por el sistema de medición 102. Como alternativa, el programa de formateado de datos 108 puede ser parte de otro componente del sistema 100, tal como el citómetro de flujo 106 o el sistema de diagnóstico 104. El programa de formateado de datos 108 puede, por ejemplo, dar formato a los datos recogidos por el citómetro de flujo 106 en un formato Flow Cytometry Standard FCS 2.0 u otro formato de archivo de datos. El sistema de medición 102 puede comprender otros componentes, tales como controladores, memorias y/o circuitería y hardware.

El sistema de diagnóstico 104 analiza los datos recibidos del sistema de medición 102, tal como se analiza con más detalle a continuación. En la realización ilustrada en la figura 1, el sistema de diagnóstico 104 comprende un controlador 110, una memoria 112, un analizador 114, una interfase de control de entrada/salida 116, una interfase de datos de entrada/salida 118, un motor gráfico 120, un motor estadístico 122, un sistema de presentación 124, una impresora 126 y una vía de transmisión del sistema de diagnóstico 130. La vía de transmisión del sistema de diagnóstico 130 puede incluir una vía de control de potencia, una vía de transmisión de control, y una vía de transmisión de señal de estado además de una vía de transmisión de datos. Sin embargo, para añadir claridad, las diversas vías de transmisión del sistema de diagnóstico se ilustran en la figura 1 como la vía de transmisión del sistema de diagnóstico 130.

El sistema de diagnóstico 104 puede estar en un emplazamiento físicamente remoto del sistema de medición 102. El sistema de medición 102 puede estar acoplado al sistema de diagnóstico 104 a través de una o más conexiones de comunicación, tales como la internet, una extranet y/o una intranet u otras redes locales o de área amplia. De modo similar, los componentes del sistema de diagnóstico 104 pueden estar en un emplazamiento físicamente remoto entre sí y pueden acoplarse juntos a través de conexiones de comunicación, tales como la internet, una extranet y/o una intranet u otras redes locales o de área amplia. Puede haber uno o más sistemas de diagnóstico y cada uno puede acoplarse a uno o más sistemas de medición.

El sistema de diagnóstico 104 puede ejecutarse en una diversidad de formas, que incluye en forma de subsistemas separados. El sistema de diagnóstico 104 puede ejecutarse como un procesador de señales digitales (DSP), un circuito integrado específico de aplicación (ASIC) o similares, o como una serie de instrucciones almacenadas en una memoria, tal como la memoria 112 y ejecutadas por un controlador, tal como el controlador 110. Así, las modificaciones de software en el hardware existente pueden permitir la ejecución del sistema de diagnóstico 104. Se identifican diversos subsistemas, tales como el analizador 114 y la interfase de control de entrada/salida 116, como bloques separados en el diagrama de bloques funcional de la figura 1 porque realizan funciones específicas que se describirán con más detalle a continuación. Estos subsistemas pueden no ser unidades discretas sino que pueden ser funciones de una rutina de software que quizás, aunque no necesariamente, puedan ser elementos rescatables y, por tanto, identificables. Puede emplearse cualquier software o combinaciones de software adecuados para ejecutar el sistema de diagnóstico 104 que incluye, por ejemplo, WinList y/o Java ejecutado con un Java Run Time Environment o un 3-D Java Run Time Environment.

Aunque la realización ilustrada indica un único controlador 110, otras realizaciones pueden comprender múltiples controladores. La memoria 112 puede comprender, por ejemplo, registros, memoria de solo lectura ("ROM"), memoria de acceso aleatorio ("RAM"), memoria flash y/o memoria de solo lectura programable y borrable eléctricamente ("EEPROM"), y puede proporcionar instrucciones y datos para su uso por parte del sistema de diagnóstico 104.

En la presente se describe una realización de la invención con respecto a un estudio que se realizó con el linaje B linfoide. Cuando resulte apropiado se incorporan las referencias a la figura 1 en la descripción del estudio. Las

realizaciones descritas en la presente pueden aplicarse para estudiar, caracterizar y diagnosticar otros linajes normales y enfermos, tales como el linaje eritroide, T linfoide y otros, que incluyen aquellos con múltiples linajes, tales como los linajes mieloides (véase Shulman H., 1999, *supra*; Wells D.A., 1998, *supra*; y Loken M.R. y Wells D.A., Normal Antigen Expression in Hematopoiesis: Basis for Interpreting Leukemia Phenotypes, en Immunophenotyping, eds. Carleton Stewart y Janel K.A. Nicholson, 2000, Wiley-Liss, Inc.).

El linaje de linfocitos B es un único linaje y está bien definido en 4 estadios de desarrollo dentro de la médula ósea con múltiples diferencias antigénicas entre estadios que han sido bien caracterizadas. El linaje B entero se identifica por la expresión de un único antígeno, CD19, que permite la detección de los 4 estadios de las células del linaje B. Las células del linaje B más temprano (estadio I) se identifican mediante la expresión de CD34, unos niveles altos de CD10 y unos niveles bajos de CD45. Durante el estadio II se pierde CD34, la intensidad de CD10 se reduce en un factor de 2, la intensidad de CD45 aumenta y CD20 comienza a ser expresado. Cuando CD20 alcanza un máximo, ya no se produce un mayor aumento en CD45 con la pérdida de CD10 que indica el estadio III. El estadio final (IV) del desarrollo de células B linfoides se caracteriza por la ausencia de CD10, la expresión de CD22 y unos niveles altos de CD45.

Tal como entenderán los expertos en la técnica, otros linajes celulares que pueden caracterizarse empleando los métodos descritos en la presente pueden comprender múltiples linajes o linajes ramificados, y los linajes pueden definirse en un número variable de estadios de desarrollo. Por ejemplo, el linaje mielóide incluye, entre otros, el linaje eritroide y el linaje de granulocitos-monocitos. El linaje de granulocitos-monocitos se ramifica en los linajes de monocitos y neutrófilos.

Los neutrófilos pueden dividirse en cinco estadios identificables. Los mieloblastos del estadio I identificados por la expresión de CD34 también muestran HLA-DR, CD13, y CD33 a altos niveles, pero no expresan CD11b, CD15, y CD16. Estos mieloblastos son de un tamaño intermedio según la dispersión de luz directa, pero tienen un SSC bajo. El avance hasta el estadio II viene indicado por la pérdida de CD34 y HLA-DR, la adquisición de unos niveles elevados de CD15, un aumento notable en la expresión de SSC, sin la expresión de CD11 b (véase Loken M.R. y Wells D.A., 2000, *supra*). Al estadio II le acompaña una ligera disminución en CD33. El estadio III del desarrollo de los neutrófilos está marcado por la adquisición de niveles intermedios de CD11 b, la pérdida de CD13, y una disminución en SSC relacionada con la aparición de gránulos secundarios. El estadio IV se advierte por el aumento correlacionado en CD13 y CD16 con otra ligera disminución en la expresión de CD33. El estadio V se corresponde con los neutrófilos maduros que se encuentran en la sangre periférica. Estas células tienen la máxima cantidad de CD16, CD13, y CD45 con un aumento en la densidad.

El linaje de los monocitos presenta tres estadios detectables basándose en la expresión de antígenos de la superficie celular. El desarrollo monocítico presenta dos estadios de maduración después del estadio de mieloblasto (indistinguible del estadio I del desarrollo de los neutrófilos). Estas células conservan el HLA-DR a lo largo de su desarrollo, por contraste con los neutrófilos que pierden rápidamente este antígeno en el estadio de promielocito. La maduración de los monocitos (estado II) se identifica primero por la aparición rápida de CD11b mientras que se mantienen unos niveles intermedios de CD45. El estadio II del desarrollo de los monocitos viene acompañado de aumentos en la expresión de CD13 y CD33, con una expresión baja de CD15. El estadio III del desarrollo se define por un aumento coordinado en CD45 y CD14 (véase Loken M.R. y Wells D.A., 2000, *supra*).

Las células eritroides solo presentan dos estadios (véase Loken M., 1992, *supra*). El compromiso con este linaje se identifica por la pérdida de CD45 y el aumento en CD71, estadio I. La expresión de la glicoforina y la aparición de la hemoglobina marcan el segundo estadio. Las etapas finales de maduración de las células eritroides pueden observarse por la pérdida del núcleo, una disminución en CD71, y la posterior pérdida de ARN en los reticulocitos (véase Loken, M.R., Shah V.O., Dattilio K.L., Civin C.I. (1987), Flow cytometric analysis of human bone marrow. I. Normal erythroid development, Blood, 69:255-263).

Tal como se describe en Loken M.R. y Wells D.A., 2000, *supra*, las células T linfoides pueden dividirse en cuatro estadios de desarrollo en el timo mediante el patrón de actividad de 10 antígenos (CD1a, CD2, CD3, CD4, CD5, CD7, CD8, CD10, CD34, y CD45). Tres estadios están claramente definidos por múltiples diferencias antigénicas, mientras que el cuarto puede distinguirse por el tamaño.

Así, tal como entenderán los expertos en la técnica tras leer la presente descripción, los métodos descritos en la presente que emplean el linaje B linfoide como ejemplo pueden utilizarse para caracterizar, en un espacio n-dimensional, otros linajes celulares tales como los descritos en la presente y los conocidos en la técnica.

En la realización de la invención descrita en la presente con respecto al linaje B linfoide, se identifican los cuatro estadios del desarrollo de células B empleando dos tubos de reactivos con cuatro colores:

Tubo 1: CD20 FITC, CD10 PE, CD45 PerCP y CD19 APC.

Tubo 2: CD22 FITC, CD34 PE, CD45 PerCP y CD19 APC.

La redundancia de los marcadores (CD19 y CD45) en ambos tubos permite la comparación de los datos entre los diferentes tubos. En el estudio, los conjuntos de datos se recogieron con 200.000 acontecimientos en un citómetro

de flujo FACS Calibur (Becton Dickinson, San José, CA). Los procedimientos para la preparación de las muestras fueron los convencionales y siguieron un protocolo fijado. Véase Am. J. Clin. Path., 110:84-94, *supra*. Se recogieron datos en modo de lista de dos pacientes fenotípicamente normales en el formato FCS para el análisis. Las agrupaciones identificadas por alguien versado en la citometría de flujo y en hemopatología, tal como un médico, se compararon con las agrupaciones identificadas mediante el sistema de diagnóstico 104 empleando algoritmos de agrupación. Los centros visuales de las agrupaciones identificados por el profesional se compararon con los generados por el sistema de diagnóstico 104. El proceso es iterativo, ya que el usuario revisa las agrupaciones identificadas basándose en los resultados de los algoritmos de agrupación y ejecuta otros algoritmos de agrupación empleando las definiciones de agrupación revisadas.

Se realizó un análisis de cuatro colores de un conjunto de células B linfoides de médula ósea normal en el tubo 1. Los especímenes se recogieron para obtener 200.000 acontecimientos para el análisis. Las células se colocaron en un tubo y se tñeron con los reactivos CD20-fluoresceína (FITC), CD10 ficoeritrina (PE), CD45 proteína de clorofila peridina (PerCP), y CD19 alofococianina (APC). Las características de las células expuestas se midieron empleando una citometría de flujo (véase el citómetro de flujo 106 de la figura 1). Un sistema, tal como el sistema 100 ilustrado en la figura 1, mide y analiza la muestra empleando una combinación de los datos recibidos de las mediciones y la entrada de un usuario, tal como un profesional o un técnico, tal como se analiza con más detalle a continuación.

Puede emplearse la especificación Flow Cytometry Standard FCS 2.0 disponible al público para almacenar las características medidas de las células en las muestras. Pueden emplearse otros formatos de datos y estructuras de datos, por ejemplo, puede emplearse un formato FCS 1.0 o FCS 3.0. Un ejemplo de una estructura de datos para almacenar un conjunto de datos se ilustra en la figura 2. Haciendo referencia a las figuras 1 y 2, el analizador 114 analiza la sección de cabecera 202, la sección de texto 204, la sección de datos 206 y la sección de análisis 208 y recoge la información, que incluye un nombre de parámetro, un número total de puntos de datos y los detalles del tipo de datos. La sección de cabecera 202 describe la localización de las otras secciones en la estructura de datos 200. La sección de cabecera 202 contiene información de desplazamiento de los puntos de inicio y de fin para las secciones de texto 204, de datos 206 y de análisis 208. La sección de texto 204 contiene una serie de parejas de valores de palabras clave codificadas por ASCII que describen diversos aspectos de la estructura de datos 200. Por ejemplo, \$TOT/5000/ es una pareja de valor de palabra clave que indica que el número total de acontecimientos en el archivo es de 5000, y \$PAR indica el número total de parámetros. La sección de datos 206 contiene los datos brutos. Estos datos se encuentran normalmente en uno de tres modos (lista, correlacionados o no correlacionados) descritos en la sección de texto 204, mediante, por ejemplo, un valor de palabra clave \$MODE. Los datos pueden estar escritos en la sección de datos 206, por ejemplo, en uno de cuatro formatos (binario, punto de flotación, punto de flotación de doble precisión o ASCII) descritos por un valor de palabra clave \$DATATYPE. Una forma habitual de almacenamiento de datos es el almacenamiento en modo de lista en la forma de números enteros binarios (\$DATATYPE// \$MODE/L/). El conjunto \$PnB de palabras clave puede especificar la anchura de bits para el almacenamiento de cada parámetro. El conjunto PnR de palabras clave puede especificar el intervalo de número de canales para cada parámetro. Por ejemplo, \$PnB/16/ \$PnR/1024/, en el que n es un número entero, puede especificar un campo de 16 bits para el parámetro n y un intervalo para los valores del parámetro n de 0 a 1023, que se corresponden con 10 bits. La sección de análisis 208 es un segmento opcional que, cuando está presente, puede contener los resultados del procesamiento de datos. El análisis también puede realizarse fuera de línea, después de que los datos hayan sido recogidos y almacenados en una estructura de datos, tal como la estructura de datos 200. En el estudio de ensayo no se utilizó una sección de análisis 208. Sin embargo, puede emplearse una sección de análisis 208 para almacenar la información que define una línea centroide y un radio para un conjunto de datos.

Los desplazamientos de los datos del formato FCS 2.0 si indican en un archivo de propiedades. Un ejemplo de archivo de propiedades 300 se ilustra en la figura 3. El archivo de propiedades contiene una sección de encabezado 302, que contiene información acerca de cómo leer el archivo de propiedades 300, una sección de formato 304, que contiene información acerca del formato de la estructura de datos 200, y una sección de filtro 306, que contiene información que el analizador 114 puede utilizar para filtrar los datos almacenados en la estructura de datos 200. El analizador 114 emplea la información extraída del archivo de propiedades 300 para analizar la estructura de datos cargada 200. El archivo de propiedades 300 puede modificarse con facilidad para permitir el uso de diversos formatos de archivos de datos, tales como los diversos formatos de Flow Cytometry Standard.

El sistema 100 puede emplear una intensidad de fluorescencia que se corresponde con CD19 como puerta inicial. Así, no es necesario analizar las 200.000 células en una lista de acontecimientos de 200.000 células, sino que puede evaluarse solo las células positivas a CD19 (que incluyen todas las células del linaje B). Esto potencia la estadística aumentando el número de células del linaje B que se van a analizar sin aumentar el tiempo de cálculo requerido para distinguir las células B linfoides de la mayoría de las otras células en la médula ósea. Sin esta puerta para las células de interés, puede ser necesario un tiempo de cálculo de 6-8 horas para identificar las agrupaciones en la lista de acontecimientos de 200.000 células. La proporción de células B linfoides inmaduras (estadios I-III) constituye un promedio menor que 2% de todas las células nucleadas en la médula ósea normal. Véase Loken, M.R., Shah, V.O., Dattilo, K.L., Civin, C.L., Flow Cytometry Analysis of Human Bone Marrow: II. Normal B Lymphoid Development, Blood, 70:1316 (1987). Por tanto, aumentando los recuentos totales hasta 200.000 y estableciendo una puerta para las células positivas a CD19, relativamente poco frecuentes, las células de interés se analizan y al mismo tiempo se mantiene el conjunto de datos completo y se evitan accidentes introducidos por el establecimiento de una puerta

electrónica para CD19 durante la recogida de datos. Sin embargo, en realizaciones alternativas, sí puede emplearse el establecimiento de una puerta electrónica para CD19 durante la recogida de datos.

Las poblaciones de interés de un ejemplo de conjunto de datos normales recogidos como se describió anteriormente con respecto al tubo 1 se ilustran en las figuras 4A a 9A como una serie de presentaciones de análisis de cuatro colores que se generaron empleando WinList. Las poblaciones de interés también pueden representarse de otras formas, tales como una correspondiente presentación de análisis de cuatro tonos, que se ilustra en las figuras 4B a 9B. Las figuras 4A a 9A y 4B a 9B se indican en la presente de forma colectiva como las figuras 4 a 9.

Las agrupaciones de acontecimientos se identifican inicialmente en múltiples proyecciones de presentaciones 2 por 2 de los 6 datos dimensionales (4 colores y 2 parámetros de dispersión de luz). Las presentaciones pueden ser, por ejemplo, representaciones de los datos en un sistema de coordenadas cartesianas. Las proyecciones de las presentaciones pueden ser generadas por el motor gráfico 120 ilustrado en la figura 1. Un usuario, tal como alguien versado en la citometría de flujo y en hematopatología, identifica una región ML en una proyección de una presentación 2 por 2 en un sistema de coordenadas con un eje horizontal que se corresponde con una dispersión de luz directa, y un eje vertical que se corresponde con una dispersión de luz lateral, tal como se ilustra en la figura 4. La región ML se corresponde con las células nucleadas. El usuario identifica las regiones de linfoides, monocitos, mieloides y blastos en una proyección de una presentación 2 por 2 en un sistema de coordenadas con un eje horizontal que se corresponde con una dispersión de luz lateral, y un eje vertical que se corresponde con un nivel de intensidad de fluorescencia para CD45, según se ilustra en la figura 5. El usuario identifica las células B linfoides en una proyección de una presentación 2 por 2 en un sistema de coordenadas con un eje horizontal que se corresponde con una dispersión de luz lateral, y un eje vertical que se corresponde con un nivel de intensidad de fluorescencia para CD19, según se ilustra en la figura 6.

El usuario identifica una agrupación del estadio I, una agrupación del estadio II y una agrupación del estadio III/IV en una proyección de una presentación 2 por 2 en un sistema de coordenadas con un eje horizontal que se corresponde con un nivel de intensidad de fluorescencia para CD19, y un eje vertical que se corresponde con un nivel de intensidad de fluorescencia para CD45, según se ilustra en la figura 7. Los estadios se corresponden con los niveles de maduración para las células B linfoides. El usuario identifica una agrupación del estadio I, una agrupación del estadio II, una agrupación del estadio III y una agrupación del estadio IV en una proyección de una presentación 2 por 2 en un sistema de coordenadas con un eje horizontal que se corresponde con un nivel de intensidad de fluorescencia para CD10, y un eje vertical que se corresponde con un nivel de intensidad de fluorescencia para CD45, según se ilustra en la figura 8. El usuario identifica una agrupación del estadio I, una agrupación del estadio II, una agrupación del estadio III y una agrupación del estadio IV en una proyección de una presentación 2 por 2 en un sistema de coordenadas con un eje horizontal que se corresponde con un nivel de intensidad de fluorescencia para CD20, y un eje vertical que se corresponde con un nivel de intensidad de fluorescencia para CD10, según se ilustra en la figura 9.

Basándose en la evaluación por parte del usuario de las figuras 4-9, las células accedidas se asignan a una agrupación inicial. Esto produce un conjunto de datos normales de siete dimensiones, y estas dimensiones se corresponden con: una dispersión de luz directa; una dispersión de luz lateral; un nivel de intensidad de fluorescencia de CD19; un nivel de intensidad de fluorescencia de CD45; un nivel de intensidad de fluorescencia de CD20; un nivel de intensidad de fluorescencia de CD10; y una agrupación, que se corresponde con un estadio de maduración dentro de la población de células B. Se asigna un color a cada identificación de la agrupación y los datos se cartografían en un espacio de seis dimensiones. Los datos son mostrados por un sistema de diagnóstico, tal como el sistema de diagnóstico 104 ilustrado en la figura 1, en una presentación pseudo-tridimensional rotatable con un código de colores basado en la identificación de agrupaciones.

El sistema de diagnóstico 104 cartografía el conjunto de datos normales en un sistema de coordenadas de tres ejes, tal como un sistema de coordenadas cartesianas, y muestra los datos para que el usuario pueda visualizarlos. Cada eje se corresponde con una de las dimensiones del conjunto de datos, cuyo color indica la agrupación a la cual se ha asignado una célula concreta. El conjunto de datos también puede representarse en una presentación tabular o en una presentación combinada. Las figuras 10A y 10B (de modo colectivo figura 10) ilustran ejemplos de presentaciones 400 que combinan una representación de una presentación pseudo-tridimensional 402 con una representación tabular 404. La figura 10A es una presentación de colores y la figura 10B es una correspondiente presentación de tonos.

La representación gráfica 402 comprende un eje x 406 que se corresponde con una intensidad de fluorescencia para CD20, un eje y 408 que se corresponde con una intensidad de fluorescencia para CD10, y un eje z 410 que se corresponde con una intensidad de fluorescencia para CD45. A los datos en una primera agrupación 412 se les asigna el color rojo y se corresponden con un nivel de maduración del estadio I. A los datos en una segunda agrupación 414 se les asigna el color verde y se corresponden con un nivel de maduración del estadio II. A los datos en una tercera agrupación 416 se les asigna el color azul y se corresponden con un nivel de maduración del estadio III. A los datos en una cuarta agrupación 418 se les asigna el color amarillo y se corresponden con un nivel de maduración del estadio IV.

La representación tabular 404 comprende una primera columna 420 que indica el número de la agrupación, una

segunda columna 422 que indica el número de puntos en la agrupación, una tercera columna 424 que indica el color o el tono asignado a la agrupación, una cuarta columna 426 que indica el radio de la agrupación, una sexta columna 428 que indica el porcentaje de puntos o acontecimientos anómalos en el conjunto total de puntos o acontecimientos, y una séptima columna 430 que indica si una distancia logarítmica entre el punto centroide para una agrupación y un punto centroide estadístico para la agrupación es mayor que un valor umbral. La presentación 400, tal como se ilustra, puede ser una presentación por ordenador interactiva. El usuario puede actualizar la información empleada para generar la presentación 400 empleando los campos de entrada de datos 432, 434. Tal como se ilustra, el valor umbral se fija a 2,5 en el campo 434.

El sistema de diagnóstico 104 permite al usuario la selección de los tres ejes en los que cartografiar los datos empleando un menú de una interfase del usuario gráfica (GUI). Las figuras 11A y 11B ilustran un ejemplo de menú 436 que puede ser empleado por un sistema de diagnóstico, tal como el sistema de diagnóstico 104 ilustrado en la figura 1. El sistema de diagnóstico 104 también permite al usuario la selección de otros ajustes a través de menús. Por ejemplo, pueden incluirse selecciones de menús para: seleccionar entre diferentes parámetros de filtración almacenados, editar los parámetros de filtración almacenados, y especificar nuevos parámetros de filtración. Por ejemplo, los datos de alta resolución pueden filtrarse para excluir los datos con un parámetro de dispersión lateral que se corresponde con más de 10^2 , y un parámetro de CD19 que se corresponde con menos de 10 al 1,6989701. Las selecciones de menús también permiten la selección de un plano en el sistema de coordenadas sobre el cual realizar la filtración. Pueden emplearse múltiples criterios de filtración, y el criterio de filtración puede ser mayor o menor que los umbrales especificados. El sistema de menús también permite la selección de una agrupación específica sobre la cual aplicar diversos criterios de filtración. Esto permite al usuario visualizar diversas presentaciones pseudo-tridimensionales del conjunto de datos normales para ayudar al usuario a seleccionar los datos iniciales para su uso por el sistema de diagnóstico 104 para definir una línea centroide y un radio para el conjunto de datos normales. El sistema de diagnóstico 104 también permite la selección de menús de un método de desviación estándar o valor fijado y la rotación de una imagen mostrada. El sistema de diagnóstico 104 también puede mostrar los límites de las agrupaciones para un conjunto de datos basándose en un centroide y un radio seleccionados.

El conjunto de datos normales también puede comprender archivos de datos distintos que se corresponden con muestras distintas. Por ejemplo, el usuario puede examinar y manipular un conjunto de datos que comprende células extraídas de un único individuo y de un único tubo, o el usuario puede combinar muestras extraídas de una pluralidad de individuos y/o tubos en un único conjunto de datos normales. Si una muestra extraída de un individuo se considera anómala, la muestra puede excluirse del conjunto normal de datos.

Haciendo referencia al estudio, en el ejemplo del conjunto de datos de células B linfoides para el tubo 1, el valor n es igual a seis. Cada punto n dimensional se cartografía en el espacio n-dimensional, que puede ser representado en una matriz de flotación por medio de n+1 parámetros de flotación. La tabla 1 ilustra la matriz de flotación para un ejemplo de un conjunto de datos de células B linfoides de seis dimensiones, en la que P_1PR_1 es el valor del primer parámetro para el primer punto, P_2PR_1 es el valor del primer parámetro para el segundo punto, ... P_nPR_1 es el valor del primer parámetro para el n° punto, etc., añadiéndose un séptimo parámetro para la agrupación a la cual se ha asignado el punto, $P_nC\#$. La matriz de flotación puede generalizarse para cualquier número de dimensiones. El sistema de diagnóstico 104 realiza uno o más algoritmos de agrupación sobre el conjunto normal de datos en el espacio n-dimensional, refinando la asignación de los puntos a una agrupación.

Tabla 1 - Matriz de flotación para un conjunto de datos de seis dimensiones

P_1PR_1	P_1PR_2	P_1PR_3	P_1PR_4	P_1PR_5	P_1PR_6	$P_1C\#$
P_2PR_1	P_2PR_2	P_2PR_3	P_2PR_4	P_2PR_5	P_2PR_6	$P_2C\#$
P_3PR_1	P_3PR_2	P_3PR_3	P_3PR_4	P_3PR_5	P_3PR_6	$P_3C\#$
...
P_nPR_1	P_nPR_2	P_nPR_3	P_nPR_4	P_nPR_5	P_nPR_6	$P_nC\#$

El sistema de diagnóstico 104 permite al usuario agrupar los datos empleando un algoritmo de agrupación seleccionado. Por ejemplo, el usuario puede especificar un número de agrupaciones, k, y emplear un algoritmo de k-means para agrupar los datos. Por ejemplo, el sistema de diagnóstico 104 puede dividir los datos en k agrupaciones y asignar un centro a cada agrupación. El centro puede asignarse aleatoriamente a uno de los puntos o introducirse basándose en observaciones del usuario. La distancia entre dos puntos en el espacio n-dimensional puede definirse como sigue:

$$D(P_1, P_2) = \text{SQRT}[(K_1(P_1PR_1 - P_2PR_1))^2 + (K_2(P_1PR_2 - P_2PR_2))^2 + (K_3(P_1PR_3 - P_2PR_3))^2 + \dots + (K_n(P_1PR_n - P_2PR_n))^2] \quad \text{Ecuación 1}$$

en la que $D(P_1, P_2)$ es la distancia entre dos puntos en el espacio n-dimensional, y P_1PR_1 es el valor del primer parámetro para el primer punto, P_2PR_1 es el valor del primer parámetro para el segundo punto, ... P_nPR_1 es el valor del primer parámetro para el n° punto, etc., y $K_1, K_2, K_3, \dots, K_n$ son las constantes de ponderación. En el estudio, las constantes de ponderación se fijan a igual a uno. En otras palabras, no se empleó ponderación en el estudio. Los centros pueden actualizarse iterativamente hasta que se cumpla un criterio de convergencia. En cada iteración, cada punto de datos se asigna a su centro más cercano, y los centros se recalculan empleando el promedio de los valores del parámetro de todos los puntos que pertenecen a una agrupación. Los criterios de convergencia típicos empleados en el estudio fueron la no reasignación (o una reasignación mínima) de puntos a nuevos centros de agrupaciones. Véase Forgy, E., Cluster Analysis of Multivariate Data: Efficiency vs. Interpretability of Classifications, Biometrics, 21:768 (1965), para un análisis de las agrupaciones de k-means.

Otro ejemplo de algoritmo de agrupación es un algoritmo de agrupación DBSCAN. Se define un radio de vecindad, Eps, y un umbral del número de puntos en la vecindad, minPts, y el sistema de diagnóstico 104 emplea un algoritmo de agrupación DBSCAN. El radio de vecindad y el umbral del número de puntos son definidos por el usuario. La agrupación basada en la densidad se basa en el hecho de que las agrupaciones tienen más densidad que sus alrededores. DBSCAN encuentra agrupaciones densas automáticamente para un umbral de densidad dado. Véase Ester, M., Kriegel, H., Sander, J., Xu, X., A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, en Proceedings of 2d International Conference on KDD (1996), para un análisis de la agrupación DBSCAN. Por definición, el umbral de densidad viene especificado por dos parámetros: el radio de vecindad (Eps) y el umbral del número de puntos en ϵ -vecindades (minPts). Un punto 'p' es directamente alcanzable por densidad desde un punto 'q', si 'p' está en la ϵ -vecindad de 'q'. Un punto 'p' es alcanzable por densidad desde 'q' si existe una cadena de puntos 'p_i', en el que $i = 1 \dots n$ y 'P_{i+1}' es directamente alcanzable por densidad desde 'p_i', 'q' es 'p₁' y 'p' es 'p_{n+1}'. Un punto 'p' está conectado por densidad con otro punto 'q' si existe un punto 'o', de modo que ambos 'p' y 'q' son alcanzables por densidad desde 'o'. En el estudio, el sistema de diagnóstico 104 comienza introduciendo un punto en un almacenamiento temporal (lista tempStore, por ejemplo) y descubriendo su ϵ -vecindad. Si la ϵ -vecindad de un punto de datos contiene menos de 'minPts' puntos, entonces se marca como ruido y se introduce otro punto en tempStore. Por lo demás, todos los puntos de ϵ -vecindarios se introducen en tempStore. El proceso completo se repite hasta que se consideran todos los puntos. En resumen, los grupos de agrupación DBSCAN conectan por densidad los puntos como una agrupación densa y eliminan los puntos que no están conectados por densidad como ruido.

El sistema de diagnóstico 104 también puede emplear, por ejemplo, la agrupación por puentes para agrupar los datos. La agrupación por puentes combina la agrupación de k-means con la agrupación DBSCAN. Véase Dash, M., Liu, H., Xu, X., '1+1>2': Merging Distance and Density Based Clustering, Proceedings of the IEEE 7th International Conference on Database Systems for Advanced Applications (DASFAA '01), 18-21 de abril, 2001, Hong Kong, China, para un análisis de la agrupación por puentes. Primero se realizó k-means, seguido de una agrupación basada en la densidad en cada agrupación de k-means y, al final, las agrupaciones de k-means se refinaron eliminando el ruido que se encuentra en la agrupación basada en la densidad. Para una fusión eficaz, cada punto de datos presenta las siguientes tres columnas para almacenar los resultados de la agrupación: <k-means_ID>, <DBSCAN ID> y <core/ ϵ -core/non-core>, en donde:

K-means-ID es la agrupación asignada a cada punto cuando se aplica k-means sobre los puntos de datos;

DBSCAN_ID es la agrupación asignada a cada punto cuando se aplica DBSCAN sobre cada agrupación de k-means; y

los valores core/ ϵ -core/non-core se asignan basándose en las siguientes definiciones:

Definición 1 (CoreDistance): Para cada agrupación, CoreDistance es la mitad de la distancia entre su centro y el centro de la agrupación más cercana.

Definición 2 (CorePoint): No está más alejado del centro de su agrupación que 'CoreDistance - ϵ '. La región central ("Core") de una agrupación es aquella en la que, en su interior, cada punto de datos es central.

Definición 3 (+ ϵ CorePoint): Su distancia desde el centro de la agrupación está entre 'CoreDistance' y 'CoreDistance + ϵ '.

Definición 4 (- ϵ CorePoint): Su distancia desde el centro de la agrupación está entre 'CoreDistance' y 'CoreDistance - ϵ '. Por conveniencia, cuando se consideran los puntos centrales + ϵ and - ϵ , juntos se indican como ϵ -central. La región ϵ -central es aquella en la cada punto es ϵ -central.

Definición 5 (punto no central): No es un punto central ni un punto ϵ -central. Una región no central es aquella en que cada punto es no central.

El sistema de diagnóstico 104 también puede emplear una agrupación por ondícula. Las transformadas de ondículas son una forma especial de las transformadas de Fourier. Véase Press, W.H., Flannery, B.P., Teukiosky, S.A., Numerical Recipes, en C: The Art of Scientific Computing, Ch. 13.10, Cambridge University Press (1992). Esta técnica está muy establecida en el procesamiento de imágenes y la áreas de búsqueda de datos para el reconocimiento de patrones y límites. Véase Sheikholeslami, G., Chatterjee, S., Zhang, A., WaveCluster: A Multi-Resolution Clustering Approach for Very Large Spatial Databases, Proceedings of the 24th VLDB Conference, Nueva York, EEUU, 1998. Por ejemplo, puede emplearse la filtración de ondículas de Daubechies convencional y la transformada de ondículas discretas n-dimensional ("N-Dimensional Discrete Wavelet Transform", NDDFT).

De una manera similar, se identifican las mismas poblaciones (estadios) en un segundo conjunto de datos normales en el segundo tubo (CD22, CD34, CD45, CD19), tal como se ilustra en color en las figuras 12A a 17A y en tonos en las figuras 12B a 17B (denominadas de modo colectivo en la presente como las figuras 12 a 17). El usuario identifica una región ML en una proyección de una presentación 2 por 2 en un sistema de coordenadas con un eje horizontal que se corresponde con una dispersión de luz directa, y un eje vertical que se corresponde con una dispersión de luz lateral, tal como se ilustra en la figura 12. La región ML se corresponde con las células nucleadas. El usuario identifica las células B linfoides en una proyección de una presentación 2 por 2 en un sistema de coordenadas con un eje horizontal que se corresponde con una dispersión de luz lateral, y un eje vertical que se corresponde con un nivel de intensidad de fluorescencia para CD19, según se ilustra en la figura 13.

El usuario identifica una agrupación del estadio I, una agrupación del estadio II y una agrupación del estadio III/IV en una proyección de una presentación 2 por 2 en un sistema de coordenadas con un eje horizontal que se corresponde con un nivel de intensidad de fluorescencia para CD19, y un eje vertical que se corresponde con un nivel de intensidad de fluorescencia para CD45, según se ilustra en la figura 14. Los estadios se corresponden con los niveles de maduración para las células B linfoides. El usuario identifica una agrupación del estadio I, una agrupación del estadio II/III y una agrupación del estadio IV en una proyección de una presentación 2 por 2 en un sistema de coordenadas con un eje horizontal que se corresponde con un nivel de intensidad de fluorescencia para CD22, y un eje vertical que se corresponde con un nivel de intensidad de fluorescencia para CD34, según se ilustra en la figura 15. El usuario identifica una agrupación del estadio I, una agrupación del estadio II/III y una agrupación del estadio IV en una proyección de una presentación 2 por 2 en un sistema de coordenadas con un eje horizontal que se corresponde con un nivel de intensidad de fluorescencia para CD34, y un eje vertical que se corresponde con un nivel de intensidad de fluorescencia para CD45, según se ilustra en la figura 16. El usuario identifica una agrupación del estadio I, una agrupación del estadio II/III y una agrupación del estadio IV en una proyección de una presentación 2 por 2 en un sistema de coordenadas con un eje horizontal que se corresponde con un nivel de intensidad de fluorescencia para CD22, y un eje vertical que se corresponde con un nivel de intensidad de fluorescencia para CD45, según se ilustra en la figura 17. Los resultados del tubo 1 y del tubo 2 se combinan para producir un único conjunto de datos normales, según se describe con más detalle a continuación.

Tras haber identificado las agrupaciones y ser refinadas por el usuario selectivamente empleando el software de agrupación, se define una línea centroide y un radio para las agrupaciones normales, en las que cada agrupación se corresponden con un nivel de maduración dentro de un linaje celular. Las figuras 18A a 18C ilustran una realización de una subrutina 500 que puede emplearse para definir una población normal de células, analizada con respecto al sistema 100 ilustrado en la figura 1 y las células B linfoides recogidas en los tubos 1 y 2, según se analizó anteriormente. El proceso completo para definir una población normal de células debe considerarse como un proceso iterativo. Otros linajes celulares, tales como un linaje mielóide, pueden comprender múltiples linajes o linajes ramificados. En estos casos, pueden definirse las múltiples líneas centroides o una línea centroide definida puede tener ramificaciones.

La subrutina 500 comienza en 502 y avanza hasta 504. En 504, el sistema 100 filtra un conjunto de datos recogidos midiendo las características de las células en el tubo 1 estableciendo una puerta para las células positivas a CD19, creando un primer conjunto de datos normales, y después avanza hasta 506. En 506, el sistema 100 filtra un conjunto de datos recogidos midiendo las características de las células en el tubo 2 estableciendo una puerta para las células positivas a CD19, creando un segundo conjunto de datos normales, y después avanza hasta 508.

En 508, el sistema 100 distingue entre las células maduras e inmaduras en el primer conjunto de datos. Esto puede realizarse, por ejemplo, representando gráficamente las intensidades de fluorescencia para CD45 frente a las intensidades de fluorescencia para CD19 y agrupando el primer conjunto de datos basándose en las entradas del usuario junto con las técnicas de agrupación automáticas. El sistema avanza hasta 510, en donde determina si debe revisar la distinción entre las células maduras e inmaduras en el primer conjunto de datos. Esta decisión puede basarse en los resultados de las técnicas de agrupación automáticas, el análisis estadístico de los datos y/o presentaciones del conjunto de datos generadas basándose en la distinción, y puede ser automática y/o basarse en entradas del usuario. Si el sistema 100 determina que la distinción debe revisarse, el sistema 100 vuelve a 508. Si el sistema 100 determina que la distinción no debe revisarse, el sistema avanza hasta 512.

En 512, el sistema 100 identifica las agrupaciones que representan los estadios I, II, III y IV en el primer conjunto de datos. Esto puede realizarse, por ejemplo, representando gráficamente las intensidades de fluorescencia para CD45 frente a las intensidades de fluorescencia para CD10 y CD20 y agrupando los datos basándose en las entradas del usuario junto con las técnicas de agrupación automáticas. El sistema 100 avanza hasta 514, en donde determina si

debe revisar la identificación de las agrupaciones en el primer conjunto de datos. Esta decisión puede basarse en los resultados de las técnicas de agrupación automáticas, el análisis estadístico de los datos y/o presentaciones del conjunto de datos generadas basándose en la identificación, y puede ser automática y/o basarse en entradas del usuario. Si el sistema 100 determina que la identificación debe revisarse, el sistema 100 vuelve a 512. Si el sistema 100 determina que la identificación debe aceptarse, el sistema avanza hasta 516.

En 516, el sistema 100 identifica una agrupación que representa el estadio I en el segundo conjunto de datos. Esto puede realizarse, por ejemplo, representando gráficamente las intensidades de fluorescencia para CD34 frente a las intensidades de fluorescencia para CD45 y agrupando los datos basándose en las entradas del usuario junto con las técnicas de agrupación automáticas. El sistema avanza hasta 518, en donde determina si debe revisar la identificación de la agrupación del estadio I en el segundo conjunto de datos. Esta decisión puede basarse en los resultados de las técnicas de agrupación automáticas, el análisis estadístico de los datos y/o presentaciones del conjunto de datos generadas basándose en la identificación, y puede ser automática y/o basarse en entradas del usuario. Si el sistema 100 determina que la identificación debe revisarse, el sistema 100 vuelve a 516. Si el sistema 100 determina que la identificación debe aceptarse, el sistema avanza hasta 520.

En 520, el sistema 100 identifica una agrupación que representa el estadio IV en el segundo conjunto de datos. Esto puede realizarse, por ejemplo, representando gráficamente las intensidades de fluorescencia para CD22 frente a las intensidades de fluorescencia para CD34 y agrupando los datos basándose en las entradas del usuario junto con las técnicas de agrupación automáticas. El sistema avanza hasta 522, en donde determina si debe revisar la identificación de la agrupación del estadio IV en el segundo conjunto de datos. Esta decisión puede basarse en los resultados de las técnicas de agrupación automáticas, el análisis estadístico de los datos y/o presentaciones del conjunto de datos generadas basándose en la identificación, y puede ser automática y/o basarse en entradas del usuario. Si el sistema 100 determina que la identificación debe revisarse, el sistema 100 vuelve a 520. Si el sistema 100 determina que la identificación debe aceptarse, el sistema avanza hasta 524.

En 524, el sistema 100 identifica una agrupación que representa los estadios II y III en el segundo conjunto de datos. Esto puede realizarse, por ejemplo, representando gráficamente las intensidades de fluorescencia para CD34 frente a las intensidades de fluorescencia para CD45 basándose en las entradas del usuario junto con las técnicas de agrupación automáticas. El sistema avanza hasta 526, en donde determina si debe revisar la identificación de la agrupación del estadio II/III en el segundo conjunto de datos. Esta decisión puede basarse en los resultados de las técnicas de agrupación automáticas, el análisis estadístico de los datos y/o presentaciones del conjunto de datos generadas basándose en la identificación, y puede ser automática y/o basarse en entradas del usuario. Si el sistema 100 determina que la identificación debe revisarse, el sistema 100 vuelve a 524. Si el sistema 100 determina que la identificación debe aceptarse, el sistema avanza hasta 528.

En 528, el sistema 100 define una línea centroide para cada agrupación identificada en los actos 512, 516, 520 y 524. Una línea centroide para una agrupación puede ser fractal y puede determinarse basándose en las entradas del usuario junto con las técnicas de agrupación automáticas. Una línea centroide para una agrupación puede definirse, por ejemplo, combinando la media geométrica en un espacio n-dimensional, siendo determinado el punto centroide por los algoritmos de agrupación. El sistema avanza hasta 530, en donde determina si debe revisar las líneas centroides definidas para las agrupaciones identificadas. Esta decisión puede basarse en los resultados de las técnicas de agrupación automáticas, el análisis estadístico de los datos y/o presentaciones del conjunto de datos generadas basándose en la identificación, y puede ser automática y/o basarse en entradas del usuario. Si el sistema 100 determina que la identificación debe revisarse, el sistema 100 vuelve a 528. Si el sistema 100 determina que la identificación debe aceptarse, el sistema avanza hasta 532.

En 532, el sistema 100 define una línea centroide normal que se corresponde con un linaje de maduración normal basándose en los conjuntos de datos combinados. Esto puede realizarse, por ejemplo, uniendo las líneas centroides definidas de las agrupaciones identificadas empleando una curvatura geométrica. El sistema 100 también puede combinar las entradas del usuario con las técnicas de agrupación automáticas para definir la línea centroide normal. La distancia a lo largo de esta línea centroide, comparada con el inicio y el fin, es una medición de la maduración de esas células para un linaje dado, según se evalúa mediante la combinación específica de reactivos monoclonales. Debe advertirse que pueden utilizarse diferentes combinaciones de anticuerpos para expandir ciertas partes de los procesos de maduración, mientras que otras combinaciones se centran en otros estadios de maduración u otros linajes.

El sistema avanza hasta 534, en donde determina si debe revisar la definición de la línea centroide normal. Esta decisión puede basarse en los resultados de las técnicas de agrupación automáticas, el análisis estadístico de los datos y/o presentaciones del conjunto de datos generadas basándose en la identificación, y puede ser automática y/o basarse en entradas del usuario. Si el sistema 100 determina que la definición debe revisarse, el sistema 100 vuelve a 532. Si el sistema 100 determina que la definición debe aceptarse, el sistema avanza hasta 536.

En 536, el sistema 100 define un límite, o un radio normal, alrededor de la línea centroide normal definida. El radio normal, o límite, puede ser un radio fijado o puede variar. Por ejemplo, puede ser una distancia fijada, tal como 10, o puede ser una función de una posición sobre la línea centroide normal definida o en el espacio n-dimensional. Puede emplearse una definición para una primera porción de la línea centroide normal definida, y puede emplearse una

segunda definición para otras porciones de la línea centroide normal definida. El radio normal puede determinarse empleando algoritmos estadísticos, tales como las técnicas de agrupación de ondículas y/o las técnicas de envuelta de límite de k-means (empleando la densidad de las agrupaciones) y/o puede basarse en entradas del usuario. También pueden emplearse algoritmos de nivelación para definir patrones tridimensionales específicos y compararse con las observaciones para un número de archivos estadísticamente determinados.

El sistema 100 avanza hasta 538, en donde determina si debe revisar el radio normal definido. Esta decisión puede basarse en los resultados de las técnicas de agrupación automáticas, el análisis estadístico de los datos y/o presentaciones del conjunto de datos generadas basándose en la identificación, y puede ser automática y/o basarse en entradas del usuario. Si el sistema 100 determina que la definición debe revisarse, el sistema 100 vuelve a 536. Si el sistema 100 determina que la definición debe aceptarse, el sistema avanza hasta 540, en donde la subrutina 500 se detiene.

En algunas realizaciones, un sistema 100 puede realizar otros actos que no se muestran en las figuras 18A a 18C, puede no realizar todos los actos mostrados en las figuras 18A a 18C, o puede realizar los actos de las figuras 18A a 18C en un orden diferente. Por ejemplo, la subrutina puede hacerse más iterativa. Por ejemplo, la subrutina 500 puede modificarse de modo que el sistema 100 determine, después del acto 538, si debe revisar la línea centroide normal definida y, si es así, volver a 532. La subrutina 500 también puede llamar a otras subrutinas para realizar diversas funciones, tales como la subrutina 600 descrita a continuación con respecto a la figura 19. La subrutina 500 también puede devolver el valor de cualquier variable deseada, tales como los datos introducidos por un usuario.

La figura 19 es un diagrama de flujo para un ejemplo de subrutina 600 que puede ser empleada por un sistema de diagnóstico, tal como el sistema de diagnóstico 104 ilustrado en la figura 1, para definir una línea centroide normal para un conjunto de agrupaciones. Las figuras 20A y 20B (de modo colectivo figura 20) ilustran representaciones gráficas de los datos, una línea centroide de referencia 702 y una línea centroide normal 704 calculada a partir del estudio.

La subrutina 600 comienza en 602 y avanza hasta 604. En 604, el sistema de diagnóstico 104 identifica un conjunto de puntos de referencia. Por ejemplo, el sistema de diagnóstico 104 puede identificar diez puntos de referencia seleccionados por un usuario después de visualizar diversas representaciones del conjunto de datos. Como alternativa, el sistema de diagnóstico 104 puede identificar una serie de puntos seleccionados estadísticamente o puede combinar las entradas de un usuario con el análisis estadístico. En el estudio, el usuario selecciona diez puntos de referencia después de visualizar diversas representaciones de los datos.

El sistema de diagnóstico 104 avanza hasta 606, en donde define una línea centroide de referencia basándose en el conjunto identificado de puntos de referencia. La figura 20 ilustra un ejemplo de una línea centroide de referencia 702 definida basándose en los diez puntos de referencia identificados por el usuario en el estudio.

El sistema de diagnóstico 104 avanza hasta 608, en donde determina el número de agrupaciones en las que agrupar los datos. Por ejemplo, en el estudio, el sistema de diagnóstico 104 agrupa los datos en cuatro agrupaciones basándose en las entradas del usuario. Como alternativa, el número de agrupaciones puede determinarse estadísticamente (empleando, por ejemplo, la agrupación con DBSCAN) o empleando las entradas de un usuario en combinación con el análisis estadístico.

El sistema de diagnóstico 104 avanza hasta 610, en donde identifica los puntos centroides para el correspondiente número de agrupaciones. Esto puede realizarse asignando cada punto a una agrupación basándose en las entradas del usuario o en algoritmos estadísticos o una combinación de ambos. Véase el análisis anterior de los algoritmos de agrupación. Los respectivos valores de parámetro para todos los puntos asignados a una agrupación se suman, y después el resultado se divide entre el número de puntos en la agrupación para obtener el valor de parámetro para el punto centroide. Por ejemplo, si el sistema de diagnóstico 104 determina, en el acto 608, que debe agrupar los datos en cuatro agrupaciones, el sistema de diagnóstico 104 identificará cuatro puntos centroides, y cada punto se corresponderá con una agrupación. La tabla 2, indicada a continuación, ilustra un ejemplo de cálculo de un punto centroide para una agrupación que contiene 5 puntos de datos en un espacio tridimensional.

Tabla 2 - Ejemplo de cálculo de puntos centroides

	Parámetro X	Parámetro Y	Parámetro Z
Punto 1	25	30	400
Punto 2	30	35	390
Punto 3	25	35	395
Punto 4	25	37	390
Punto 5	20	33	392

El número de puntos, el número de dimensiones y los valores de parámetro para la tabla 2 se seleccionaron para facilitar el ejemplo.

5 El sistema de diagnóstico 104 avanza hasta 612, en donde determina un correspondiente punto más cercano en la línea centroide de referencia para cada punto centroide identificado.

El sistema de diagnóstico 104 avanza hasta 614, en donde calcula la diferencia entre cada punto centroide y el punto más cercano en la línea centroide de referencia. En el estudio, esto se realizó empleando la fórmula de la distancia cuadrada analizada anteriormente, sin ponderación. Véase la ecuación 1.

10 El sistema de diagnóstico avanza hasta 616, en donde ajusta los puntos de referencia basándose en los puntos centroides y los puntos de referencia más cercanos. En el estudio, esto se realizó sumando la diferencia entre el punto centroide y el punto más cercano de una agrupación a los puntos de referencia en esa agrupación.

15 El sistema de diagnóstico avanza hasta 618, en donde redefine la línea centroide de referencia usando los puntos de referencia ajustados y los puntos centroides para cada agrupación. En el estudio, esto se realizó conectando las líneas centroides para cada agrupación empleando una curvatura geométrica. Un ejemplo de una línea centroide de referencia redefinida se ilustra en la figura 20 como la línea 704. La línea centroide de referencia puede refinarse aún más empleando un análisis estadístico. Por ejemplo, los puntos estadísticamente insignificantes o los puntos fuera de un radio definido pueden ser eliminados del conjunto de datos. Los cálculos realizados por el sistema de diagnóstico 104 empleando la subrutina 600 pueden almacenarse para un uso posterior. Por ejemplo, en la agrupación de los datos durante el estudio, el sistema de diagnóstico 104 determina la distancia cuadrada entre los puntos centroides y los puntos de referencia. Estos datos se almacenaron para su uso para calcular los valores de desviación estándar.

20 El sistema de diagnóstico 104 avanza hasta 620, en donde devuelve la línea centroide redefinida y el valor de cualquiera de las variables deseada, tales como las entradas de usuario. El sistema de diagnóstico avanza hasta 622, en donde se detiene.

25 En algunas realizaciones, un sistema 100 puede realizar otros actos que no se muestran en la figura 19, puede no realizar todos los actos mostrados en la figura 19, o puede realizar los actos de la figura 19 en un orden diferente. Por ejemplo, la subrutina puede hacerse más iterativa. Por ejemplo, la subrutina 600 puede modificarse de modo que el sistema 100 determina, después del acto 616, si debe modificar el número de agrupaciones y, si es así, volver al acto 608. La subrutina 600 también puede llamar a otras subrutinas para realizar diversas funciones.

30 La figura 21 es un diagrama de flujo que ilustra un ejemplo de subrutina 800 que puede ser empleada por un sistema de diagnóstico, tal como el sistema de diagnóstico 104 ilustrado en la figura 1, para definir una línea centroide normal y un radio normal para un conjunto de agrupaciones.

35 La subrutina 800 comienza en 802 y avanza hasta 804. En 804, el sistema de diagnóstico 104 identifica un conjunto de puntos de referencia. Por ejemplo, el sistema de diagnóstico 104 puede identificar diez puntos de referencia seleccionados por un usuario después de visualizar diversas representaciones del conjunto de datos. Como alternativa, el sistema de diagnóstico 104 puede identificar una serie de puntos de referencia seleccionados estadísticamente o identificar los puntos de referencia basándose en el análisis estadístico combinado con las entradas de un usuario. En el estudio, un usuario selecciona los puntos de referencia después de visualizar diversas representaciones gráficas de los datos.

40 El sistema de diagnóstico 104 avanza hasta 806, en donde define una línea centroide de referencia basándose en el conjunto identificado de puntos de referencia. La figura 20 ilustra un ejemplo de una línea centroide de referencia 702 definida basándose en los diez puntos identificados por el usuario en el estudio.

45 El sistema de diagnóstico 104 avanza hasta 808, en donde determina el número de agrupaciones en las que agrupar los datos. Por ejemplo, en el estudio, el sistema de diagnóstico 104 agrupa los datos en cuatro agrupaciones basándose en las entradas del usuario.

50 El sistema de diagnóstico 104 avanza hasta 810, en donde identifica los puntos centroides para el correspondiente número de agrupaciones. Esto puede realizarse asignando cada punto a una agrupación basándose en las entradas del usuario o en algoritmos estadísticos o, como en el estudio, en una combinación de ambos. Véase el análisis anterior de los algoritmos de agrupación. Los respectivos valores de parámetro para todos los puntos asignados a una agrupación se suman, y después el resultado se divide entre el número de puntos en la agrupación para obtener el valor de parámetro para el punto centroide. Por ejemplo, si el sistema de diagnóstico 104 determina, en el acto 808, que debe agrupar los datos en cuatro agrupaciones, el sistema de diagnóstico 104 identificará cuatro puntos centroides, y cada punto se corresponderá con una agrupación.

55 El sistema de diagnóstico 104 avanza hasta 812, en donde determina un correspondiente punto más cercano en la línea centroide de referencia para cada punto centroide identificado.

El sistema de diagnóstico 104 avanza hasta 814, en donde calcula la diferencia entre cada punto centroide y el punto más cercano en la línea centroide de referencia. En el estudio, esto se realizó empleando la fórmula de la distancia cuadrada analizada anteriormente, sin ponderación. Véase la ecuación 1.

5 El sistema de diagnóstico 104 avanza hasta 816, en donde ajusta los puntos de referencia basándose en los puntos centroides y los puntos de referencia más cercanos empleando las entradas del usuario, el análisis estadístico o una combinación de ambos. En el estudio, la diferencia entre el punto centroide y el punto más cercano de una agrupación se sumaron a los puntos de referencia en esa agrupación.

10 El sistema de diagnóstico 104 avanza hasta 818, en donde redefine la línea centroide de referencia usando los puntos de referencia ajustados y los puntos centroides para cada agrupación. En el estudio, esto se realizó conectando las líneas centroides para cada agrupación empleando una curvatura geométrica. Un ejemplo de una línea centroide de referencia redefinida se ilustra en la figura 20 como la línea 704.

15 El sistema de diagnóstico 104 avanza hasta 820, en donde define un radio para el conjunto de agrupaciones. Tal como se indicó anteriormente, el radio puede ser una función de la posición sobre la línea centroide de referencia o en el espacio n-dimensional. La línea centroide de referencia y el radio pueden formar diversas conformaciones de agrupaciones. Por ejemplo, la línea centroide de referencia y el radio pueden definir agrupaciones esféricas, hiperesferas o hiperelipsoides. Las agrupaciones pueden tomar la forma de salchichas o barbas o diversas otras formas. En el estudio, el usuario introduce un radio para cada agrupación en el conjunto de agrupaciones normales, y el radio es una distancia desde el punto más cercano sobre la línea centroide de referencia.

20 El sistema de diagnóstico 104 avanza hasta 822, en donde determina si se cumple un criterio de error. Por ejemplo, el sistema de diagnóstico 104 puede determinar si un número estadísticamente insignificante de puntos están fuera de las agrupaciones definidas por la línea centroide de referencia y el radio. Si el criterio de error se cumple, el sistema de diagnóstico 104 avanza hasta 824, en donde la subrutina devuelve la línea centroide definida y el radio para el conjunto de datos, así como cualquier otra variable deseada. Si el criterio de error no se cumple, el sistema de diagnóstico 104 avanza hasta 826, en donde ajusta el conjunto de datos. Por ejemplo, el sistema de diagnóstico 25 104 puede determinar que los puntos estadísticamente insignificantes en el conjunto de datos deben descartarse. El sistema de diagnóstico 104 vuelve a 810 para volver a procesar el conjunto de datos ajustado.

30 En algunas realizaciones de un sistema 100, este puede realizar otros actos no mostrados en la figura 21, puede no realizar todos los actos mostrados en la figura 21, o puede realizar los actos de la figura 21 en un orden diferente. Por ejemplo, la subrutina puede hacerse más iterativa. La subrutina 800 también puede llamar a otras subrutinas para realizar diversas funciones. Por ejemplo, la subrutina 800 puede llamar a un subrutina para determinar si las agrupaciones identificadas deben volver a agruparse, tal como la subrutina 900 ilustrada en la figura 22.

35 La figura 22 es un diagrama de flujo que ilustra un ejemplo de subrutina 900 que puede ser empleada por un sistema de diagnóstico, tal como el sistema de diagnóstico 104 ilustrado en la figura 1, para determinar si los puntos en un conjunto de datos están contenidos dentro de un conjunto de agrupaciones definidas por una línea centroide y un radio. Esta información puede ser utilizada por el sistema de diagnóstico 104 para, por ejemplo, determinar si un conjunto definido de agrupaciones normales debe redefinirse porque demasiadas células se han clasificado como anómalas, o para detectar células anómalas en un conjunto de células de ensayo.

40 La subrutina comienza en 902 y avanza hasta 904. En 904, el sistema de diagnóstico 104 recupera el conjunto de datos y avanza hasta 906. En 906, el sistema de diagnóstico 104 fija un campo de datos asociado con cada punto del conjunto de datos para indicar que la subrutina aún no ha clasificado el punto, y avanza hasta 908.

45 En 908, el sistema de diagnóstico 104 recupera los puntos asociados con una agrupación seleccionada del conjunto de datos y avanza hasta 910. En 910, el sistema de diagnóstico 104 determina si un punto no clasificado asociado con la agrupación seleccionada está dentro de la línea centroide y el radio para la agrupación seleccionada. Esto puede realizarse, por ejemplo, calculando la distancia entre el punto no clasificado y el punto más cercano sobre la línea centroide para la agrupación, clasificando el punto como normal si la distancia es menor que el radio de la agrupación en el punto más cercano sobre la línea centroide, y clasificando el punto como anómalo si la distancia no es menor que el radio de la agrupación en el punto más cercano sobre la línea centroide.

50 Si el sistema de diagnóstico 104 determina, en 910, que el punto está dentro de la agrupación seleccionada, el sistema de diagnóstico 104 avanza hasta 912, en donde clasifica la célula como normal e indica que la célula ha sido clasificada. Si el sistema de diagnóstico 104 determina, en 910, que el punto no está dentro de la agrupación seleccionada, el sistema de diagnóstico 104 avanza hasta 914, en donde clasifica la célula como anómala e indica que la célula ha sido clasificada. Puede emplearse el mismo campo de datos para indicar si una célula no ha sido clasificada, si se clasificado como normal o si se ha clasificado como anómala. Como alternativa, pueden emplearse dos o más campos de datos para indicar, respectivamente, si una célula ha sido clasificada y, si es así, si la célula es normal o anómala.

55 El sistema de diagnóstico 104 avanza de 912 o 914 hasta 916, en donde determina si todas las células asociadas con la agrupación seleccionada han sido clasificadas. Si la respuesta en 916 es NO, el sistema de diagnóstico 104 vuelve a 910. Si la respuesta en 916 es SÍ, el sistema de diagnóstico 104 avanza hasta 918. En 918, el sistema de

diagnóstico 104 determina si todas las agrupaciones en el conjunto de agrupaciones han sido procesadas. Si la respuesta en 918 es NO, el sistema de diagnóstico 104 vuelve a 908. Si la respuesta en 918 es SÍ, el sistema de diagnóstico 104 avanza hasta 920, en donde la subrutina 900 se detiene.

5 En algunas realizaciones de un sistema 100, este puede realizar otros actos no mostrados en la figura 22, puede no realizar todos los actos mostrados en la figura 22, o puede realizar los actos de la figura 22 en un orden diferente. Por ejemplo, la subrutina 900 puede modificarse para que procese un conjunto de datos de modo secuencial, en lugar de procesar los datos agrupación a agrupación y sin fijar un indicador para saber si un punto de datos ha sido clasificado. La subrutina 900 también puede llamar a otras subrutinas, por ejemplo, la subrutina 900 puede llamar a un subrutina para calcular la distancia entre un punto y el punto más cercano en una línea centroide.

10 Los datos generados por el sistema 100, que incluyen los datos generados para definir un linaje celular normal y los datos procedentes de conjuntos de células de ensayo, pueden representarse en diversos formatos y emplearse para diversos objetivos. Por ejemplo, tal como se analizó anteriormente, los datos pueden presentarse como múltiples proyecciones de 2 por 2 de los datos multidimensionales en un sistema de coordenadas cartesianas o como proyecciones pseudo-tridimensionales de los datos multidimensionales en un sistema de coordenadas cartesianas.
15 Véanse las figuras 4-10 y 12-17, y 20, analizadas anteriormente. Pueden utilizarse colores o tonos para mostrar más dimensiones. Estos métodos para mostrar los datos son particularmente útiles para el usuario para definir y redefinir un centroide y un radio normales para un linaje de maduración dado.

Los datos también pueden presentarse como una gráfica bidimensional de frecuencia celular continua a lo largo de una línea centroide definida. La posición a lo largo de la línea centroide se corresponde con una medida del tiempo dentro del proceso de maduración. Así, puede generarse un histograma que muestre la distribución de grupo de las células a lo largo del proceso de maduración. Las figuras 23A y 23B (de modo colectivo figura 23) ilustran gráficas de la frecuencia celular continua muestreada a lo largo de una línea centroide normal definida para un linaje de células B linfoides. Un eje horizontal 1 se corresponde con la posición a lo largo de la línea centroide definida. Se identifican cuatro agrupaciones 2, 3, 4, 5, que se corresponden con estadios de maduración, a lo largo del eje horizontal 1. Un eje vertical 6 se corresponde con el número de puntos en el conjunto de datos en diversos puntos de muestras a lo largo de la línea centroide. En la figura 23, se seleccionaron 108 puntos de muestras para la línea centroide como sigue. Se identificaron diez puntos de referencia a lo largo de la línea centroide. Se calcularon los puntos medios a lo largo de la línea centroide para los diez puntos de referencia, que produjeron 19 puntos. Después se calcularon seis puntos medios para los 19 puntos, que produjeron 108 puntos. También se muestra el porcentaje del total de puntos de datos muestreados para cada agrupación.

Pueden utilizarse más especímenes para definir el centroide normal y el radio. Por ejemplo, el proceso de paneles de cuatro colores y dos tubos descrito anteriormente puede utilizarse para teñir un número mayor de especímenes de médula ósea que muestren una expresión normal del antígeno. Estos especímenes pueden seleccionarse del circuito de producción habitual y pueden incluir especímenes de donantes de médula ósea, pacientes sin neoplasmas hematológicos, y pacientes que han recibido un trasplante con 100% de quimerismo del donante que fueron trasplantados debido a enfermedades que no son ALL. Los especímenes pueden incluir especímenes pediátricos y de adulto. Estos otros especímenes pueden ser aleatorios, o pueden seleccionarse con respecto a criterios deseados, tales como sexo, edad o grupo minoritario. Se espera que la selección por sexo, edad o grupo minoritario no produzca diferencias significativas en el centroide y el radio normales definidos para el linaje de maduración de células B linfoides.

El conjunto de datos expandido puede utilizarse para evaluar la variabilidad de las posiciones de la agrupación para los individuos de los que se recogieron especímenes, así como las diferencias en la composición que se prevén en un análisis habitual de especímenes. El conjunto de datos también puede incluir y/o compararse con los datos de pacientes con especímenes de médula ósea anómalas que no son resultado de un proceso clonal o neoplásico, tales como especímenes procedentes de pacientes en un momento cercano al trasplante de células pluripotenciales que contienen solo las células más inmaduras o pacientes tratados con rituxano (anti-CD20). En estos pacientes, el desarrollo de células B linfoides en la médula ósea se ve truncado en el comienzo del estadio II, siendo eliminadas por el fármaco las células que expresan CD20. El conjunto de datos también puede compararse con especímenes de sangre periférica que contienen solo células de estadio IV.

50 La figura 24 es un diagrama de flujo para un ejemplo de subrutina 1000 que puede ser empleada por un sistema de diagnóstico, tal como el sistema de diagnóstico 104 ilustrado en la figura 1, para comparar un conjunto de datos de ensayo con un conjunto normal de agrupaciones definidas por los puntos centroides. Esta información puede ser utilizada por el sistema de diagnóstico para determinar, por ejemplo, si un conjunto normal de agrupaciones debe redefinirse.

55 La subrutina comienza en 1002 y avanza hasta 1004. En 1004, el sistema de diagnóstico 104 recupera el conjunto de datos y avanza hasta 1006. En 1006, el sistema de diagnóstico 104 asigna los puntos en el conjunto de datos de ensayo a agrupaciones, tal como se analizó anteriormente, y avanza hasta 1008. En 1008, el sistema de diagnóstico 104 determina un punto centroide para cada agrupación en el conjunto de datos de ensayo, tal como se analizó anteriormente. Por ejemplo, el sistema de diagnóstico puede determinar los valores de parámetro para el punto centroide de una agrupación sumando los correspondientes valores de parámetro para cada punto en la agrupación
60

y dividiendo el resultado entre el número de puntos en la agrupación. Como alternativa, el sistema de diagnóstico puede emplear un punto centroide estadísticamente ajustado para el conjunto de datos de ensayo. El sistema de diagnóstico 104 avanza desde 1008 hasta 1010.

5 En 1010, el sistema de diagnóstico 104 determina los correspondientes puntos centroides estadísticos para cada agrupación basándose en conjuntos de datos previamente analizados. Por ejemplo, pueden determinarse los valores de parámetro para un punto centroide estadístico sumando los correspondientes valores de parámetro para puntos centroides definidos para un conjunto de conjuntos de datos previamente analizados y dividiendo el resultado entre el número de conjuntos de datos. El sistema de diagnóstico 104 avanza desde 1010 hasta 1012.

10 En 1012, el sistema de diagnóstico 104 determina si se cumple un criterio de error para una agrupación en el conjunto de datos de ensayo. Por ejemplo, el sistema de diagnóstico 104 puede comparar el logaritmo de la distancia entre el punto centroide de la agrupación y el correspondiente punto centroide estadístico con un valor umbral, tal como 2,5. Si el logaritmo de la distancia es mayor que el valor umbral, el sistema de diagnóstico 104 puede determinar que no se ha cumplido el criterio de error. Pueden emplearse otros criterios de error.

15 Si el sistema de diagnóstico 104 determina, en 1012, que el criterio de error para una agrupación en el conjunto de datos de ensayo no se cumple, el sistema de diagnóstico 104 avanza hasta 1014, en donde se fija una indicación de un error para la agrupación en el conjunto de datos de ensayo. Si el sistema de diagnóstico 104 determina, en 1012, que el criterio de error para una agrupación en el conjunto de datos de ensayo se cumple, el sistema de diagnóstico 104 avanza hasta 1016, en donde se fija una indicación de no error para la agrupación en el conjunto de datos de ensayo.

20 El sistema de diagnóstico 104 avanza desde 1014 o 1016 hasta 1018, en donde determina si todas las agrupaciones en el conjunto de datos de ensayo han sido evaluadas. Si el sistema de diagnóstico 104 determina, en 1018, que no todas las agrupaciones han sido procesadas, entonces el sistema de diagnóstico 104 vuelve a 1012. Si el sistema de diagnóstico 104 determina, en 1018, que todas las agrupaciones han sido evaluadas, el sistema de diagnóstico 104 avanza hasta 1020, en donde la subrutina 1000 se detiene.

25 En algunas realizaciones de un sistema 100, este puede realizar otros actos no mostrados en la figura 24, puede no realizar todos los actos mostrados en la figura 24, o puede realizar los actos de la figura 24 en un orden diferente. Por ejemplo, la subrutina 1000 puede modificarse para comparar secuencialmente todos los conjuntos de datos en un conjunto normal de conjuntos de datos para determinar cuáles conjuntos de datos deben retirarse del conjunto normal de conjuntos de datos.

30 Tras haber definido los límites de la agrupación (centroide normal y radio) para un linaje de maduración normal, una muestra de ensayo puede analizarse sometiéndola a la exposición de los mismos reactivos y protocolos de medición empleados con los conjuntos de datos utilizados para definir el linaje de maduración normal. Los resultados para la muestra de datos de ensayo entonces pueden compararse con el linaje de maduración normal definido, permitiendo que la muestra de ensayo sea caracterizada y diagnosticada. Solo es necesario proporcionar a un sistema, tal como el sistema 100 ilustrado en la figura 1, la definición de los límites de agrupaciones normales para diagnosticar una muestra de ensayo. Como alternativa, al sistema 100 se le puede proporcionar el conjunto de datos normales definidos y la línea centroide y el radio definidos, o al sistema 100 se le puede proporcionar el conjunto de datos normales definidos y puede determinar la definición de los límites de las agrupaciones normales.

35 La figura 25 ilustra una estructura de datos 1100 adecuada para proporcionar las definiciones para los límites normales definidos para un linaje celular. La estructura de datos 1100 y las correspondientes instrucciones pueden ser almacenadas en un medio de lectura por ordenador, tal como una memoria, que puede incluir la memoria 112 ilustrada en la figura 1, o memorias portátiles, tales como CD ROM, disquetes y/o memorias flash, y/o pueden transmitirse como un señal en un medio de transmisión de señales, tales como un medio cableado o sin cables. La estructura de datos 1100 tiene una sección de cabecera 1102 que describe las localizaciones de las otras secciones de la estructura de datos 1100. Una sección de texto 1104 contiene información que describe diversos aspectos de la estructura de datos 1100, tales como el número de agrupaciones y la forma en que se definen la línea centroide y el radio. Por ejemplo, la línea centroide puede definirse proporcionando parámetros para la inserción en una ecuación o proporcionando puntos de referencia que se van a conectar entre sí, o una de sus combinaciones. De modo similar, el radio puede definirse proporcionando parámetros para la inserción en una ecuación o valores de radio fijados para una agrupación, o una de sus combinaciones. Por ejemplo, el radio puede tener un valor fijado dentro de una agrupación y puede ser una función de la posición dentro de una segunda agrupación. Una sección de datos de centroide 1106 de la estructura de datos 1100 contiene información que define la línea centroide, y una sección de datos de radio 1108 contiene información que define el radio. Si se desea, puede proporcionarse un conjunto de datos normales para definir la línea centroide y el radio normales, como un campo de datos adicional en la estructura de datos 1100 o en una estructura de datos separada, tal como la estructura de datos 200 ilustrada en la figura 2.

Las agrupaciones individuales también pueden dividirse en subagrupaciones, que pueden ser definidas y analizadas empleando procesos similares a los analizados anteriormente. Por ejemplo, la subrutina 800 ilustrada en la figura 21 puede modificarse para definir una línea o punto centroide y un radio para una subagrupación, y la subrutina 900

ilustrada en la figura 22 puede modificarse para determinar si un conjunto de células de ensayo contiene una subagrupación que se corresponde con una subagrupación definida. Se espera que la agrupación con DBSCAN sea particularmente útil para identificar subagrupaciones que se corresponden con un nivel de submaduración dentro de una agrupación que se corresponde con un nivel de maduración dentro de un linaje celular.

5 Un sistema 100 puede utilizarse para diagnosticar un conjunto de datos de ensayo comparando el conjunto de datos de ensayo con la línea centroide y el radio normales definidos para el linaje celular. El conjuntos de datos de ensayo completo puede compararse con el normal definido y mostrarse por medio de un sistema de diagnóstico, tal como el sistema de diagnóstico 100 en la figura 1, sobre un medio o dispositivo de presentación adecuado, tal como un barrido de monitor, una presentación de matriz activa o pasiva, o sobre un medio pasivo, tal como papel o papel vitela. Como alternativa, los datos de los acontecimientos en el conjunto de datos de ensayo dentro de posiciones "normales", específicamente linfoblastos de linaje B, pueden restarse del conjunto de datos de ensayo, dejando un conjunto de datos "anómalo" que se corresponde con poblaciones residuales de células "anómalas" potenciales (linfoblastos leucémicos). El resto de los acontecimientos anómalos después puede analizarse y mostrarse por medio de un sistema de diagnóstico, tal como el sistema de diagnóstico 104 en la figura 1, y el usuario. El resto de los acontecimientos anómalos puede definir un subconjunto anómalo del conjunto de datos de ensayo. Las técnicas de agrupación, tales como las analizadas anteriormente, pueden emplearse para identificar agrupaciones con el subconjunto anómalo del conjunto de datos de ensayo, y puede emplearse un análisis estadístico para determinar si cualquiera de las agrupaciones identificadas dentro del subconjunto anómalo es significativa.

20 El sistema 100 puede ensayarse antes de ser empleado para diagnosticar cánceres. Por ejemplo, una serie de especímenes procedentes de paciente con ALL manifiesta puede teñirse y los datos pueden recogerse para la comparación con los especímenes normales. Se espera que estos especímenes presenten células normales identificables que el sistema 100 pueda identificar, así como células leucémicas positivas a CD19 que no se encuentren dentro de los límites definidos por la línea centroide y el radio normales. Debe advertirse que todas las células leucémicas ALL del linaje B expresan CD19 y, por tanto, se incluirán dentro de la estrategia de establecimiento de una puerta original.

30 El ensayo del sistema 100 puede incluir mezclar diferentes proporciones de datos de todos los pacientes con ALL con especímenes normales para imitar la detección de la enfermedad residual. Por ejemplo, el sistema 100 puede procesar 25 especímenes normales y generar una línea centroide y un radio definidos para un linaje de maduración normal, que el sistema 100 puede almacenar en la memoria 112 como objetos digitales. Esta información puede autoconectarse con los algoritmos estadísticos en el archivo de datos que contiene una agrupación de células aberrantes. Los acontecimientos celulares confinados a la región de las agrupaciones normales pueden eliminarse y el resto de los acontecimiento representarán una agrupación "anómala". El número y la localización de las células tumorales previstas en la mezcla pueden compararse con las que ya se han identificado. Esto puede realizarse antes y después de haber restado las células "normales" del conjunto de datos de ensayo.

35 Pueden emplearse algoritmos de nivelación, que incluyen algoritmos de promediado y de filtración, para nivelar la representación de los datos. Por ejemplo, puede promediarse una porción de una agrupación. Por ejemplo, puede saberse que el nivel de maduración promedio para una porción de una agrupación concreta es un indicador significativo de que la muestra es normal, pero que las varianzas individuales en esa porción de la agrupación no son significativas.

40 Los datos para dos conjuntos de datos pueden presentarse simultáneamente de esta manera. Por ejemplo, los datos de una muestra de ensayo pueden superponerse a los datos empleados para definir la línea centroide normal. Puede emplearse un primer color u otro indicador para ilustrar la distribución normal, y puede empelarse un segundo color u otro indicador para ilustrar la distribución del espécimen de ensayo.

45 Pueden emplearse presentaciones de los datos más simplificadas y compararse para el impacto visual y la facilidad de interpretar el desarrollo normal y/o anómalo. Por ejemplo, las proporciones de células en cada uno de los cuatro estadios del linaje de células B linfoides puede representarse gráficamente para representar las agrupaciones identificables en el espacio de datos. Los acontecimientos totales en cada una de las cuatro agrupaciones pueden presentarse para que representen la maduración de las células dentro de la médula ósea normal y/o para una muestra de ensayo frente a una representación normal. Los parámetros de las células anómalas que pueden representarse incluyen: número de acontecimiento anómalos, distancia desde las normales, dispersión dentro de la población anómala, y marcadores celulares que distinguen las células aberrantes de las normales.

55 La figura 26 ilustra un ejemplo de una representación simplificada de datos obtenidos de una muestra de ensayo superpuesta a una representación de un conjunto de datos normales definidos. Un eje horizontal 1 se corresponde con una indicación de un nivel de maduración de un linaje celular e indica cuatro estadios 2, 3, 4, 5, que se corresponden con agrupaciones de nivel de maduración dentro del linaje celular. Un eje vertical 6 se corresponde con una indicación de un número de células a diversos niveles de maduración. La indicación puede ser, por ejemplo, un porcentaje del número total de células dentro de un estadio o un indicador logarítmico. Una banda 7 ilustra un intervalo normal definido para una muestra. La banda 7 puede corresponder, por ejemplo, a una desviación estándar para un conjunto normal de células, o puede corresponderse con una línea centroide y un radio definidos para un conjunto de células normal. Una línea discontinua 8 ilustra los resultados para una muestra de ensayo.

60

Puede emplearse un proceso de control de calidad. Por ejemplo, pueden emplearse preparaciones de esferas para evaluar la actuación del instrumento, tales como esferas Rainbow (RCP y RFP, Spherotech, Libertyville, IL) que son microesferas de plástico con un tinte embebido en el interior de la partícula que asegura la estabilidad de la fluorescencia. Las esferas RFP solo presentan un único pico en cada uno de los cuatro canales de fluorescencia y se emplean como patrón primario. Las esferas RCP, una mezcla de esferas de seis intensidades observadas en todos los canales, actúan como patrón secundario y proporcionan datos con respecto a la linealidad de cada uno de los detectores de fluorescencia. Se establece la compensación espectral de emisión de fluorescencia y se controla mediante la tinción de sangre normal con un anticuerpo anti-CD4 conjugado con cada uno de los cromóforos empleados (FITC, PE, PerCP, y APC). Las células teñidas con estos anticuerpos por separado se analizan para asegurar que se detecta la fluorescencia del cromóforo esperado solo en el canal de fluorescencia apropiado (24). Cada lote de reactivo empleado en la evaluación de las células se titula antes de introducirse en el recuento. Se selecciona la titulación del anticuerpo que produzca la fluorescencia máxima y se comprueba la especificidad del reactivo para cada nuevo lote de anticuerpos.

Empleando estos procedimientos de control de calidad, dos citómetros de flujo generaron experimentalmente resultados idénticos para el mismo espécimen. En un estudio de sangre de adulto normal empleando estos procedimientos de control de calidad, se descubrió que la intensidad de CD4 sobre los linfocitos casi no varía para los 21 individuos ensayados en los dos instrumentos, recogida a lo largo de un periodo de ocho meses. El promedio de intensidad de fluorescencia de CD4 para estos 21 individuos fue de 1596 +/-116 desviación estándar en unidades de fluorescencia, que produce una CV de 7%. Estos resultados demuestran que, en un espacio de datos con un intervalo dinámico de cuatro décadas, la variación biológica de un individuo a otro para este antígeno es fundamentalmente cero. La cantidad de CD4 expresado sobre los linfocitos es, en sí misma, un patrón biológico. La cuantificación de la posición de la línea centroide position (medida en células de médula ósea inmaduras) puede compararse con la variabilidad de los antígenos expresados sobre células de sangre maduras normales, lo cual proporcionará una base para comprender la variación biológica entre individuos con respecto a la intensidad de expresión del antígeno durante la maduración de las células sanguíneas, no solo en células maduras.

La tolerancia de un sistema, tal como el sistema 100 ilustrado en la figura 1, puede determinarse cambiando el valor diana para las esferas de control de calidad fluorescentes de patrón primario en una cantidad conocida (factores de 2, y 4). En otras palabras, un sistema puede desafinarse por medio de cantidades conocidas. Cada canal puede ensayarse por separado y los canales pueden ensayarse juntos, después de establecer la compensación apropiada. Por ejemplo, pueden recogerse células de la médula ósea teñidas con las combinaciones de cuatro colores para cada escenario y los datos pueden analizarse empleando el sistema que se va a ensayar. Esto evalúa hasta dónde puede alejarse de la conformación óptima del patrón un sistema y seguir realizando la identificación correcta de las células en los estadios del desarrollo por parte del sistema. Esta actuación después define la tolerancia requerida para un programa de control de la calidad basándose en la capacidad del sistema para identificar las poblaciones celulares apropiadas.

Tal como reconocerán los expertos en la técnica, los anteriores métodos pueden emplearse en una serie de entornos que incluyen, pero no se limitan al diagnóstico y el control de la enfermedad y del tratamiento.

Los anteriores ejemplos se ofrecen como ilustración y no como limitación. Por consiguiente, la invención no está limitada, excepto por las reivindicaciones adjuntas.

REIVINDICACIONES

1.- Un sistema de diagnóstico que comprende un medio para llevar a cabo un método de caracterización de un conjunto de ensayo de células biológicas, que comprende:

5 a) cartografiar cada célula en el conjunto de ensayo de células biológicas a un punto correspondiente en un espacio n-dimensional, en el que $n \geq 3$ empleando un protocolo definido en el que al menos algunas de las n-dimensiones son datos de intensidad fluorescente de las células, y los puntos correspondientes forman un conjunto de ensayo de puntos;

10 b) comparar el conjunto de ensayo de puntos con un conjunto definido de agrupaciones de células biológicas normales, es decir, un conjunto definido de agrupaciones normales, en el espacio n-dimensional, en el que cada agrupación en el conjunto definido de agrupaciones normales se corresponde con un nivel de maduración dentro de un linaje celular, y en el que, para el conjunto de agrupaciones normales, se define una línea centroide de referencia y un radio, y la línea centroide de referencia y el radio forman diversas conformaciones de agrupaciones, en el que la línea centroide de referencia para el conjunto de agrupaciones se basa en un conjunto de puntos de referencia, y en el que el radio es una función de una posición en la línea centroide de referencia, en el que la línea centroide de referencia y el radio se ajustan de una manera iterativa empleando al menos un análisis estadístico y la entrada de un usuario, en el que el ajuste incluye determinar una distancia desde un punto centroide identificado de cada agrupación y un punto más cercano en la línea centroide de referencia;

15 c) determinar, para cada punto del conjunto de ensayo de puntos, basándose en la línea centroide de referencia y el radio determinados, si se encuentra fuera del conjunto definido de agrupaciones normales basándose en la comparación; y

20 d) caracterizar el conjunto de ensayo de células biológicas basándose al menos en parte en una determinación de una cantidad de células anómalas en el conjunto de ensayo de células biológicas basándose en la determinación de la etapa c).

25 2.- El sistema de diagnóstico de la reivindicación 1, en el que la línea centroide de referencia para el conjunto de agrupaciones se basa en un conjunto de puntos de referencia y en los puntos centroides de las agrupaciones en el conjunto de agrupaciones normales.

3.- El sistema de diagnóstico de la reivindicación 1, que es para diagnosticar el cáncer y comprende, además de las etapas de la reivindicación 1, las siguientes etapas:

30 exponer cada célula en un conjunto normal de células biológicas a una pluralidad de cuatro o más reactivos empleado un primer protocolo;

medir una correspondiente pluralidad de intensidades de fluorescencia de cada célula en el conjunto normal de células biológicas empleando un segundo protocolo;

35 cartografiar cada célula en el conjunto normal de células biológicas a un punto correspondiente en un espacio n-dimensional, basándose al menos en parte en la pluralidad de intensidades de fluorescencia medidas en la célula en el conjunto normal de células biológicas, en el que los puntos correspondientes forman un conjunto normal de puntos, en el que la definición de la línea centroide de referencia y el radio para el conjunto de agrupaciones normales se basa en el cartografiado del conjunto normal de puntos en el espacio n-dimensional;

exponer cada célula en el conjunto de ensayo de células biológicas a la pluralidad de reactivos empleando el primer protocolo;

40 medir una correspondiente pluralidad de intensidades de fluorescencia de cada célula en el conjunto de ensayo de células biológicas empleando el segundo protocolo;

45 cartografiar cada célula en el conjunto de ensayo de células biológicas a un punto correspondiente en un espacio n-dimensional, basándose al menos en parte en la pluralidad de intensidades de fluorescencia medidas en la célula en el conjunto de ensayo de células biológicas, en el que los puntos correspondientes forman un conjunto de ensayo de puntos;

comparar el conjunto de ensayo de puntos con el conjunto de agrupaciones normales; y

caracterizar y diagnosticar el conjunto de ensayo de células biológicas basándose en la comparación.

4.- El sistema de diagnóstico de la reivindicación 1, que comprende además las etapas de:

50 cartografiar cada célula en el conjunto normal de células biológicas a un punto correspondiente en un espacio n-dimensional utilizando el protocolo definido, en el que los puntos correspondientes forman un conjunto normal de puntos, en el que la definición de la línea centroide de referencia y el radio para el conjunto de agrupaciones normales en el espacio n-dimensional se basa en el cartografiado del conjunto normal de puntos en el espacio n-

dimensional.

5.- El sistema de diagnóstico de la reivindicación 1, que comprende además las etapas de:

5 representar el conjunto de ensayo en un diagrama de coordenadas cartesianas que comprende un primer eje que se corresponde con una maduración de la célula dentro de un linaje celular, y un segundo eje que se corresponde con una frecuencia de aparición; y

representar, en el diagrama de coordenadas cartesianas, el conjunto de agrupaciones normales en el espacio n-dimensional.

6.- El sistema de diagnóstico de cualquiera de las reivindicaciones anteriores para diagnosticar el cáncer, en el que al menos una agrupación en el conjunto de agrupaciones es un hiperelipsoide.

10

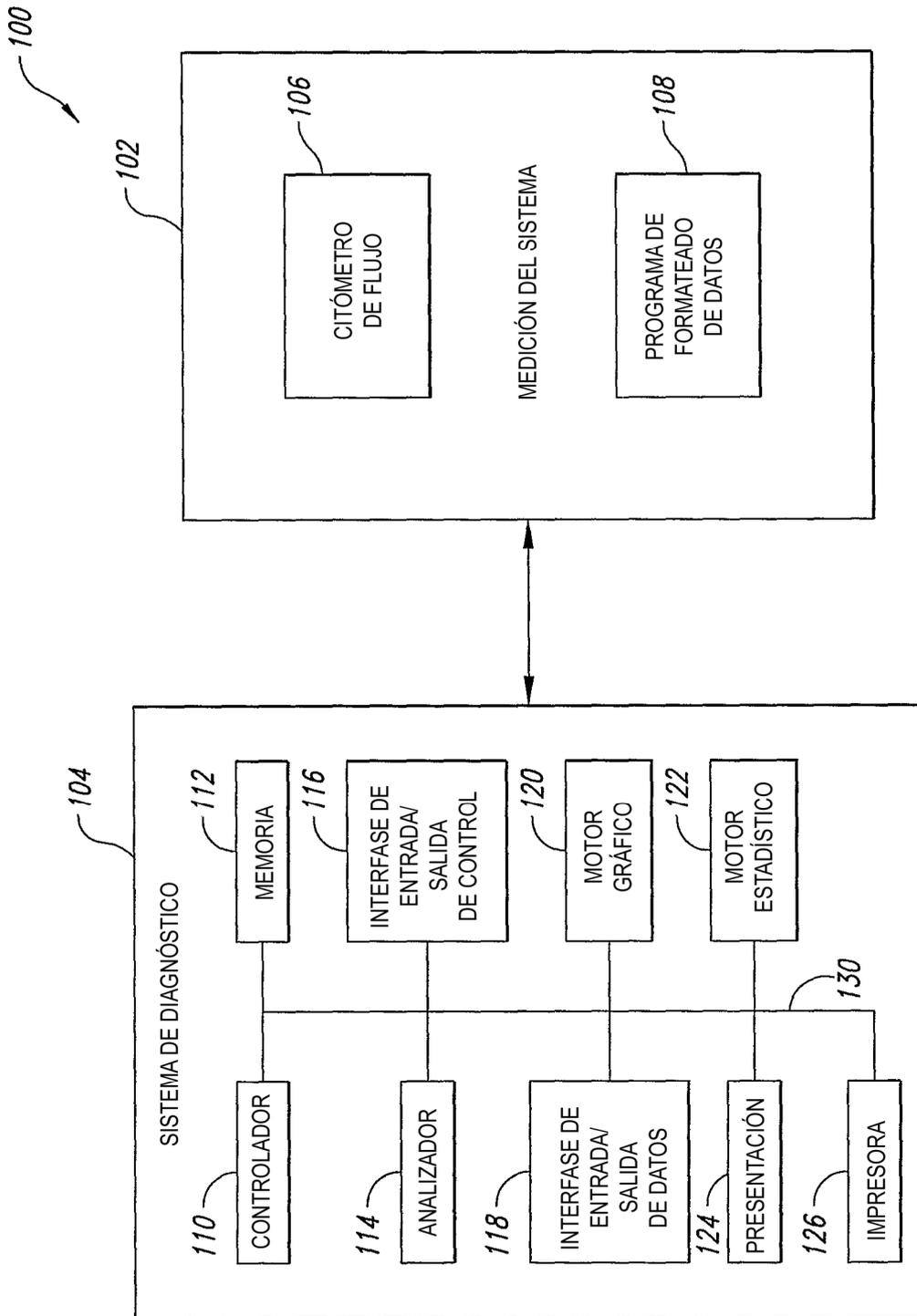


FIG. 1

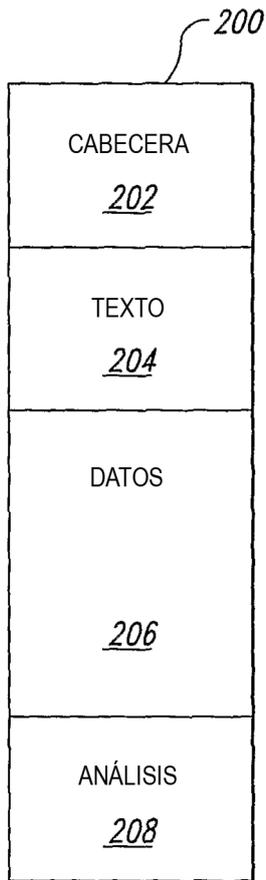


FIG. 2

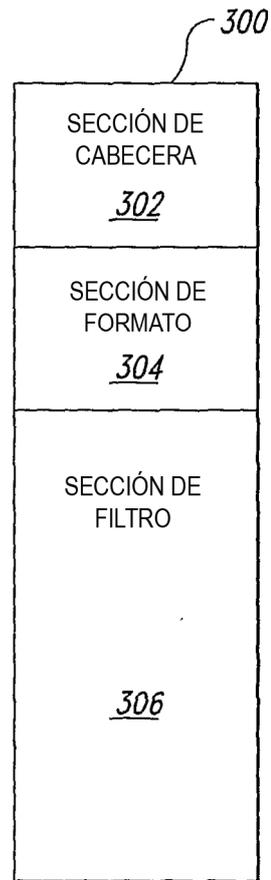


FIG. 3

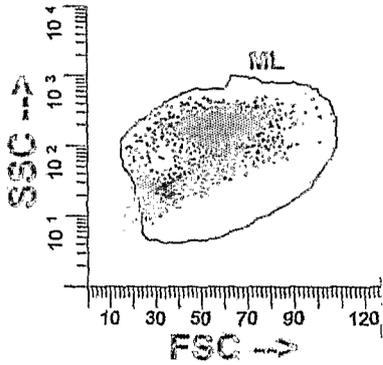


FIG. 4A

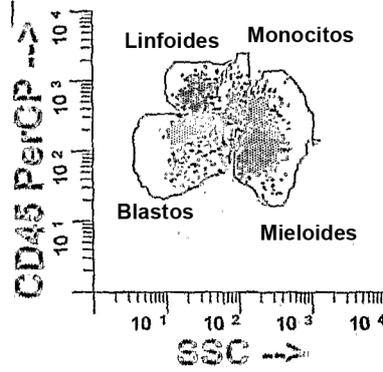


FIG. 5A

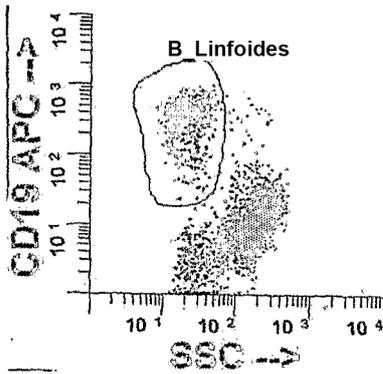


FIG. 6A

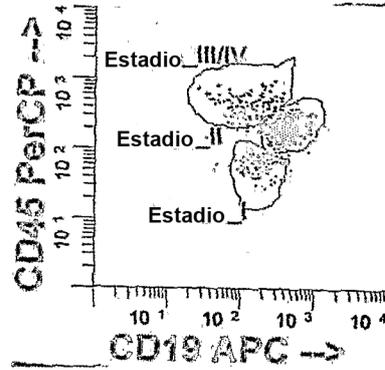


FIG. 7A

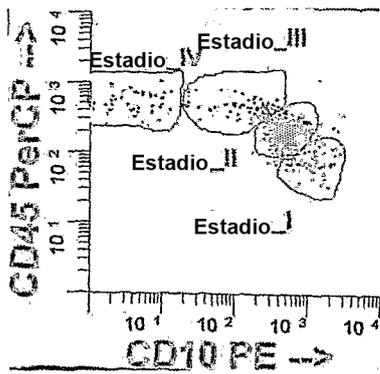


FIG. 8A

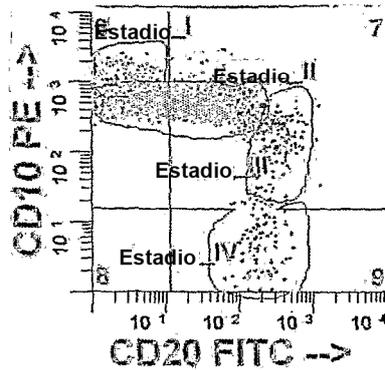


FIG. 9A

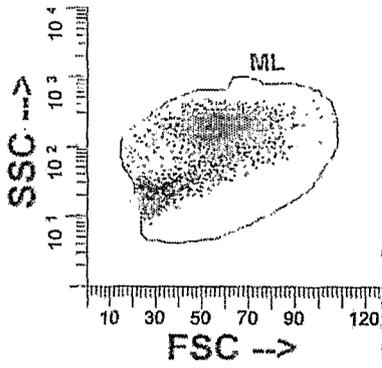


FIG. 4B

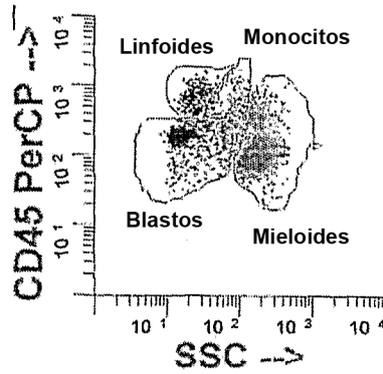


FIG. 5B

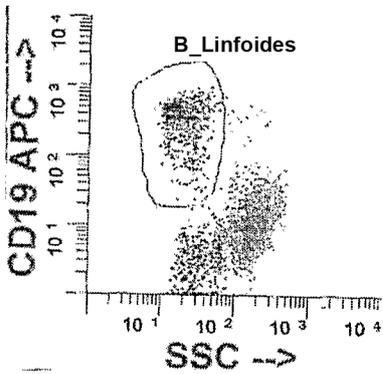


FIG. 6B

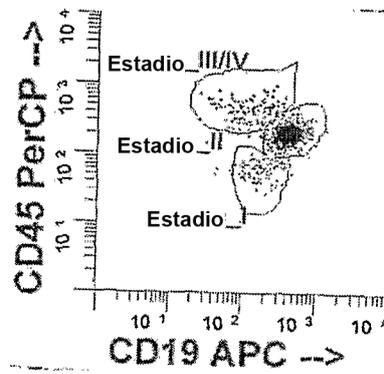


FIG. 7B

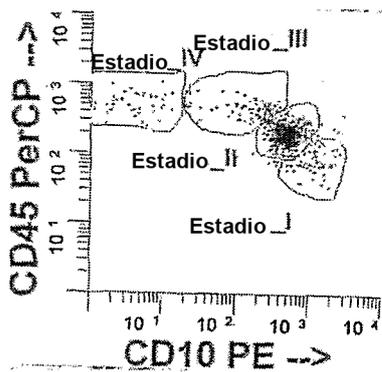


FIG. 8B

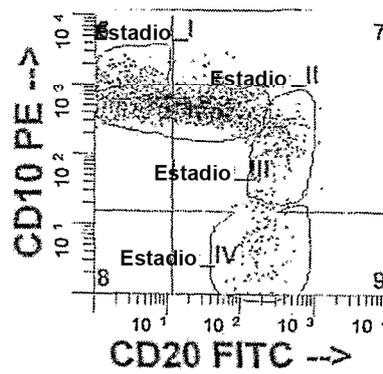


FIG. 9B

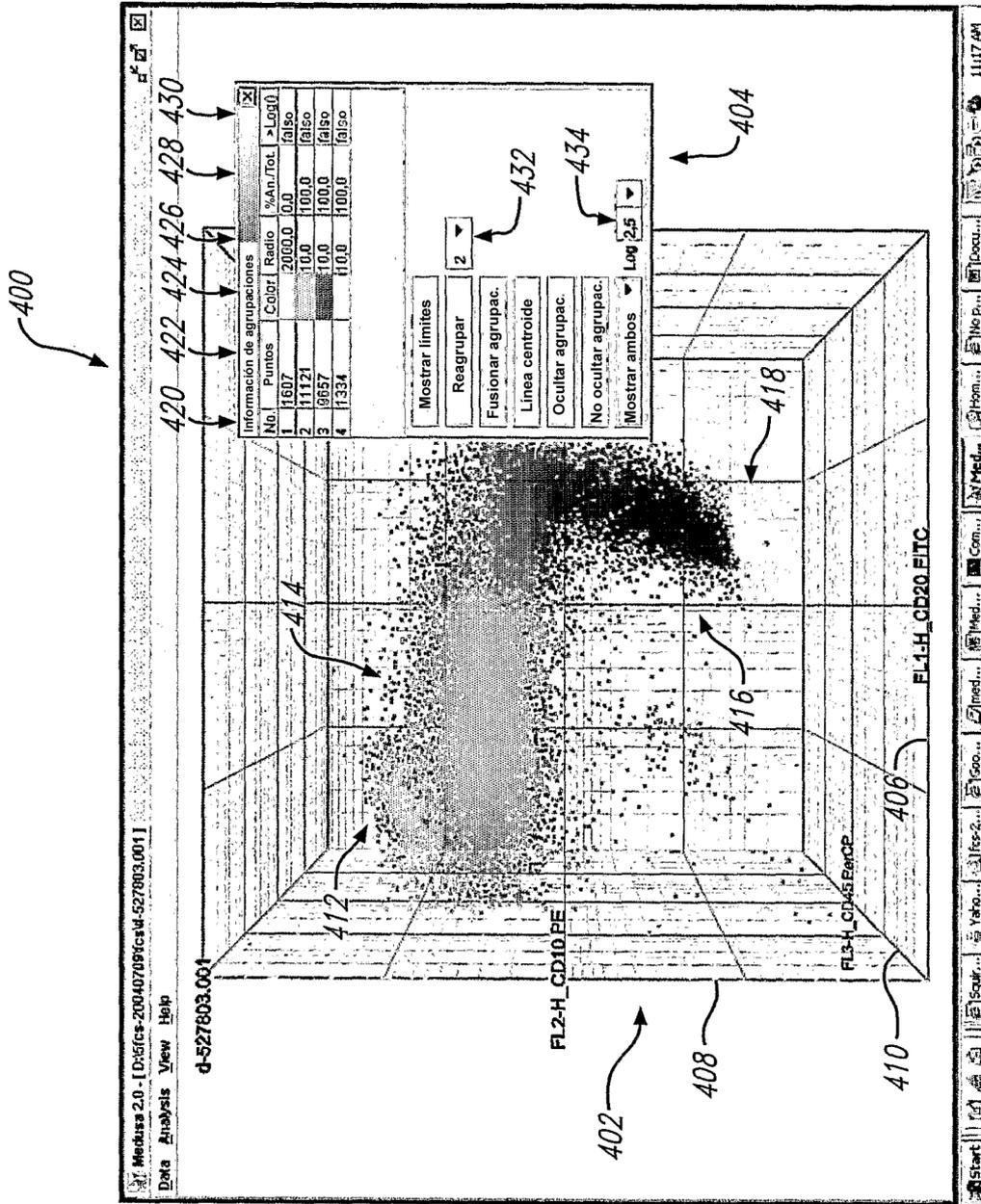


FIG. 10A

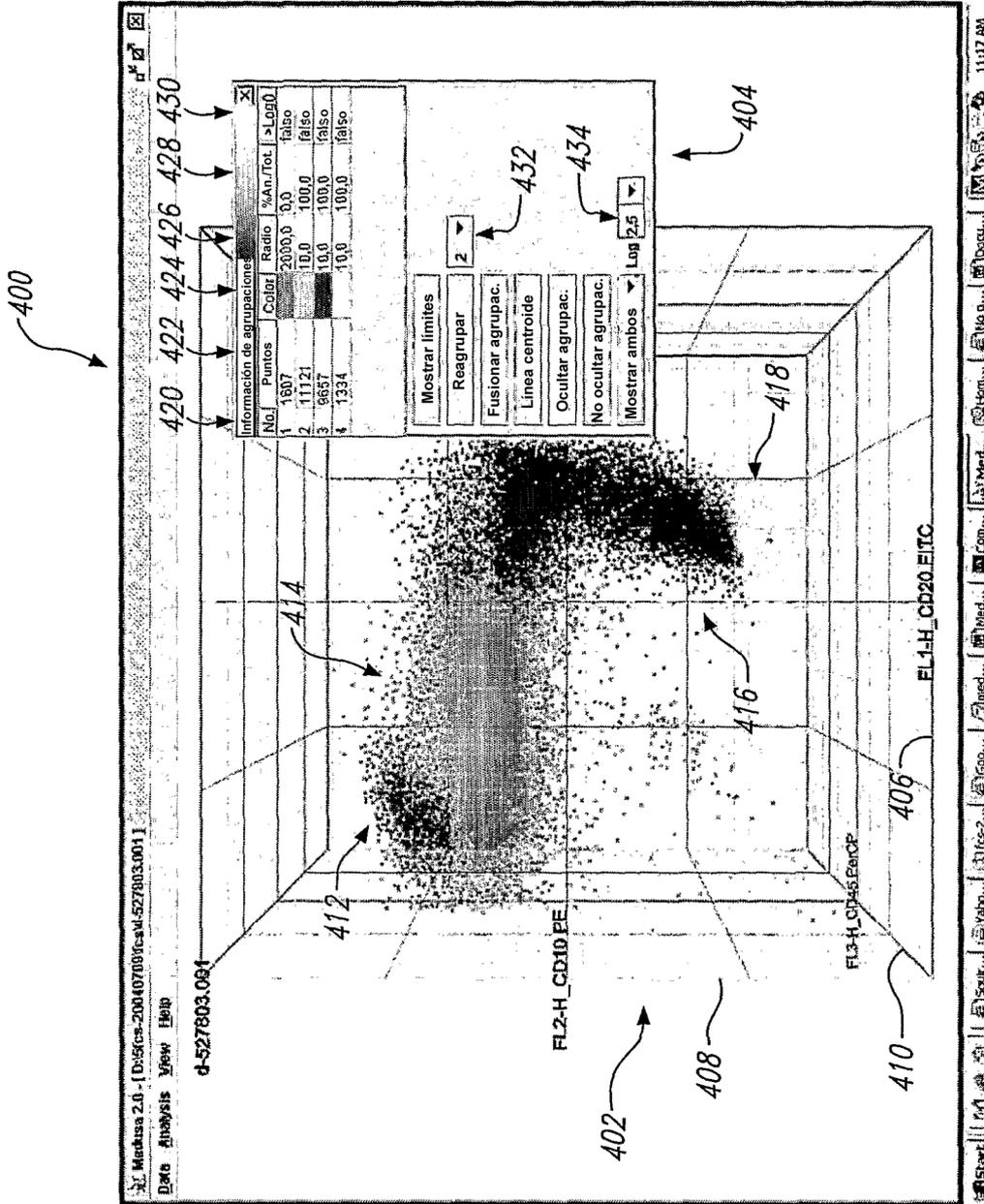


FIG. 10B

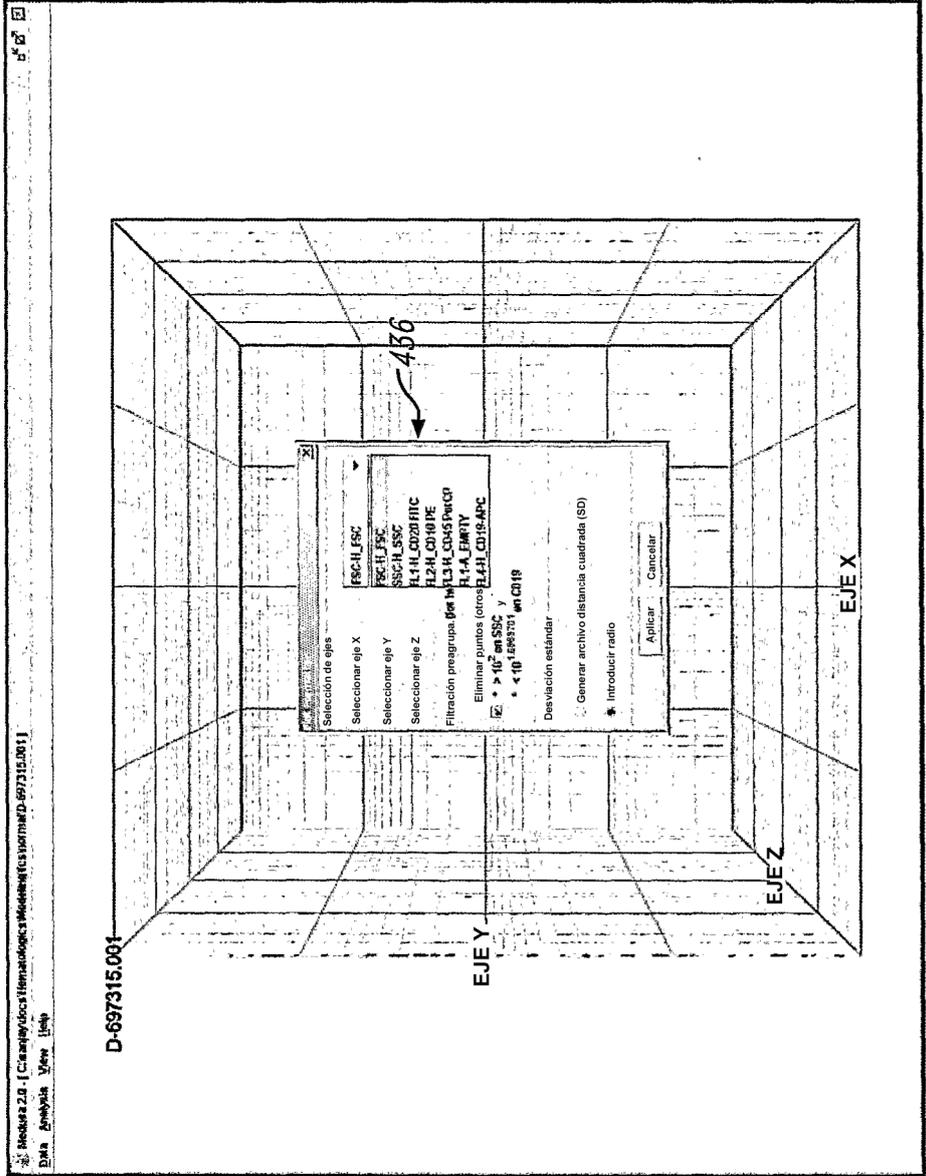


FIG. 11A

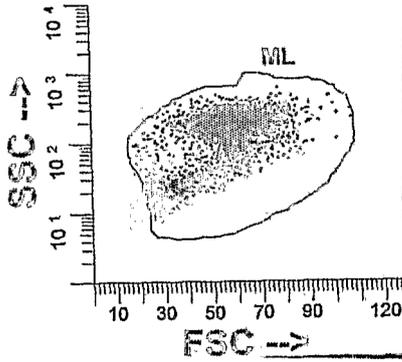


FIG. 12A

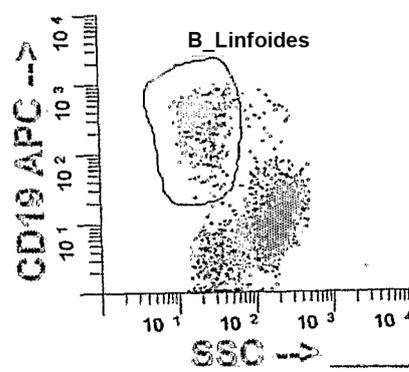


FIG. 13A

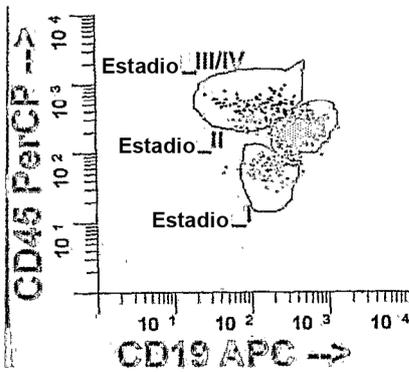


FIG. 14A

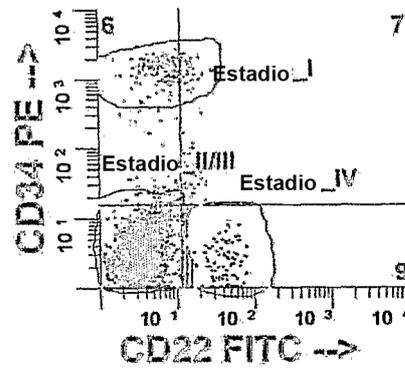


FIG. 15A

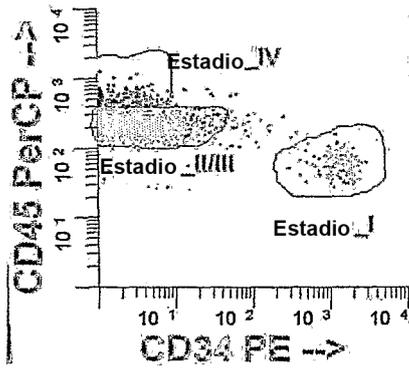


FIG. 16A

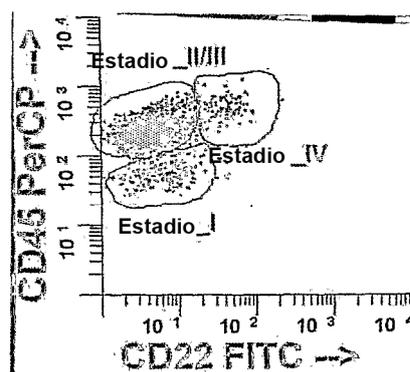


FIG. 17A

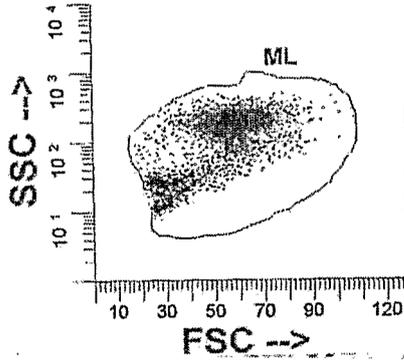


FIG. 12B

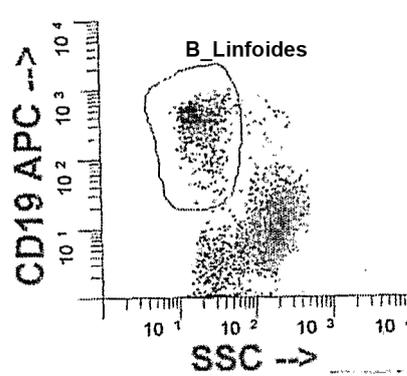


FIG. 13B

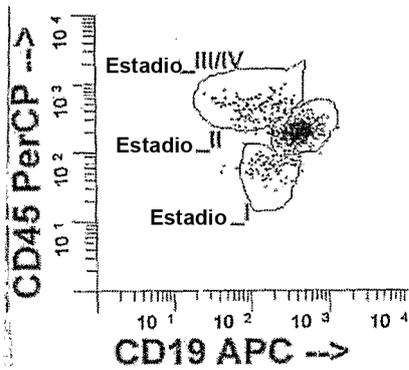


FIG. 14B

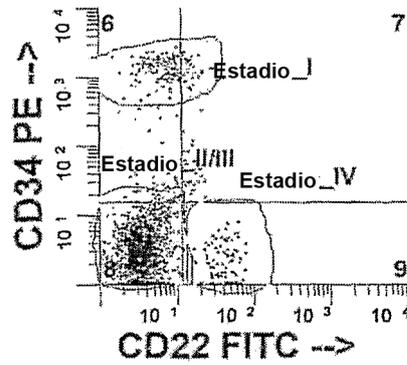


FIG. 15B

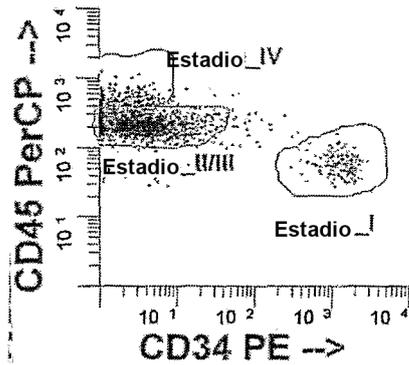


FIG. 16B

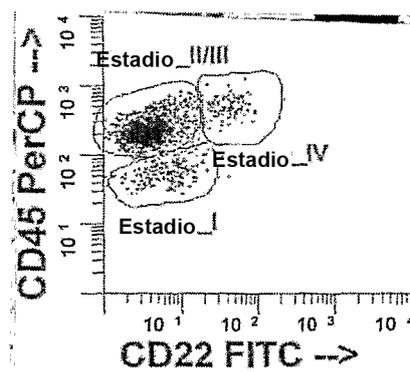


FIG. 17B

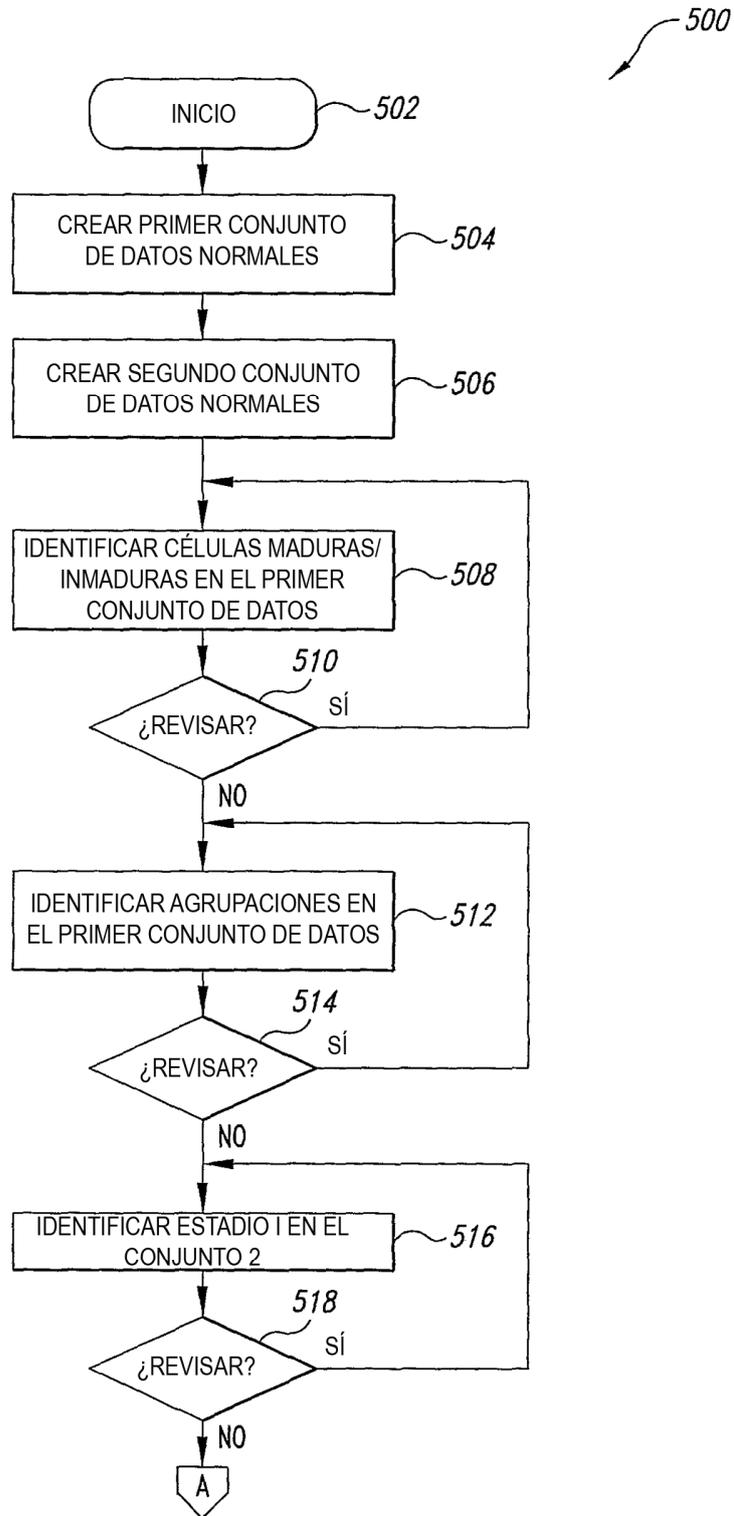


FIG. 18A

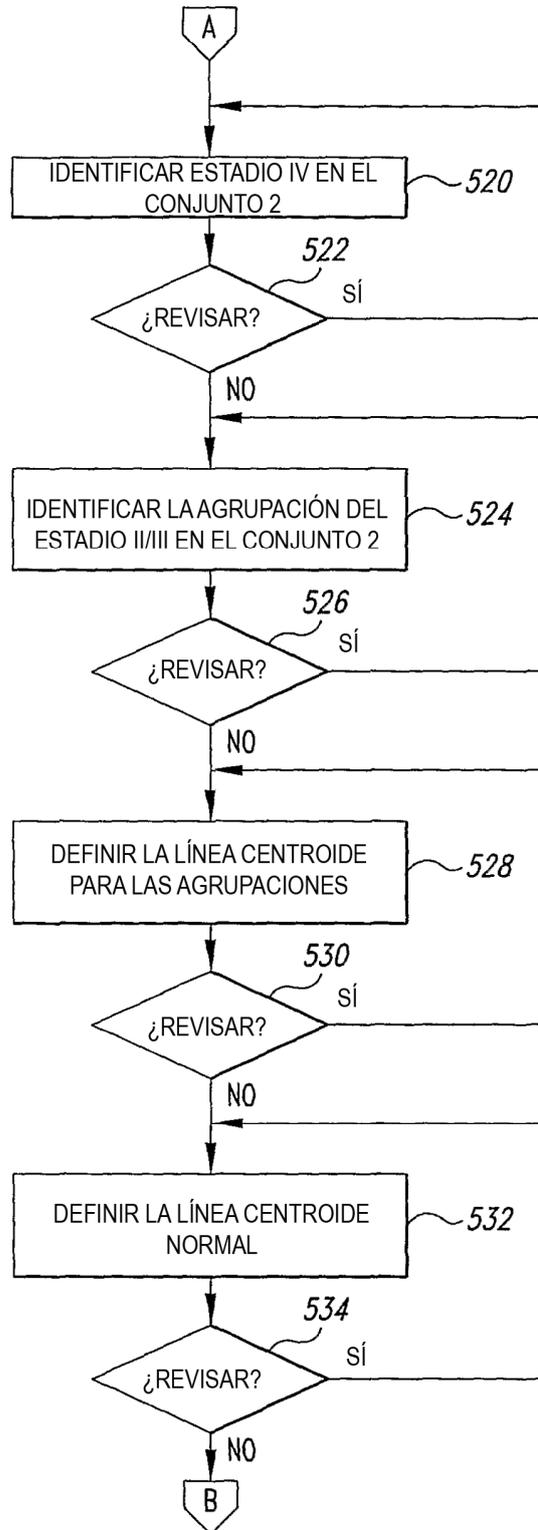


FIG. 18B

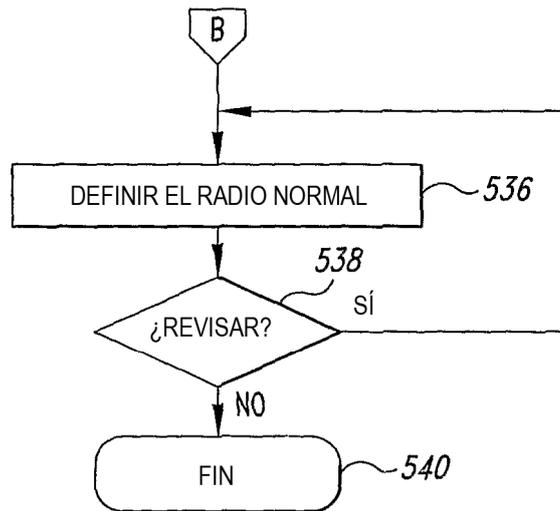


FIG. 18C

600

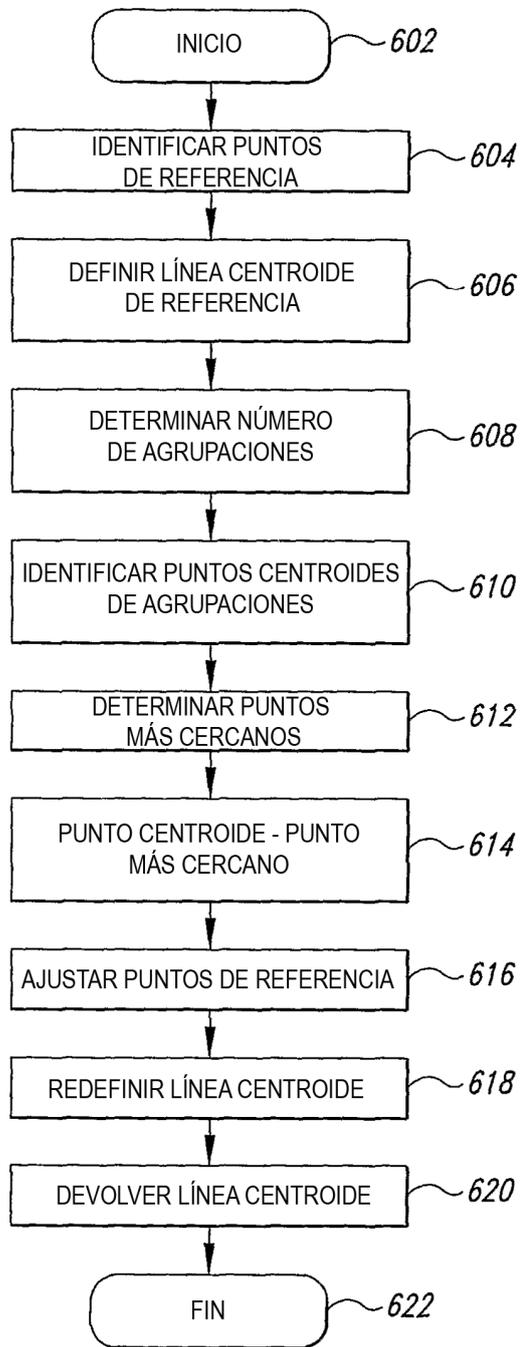


FIG. 19

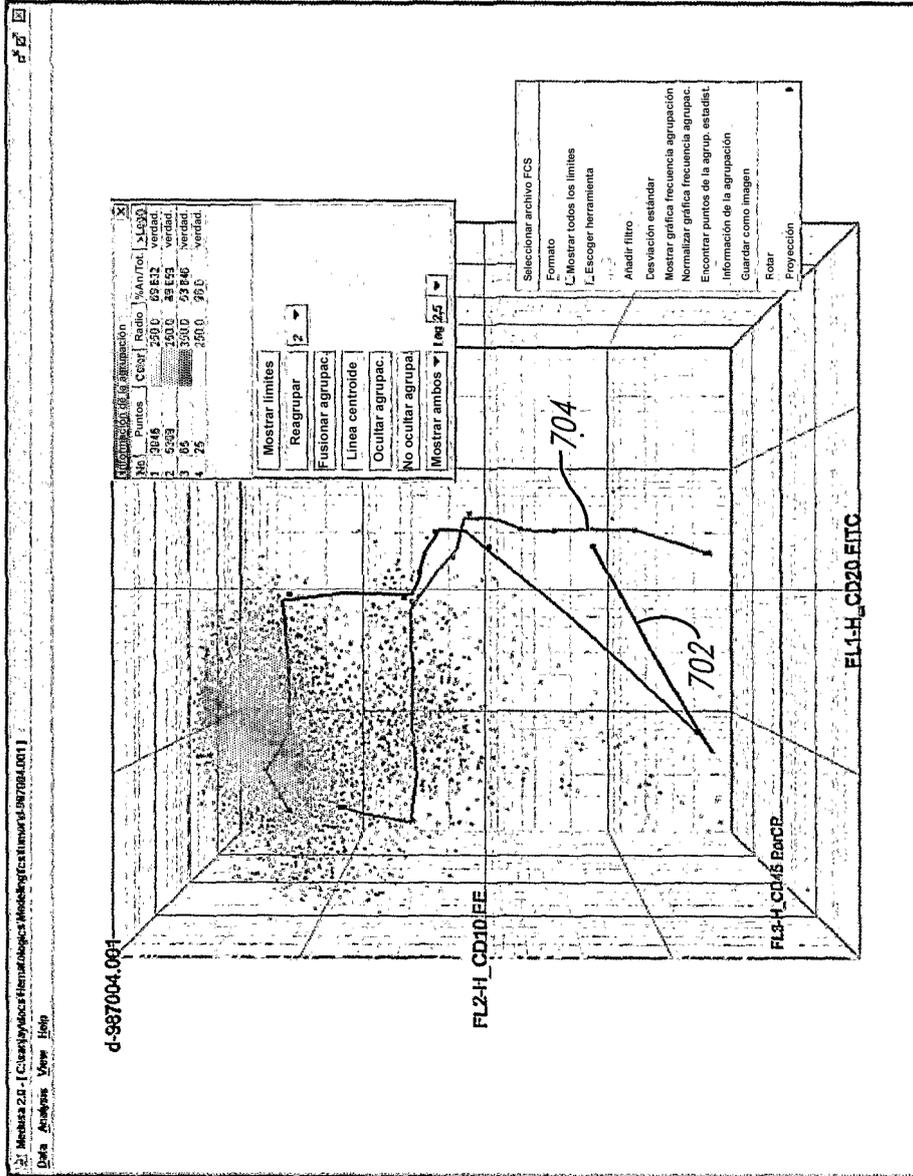


FIG. 20A

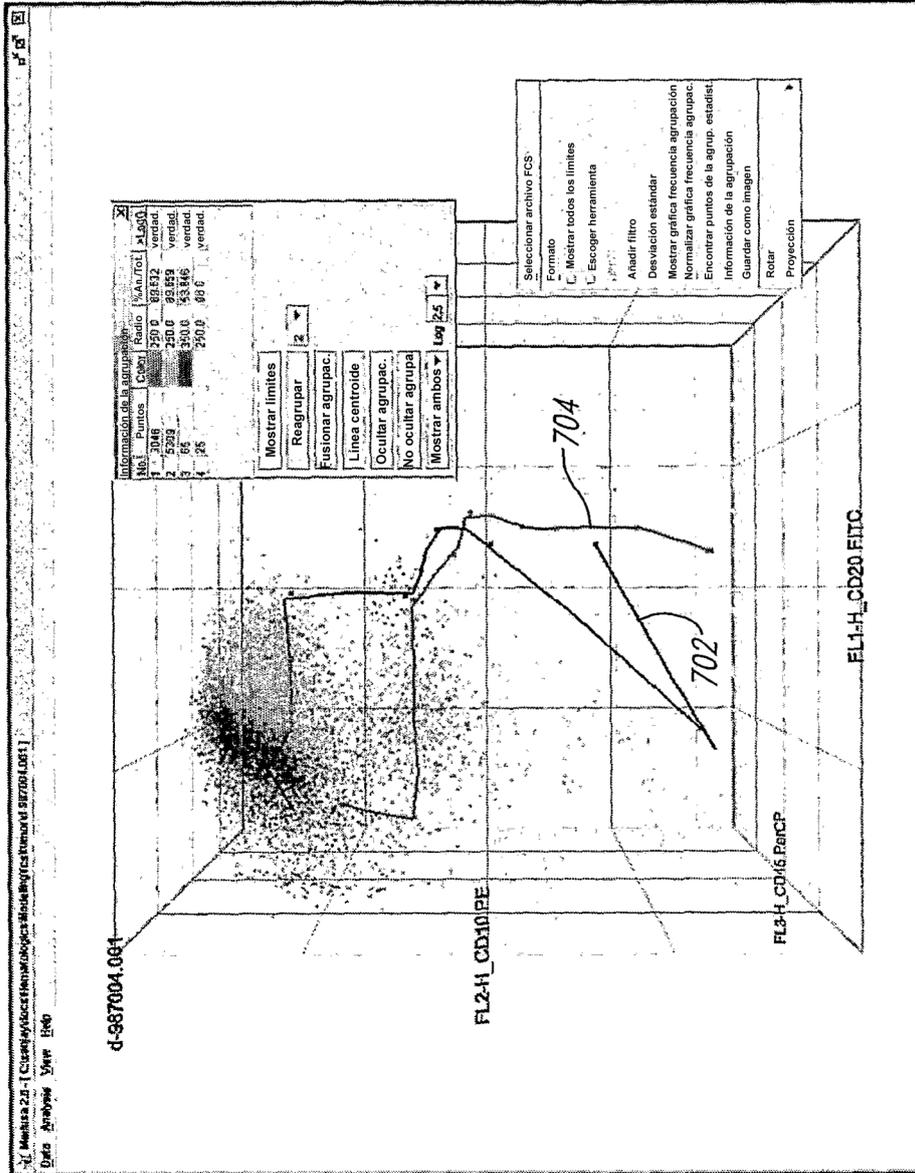


FIG. 20B

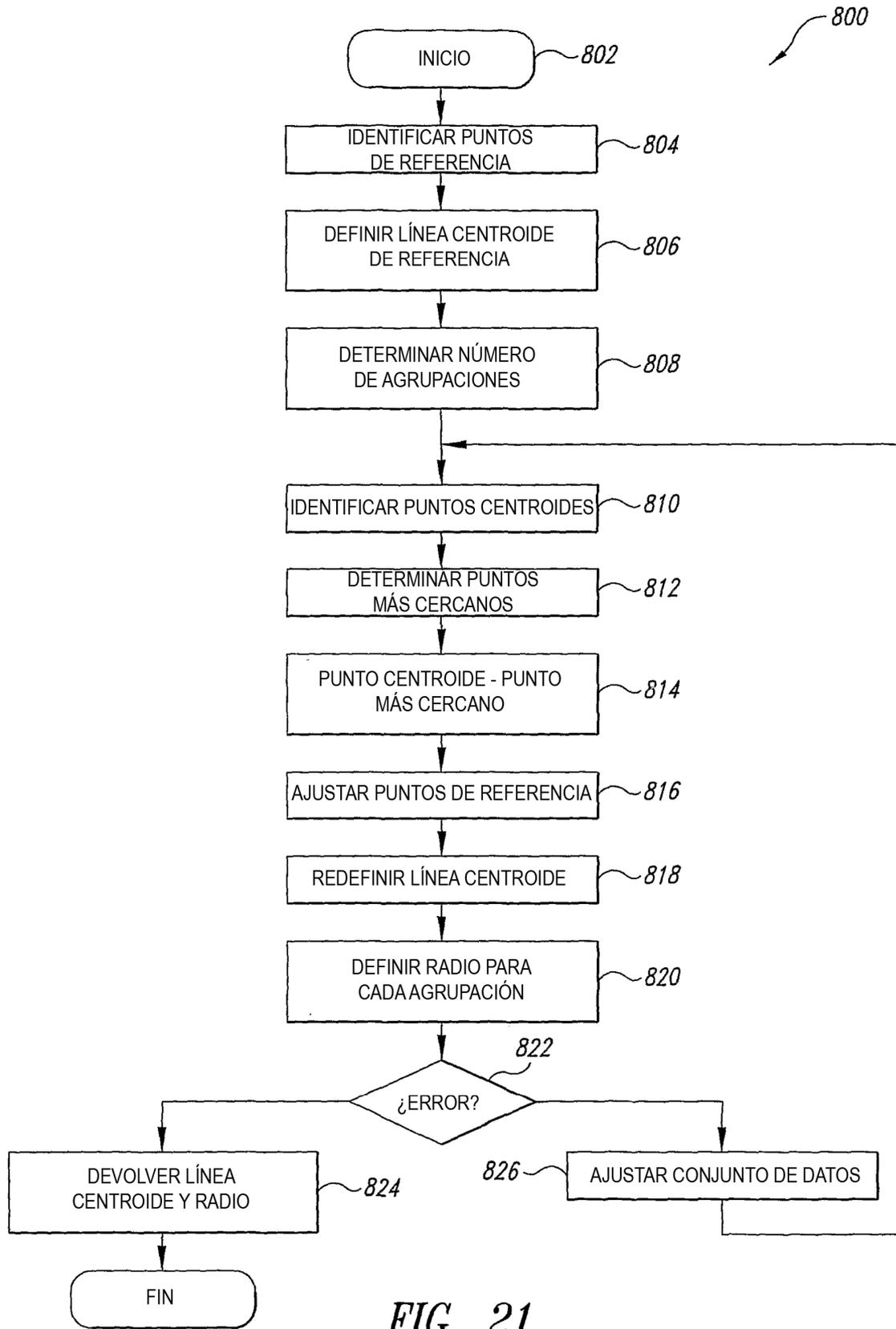


FIG. 21

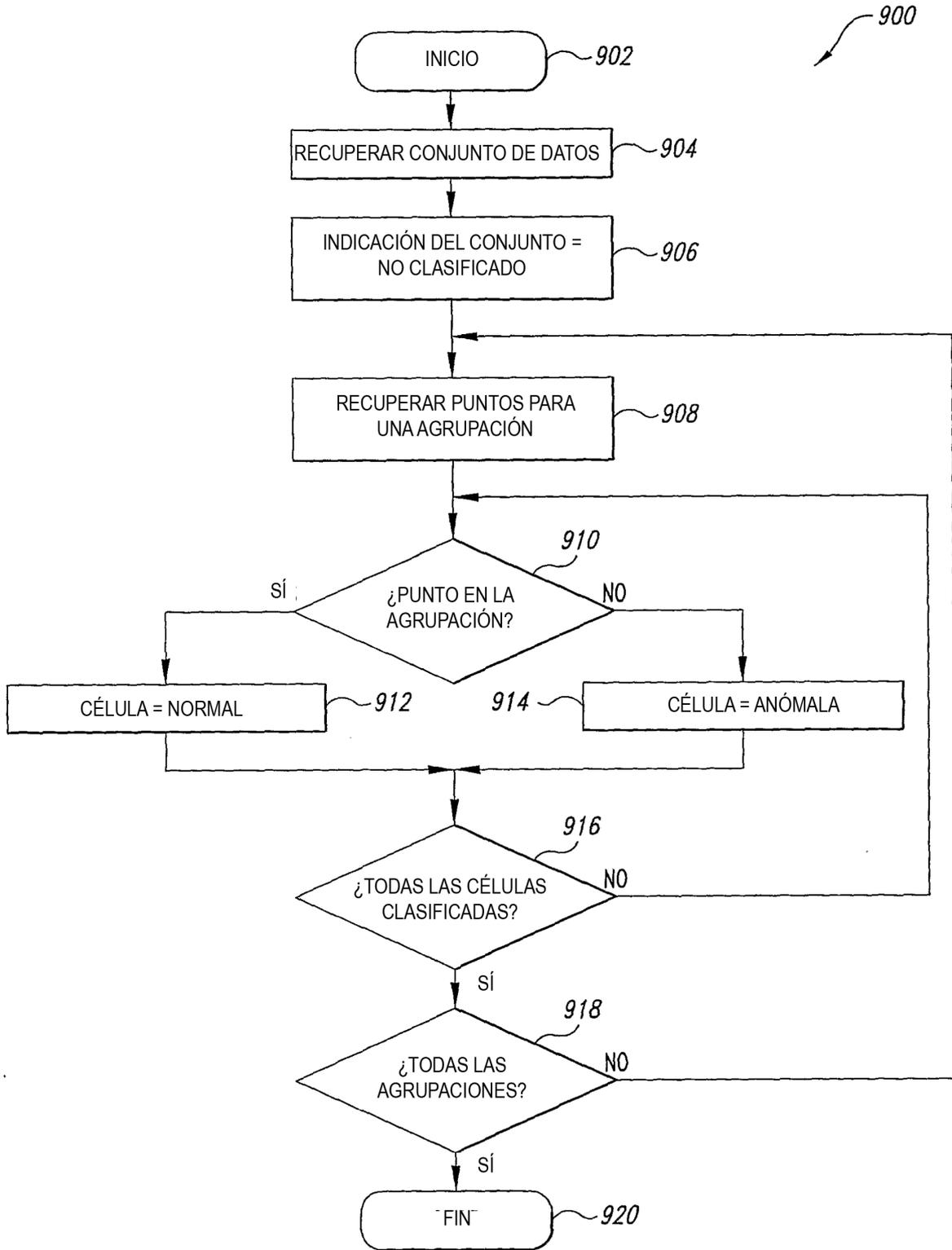


FIG. 22

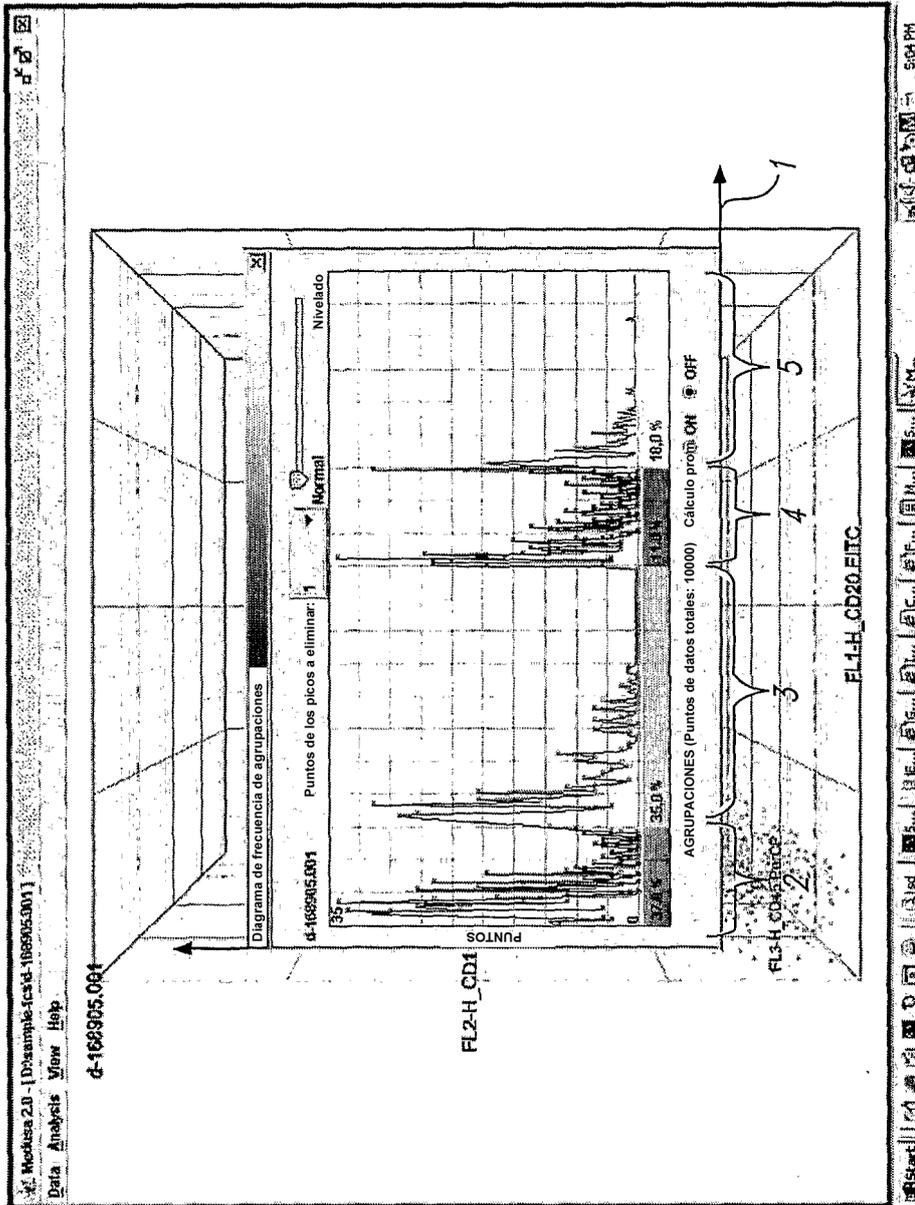


FIG. 23B

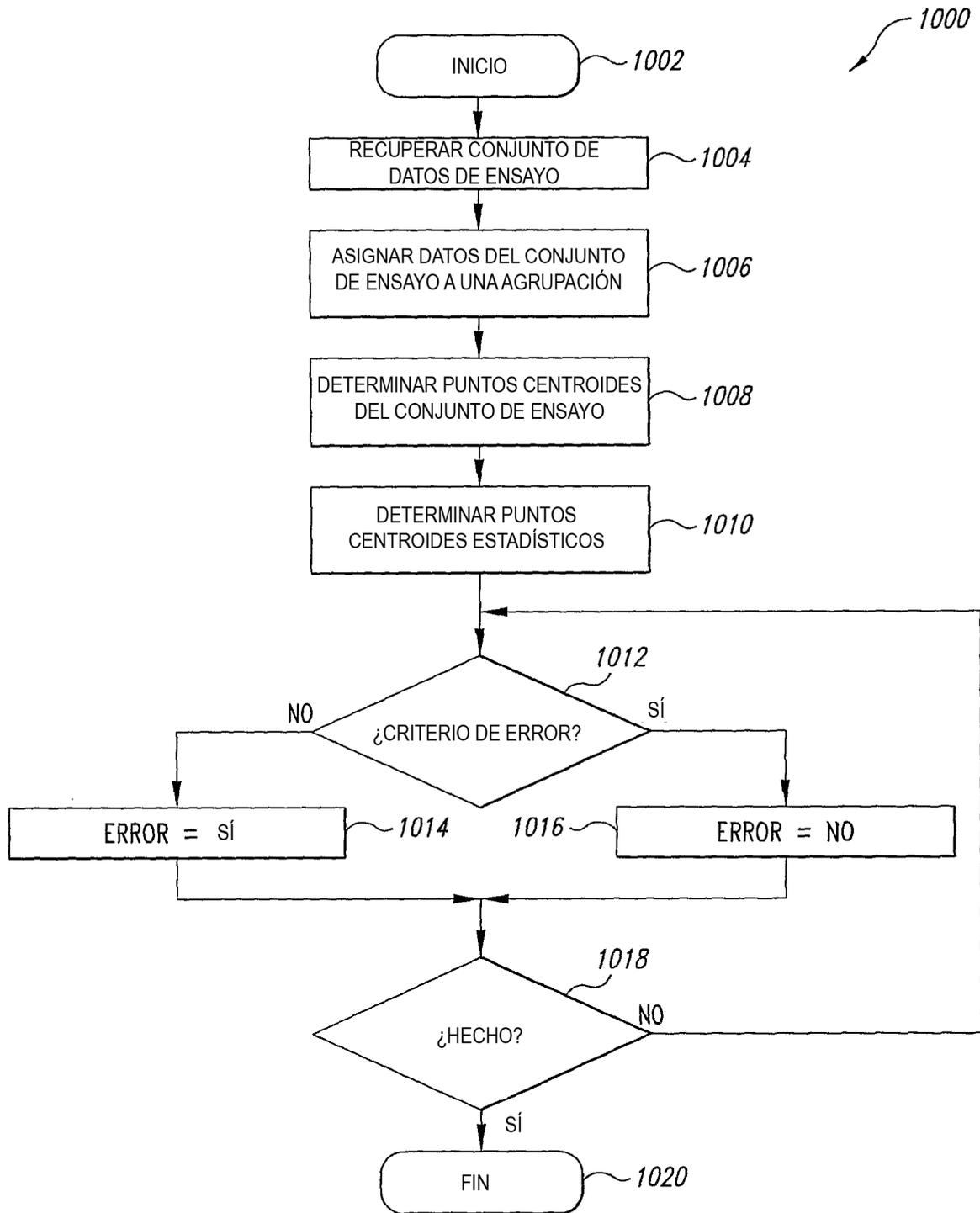


FIG. 24

CABECERA <u>1102</u>
TEXTO <u>1104</u>
DATOS DE CENTROIDE <u>1106</u>
DATOS DE RADIO <u>1108</u>

FIG. 25

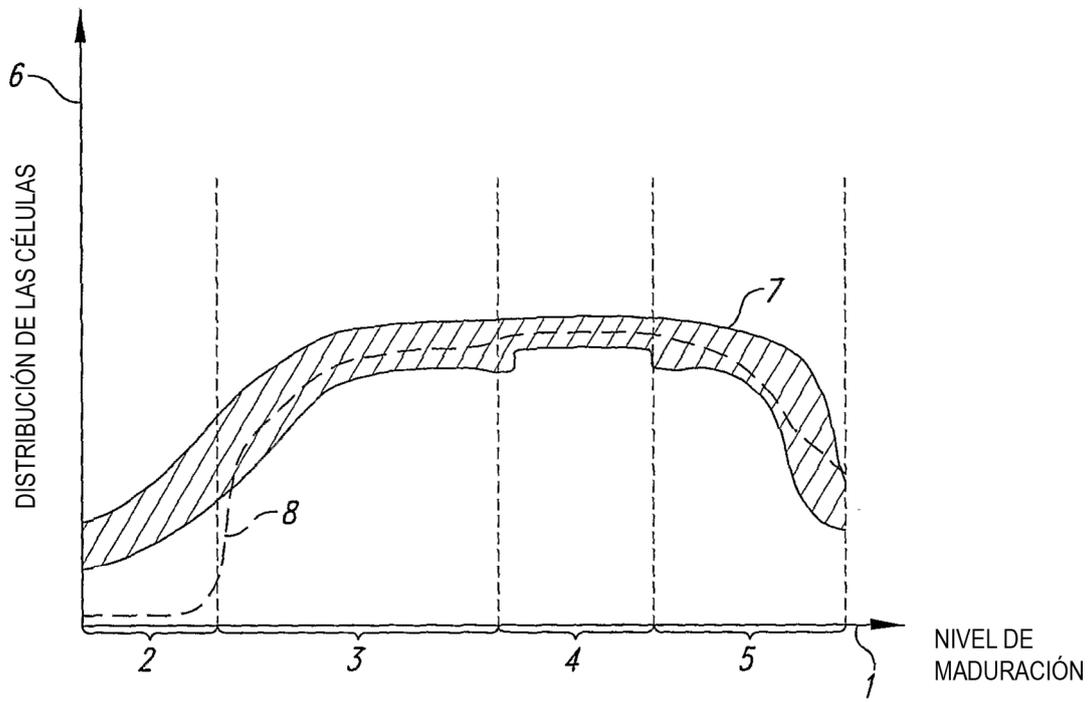


FIG. 26