(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **3 016 786**

(51) Int. Cl.:

C12Q 1/6886 (2008.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **31.07.2019 PCT/US2019/044497**

(87) Fecha y número de publicación internacional: **06.02.2020 WO20028563**

(96) Fecha de presentación y número de la solicitud europea: **31.07.2019 E 19753542 (0)**

(97) Fecha y número de publicación de la concesión europea: **19.02.2025 EP 3830299**

(54) Título: **Predicción y caracterización de subtipos de célula de origen de DLBCL**

(30) Prioridad:

01.08.2018 US 201862713434 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
09.05.2025

(73) Titular/es:

GENENTECH, INC. (33.33%)
1 DNA Way
South San Francisco, CA 94080, US;
FOUNDATION MEDICINE, INC. (33.33%) y
F. HOFFMANN-LA ROCHE AG (33.33%)

(72) Inventor/es:

TRABUCCO, SALLY, ELIZABETH;
BOLEN, CHRISTOPHER, R.;
OESTERGAARD, MIKKEL, ZAHLE;
SOKOL, ETHAN, SAMUEL y
ALBACKER, LEE, ALAN

(74) Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

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

Predicción y caracterización de subtipos de célula de origen de DLBCL

Antecedentes de la invención

El linfoma difuso de linfocitos B grandes (DLBCL) es la forma más común de linfoma no hodgkiniano, que representan más de 25.000 nuevos casos al año en Estados Unidos (R., T.L., *et al.* 2016 US lymphoid malignancy statistics by World Health Organization subtypes. CA: A Cancer Journal for Clinicians 66, 443-459 (2016)). El pronóstico de los pacientes con DLBCL es bastante bueno, con tasas de supervivencia a cinco años de entre el 55-62 % (R., T.L., *et al.* 2016 US lymphoid malignancy statistics by World Health Organization subtypes. CA: A Cancer Journal for Clinicians 66, 443-459 (2016)). Sin embargo, los pacientes recidivantes y los que no responden inicialmente al tratamiento estándar siguen teniendo opciones limitadas, que consisten principalmente en trasplantes autólogos de células madre o ensayos clínicos. De manera adicional, aunque el 60-65 % de los pacientes responden plenamente a terapia con R-CHOP (rituximab, ciclofosfamida, doxorrubicina, vincristina y prednisona), existe una diferencia significativa en las tasas de respuesta según la célula de origen (COO) del DLBCL.

El DLBCL tiene dos subtipos de COO: linfocito B activado (ABC) y linfocito B del centro germinal (GCB). El subtipo ABC tiene un mal pronóstico en comparación con el GCB, y la COO puede predecir la respuesta a algunos agentes terapéuticos nuevos. Tradicionalmente, el subtipo de COO se ha determinado mediante la expresión evaluada por micromatriz (ABC, GCB, no clasificado), algoritmos basados en la inmunohistoquímica (IHC) (GCB o no GCB), y ensayos basados en la expresión, como el ensayo de subtipificación de linfomas (LST) sólo para uso de investigación Nanostring (ABC, GCB, no clasificado) de Nanostring Technologies (también denominado en el presente documento ensayo Nanostring o Nanostring). Desafortunadamente, la micromatriz no suele ser factible y algunos informes no han logrado mostrar una diferencia pronóstica entre GCB y no GCB mediante algoritmos basados en IHC. Esto ha llevado a algunos a adoptar el ensayo Nanostring como método preferido para evaluar la COO, pero en algunos casos el contenido tumoral o la calidad del ARN no son suficientes para realizar este ensayo. Los subtipos de COO presentan diferentes mutaciones génicas, con GCB normalmente caracterizado por alteraciones de *EZH2* y translocaciones *IGH:BCL2*, mientras que ABC está dominado por alteraciones de la señalización NF-KB y BCR como *MYD88* y *CD79B*. En el presente documento se utiliza las diferencias mutacionales en los subtipos de COO para desarrollar un modelo de clasificación de ADN de COO (COODC) para predecir COO a partir de características basadas en el ADN en una plataforma utilizada clínicamente.

La subtipificación del origen de COO se desarrolló originalmente utilizando micromatrices de expresión de ARN (Alizadeh, A.A., *et al.* Distinct types of diffuse large B Cell lymphoma identified by gene expression profiling. Nature 403, 503 (2000)) y es un componente necesario de la atención clínica de DLBCL según las clasificaciones de neoplasias linfoides de la OMS actualizadas de 2016 (Swerdlow, S.H., *et al.* The 2016 revision of the World Health Organization classification of lymphoid neoplasms. Blood 127, 2375-2390 (2016)). Los clasificadores basados en el ARN suelen dividir DLBCL en los subtipos de linfocitos B activados (ABC) y linfocitos B del centro germinal (GC) (GCB). El DLBCL que no entra en ninguna de estas categorías se denomina no clasificado. Estos subtipos son tanto pronósticos como predictivos. Entre los pacientes tratados con terapias modernas anti-CD20, el subtipo ABC tiene un peor pronóstico, con una supervivencia sin progresión a los tres años del 59 %, en comparación con el 75 % de los pacientes con GCB (Vitolo, U., *et al.* Obinutuzumab or Rituximab Plus Cyclophosphamide, Doxorubicin, Vincristine, and Prednisone in Previously Untreated Diffuse Large B Cell Lymphoma. Journal of Clinical Oncology 35, 3529-3537 (2017)). Sin embargo, se ha desarrollado recientemente que los agentes terapéuticos dirigidos, incluido el inhibidor de la señalización BCR ibrutinib, se dirigen a las vías comúnmente alteradas en ABC. Ibrutinib ha demostrado un beneficio clínico significativo entre los pacientes con mutaciones en los genes de señalización BCR, que son comunes en el subconjunto ABC (Wilson, W.H., *et al.* Targeting B cell receptor signaling with ibrutinib in diffuse large B cell lymphoma. Nature Medicine 21, 922 (2015)). Está claro que la evaluación precisa del subtipo de COO es fundamental para el tratamiento clínico adecuado del DLBCL.

En la mayoría de los entornos clínicos no es factible realizar análisis de micromatrices para cada paciente. Esto ha llevado al desarrollo y la adopción de algoritmos basados en la inmunohistoquímica (IHC) para aproximar la COO. El más utilizado de éstos es el algoritmo de Hans (Hans, C.P., *et al.* Confirmation of the molecular classification of diffuse large B Cell lymphoma by immunohistochemistry using a tissue microarray. Blood 103, 275-282 (2003)), que es capaz de clasificar DLBCL de GCB y no GC, este último incluye tanto los subtipos ABC como no clasificado. Desafortunadamente, estos algoritmos no recapitulan de forma fiable los hallazgos pronósticos originales y, en algunos casos, los algoritmos basados en IHC no muestran diferencias significativas de supervivencia entre los tipos GCB y no GCB (Gribben, R.C., *et al.* Poor Concordance among Nine Immunohistochemistry Classifiers of Cell-of-Origin for Diffuse Large B Cell Lymphoma: Implications for Therapeutic Strategies. (2013); Meyer, P. N., *et al.* Immunohistochemical Methods for Predicting Cell of Origin and Survival in Patients With Diffuse Large B Cell Lymphoma Treated With Rituximab. Journal of Clinical Oncology 29, 200-207 (2011)).

Dada la necesidad de métodos fiables, clínicamente útiles para determinar la COO para el DLBCL, los esfuerzos recientes se han centrado en métodos más simples, más ágiles, incluido el ensayo Nanostring (Scott, D.W., *et al.* Determining cell-of-origin subtypes of diffuse large B Cell lymphoma using gene expression in formalin-fixed paraffin-

embedded tissue. Blood 123, 1214-1217 (2014)). Este método utiliza la expresión de ARN de 20 genes para determinar la COO con un 98 % de concordancia con la micromatriz y un 8,8 % de denominaciones no clasificado en comparación con el 14,7 % mediante la micromatriz de método de referencia. Aunque esto es muy prometedor para la utilidad clínica, requiere ARN y muestras con al menos un 60 % de pureza tumoral (es decir, contenido tumoral) para alcanzar esta precisión (Scott, D.W., *et al.* Determining cell-of-origin subtypes of diffuse large B Cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. Blood 123, 1214-1217 (2014)), que no es accesible de forma rutinaria.

Los subtipos de COO presentan diferentes mutaciones génicas, con GCB normalmente caracterizado por alteraciones de *EZH2* y translocaciones *IGH:BCL2*, mientras que ABC está dominado por alteraciones de la señalización NF-KB y BCR como *MYD88* y *CD79B*. Chapuy y sus colegas mostraron recientemente firmas mutacionales distintas en muestras de DLBCL, incluyendo una firma canónica de citidina desaminasa inducida por activación (AID) (Chapuy, B., *et al.* Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nature Medicine 24, 679-690 (2018)). Aunque los informes anteriores no identifican diferencias en la firma mutacional entre subtipos, se plantea la hipótesis de que existen diferencias en los perfiles mutacionales de los tumores de GCB y ABC, y lo más probable es que sean el resultado de historias distintas de las células precursoras. Se cree que los DLBCL de GCB surgen de los linfocitos B del centro germinal (GC), mientras que se cree que el subtipo ABC surge de los linfocitos B post-GC. Durante el proceso de maduración de los anticuerpos, los linfocitos B del GC están sujetos a recombinación por cambio de clase (CSR) e hipermutación somática (SHM) inducida por la proteína AID que induce roturas de doble cadena (DSB) y mutaciones de una sola base. Estudios recientes han demostrado que ciertos genes en tumores de DLBCL presentan evidencias de alteraciones frecuentes inducidas por AID (Chapuy, B., *et al.* Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nature Medicine 24, 679-690 (2018); Schmitz, R., *et al.* Genetics and Pathogenesis of Diffuse Large B Cell Lymphoma. New England Journal of Medicine 378, 1396-1407 (2018)) y ciertos genes relacionados con el DLBCL, tal como PIM1, se cree que son puntos calientes para la actividad de AID fuera de diana.

Con el fin de resolver el problema de encontrar un diagnóstico fiable para una COO de DLBCL que no requiera ARN y que no necesite muestras con al menos un 60 % de pureza tumoral, se inventó un clasificador sólo de ADN utilizando la plataforma FoundationOne®Heme de Foundation Medicine y se divulga en el presente documento. Se demuestra que este ensayo mantiene una concordancia del 89 % (p. ej., aproximadamente el 90 %) con Nanostring, y requiere muestras que tengan sólo un 20 % de contenido tumoral (p. ej., aproximadamente un 20 % de contenido tumoral). Además, se demuestra que las clasificaciones de COO de este ensayo recapitulan de forma fiable las diferencias pronósticas observadas en clasificaciones basadas en ARN como Nanostring.

Breve sumario de la invención

Se ha inventado un método basado en ADN para determinar la COO para DLBCL y se divulga en el presente documento. En determinadas realizaciones, este ensayo mantiene una concordancia del 89 % con Nanostring. En determinadas realizaciones, este ensayo requiere muestras que tengan un 20 % de contenido tumoral en comparación con el 60 % necesario para Nanostring.

En algunas realizaciones, en el presente documento se proporciona un método para determinar la célula de origen (COO) del linfoma difuso de linfocitos B grandes (DLBCL) utilizando una clasificación del ADN de la célula de origen (COODC). En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la COODC comprende: (a) adquirir, p. ej., recoger, una muestra, p. ej., una muestra clínica, de un paciente diagnosticado con DLBCL, (b) realizar la secuenciación de ADN en la muestra, p. ej., muestra clínica, y (c) aplicar un clasificador COODC predefinido a una lista de características genómicas (p. ej., una o más características descritas en la Tabla 1) para calcular una puntuación predictora. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la secuenciación de ADN se realiza utilizando el componente de ADN de la plataforma FoundationOne®Heme. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la secuenciación de ADN comprende la secuenciación dirigida del ADN de aproximadamente 300-500 genes. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la detección de alteraciones en los genes BCL2, EZH2 y TNSFRSF14 utilizando la COODC predice que la COO del DLBCL es un linfocito B del centro germinal (GCB). En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, se detectan alteraciones en los genes BCL2, EZH2 y TNSFRSF14. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la detección de alteraciones en el número de copias del cromosoma 3q, los genes MYD88 y CD79B utilizando la COODC predice que la COO del DLBCL es un linfocito B activado (ABC). En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, se detectan alteraciones en el número de copias del cromosoma 3q, los genes MYD88 y CD79B. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la detección de alteraciones en los genes NOTCH1, NOTCH2 y BCL6 utilizando la COODC predice que la COO del DLBCL es un linfocito B activado (ABC). En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, se detectan alteraciones en los genes NOTCH1, NOTCH2 y BCL6. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la detección de reordenamientos de IGH:BCL2, alteraciones de CREBBP y alteraciones de TNFRSF14 predicen que la COO del DLBCL es un GCB. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, se detectan reordenamientos de IGH:BCL2,

alteraciones de CREBBP y alteraciones de TNFRSF14. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la detección de deleciones de CDKN2A/B, alteraciones de CD79B en el aminoácido 196 y alteraciones de MYD88 en el aminoácido 265 predicen que la COO del DLBCL es un ABC. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, se detectan una deleción de CDKN2A/B, una alteración de CD79B en el aminoácido 196 y una alteración de MYD88 en el aminoácido 265. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, una firma COSMIC 3 predice que la COO del DLBCL es GCB. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, se detecta una firma COSMIC 3. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, el método comprende además la etapa de (d) clasificar a un paciente que tiene una puntuación predictora por debajo de un punto de corte bajo (p. ej., un punto de corte bajo predefinido) como GCB, y a un paciente que tiene una puntuación predictora por encima de un punto de corte alto (p. ej., un punto de corte alto predefinido) como ABC. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, el método comprende además la etapa de (e) clasificar como no clasificado a un paciente que tiene una puntuación predictora superior o igual al punto de corte bajo (p. ej., el punto de corte bajo predefinido) e inferior al punto de corte alto (p. ej., el punto de corte alto predefinido). En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la COODC se desarrolló utilizando los métodos descritos en los Ejemplos 3-9. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, no se realiza ningún análisis de ARN. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, el análisis de RNA no contribuye sustancialmente a la determinación de la COO. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, no se realiza ningún análisis de inmunohistoquímica (IHC). En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, el análisis de IHC no contribuye sustancialmente a la determinación de la COO.

En algunas realizaciones, en el presente documento se proporciona un método para determinar si la célula de origen (COO) del linfoma difuso de linfocitos B grandes (DLBCL) es un linfocito B activado (ABC) o un linfocito B del centro germinal (GCB) utilizando un modelo de clasificación del ADN de la célula de origen (COODC). En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, el método comprende (a) adquirir, p. ej., recoger, una muestra, p. ej., una muestra clínica, de un paciente diagnosticado con DLBCL, (b) realizar la secuenciación de ADN en la muestra, p. ej., muestra clínica, (c) aplicar el clasificador COODC predefinido a la lista de características genómicas (p. ej., una o más características descritas en la Tabla 1) para calcular una puntuación predictora, y (d) clasificar a un paciente que tiene una puntuación predictora inferior a un punto de corte bajo (p. ej., un punto de corte predefinido bajo) como GCB, y a un paciente que tiene una puntuación predictora superior o igual a un punto de corte alto (p. ej., un punto de corte predefinido alto) como ABC. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, el método comprende además la etapa de (e) clasificar como no clasificado a un paciente que tiene una puntuación predictora superior o igual al punto de corte bajo (p. ej., el punto de corte bajo predefinido) e inferior al punto de corte alto (p. ej., el punto de corte alto predefinido). En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la secuenciación de ADN se realiza utilizando el componente de ADN de la plataforma FoundationOne®Heme. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la secuenciación de ADN comprende la secuenciación dirigida del ADN de aproximadamente 300-500 genes. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la detección de alteraciones en los genes BCL2, EZH2 y TNSFRSF14 utilizando la COODC predice que la COO del DLBCL es linfocito B del centro germinal (GCB). En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, se detectan alteraciones en los genes BCL2, EZH2 y TNSFRSF14. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la detección de alteraciones en el número de copias del cromosoma 3q, los genes MYD88 y CD79B utilizando la COODC predice que la COO del DLBCL es un linfocito B activado (ABC). En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, se detectan alteraciones en el número de copias del cromosoma 3q, los genes MYD88 y CD79B. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la detección de alteraciones en los genes NOTCH1, NOTCH2 y BCL6 utilizando la COODC predice que la COO del DLBCL es un linfocito B activado (ABC). En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, se detectan alteraciones en los genes NOTCH1, NOTCH2 y BCL6. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la detección de reordenamientos de IGH:BCL2, alteraciones de CREBBP y alteraciones de TNFRSF14 predicen que la COO del DLBCL es GCB. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, se detectan reordenamientos de IGH:BCL2, alteraciones de CREBBP y alteraciones de TNFRSF14. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la detección de deleciones de CDKN2A/B, alteraciones de CD79B en el aminoácido 196 y alteraciones de MYD88 en el aminoácido 265 predicen que la COO del DLBCL es ABC. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, se detectan deleciones de CDKN2A/B, alteraciones de CD79B en el aminoácido 196 y alteraciones de MYD88 en el aminoácido 265. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, una firma COSMIC 3 predice que la COO del DLBCL es GCB. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, se detecta una firma COSMIC 3. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la COODC se desarrolló utilizando los métodos descritos en los Ejemplos 3-9. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, no se realiza ningún análisis de ARN. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, el análisis del ARN no contribuye sustancialmente a la determinación de

COO. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, no se realiza ningún análisis de inmunohistoquímica (IHC). En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la IHC no contribuye sustancialmente a la determinación de COO.

En algunas realizaciones, en el presente documento se proporciona un método para diseñar una terapia para el tratamiento del linfoma difuso de linfocitos B grandes (DLBCL) utilizando un modelo de clasificación del ADN de la célula de origen (COODC) que comprende: (a) adquirir, p. ej., recoger, una muestra, p. ej., una muestra clínica, de un paciente diagnosticado con DLBCL, (b) realizar la secuenciación de ADN en la muestra, p. ej., muestra clínica, (c) aplicar el clasificador COODC predefinido a la lista de características genómicas (p. ej., una o más características descritas en la Tabla 1) para calcular una puntuación predictora, y (d) clasificar a un paciente que tiene una puntuación predictora inferior a un punto de corte bajo (p. ej., un punto de corte predefinido bajo) como GCB, y a un paciente que tiene una puntuación predictora superior o igual a un punto de corte alto (p. ej., un punto de corte predefinido alto) como ABC, en donde se administra al paciente una terapia recomendada para su COO. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, el método comprende además la etapa de (e) clasificar como no clasificado a un paciente que tiene una puntuación predictora superior o igual al punto de corte bajo (p. ej., el punto de corte bajo predefinido) e inferior al punto de corte alto (p. ej., el punto de corte alto predefinido). En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, al paciente con COO de GCB se le administra una terapia eficaz para la COO de GCB. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, al paciente con COO de ABC se le administra una terapia eficaz para la COO de ABC.

En algunas realizaciones, en el presente documento se proporciona un método para predecir la respuesta a la terapia para el tratamiento del linfoma difuso de linfocitos B grandes (DLBCL) utilizando un modelo de clasificación del ADN de la célula de origen (COODC) que comprende (a) adquirir, p. ej., recoger, una muestra, p. ej., una muestra clínica, de un paciente diagnosticado con DLBCL, (b) realizar la secuenciación de ADN en la muestra, p. ej., muestra clínica, (c) aplicar el clasificador COODC predefinido a la lista de características genómicas (p. ej., una o más características descritas en la Tabla 1) para calcular una puntuación predictora, y (d) clasificar a un paciente que tiene una puntuación predictora inferior a un punto de corte bajo (p. ej., un punto de corte predefinido bajo) como GCB, y a un paciente que tiene una puntuación predictora superior o igual a un punto de corte alto (p. ej., un punto de corte predefinido alto) como ABC, en donde la respuesta a una terapia para el tratamiento de DLBCL depende de la COO. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, el método comprende además la etapa de (e) clasificar como no clasificado a un paciente que tiene una puntuación predictora superior o igual al punto de corte bajo (p. ej., el punto de corte bajo predefinido) e inferior al punto de corte alto (p. ej., el punto de corte alto predefinido). En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la COO de ABC es predictivo para la respuesta a terapias conocidas por su eficacia en el subtipo ABC. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la terapia comprende ibrutinib y/o lenalidomida. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la COO de GCB es predictivo para la respuesta a terapias conocidas por su eficacia en el subtipo GCB. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la terapia comprende ibrutinib y/o lenalidomida. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la célula de origen (COO) del linfoma difuso de linfocitos B grandes (DLBCL) se determina utilizando un método de plataforma basado en el ADN en donde ni el ARN (p. ej., el análisis del ARN) ni la inmunohistoquímica son componentes del método. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la plataforma basada en el ADN es la COODC descrita en cualquiera de las realizaciones descritas en el presente documento.

En algunas realizaciones, en el presente documento se proporciona un diagnóstico de clasificación del ADN de la célula de origen (COODC) utilizado para determinar si la célula de origen (COO) del linfoma difuso de linfocitos B grandes (DLBCL) es un linfocito B activado (ABC) o un linfocito B del centro germinal (GCB). En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la COODC es la COODC descrita en cualquiera de las realizaciones descritas en el presente documento. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la COODC es un kit.

En algunas realizaciones, en el presente documento se proporciona un método para utilizar un diagnóstico de clasificación del ADN de la célula de origen (COODC) para determinar si la célula de origen (COO) del linfoma difuso de linfocitos B grandes (DLBCL) es un linfocito B activado (ABC) o un linfocito B del centro germinal (GCB) que comprende: (a) adquirir, p. ej., recoger, una muestra, p. ej., una muestra clínica, de un paciente diagnosticado con DLBCL, (b) realizar la secuenciación de ADN en la muestra, p. ej., muestra clínica, (c) aplicar un clasificador COODC (p. ej., un clasificador COODC predefinido) a la lista de características genómicas (p. ej., una o más características descritas en la Tabla 1) para calcular una puntuación predictora, y (d) clasificar a un paciente que tiene una puntuación predictora inferior a un punto de corte bajo (p. ej., un punto de corte predefinido bajo) como GCB, y a un paciente que tiene una puntuación predictora superior o igual a un punto de corte alto (p. ej., un punto de corte predefinido alto) como ABC. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, el método comprende además la etapa de (e) clasificar como no clasificado a un paciente que tiene una puntuación predictora superior o igual al punto de corte bajo (p. ej., el punto de corte bajo predefinido) e inferior al punto de corte alto (p. ej., el punto de corte alto predefinido). En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, el diagnóstico es un kit. En determinadas realizaciones según (o aplicadas a) cualquiera de las

realizaciones anteriores, se determina que la COO es ABC o GCB según cualquiera de las realizaciones anteriores.

En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la muestra es una muestra clínica. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la muestra clínica es una biopsia de tumor, sangre, aspirado de médula ósea o un ácido nucleico extraído. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la biopsia de tumor se prepara en un portaobjetos fijado en formol e incluido en parafina (FFPE). En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la secuenciación de ADN comprende la secuenciación dirigida de ADN de aproximadamente 465 genes. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, una pluralidad de muestras (p. ej., una pluralidad de muestras clínicas), p. ej., de una pluralidad de pacientes, se adquieren (p. ej., se recogen).

En algunas realizaciones, en el presente documento se proporciona un método para seleccionar, clasificar o tratar a un sujeto que comprende adquirir o proporcionar un valor para la célula de origen (COO) y en respuesta al valor seleccionar, clasificar, evaluar o tratar al sujeto. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, el valor se adquiere de otra entidad, p. ej., un laboratorio. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, el valor es sustancialmente el mismo que el valor que proporcionaría el método o diagnóstico de cualquiera de las realizaciones descritas en el presente documento. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, el valor se determina mediante cualquiera de las realizaciones descritas en el presente documento. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, el sujeto tiene DLBCL.

En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, el punto de corte bajo (p. ej., el punto de corte bajo predefinido) para clasificar patentes o muestras como GCB está entre 0,2-0,3, p. ej., entre 0,21-0,29, 0,22-0,28, 0,23-0,27, 0,24-0,26, 0,21-0,3, 0,22-0,3, 0,23-0,3, 0,24-0,3, 0,25-0,3, 0,26-0,3, 0,27-0.3, 0,28-0,3, 0,29-0,3, 0,2-0,29, 0,2-0,28, 0,2-0,27, 0,2-0,26, 0,2-0,25, 0,2-0,24, 0,2-0,23, 0,2-0,22 o 0,2-0,21, p. ej., 0,200, 0,201, 0,202, 0,203, 0,204, 0,205, 0,206, 0,207, 0,208, 0,209, 0,210, 0,211, 0,212, 0,213, 0,214, 0,215, 0,216, 0,217, 0,218, 0,219, 0,220, 0,221, 0,222, 0,223, 0,224, 0,225, 0,226, 0,227, 0,228, 0,229, 0,230, 0,231, 0,232, 0,233, 0,234, 0,235, 0,236, 0,237, 0,238, 0,239, 0,240, 0,241, 0,242, 0,243, 0,244, 0,245, 0,246, 0,247, 0,248, 0,249, 0,250, 0,251, 0,252, 0,253, 0,254, 0,255, 0,256, 0,257, 0,258, 0,259, 0,260, 0,261, 0,262, 0,263, 0,264, 0,265, 0,266, 0,267, 0,268, 0,269, 0,270, 0,271, 0,272, 0,273, 0,274, 0,275, 0,276, 0,277, 0,278, 0,279, 0,280, 0,281, 0,282, 0,283, 0,284, 0,285, 0,286, 0,287, 0,288, 0,289, 0,290, 0,291, 0,292, 0,293, 0,294, 0,295, 0,296, 0,297, 0,298, 0,299 o 0,300. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, el punto de corte alto (p. ej., el punto de corte alto predefinido) para clasificar las muestras como ABC está entre 0,4-0,6, p. ej., 0,41-0,59, 0,42-0,58, 0,43-0,57, 0,44-0,56, 0,45-0,55, 0,46-0,54, 0,47-0,53, 0,48-0,52, 0,49-0,51, 0,4-0,59, 0,4-0,58, 0,4-0,57, 0,4-0,56, 0,4-0,55, 0,4-0,54, 0,4-0,53, 0,4-0,52, 0,4-0,51, 0,4-0,50. 0,4-0,49, 0,4-0,48, 0,4-0,47, 0,4-0,46, 0,4-0,45, 0,4-0,44, 0,4-0,42, 0,4-0,41, 0,41-0,6, 0,42-0,6, 0,43-0,6, 0,44-0,6, 0,45-0,6, 0,46-0,6, 0,47-0,6, 0,48-0,6, 0,49-0,6, 0,5-0,6, 0,51-0,6, 0,52-0,6, 0,53-0,6, 0,54-0,6, 0,55-0,6, 0,56-0,6, 0,57-0,6, 0,58-0,6 o 0,59-0,6, p. ej., 0,400, 0,401, 0,402, 0,403, 0,404, 0,405, 0,406, 0,407, 0,408, 0,409, 0,410, 0,411, 0,412, 0,413, 0,414, 0,415, 0,416, 0,417, 0,418, 0,419, 0,420, 0,421, 0,422, 0,423, 0,424, 0,425, 0,426, 0,427, 0,428, 0,429, 0,430, 0,431, 0,432, 0,433, 0,434, 0,435, 0,436, 0,437, 0,438, 0,439, 0,440, 0,441, 0,442, 0,443, 0,444, 0,445, 0,446, 0,447, 0,448, 0,449, 0,450, 0,451, 0,452, 0,453, 0,454, 0,455, 0,456, 0,457, 0,458, 0,459, 0,460, 0,461, 0,462, 0,463, 0,464, 0,465, 0,466, 0,467, 0,468, 0,469, 0,470, 0,471, 0,472, 0,473, 0,474, 0,475, 0,476, 0,477, 0,478, 0,479, 0,480, 0,481, 0,482, 0,483, 0,484, 0,485, 0,486, 0,487, 0,488, 0,489, 0,490, 0,491, 0,492, 0,493, 0,494, 0,495, 0,496, 0,497, 0,498, 0,499, 0,500, 0,501, 0,502, 0,503, 0,504, 0,505, 0,506, 0,507, 0,508, 0,509, 0,510, 0,511, 0,512, 0,513, 0,514, 0,515, 0,516, 0,517, 0,518, 0,519, 0,520, 0,521, 0,522, 0,523, 0,524, 0,525, 0,526, 0,527, 0,528, 0,529, 0,530, 0,531, 0,532, 0,533, 0,534, 0,535, 0,536, 0,537, 0,538, 0,539, 0,540, 0,541, 0,542, 0,543, 0,544, 0,545, 0,546, 0,547, 0,548, 0,549, 0,550, 0,551, 0,552, 0,553, 0,554, 0,555, 0,556, 0,557, 0,558, 0,559, 0,560, 0,561, 0,562, 0,563, 0,564, 0,565, 0,566, 0,567, 0,568, 0,569, 0,570, 0,571, 0,572, 0,573, 0,574, 0,575, 0,576, 0,577, 0,578, 0,579, 0,580, 0,581, 0,582, 0,583, 0,584, 0,585, 0,586, 0,587, 0,588, 0,589, 0,590, 0,591, 0,592, 0,593, 0,594, 0,595, 0,596, 0,597, 0,598, 0,599 o 0,600.

En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, si la puntuación predictora es inferior a 0,3, p. ej., inferior a 0,299, 0,298, 0,297, 0,296, 0,295, 0,294, 0,293, 0,292, 0,291, 0,29, 0,289, 0,288, 0,287, 0,286, 0,285, 0,284, 0,283, 0,282, 0,281, 0,28, 0,279, 0,278, 0,277, 0,276, 0,275, 0,274, 0,273, 0,272, 0,271, 0,27, 0,269, 0,268, 0,267, 0,266, 0,265, 0,264, 0,263, 0,262, 0,261, 0,26, 0,259, 0,258, 0,257, 0,256, 0,255, 0,254, 0,253, 0,252, 0,251, 0,25, 0,249, 0,248, 0,247, 0,246, 0,245, 0,244, 0,243, 0,242, 0,241, 0,24, 0,239, 0,238, 0,237, 0,236, 0,235, 0,234, 0,233, 0,232, 0,231, 0,23, 0,229, 0,228, 0,227, 0,226, 0,225, 0,224, 0,223, 0,222, 0,221, 0,22, 0,219, 0,218, 0,217, 0,216, 0,215, 0,214, 0,213, 0,212, 0,211, 0,21, 0,209, 0,208, 0,207, 0,206, 0,205, 0,204, 0,203, 0,202, 0,201, 0,200, 0,19, 0,18, 0,17, 0,16, 0,15, 0,14, 0,13, 0,12, 0,11, 0,1, 0,09, 0,08, 0,07, 0,06, 0,05, 0,04, 0,03, 0,02 o 0,01, el paciente o la muestra se clasifica como GCB. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, si la puntuación predictora es superior o igual a 0,4, p. ej., superior o igual a 0,401, 0,402, 0,403, 0,404, 0,405, 0,406, 0,407, 0,408, 0,409, 0,410, 0,411, 0,412, 0,413, 0,414, 0,415, 0,416, 0,417, 0,418, 0,419, 0,420, 0,421, 0,422, 0,423, 0,424, 0,425, 0,426, 0,427, 0,428, 0,429, 0,430, 0,431, 0,432, 0,433, 0,434, 0,435, 0,436, 0,437, 0,438, 0,439, 0,440, 0,441, 0,442, 0,443, 0,444, 0,445, 0,446, 0,447, 0,448, 0,449, 0,450, 0,451, 0,452, 0,453, 0,454, 0,455, 0,456, 0,457, 0,458, 0,459, 0,460, 0,461, 0,462, 0,463, 0,464, 0,465, 0,466, 0,467, 0,468, 0,469, 0,470, 0,471, 0,472, 0,473, 0,474, 0,475, 0,476, 0,477, 0,478, 0,479, 0,480, 0,481, 0,482, 0,483, 0,484,

0,485, 0,486, 0,487, 0,488, 0,489, 0,490, 0,491, 0,492, 0,493, 0,494, 0,495, 0,496, 0,497, 0,498, 0,499, 0,500, 0,501, 0,502, 0,503, 0,504, 0,505, 0,506, 0,507, 0,508, 0,509, 0,510, 0,511, 0,512, 0,513, 0,514, 0,515, 0,516, 0,517, 0,518, 0,519, 0,520, 0,521, 0,522, 0,523, 0,524, 0,525, 0,526, 0,527, 0,528, 0,529, 0,530, 0,531, 0,532, 0,533, 0,534, 0,535, 0,536, 0,537, 0,538, 0,539, 0,540, 0,541, 0,542, 0,543, 0,544, 0,545, 0,546, 0,547, 0,548, 0,549, 0,550, 0,551, 0,552, 0,553, 0,554, 0,555, 0,556, 0,557, 0,558, 0,559, 0,560, 0,561, 0,562, 0,563, 0,564, 0,565, 0,566, 0,567, 0,568, 0,569, 0,570, 0,571, 0,572, 0,573, 0,574, 0,575, 0,576, 0,577, 0,578, 0,579, 0,580, 0,581, 0,582, 0,583, 0,584, 0,585, 0,586, 0,587, 0,588, 0,589, 0,590, 0,591, 0,592, 0,593, 0,594, 0,595, 0,596, 0,597, 0,598, 0,599, 0,60, 0,61, 0,62, 0,63, 0,64, 0,66, 0,66, 0,67, 0,68, 0,69, 0,70, 0,71, 0,72, 0,73, 0,74, 0,77, 0,76, 0,77, 0,78, 0,79, 0,80, 0,81, 0,82, 0,83, 0,84, 0,88, 0,86, 0,87, 0,88, 0,89, 0,90, 0,90, 0,91, 0,92, 0,93, 0,94, 0,99, 0,96, 0,97, 0,98 o 0,99, el paciente o la muestra se clasifica como ABC.

En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, si la puntuación predictiva es inferior al punto de corte alto predefinido, pero superior o igual al punto de corte predefinido bajo, entonces el paciente o la muestra no están clasificados. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, si la puntuación predictiva es superior o igual a un punto de corte predefinido bajo descrito en el presente documento (p. ej., entre 0,2-0,3) e inferior a un punto de corte predefinido alto descrito en el presente documento (p. ej., entre 0,4-0,6), p. ej., superior o igual a (a) uno cualquiera de 0,299, 0,298, 0,297, 0,296, 0,295, 0,294, 0,293, 0,292, 0,291, 0,29, 0,289, 0,288, 0,287, 0,286, 0,285, 0,284, 0,283, 0,282, 0,281, 0,28, 0,279, 0,278, 0,277, 0,276, 0,275, 0,274, 0,273, 0,272, 0,271, 0,27, 0,269, 0,268, 0,267, 0,266, 0,265, 0,264, 0,263, 0,262, 0,261, 0,26, 0,259, 0,258, 0,257, 0,256, 0,255, 0,254, 0,253, 0,252, 0,251, 0,25, 0,249, 0,248, 0,247, 0,246, 0,245, 0,244, 0,243, 0,242, 0,241, 0,24, 0,239, 0,238, 0,237, 0,236, 0,235, 0,234, 0,233, 0,232, 0,231, 0,23, 0,229, 0,228, 0,227, 0,226, 0,225, 0,224, 0,223, 0,222, 0,221, 0,22, 0,219, 0,218, 0,217, 0,216, 0,215, 0,214, 0,213, 0,212, 0,211, 0,21, 0,209, 0,208, 0,207, 0,206, 0,205, 0,204, 0,203, 0,202, 0,201 o 0,200 y (b) inferior a uno cualquiera de 0,401, 0,402, 0,403, 0,404, 0,405, 0,406, 0,407, 0,408, 0,409, 0,410, 0,411, 0,412, 0,413, 0,414, 0,415, 0,416, 0,417, 0,418, 0,419, 0,420, 0,421, 0,422, 0,423, 0,424, 0,425, 0,426, 0,427, 0,428, 0,429, 0,430, 0,431, 0,432, 0,433, 0,434, 0,435, 0,436, 0,437, 0,438, 0,439, 0,440, 0,441, 0,442, 0,443, 0,444, 0,445, 0,446, 0,447, 0,448, 0,449, 0,450, 0,451, 0,452, 0,453, 0,454, 0,455, 0,456, 0,457, 0,458, 0,459, 0,460, 0,461, 0,462, 0,463, 0,464, 0,465, 0,466, 0,467, 0,468, 0,469, 0,470, 0,471, 0,472, 0,473, 0,474, 0,475, 0,476, 0,477, 0,478, 0,479, 0,480, 0,481, 0,482, 0,483, 0,484, 0,485, 0,486, 0,487, 0,488, 0,489, 0,490, 0,491, 0,492, 0,493, 0,494, 0,495, 0,496, 0,497, 0,498, 0,499, 0,500, 0,501, 0,502, 0,503, 0,504, 0,505, 0,506, 0,507, 0,508, 0,509, 0,510, 0,511, 0,512, 0,513, 0,514, 0,515, 0,516, 0,517, 0,518, 0,519, 0,520, 0,521, 0,522, 0,523, 0,524, 0,525, 0,526, 0,527, 0,528, 0,529, 0,530, 0,531, 0,532, 0,533, 0,534, 0,535, 0,536, 0,537, 0,538, 0,539, 0,540, 0,541, 0,542, 0,543, 0,544, 0,545, 0,546, 0,547, 0,548, 0,549, 0,550, 0,551, 0,552, 0,553, 0,554, 0,555, 0,556, 0,557, 0,558, 0,559, 0,560, 0,561, 0,562, 0,563, 0,564, 0,565, 0,566, 0,567, 0,568, 0,569, 0,570, 0,571, 0,572, 0,573, 0,574, 0,575, 0,576, 0,577, 0,578, 0,579, 0,580, 0,581, 0,582, 0,583, 0,584, 0,585, 0,586, 0,587, 0,588, 0,589, 0,590, 0,591, 0,592, 0,593, 0,594, 0,595, 0,596, 0,597, 0,598, 0,599 o 0,60, p. ej., si la puntuación predictiva está entre 0,2-0,6, 0,22-0,58, 0,24-0,56, 0,26-0,54, 0,28-0,52, 0,3-0,5, 0,32-0,48, 0,34-0,46, 0,36-0,44, 0,38-0.42, 0,2-0,4, 0,22-0,38, 0,24-0,36, 0,26-0,34, 0,28-0,32, 0,3-0,6, 0,32-0,58, 0,34-0,56, 0,36-0,54, 0,38-0,52, 0,4-0,5, 0,42-0,48 o 0,44-0,46, p. ej., entre 0,3-0,4, 0,31-0,39, 0,32-0,38, 0,33-0,37, 0,34-0,36, 0,3-0,39, 0,3-0,38, 0,3-0,37, 0,3-0,36, 0,3-0,35, 0,3-0,34, 0,3-0,33, 0,3-0,32, 0,3-0,31, 0,31-0,4, 0,32-0,4, 0,33-0,4, 0,34-0,4, 0,35-0,4, 0,36-0,4, 0,37-0,4, 0,38-0,4, 0,39-0,4, p. ej., 0,3, 0,301, 0,302, 0,303, 0,304, 0,305, 0,306, 0,307, 0,308, 0,309, 0,31, 0,311, 0,312, 0,313, 0,314, 0,315, 0,316, 0,317, 0,318, 0,319, 0,32, 0,321, 0,322, 0,323, 0,324, 0,325, 0,326, 0,327, 0,328, 0,329, 0,33, 0,331, 0,332, 0,333, 0,334, 0,335, 0,336, 0,337, 0,338, 0,339, 0,34, 0,341, 0,342, 0,343, 0,344, 0,345, 0,346, 0,347, 0,348, 0,349, 0,35, 0,351, 0,352, 0,353, 0,354, 0,355, 0,356, 0,357, 0,358, 0,359, 0,36, 0,361, 0,362, 0,363, 0,364, 0,365, 0,366, 0,367, 0,368, 0,369, 0,37, 0,371, 0,372, 0,373, 0,374, 0,375, 0,376, 0,377, 0,378, 0,379, 0,38, 0,381, 0,382, 0,383, 0,384, 0,385, 0,386, 0,387, 0,388, 0,389, 0,39, 0,391, 0,392, 0,393, 0,394, 0,395, 0,396, 0,397, 0,398, 0,399, o 0,4, el paciente o la muestra no están clasificados.

Sin desear quedar ligado a teoría alguna, se cree que en algunas realizaciones, el modelo descrito en el presente documento, que se describe, p. ej., mediante las características y los pesos, puede utilizarse para generar puntuaciones predictoras ajustadas (p. ej., probabilidades) para el subtipo ABC y estas puntuaciones predictoras (p. ej., probabilidades) pueden utilizarse a continuación para determinar qué subtipo es el paciente o la muestra basándose en puntos de corte predeterminados. Por ejemplo, los puntos de corte óptimos (p. ej., los puntos de corte descritos en el presente documento) para la puntuación de probabilidad continua pueden determinarse maximizando el estadístico J de Youden (el punto con la mayor sensibilidad y especificidad totales) en el conjunto de entrenamiento y, a continuación, añadiendo un amortiguador de probabilidad de 0,1 a ambos lados del punto de corte inicial para definir una región no clasificada, p. ej., como se describe en el Ejemplo 3 (p. ej., Figuras 1A y 1B). En determinadas realizaciones, el punto de corte, p. ej., el punto de corte predefinido bajo descrito en el presente documento o el punto de corte predefinido alto descrito en el presente documento, se determina como se describe en el Ejemplo 3 (p. ej., Figuras 1A y 1B).

En algunas realizaciones, se proporciona en el presente documento un método para tratar a un sujeto que tiene DLBCL que comprende determinar si el sujeto tiene una COO de ABC o GCB, que comprende i) adquirir, p. ej., recoger, una muestra, p. ej., una muestra clínica, y ii) realizar un ensayo de genotipado en la muestra para determinar si el sujeto tiene una COO de ABC o GCB, y si el sujeto tiene DLBCL de COO de ABC, a continuación administrar ibrutinib y/o lenalidomida y si el sujeto tiene DLBCL de la COO de GC, a continuación administrar ibrutinib y/o lenalidomida. En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, la COO se determina

utilizando un método o diagnóstico de cualquiera de las realizaciones descritas en el presente documento.

En algunas realizaciones, en el presente documento se proporciona un sistema informático configurado para realizar cualquiera de los métodos, diagnósticos, kits o clasificadores de cualquiera de las realizaciones descritas en el presente documento.

Breve descripción de los dibujos

FIG. 1A y **FIG. 1B** muestran curvas de características operativas del receptor (ROC) para el modelo COODC. **FIG. 1A** muestra curvas ROC que indican la sensibilidad y especificidad del modelo para ABC y GCB para el conjunto de entrenamiento GOYA y **FIG. 1B** muestra las curvas ROC que indican la sensibilidad y especificidad del modelo para ABC y GCB para el conjunto de validación retenido GOYA.

FIGS. 2A-2F muestran la concordancia de COO determinada por COODC con COO determinada por Nanostring para las muestras GOYA. **FIG. 2A** muestra gráficos circulares que muestran el desglose general de COO para todas las muestras GOYA determinado por Nanostring o COODC. **FIG. 2B** muestra gráficos circulares que muestran el desglose de denominaciones de COO por COODC dentro de cada grupo de COO de Nanostring. Curvas de supervivencia que indican la supervivencia sin progresión para los grupos de COO determinada por COODC **(FIG. 2C)** y COO determinada por COODC **(FIGs. 2D, 2E, 2F)** en comparación con COO determinada por Nanostring.

FIGS. 3A-3C muestran la concordancia de COO determinada por COODC con COO determinada por Nanostring para las muestras MAIN. **FIG. 3A** muestra gráficos circulares que muestran el desglose general de COO para todas las muestras MAIN determinado por Nanostring o COODC. **FIG. 3B muestra** gráficos circulares que muestran el desglose de denominaciones de COO por COODC dentro de cada grupo de COO de Nanostring para MAIN. **FIG. 3C** muestra un gráfico circular que muestra el desglose general de COO para muestras de la base de datos de investigación genómica desidentificada de Foundation Medicine ("FM-clinical") determinado por COODC.

FIG. 4A - 4C muestran el enriquecimiento de las características del modelo COODC por COO. El enriquecimiento se evalúa usando la COO evaluado por Nanostring. **FIG. 4A** muestra el enriquecimiento para todas las características binarias con pesos distintos de cero en el modelo. El enriquecimiento en ABC se indica mediante los puntos grises más oscuros (Log2 de razón de posibilidades >0), el enriquecimiento en GCB se indica mediante los puntos naranjas (Log2 de razón de posibilidades <0). El tamaño de los puntos indica la frecuencia de la característica en el grupo enriquecido. Las etiquetas indican características con un enriquecimiento significativo ($p < 0,05$). **FIG. 4B** muestra el enriquecimiento para características continuas con pesos distintos de cero en el modelo. GCB está siempre a la izquierda de cada panel y ABC a la derecha. Las líneas horizontales de cada caja representan la mediana de la característica, mientras que los límites inferior y superior de cada caja representan el percentil 25 y el percentil 75, respectivamente. Los bigotes asociados a cada caja se extienden hasta valores extremos no superiores a 1,5 veces el rango intercuartílico más allá de la caja correspondiente. Los puntos situados más allá de los bigotes se consideran valores atípicos y se representan individualmente. *** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$, ns = no significativo. **FIG 4C** muestra el enriquecimiento de las características binarias de COODC en COO no clasificada frente a no no clasificada. El enriquecimiento se evalúa usando la COO evaluado por Nanostring. Muestra el enriquecimiento para todas las características binarias con pesos distintos de cero en el modelo. El enriquecimiento en no clasificada se indica con puntos grises (Log2 de razón de posibilidades >0) y el enriquecimiento en no no clasificada se indica con puntos morados (Log2 de razón de posibilidades <0). El tamaño de los puntos indica la frecuencia de la característica en los grupos enriquecidos. Las etiquetas indican características con un enriquecimiento significativo en muestras no clasificadas ($p < 0,05$). Las características enriquecidas en no no clasificadas no se muestran, sino características ABC frente a GCB se indican en la **FIG 4A.**

FIGS. 5A-5D muestran las firmas de mutación por COO. **FIGS. 5A-1-5A-3** muestran gráficos del contexto de trinucleótidos para la firma 3 (firma "BRCA") (parte superior) y la firma 23 (parte inferior). **FIG. 5A** describe la disposición posicional entre las **FIGs. 5A-1, 5A-2** y **5A-3. FIG. 5B** muestra la frecuencia de firmas mutacionales por tipo de COODC para muestras GOYA en regiones genómicas cebadas como se describe en los métodos. **FIG. 5C** muestra la frecuencia de firmas mutacionales por tipo de COODC para muestras FM-Clinical en regiones genómicas cebadas como se describe en los métodos. Sólo se incluyen las muestras con >=20 mutaciones evaluables. Si no se detecta ninguna firma dominante (es decir, ninguna firma con una puntuación >0,4), la firma se notifica como "Ninguna". **FIG. 5D** muestra las cajas (con bigotes) de todas las alteraciones evaluadas para el contexto de trinucleótido (izquierda) y para TMB (derecha), con la mediana indicada por la línea horizontal en cada caja. Los límites inferior y superior de cada caja representan el percentil 25 y el percentil 75, respectivamente. Los bigotes asociados a cada caja se extienden hasta valores extremos no superiores a 1,5 veces el rango intercuartílico más allá de la caja. Los puntos situados más allá de los bigotes se consideran valores atípicos y se representan individualmente. *** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$, ns = no significativo.

FIG. 6A y **FIG. 6B** muestran otros subconjuntos genéticamente definidos de DLBCL en muestras GOYA. **FIG. 6A** muestra la frecuencia de subconjuntos que se aproximan a grupos BN2, N1, MCD y EZB identificados por Schmitz

y sus colegas (Schmitz, R., *et al.* Genetics and Pathogenesis of Diffuse Large B Cell Lymphoma. New England Journal of Medicine 378, 1396-1407 (2018)). "Ajuste múltiple" indica las muestras que cumplían los requisitos para más de un subconjunto definido genéticamente. Las muestras que no cumplían los requisitos de ningún subconjunto definido genéticamente se agruparon en función del subtipo por COODC. **FIG. 6B** muestra la frecuencia de los subconjuntos que se aproximan a los grupos C1-C5 identificados por Chapuy y sus colegas (Chapuy, B., *et al.* Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nature Medicine 24, 679-690 (2018)). "Ajuste múltiple" indica las muestras que cumplían los requisitos para más de un subconjunto definido genéticamente. Las muestras que no cumplían los requisitos de ningún subconjunto definido genéticamente se agruparon en función del subtipo por COODC.

FIG. 7A y **FIG. 7B** muestran otros subconjuntos genéticamente definidos de DLBCL en muestras FM-clinical. **FIG. 7A** muestra la frecuencia de subconjuntos que se aproximan a grupos BN2, N1, MCD y EZB identificados por Schmitz y sus colegas (Schmitz, R., *et al.* Genetics and Pathogenesis of Diffuse Large B Cell Lymphoma. New England Journal of Medicine 378, 1396-1407 (2018)). "Ajuste múltiple" indica las muestras que cumplían los requisitos para más de un subconjunto definido genéticamente. Las muestras que no cumplían los requisitos de ningún subconjunto definido genéticamente se agruparon en función del subtipo por COODC. **FIG. 7B** muestra la frecuencia de los subconjuntos que se aproximan a los grupos C1-C5 identificados por Chapuy y sus colegas (Chapuy, B., *et al.* Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nature Medicine 24, 679-690 (2018)). "Ajuste múltiple" indica las muestras que cumplían los requisitos para más de un subconjunto definido genéticamente. Las muestras que no cumplían los requisitos de ningún subconjunto definido genéticamente se agruparon en función del subtipo por COODC.

Descripción detallada de la invención

En el presente documento se describe un nuevo método, clínicamente relevante para determinar la COO de muestras de DLBCL con aproximadamente un 20 % de pureza tumoral (es decir, contenido tumoral) y sin necesidad de ARN o tejido normal emparejado. Este método concuerda en ~90 % con el ensayo Nanostring en comparación tanto con un conjunto de validación retenido como con una cohorte independiente. Se observó que las alteraciones del número de copias de los cromosomas 6p y 9p, así como la frecuencia de las transversiones T>A y T>G, eran distintas entre los subtipos ABC y GCB. Finalmente, se observó que las firmas mutacionales subyacentes al ABC son diferentes de las encontradas en el GCB, lo que sugiere diferentes funciones de la AID (citidina desaminasa inducida por activación) en el desarrollo de estos dos subtipos.

Las directrices actuales de la OMS exigen la determinación de la COO para los DLBCL recién diagnosticados, pero es un reto clínico obtener una determinación estandarizada, rigurosa para la mayoría de las muestras (He, J., *et al.* Integrated genomic DNA/RNA profiling of hematologic malignancies in the clinical setting. Blood 127, 3004-3014 (2016)). Los algoritmos de IHC varían tanto en viabilidad como en utilidad de los resultados (Hans, C.P., *et al.* Confirmation of the molecular classification of diffuse large B Cell lymphoma by immunohistochemistry using a tissue microarray. Blood 103, 275-282 (2003); Gribben, R.C., *et al.* Poor Concordance among Nine Immunohistochemistry Classifiers of Cell-of-Origin for Diffuse Large B Cell Lymphoma: Implications for Therapeutic Strategies. (2013); Meyer, P. N., *et al.* Immunohistochemical Methods for Predicting Cell of Origin and Survival in Patients With Diffuse Large B Cell Lymphoma Treated With Rituximab. Journal of Clinical Oncology 29, 200-207 (2011)). El ensayo Nanostring ha empezado a perfilarse como una alternativa más factible desde el punto de vista clínico que evalúa la expresión de un pequeño subconjunto de genes. Sin embargo, la validación publicada para este ensayo, que concuerda en un 98 % con el ensayo de micromatrices, requiere >60 % de contenido tumoral, así como ARN de alta calidad, que a menudo no está disponible en el entorno clínico (Scott, D.W., *et al.* Determining cell-of-origin subtypes of diffuse large B Cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. Blood 123, 1214-1217 (2014)). En el presente documento se describe un nuevo clasificador que concuerda en ~90 % con Nanostring, requiere sólo un 20 % de contenido tumoral (p. ej., aproximadamente un 20 % de contenido tumoral), y no requiere ARN, que es más susceptible a la degradación o de menor calidad que el ADN. Basándose en datos de muestras FM-clinical, se estima que el 50 % de las muestras de DLBCL no serían elegibles o los resultados podrían no ser fiables para la determinación de la COO en paneles de expresión como Nanostring debido a la falta de ARN o a tener una pureza tumoral <60 %. Por lo tanto, este método aplicado clínicamente aumentará en gran medida el número de muestras con una determinación fiable de la COO.

Se ha demostrado que una clasificación más granular del DLBCL utilizando alteraciones genómicas puede proporcionar más información sobre el pronóstico (Chapuy, B., *et al.* Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nature Medicine 24, 679-690 (2018); Schmitz, R., *et al.* Genetics and Pathogenesis of Diffuse Large B Cell Lymphoma. New England Journal of Medicine 378, 1396-1407 (2018)). Las mutaciones utilizadas en el modelo de predicción de la COO divulgado en este documento compartían un solapamiento significativo con las mutaciones destacadas por estos artículos. Como se ha descubierto y divulgado en el presente documento, las alteraciones en *BCL2, EZH2* y *TNFRSF14* son todas importantes predictores del subtipo GCB, coherente con el enriquecimiento de muestras de GCB entre el clúster EZB en Schmitz *et al.,* o los DLBCL del grupo C3 en Chapuy *et al.* (véanse las referencias anteriores). De manera similar, las mutaciones en *MYD88* y *CD79B* fueron altamente predictivas de un subtipo ABC. De manera interesante, se descubrió y divulgó en el presente documento que las mutaciones en *NOTCH1, NOTCH2* y *BCL6* son todas ligeramente

predictivas del subtipo ABC, a pesar de los subtipos mixtos encontrados en los grupos BN2, N1 y C1. Sin embargo, a estas alteraciones se les dio menos importancia que a otras alteraciones específicas de ABC, lo que significa que las muestras mutadas tendrían más probabilidades de caer en el subgrupo "no clasificado".

También se descubrió y divulgó en el presente documento que muchas de las alteraciones del número de copias a nivel del brazo en el modelo COODC son similares al DLBCL del grupo C2 en Chapuy *et al.* (véase anteriormente), incluido 17p (tanto el número de copias promedio como la fracción con pérdida de heterocigosidad (LOH)), 6p (número de copias promedio) y 5p (número de copias promedio). Además de estos eventos a nivel de brazo completo, varias alteraciones parciales a nivel de brazo se correlacionan con alteraciones completas a nivel de brazo incluidas en el modelo COODC, como 4q35.1 y 4q21.22 en el grupo C2 y la fracción 4q bajo LOH en COODC. Esto sugiere que muchos de los grupos genómicos identificados por Schmitz y Chapuy se solapan con la COODC definida genómicamente divulgada en el presente documento (véanse Schmitz *et al.,* y Chapuy *et al*).

Los subconjuntos aproximados similares a los grupos BN2, N1, MCD y EZB identificados por Schmitz y los grupos C1-C5 identificados por Chapuy y sus colegas se usaron en el presente documento, lo que ilustra que una mayor subdivisión en aproximaciones de estas poblaciones debería ser factible en un ensayo clínico a gran escala, que podría proporcionar información pronóstica más detallada. De manera adicional, en el presente documento se proporcionan biomarcadores terapéuticamente relevantes, incluida la dependencia de BCR, *TP53* de tipo natural, alteraciones de *EZH2* y alteraciones de *BCL2*. Dada la alta frecuencia de muestras en los conjuntos de datos de DLBCL utilizados para las invenciones y descubrimientos divulgados en el presente documento con biomarcadores asociados a terapias dirigidas (85 % en las muestras del estudio GOYA y 81 % en las muestras FM-clinical), está claro que la adición de capacidades de llamada de COO a la secuenciación de nueva generación (NGS) proporciona información tanto pronóstica como terapéuticamente relevante sobre la capacidad de actuación en pacientes con DLBCL.

También se descubrió, utilizando los conjuntos de datos de DLBCL divulgados en el presente documento, que existen similitudes en las firmas mutacionales con las identificadas por Chapuy.

Se descubrieron dos firmas mutacionales principales en muestras de DLBCL: firma COSMIC 3, anotada como firma BRCA, y firma COSMIC 23. La firma COSMIC 23, que se encontró principalmente en el subtipo ABC, presenta frecuentes alteraciones en el contexto R[C]Y, lo que sugiere una firma canónica de AID. De manera interesante, la firma COSMIC 3 es común en todos los subtipos de DLBCL, hallada en el 26 % de los GCB, el 16 % de los ABC y el 25 % de las no clasificadas de las muestras GOYA por COODC cuando se pudo determinar una firma. Dado que la firma COSMIC 3 parece capturar los efectos de DSB resueltos por unión de extremos no homólogos (NHEJ) cuando BRCA1/2 está ausente, es probable que esta firma capte la resolución de DSB inducida por AID, que también utiliza NHEJ (Kotnis, A., Du, L., Liu, C., Popov, S.W. & Pan-Hammarström, Q. Non-homologous end joining in class switch recombination: the beginning of the end. in Philos Trans R Soc Lond B Biol Sci, vol. 364 653-665 (2009)), y es esencial para CSR.

De manera adicional, se descubrió que casi exclusivamente la firma COSMIC 3 se encuentra en toda la región de los subtipos del locus IGH, lo que sugiere un proceso mutacional compartido.

Estos hallazgos encajan en un modelo según el cual los GCB se originan en una célula que ha experimentado el estrés mutacional de DSB mediada por AID, mientras que los ABC se originan en una célula que ha sufrido estrés mutacional relacionado tanto con DSB mediada por AID como con alteraciones de una sola base mediadas por AID. Esto encaja con el modelo actual en el que los GCB se originan a partir de un linfocito B del centro germinal anterior, mientras que los ABC se originan a partir de una célula del centro germinal posterior o tardía, más avanzada desde el punto de vista del desarrollo. Esto es especialmente interesante dado que sólo hay una pequeña diferencia en la carga mutacional tumoral (TMB) entre los subtipos ABC y GCB (mediana de ABC 8,9; GCB 12,1), p=0,011), y que hay un número ligeramente inferior de mutaciones evaluables para la llamada firma mutacional (incluyendo alteraciones silenciosas y no codificantes) en ABC que en GCB (mediana de ABC 23,5, GCB 31, p=0,0019). Esto sugiere que, aunque algunas células ABC experimentan tanto DSB mediada por AID como alteraciones de una sola base mediadas por AID, y aunque las células GCB parecen experimentar principalmente DSB mediada por AID, los procesos mutacionales que conducen a la transformación dan lugar a una carga mutacional global similar.

En determinadas realizaciones según (o aplicadas a) cualquiera de las realizaciones anteriores, "muestra" se refiere a una muestra biológica obtenida o derivada de una fuente de interés, como se describe en el presente documento. En algunas realizaciones, una fuente de interés comprende un organismo, tal como un animal o un ser humano. La fuente de la muestra puede ser tejido sólido como el de un órgano fresco, congelado o conservado, muestra de tejido, biopsia, resección, frotis o aspirado; sangre o cualquier constituyente de la sangre; líquidos corporales, tales como líquido cefalorraquídeo, líquido amniótico, líquido peritoneal o líquido intersticial; o células de cualquier momento durante la gestación o el desarrollo del sujeto. En algunas realizaciones, la fuente de la muestra es la sangre o componentes de la sangre.

En algunas realizaciones, la muestra es o comprende tejido o fluido biológico. La muestra puede contener compuestos que no están naturalmente mezclados con el tejido en la naturaleza, tal como conservantes, anticoagulantes,

tampones, fijadores, nutrientes, antibióticos o similares. En una realización, la muestra se conserva como muestra congelada o como preparación de tejido fijado con formaldehído o paraformaldehído e incluido en parafina (FFPE). Por ejemplo, la muestra puede incluirse en una matriz, p. ej., un bloque de FFPE o una muestra congelada. En otra realización, la muestra es una muestra de sangre o de constituyentes de la sangre. En otra realización más, la muestra es una muestra de aspirado de médula ósea. En otra realización, la muestra comprende ADN extracelular circulante (ADNlc). Sin desear quedar ligado a teoría alguna, se cree que en algunas realizaciones, el ADNlc es ADN procedente de células apoptósicas o necróticas. Normalmente, el ADNlc está unido por proteínas (p. ej., histona) y protegido por nucleasas. El ADNlc puede utilizarse como biomarcador para pruebas prenatales no invasivas (NIPT), trasplante de órgano, miocardiopatía, microbioma y cáncer. En otra realización, la muestra comprende ADN tumoral circulante (ADNtc). Sin desear quedar ligado a teoría alguna, se cree que en algunas realizaciones, el ADNtc es ADNlc con una alteración genética o epigenética (p. ej., una alteración somática o una firma de metilación) que puede discriminar si procede de una célula tumoral o de una célula no tumoral. En otra realización, la muestra comprende células tumorales circulantes (CTC). Sin desear quedar ligado a teoría alguna, se cree que en algunas realizaciones, las CTC son células que se desprenden de un tumor primario o metastásico y pasan a la circulación. En algunas realizaciones, la apoptosis de CTC es una fuente de ADNtc en la sangre/linfa.

En algunas realizaciones, una muestra biológica puede ser o comprender médula ósea; sangre; células sanguíneas; ascitis; muestras de biopsia de tejido o aguja fina; líquidos corporales que contienen células; ácidos nucleicos flotantes libres; esputo; saliva; orina; líquido cefalorraquídeo, líquido peritoneal; líquido pleural; heces; linfa; líquido ginecológico; frotis de piel; frotis vaginales; frotis orales; frotis nasales; lavados tales como lavados ductales o lavados broncoalveolares; aspirados; raspados; especímenes de médula ósea; especímenes de biopsia de tejido; especímenes quirúrgico; heces, otros líquidos corporales, secreciones y/o excreciones; y/o células de los mismos, etc.

En algunas realizaciones, una muestra biológica es o comprende células obtenidas de un individuo. En algunas realizaciones, las células obtenidas son o incluyen células de un individuo del que se obtiene la muestra. En algunas realizaciones, una muestra es una "muestra primaria" obtenida directamente de una fuente de interés por cualquier medio apropiado. Por ejemplo, en algunas realizaciones, una muestra biológica primaria se obtiene mediante un método elegido entre biopsia (p. ej., aspiración con aguja fina o biopsia de tejido), cirugía, recogida de fluido corporal (p. ej., sangre, linfa o heces), etc. En algunas realizaciones, como quedará claro por el contexto, el término "muestra" se refiere a una preparación que se obtiene procesando (p. ej., eliminando uno o más componentes de y/o añadiendo uno o más agentes a) una muestra primaria, p. ej., filtrando mediante una membrana semipermeable. Dicha "muestra procesada" puede comprender, por ejemplo, ácidos nucleicos o proteínas extraídos de una muestra u obtenidos sometiendo una muestra primaria a técnicas como la amplificación o la transcripción inversa de ARNm, aislamiento y/o purificación de determinados componentes, etc.

En algunas realizaciones, la muestra es una célula asociada a un tumor, p. ej., una célula tumoral o un linfocito infiltrante tumoral (TIL). En algunas realizaciones, la muestra incluye una o más células premalignas o malignas. En algunas realizaciones, la muestra se adquiere de una neoplasia (o preneoplasia) hematológica, p. ej., una neoplasia (o preneoplasia) hematológica descrita en el presente documento, p. ej., un linfoma difuso de linfocitos B grandes (DLBCL). En otras realizaciones, la muestra incluye una o más células tumorales circulantes (CTC) (p. ej., una CTC obtenida de una muestra de sangre). En algunas realizaciones, la muestra es una célula no asociada a un tumor, p. ej., una célula no tumoral o un linfocito de sangre periférica.

Ejemplos

Ejemplo 1: Muestras

Se disponía de datos clínicos y genómicos de los pacientes tratados en el ensayo clínico GOYA (NCT01287741; R-CHOP frente a G-CHOP) (Vitolo, U., *et al.* Obinutuzumab or Rituximab Plus Cyclophosphamide, Doxorubicin, Vincristine, and Prednisone in Previously Untreated Diffuse Large B Cell Lymphoma. Journal of Clinical Oncology 35, 3529-3537 (2017)) y el ensayo clínico MAIN (NCT00486759; R-CHOP +/- Bevacizumab) (Seymour, J.F., *et al.* R-CHOP with or without bevacizumab in patients with previously untreated diffuse large B Cell lymphoma: final MAIN study outcomes. Haematologica 99, 1343-1349 (2014)). Los protocolos de los ensayos originales fueron aprobados por comités éticos locales o nacionales de acuerdo con las leyes de cada país, y los estudios se realizaron de acuerdo con la Declaración de Helsinki. Los subtipos pronósticos de DLBCL de linfocitos B activados (ABC)/linfocitos B del centro germinal (GCB)/no clasificados se determinaron mediante el ensayo Nanostring. Las medianas de seguimiento fueron de 24 y 29 meses para MAIN y GOYA, respectivamente (≈ 30 % de acontecimientos de supervivencia sin progresión (PFS) en ambas cohortes). Las muestras de atención clínica rutinaria (a continuación en el presente documento denominadas muestras FM-clinical) se secuenciaron como parte de la atención clínica rutinaria y no se disponía de los datos clínicos correspondientes.

Ejemplo 2: Secuenciación de ADN

Todas las muestras se secuenciaron utilizando el componente de ADN de la plataforma FoundationOne®Heme, que incluye la secuenciación dirigida de ADN de aproximadamente 465 genes, como se describe en He, J., *et al.* Integrated genomic DNA/RNA profiling of hematologic malignancies in the clinical setting. Blood 127, 3004-3014 (2016). Además

de las variantes cortas validadas, alteraciones del número de copias (amplificaciones de alto nivel y deleciones profundas), reordenamientos, carga mutacional tumoral (TMB) y estado de inestabilidad de microsatélites (MSI), sólo se utilizaron las características de la plataforma para uso en investigación, incluidas las métricas del número de copias a nivel de brazo cromosómico y de pérdida de heterocigosidad (LOH).

Ejemplo 3: Aprendizaje automático

El modelo COODC se desarrolló mediante una regresión Lasso penalizada utilizando una validación cruzada interna de 25 veces implementada a partir del paquete glmnet (versión 2.0-10) en la versión 3.3.2 de R y utilizando la versión 1.0.136 de RStudio. Se dividieron 482 muestras GOYA con datos de Nanostring en un conjunto de entrenamiento (70 % de las muestras) y un conjunto de validación (30 % de las muestras). El conjunto de entrenamiento inicial se refinó aún más eliminando las muestras no clasificadas por Nanostring para centrar el entrenamiento en realizar denominaciones ABC o GCB. Se utilizaron 296 muestras para el conjunto de entrenamiento final, mientras que para el conjunto de validación se utilizaron 139 muestras. En el modelo se utilizaron 592 características. Las características continuas se puntuaron con z para mantener una escala coherente entre características continuas y binarias. El modelo COODC final incluía 74 características distintas de cero **(Tabla 1)**. Las probabilidades por muestra extraídas del modelo se utilizaron para determinar los puntos de corte ideales. Se generaron curvas ROC **(Figuras 1A y 1B)** y se eligió el "mejor" punto de corte, que optimiza la especificidad y la sensibilidad. A continuación, se añadió un 10 % a cada lado del punto de corte óptimo para generar una zona no clasificada. A continuación, se extrajeron las probabilidades para el conjunto de validación retenido y 44 muestras de validación independientes procedentes del estudio MAIN. Estos dos conjuntos de validación se utilizaron para determinar la precisión del modelo.

Tabla 1: Características del modelo COODC

Características	Coef.
(Intersección)	0,74622792
T.A	0,1466164
T.G	0,19169638
Núm de segmentos de chr2p	0,12336813
NC promedio de chr3q	-0,5625825
Fracción de LOH de chr4q	0,00273763
NC promedio de chr5q	0,00758733
NC promedio de chr6p	0,00369147
NC promedio de chr7q	0,22046432
NC promedio de chr8q	0,00869661
Núm de segmentos de chr9p	-0,0870849
Núm de segmentos de chr10p	-0,0132862
NC modal de chr10p	0,04487006
NC promedio de chr12p	0,10812066
NC promedio de chr12q	0,33833335
NC promedio de chr17p	0,06825782
Fracción de LOH de chr17p	-0,0668038
Fracción de LOH de chr17q	0,0772327
NC promedio de chr18q	-0,0135074
NC promedio de chr19q	-0,2309455
NC modal de chr19q	-0,0159153
Caract. : PIM1 | isKL - iguales - KL	-0,2789378
Caract. : NOTCH1 | tipo de variante - iguales - SV + codón - iguales - 2514	-0,7118127

(continuación)

Características	Coef.
Caract. : BCOR | tipo de variante - iguales -SV + isKL - iguales - KL	-0,4570838
Caract. : TNFRSF14 | isKL - iguales - KL	0,30795551
Caract. : BCL2 | isKL - iguales - KT	0,91799866
Caract. : HIST1H2AM | isKT - iguales - KL	-1,1452282
Caract. : CREBBP | isKL - iguales - KL	0,39528297
Caract. : GNA13 | tipo de variante - iguales - SV + codón - iguales - 28	0,04776333
Caract. : BCL10 | tipo de variante - iguales - SV + isKL - iguales - KL	-0,657969
Caract. : KMT2D | tipo de variante - iguales - SV + isKL - iguales - KL	-0,0542613
Caract. : BRCA1 | isKTL - iguales - KL	-0,2540207
Caract. : DUSP2 | tipo de variante - iguales - SV + codón - iguales - 134	-0,8152089
Caract. : PIM1 | tipo de variante - iguales - SV + codón - iguales - 135	-0,1497665
Caract. : PTPN11 | isKTL - iguales - KL	-0,1167284
Caract. **:** BTG2 | tipo de variante - iguales - SV + codón - iguales - 48	-0,0152634
Caract. : BTG2 | tipo de variante - iguales - SV + codón - iguales - 45	0,02565534
Caract. : CCND3 | tipo de variante - iguales - SV + codón - iguales - 190	0,57357736
Caract. **:** DAXX | tipo de variante - iguales - SV + codón - iguales - 455	-0,3863176
Caract. : FBXO11 | isKL - iguales - KL	0,12557906
Caract. : MAPK1 | tipo de variante - iguales - SV + codón - iguales - 1	0,020638698
Caract. : IGH | tipo de variante - iguales - RF + gen2 - iguales a - BCL2 + isKL -	0,42225842
Caract. : KDM6A | tipo de variante - iguales - SV + isKL- iguales - KL	-0,382399
puntos calientes de STAT6 417 419	0,42139716
Caract. : BCL6 | tipo de variante - iguales - RF + gen2 - iguales - IGH + isKL -	-0,1259457
Caract. : B2M | tipo de variante - iguales - SV + codón - iguales - 1	0,74380662
Caract. : PTPN6 | tipo de variante - iguales - SV + isKL - iguales - KL	-0,1007704
Caract. : TP53 | tipo de variante - iguales - SV + codón - iguales - 273	0,00758395
Caract. **:** KMT2D | isKTL - iguales - KL	-0,0167986
Caract. : HTST1H1E | tipo de variante - iguales - SV + isKL - iguales - KL	0,18457867
Caract. : CARD11 | tipo de variante - iguales - SV + codón - iguales - 230	-0,0008312
del de CDKN2A o B	-0,6592701
Caract. : PTPN11 | tipo de variante - iguales - SV + codón - iguales - 197	-0,0001514
Caract. : PIM1 | tipo de variante - iguales - SV + codón - iguales - 97	-0,8007032
Caract. : CD79B | tipo de variante - iguales - SV + codón - iguales - 196	-0,0960342
Caract. : CDKN2B | tipo de variante - iguales - RF + isKL - iguales - KL	-0,7328949
Caract. : MYC | tipo de variante - iguales - RF + gen2 - iguales - IGH + isKL - iguales -	-0,0427746
Caract. : MYD88 | tipo de variante - iguales - SV + codón - iguales - 265	-0,359751
Caract. : DDX3X | tipo de variante - iguales - SV + isKL - iguales - KL	0,06998015

(continuación)

Características	Coef.
Caract. : PIM1 | tipo de variante - iguales - SV + codón - iguales - 30	-0,0375632
Caract. : BCL10 | isKTL - iguales - KL	-0,1240489
Caract. : TBL1XR1 | tipo de variante - iguales - SV + isKL - iguales - KL	-0,4020574
puntos calientes de KLHL6 547 568 65 90	0,0732899
Caract. : IGH | tipo de variante - iguales - RF + gen2 - iguales a - BCL6 + isKL -	-0,0026765
Caract. : IRF4 | tipo de variante - iguales - SV + codón - iguales - 70	-0,2206269
Caract. : PRDM1 | isKL - iguales - KL	-0,3567579
Caract. : COND3 | tipo de variante - iguales - SV + isKL - iguales - KL	-0,0157609
Caract. : BCL2 | tipo de variante - iguales - RF + isKL - iguales - KL	0,02093477
Caract. : HTST1H2AM | tipo de variante - iguales - SV + isKL - iguales - KL	-0,0517756
Caract. : KLHL6 | tipo de variante - iguales - SV + codón - iguales -568	0,03590108
Caract. : KDM6A | isKTL - iguales - KL	-0,0095082
Caract. : PTPN6 | isKL - iguales - KL	-0,0194835
Caract. : PTPN11 | tipo de variante - iguales - SV + isKL - iguales - KL	-0,0014204
Caract. : ETS1 | tipo de variante - iguales - SV + codón - iguales - 17	-0,2238722
Caract. : PIM1 | tipo de variante - iguales - SV + codón - iguales - 10	-0,0111813

Ejemplo 4: Estadísticas

El enriquecimiento de características se evaluó utilizando la COO asignada a Nanostring. El enriquecimiento de características binarias se determinó utilizando una prueba exacta de Fisher. El enriquecimiento de características continuo se determinó utilizando una prueba de Mann-Whitney-Wilcox. Se calcularon las razones de riesgos univariantes y los valores de p para la asociación de los subtipos de COO con PFS usando la regresión de Cox.

Ejemplo 5: Firmas mutacionales

Las firmas mutacionales se determinaron según lo descrito por Zehir *et al.* (Zehir, A., *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med 23, 703-713 (2017)). Resumiendo, se descompusieron matrices de trinucleótidos en las 30 firmas COSMIC (Alexandrov, L.B., *et al.* Signatures of mutational processes in human cancer. Nature 500, 415-421 (2013)). Las firmas se agregaron a APOBEC (firmas 2 y 13); tabaquismo (firma 4), BRCA (firma 3); MMR (firmas 1, 6, 15, 20 y 26); UV (firma 7); POLE (firma 10); y alquilantes (firma 11). Todas las mutaciones puntuales se incluyeron en el análisis, a excepción de las alteraciones impulsoras y las alteraciones de la línea germinal previstas. Se consideró que una muestra tenía una firma dominante si una clase mutacional albergaba una puntuación de 0,4 o superior. Para el cálculo del genoma completo, se consideraron mutaciones en todas las regiones cebadas para la detección de variantes cortas, excluyendo las regiones con artefactos de secuenciación conocidos. Sólo se tuvieron en cuenta las muestras que presentaban 20 o más mutaciones evaluables. Para el análisis de IGH, se evaluaron mutaciones en regiones seleccionadas del locus genómico IGH (chr14:106211312-106382700 hg19).

Ejemplo 6: Clasificador de ADN de COO: un modelo definido genómicamente para célula de origen

Se dividieron 482 muestras de DLBCL del estudio GOYA cuyo ADN se secuenció mediante FoundationOne®Heme en dos conjuntos: 2/3 en un conjunto de entrenamiento y 1/3 en un conjunto de validación retenido. El conjunto de entrenamiento se subdividió adicionalmente para incluir únicamente muestras con una llamada ABC o GCB determinada por Nanostring; se excluyeron las muestras no clasificadas o muestras sin llamada. Una vez completado el modelo, se utilizó el conjunto de validación retenido para determinar la concordancia. A continuación, los datos del conjunto de entrenamiento se utilizaron para entrenar un modelo de regresión logística penalizada para identificar muestras de ABC o GCB utilizando características basadas en el ADN sin necesidad de ARN. Estaban disponibles 594 características para entrenar el modelo, incluidas las características binarias de cualquier alteración en un gen, alteraciones específicas (codones) y alteraciones de punto caliente (uno cualquiera de los múltiples codones) que se produjeron al menos 5 veces en el conjunto de datos GOYA, así como características derivadas basadas en el ADN

como la carga mutacional tumoral (TMB), métricas de número de copias a nivel de brazo cromosómico y de cigosidad, y frecuencia de clases de alteración (p. ej., T mutada a A). Las probabilidades por muestra se extrajeron del modelo y se eligió un par de puntos de corte para optimizar la sensibilidad y la especificidad, con especial atención a la optimización de la precisión de ABC. Este modelo se denomina en el presente documento clasificador de ADN de la célula de origen (COODC; también se hace referencia en el presente documento al clasificador, p. ej., como método, modelo y ensayo). El modelo COODC contenía un total de 74 características genómicas y generaba una puntuación de probabilidad continua de que una muestra fuera ABC que oscilaba entre 0 y 0,999. Las 74 características genómicas incluían 18 características de alteración a nivel de brazo, incluidas las características de número de copias y pérdida de heterocigosidad, 32 características de las variantes cortas de genes, 6 características basadas en reordenamiento, 13 características a nivel de genes (incluido el número de copias, reordenamiento y alteraciones de variantes cortas) y diversas otras características resumidas (incluida la prevalencia de mutaciones T>A).

Para definir puntos de corte óptimos para la puntuación de probabilidad continua, se determinó el "mejor" punto de corte del modelo ROC **(Figuras 1A** y **1B)**, es decir, el punto con la mayor sensibilidad y especificidad totales. A continuación, se añadió un búfer de probabilidad de 0,1 a cada lado del punto de corte óptimo para definir una región no clasificada. Esto dio lugar a desgloses de subtipos similares en todo el conjunto de datos GOYA, con un 57,9 % de GCB, un 30,3 % de ABC y un 11,8 % de no clasificado según lo determinado por la COODC en comparación con el 54,5 % de GCB, el 26,5 % de ABC y el 15,6 % de no clasificado (el 3,4 % de los casos no se sometió a la tipificación de COO por Nanostring), según lo determinado por Nanostring **(Figura 2A)**. Esto da como resultado 85 (88,5 %) denominaciones concordantes dentro del conjunto de validación retenido cuando se consideran 96 muestras con una denominación ABC o GCB de ambos métodos COODC y Nanostring (se excluyeron las muestras no clasificadas por cualquiera de los métodos) **(Tabla** 2). Para esta evaluación, se excluyeron todas las denominaciones no clasificadas. Sin embargo, en consonancia con la evaluación proporcionada por Scott *et al.* (Scott, D.W., *et al.* Determining cell-of-origin subtypes of diffuse large B Cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. Blood 123, 1214-1217

(2014)), donde sólo se excluyeron los no clasificados del "método de referencia", se encontró un 90 % (97/108) de concordancia con Nanostring **(Tabla 2).** De 75 muestras consideradas GCB por Nanostring, 59 (78,7 %) también fueron denominados GCB por COODC, 9 (12 %) se denominaron no clasificadas y 7 (9,3 %) se denominaron incorrectamente ABC **(Figura 2B** y **Tabla 2).** De manera similar, de 33 denominaciones ABC por Nanostring, el 78,8 % (26) se denominaron ABC por COODC, el 9,1 % (3) se denominaron no clasificadas, y el 12,1 % (4) se denominaron incorrectamente GCB **(Figura 2B** y **Tabla 2)**. Además, la concordancia con Nanostring se evaluó en una cohorte de estudio independiente (estudio MAIN) de 44 muestras. En esta cohorte, la COODC demostró una alta concordancia continuada con una precisión del 91,9 % **(Figuras 3A, 3B,** y **3C** y **Tabla 3).** El modelo COODC se aplicó además a un conjunto independiente de 597 muestras FM-clinical **(Figuras 3A, 3B** y **3C).** Aunque no existe un método de referencia de evaluación de COO con el que comparar, se encontraron desgloses similares del tipo de COO por COODC que en las muestras GOYA, con un 60 % de GCB, un 30 % de ABC y un 10 % de no clasificado.

Tabla 2: Conjunto de validación retenido

Conjunto de validación retenido		Nanostring		
		ABC	GCB	NO CLASIFICADO
COODC	ABC	26	7	15
	GCB	4	59	8
	NO CLASIFICADO	3	9	8

Tabla 3: Conjunto de validación retenido

		Nanostring		
		ABC	GCB	NO CLASIFICADO
COODC	ABC	16	1	3
	GCB	2	18	1
	NO CLASIFICADO	1	2	0

En términos generales, las estimaciones de supervivencia sin progresión (PFS) para las clasificaciones por COODC fueron muy coherentes con las denominaciones de COO por Nanostring **(Figura** 2D), y se observó una PFS significativamente reducida entre las muestras clasificadas como ABC (CRI: 1,6; intervalo de confianza (IC) del 95 % [1,1-2,4]; p=0,011) o no clasificadas (CRI: 1,9; IC del 95 % CI [1,1-3,2]; p=0,021) en comparación con las muestras GCB. Esto indica que es posible estratificar el pronóstico de los pacientes utilizando un modelo de COO basado únicamente en el ADN, como la COODC.

Ejemplo 7: Importancia biológica de las características incluidas en el modelo COODC

Debido a que se adoptó un enfoque no sesgado para la selección de características, la importancia biológica potencial de las características incluidas en el modelo COODC era de interés. Se comprobó que una serie de características esperadas, incluyendo reordenamientos de *IGH:BCL2*, alteraciones de *CREBBP* y alteraciones de *TNFRSF14* están todas muy enriquecidas en las muestras GCB. El aumento del número de copias del cromosoma 3q, las deleciones de *CDKN2A/B*, las alteraciones de *CD79B* en el aminoácido 196 y las alteraciones de *MYD88* en el aminoácido 265 estaban muy enriquecidas en las muestras ABC **(Figura 4A)**. También se identificaron algunas características novedades de especial interés, incluida la frecuencia de las alteraciones T>A y T>G, el número de copias promedio del cromosoma 6p y el número de segmentos de número de copias del cromosoma 9p **(Figura 4B)**. Tanto las frecuencias de alteración T>A como T>G en las alteraciones no impulsoras son mayores en GCB ($p<0,0001$ para ambas) **(Figura 4B)**. Un menor número de copias promedio del cromosoma 6p, que codifica el locus HLA, se encontró en ABC ($p=0,0064$). El número de segmentos modelados en el cromosoma 9p, que incluye *CD274* y *PDCD1LG2,* es mayor en ABC ($p<0,0001$), lo que sugiere que las regiones de este brazo están fracturadas, dando lugar a diferencias en el número de copias. Estos dos componentes, que incluyen el número de segmentos modelados en el cromosoma 9p y el menor número de copias del cromosoma 6p, apuntan a una posible importancia biológica para la evasión inmunitaria.

Además de las características específicamente enriquecidas en ABC o GCB, había pruebas del carácter verdaderamente intermedio de las muestras no clasificadas. Se identificó un enriquecimiento de características novedosas, como reordenamientos de *IGH:BCL6* (Figura 4C), así como un enriquecimiento de características similares a ABC o GCB, como el número de copias promedio del cromosoma 6p en el que no clasificado es similar a GCB, y la frecuencia de alteración T>A en la que no clasificado es similar a ABC (Figura 4B). Muchas de las características incluidas en el modelo no estaban significativamente enriquecidas en los subtipos ABC o GCB. Esto sugiere que algunas características pueden ser aspectos importantes de la predicción sólo en combinación con otras características o en ausencia de otras características.

Ejemplo 8: Firmas mutacionales

La diferencia observada en las alteraciones T>A y T>G dio lugar a una investigación de las posibles diferencias en las firmas mutacionales. Dado el papel conocido de AID en SHM y CSR en linfocitos B normales del centro germinal, el papel propuesto en la transformación a linfoma, y los datos recientes de Chapuy que muestran firmas mutacionales distintas en DLBCL, se investigó si había evidencias de los procesos mutacionales de AID en todas las alteraciones del ADN. Utilizando firmas mutacionales de trinucleótidos identificadas en COSMIC, se observó que el 26,3 % de las muestras COODC-GCB, en comparación con el 15,8 % de las muestras ABC ($p = 0,13$), presentan una firma COSMIC 3 dominante, mientras que el 14 % de las muestras ABC en comparación con el 1,7 % de las muestras GCB ($p = 0,002$) tienen una firma 23 dominante **(Figura 5B)**.

La firma 23 está dominada por el contexto de trinucleótido en el que C se altera para dar T **(Figura 5A)**. De los principales contextos de trinucleótido, 3/5 se encuentran en el contexto R[C]Y **(Figura 5A)**. Esta firma es similar a WR[C]Y, un contexto común para AID dirigido a la hipermutación somática, lo que sugiere una fuerte firma mutacional de AID en ABC. De manera notable, esto no fue el resultado de un menor número de mutaciones en general en el subconjunto GCB; el recuento total de mutaciones fue mayor en promedio en los subconjuntos GCB en comparación con ABC (31 en GCB frente a 23,5 en ABC, $p=0,002$) **(Figura 5D)**. De manera interesante, la firma COSMIC predominante en las muestras GCB es la firma 3, que ha sido anotada como firma "BRCA". Dadas las similitudes entre la resolución de DSB en ausencia de BRCA, que utiliza la unión de extremos no homólogos (NHEJ) y la resolución de DSB inducida por AID como parte de CSR, que también utiliza principalmente NHEJ, se presume que la firma 3 puede identificar las cicatrices mutacionales de la reparación de DSB, y en este contexto, puede representar una firma AID-DSB.

Ejemplo 9: Otros subconjuntos de DLBCL definidos genéticamente

A la luz de las recientes publicaciones que destacan el uso de subconjuntos de DLBCL definidos genéticamente para estratificar más grupos de buen y mal pronóstico, se sometieron a investigación subconjuntos en los datos de este panel NGS dirigido. Dadas las similitudes entre los subconjuntos identificados por Schmitz *et al.* y Chapuy *et al.* (MCD es similar a C5, EZB es similar a C3 y BN2 es similar a C1 según la clasificación de Schmitz y Chapuy, respectivamente), definiciones aproximadas correspondientes a la "semilla" de estos subconjuntos específicos, incluyendo alteraciones de CD79B o MYD88 L265P para MCD/C5, reordenamiento de BCL6 o alteración de NOTCH2 para BN2/C1, se identificaron la alteración de EZH2 o el reordenamiento de BCL2 para EZB/C3 y la alteración de NOTCH1 para N1.

Utilizando estos grupos genéticamente definidos simplificados, el 46 % de las muestras GOYA y el 57 % de muestras FM-clinical pudieron clasificarse en uno de estos grupos, mientras que el 49 % de las muestras GOYA y el 37 % de las muestras FM-clinical no pudieron clasificarse, y el 5 % restante de las muestras GOYA y el 6 % de las muestras FM-clinical pudieron clasificarse en más de uno de estos grupos genéticamente definidos **(Figuras 6A** y **6B, Figuras 7A** y **7B).**

Además de estos grupos que son similares a los recientemente definidos como potencialmente de valor pronóstico por Schmitz y sus colegas y Chapuy y sus colegas, se identificaron subgrupos con valor terapéutico **(Tabla 4** y **Tabla 5)** utilizando la COODC divulgada en el presente documento. Fue posible identificar subgrupos independientes de BCR y dependientes de BCR utilizando alteraciones de *CD79B* con ABC, con implicaciones predictivas para ibrutinib, como demostraron Wilson y sus colegas (Wilson, W.H., *et al.* Targeting B cell receptor signaling with ibrutinib in diffuse large B cell lymphoma. Nature Medicine 21, 922 (2015)). También se encontró que un subgrupo dependiente de BCR, definido por la presencia de una alteración deletérea predicha de *CD79B*, está presente en el 24,5 % de las muestras del estudio GOYA ABC (9,4 % en general en GOYA) y en el 21,9 % de las muestras FM-clinical ABC (7,9 % en general en las muestras FM-clinical). Adicionalmente, se observó que el 79,2 % de las muestras GOYA y el 63,8 % de las muestras FM-clinical eran *TP53* de tipo natural, lo que sugiere una posible relevancia de los inhibidores de MDM2 en estas cohortes. Los inhibidores de EZH2 pueden ser clínicamente relevantes en el 10 % de las muestras GOYA y el 12,7 % de las muestras FM-clinical con alteraciones de EZH2, mientras que los inhibidores de BCL2 podrían considerarse en el 23,2 % de las muestras GOYA y el 32,5 % de las muestras FM-clinical con reordenamientos de *IGH:BCL2*, amplificaciones de *BCL2* o variantes cortas de *BCL2* **(Tabla 4** y **Tabla 5)**. En términos generales, esto sugiere que hasta el 85 % de las muestras de DLBCL en las cohortes sometidas a prueba podrían ser elegibles para terapia dirigida utilizando un inhibidor de MDM2, un inhibidor de EZH2, un inhibidor de BCL2 o un inhibidor de la vía BCR.

Tabla 4: Subgrupos adicionales definidos molecularmente para GOYA

Subtipo por COODC		**GCB**	**ABC**	**NO CLASIFICADO**
Similar a MCD de Schmitz	12,4	2,1	34,4	6,8
Similar a BN2 de Schmitz	14,4	13,8	12,6	22,0
Similar a EZB de Schmitz	19,2	31,1	0,7	8,5
Similar a N1 de Schmitz	2,4	1,0	4,0	5,1
Schmitz_múltiple	6,2	5,5	7,9	5,1
Similar a C1 de Chapuy	3,2	3,8	2,0	3,4
Similar a C2 de Chapuy	1,8	1,0	2,6	3,4
Similar a C3 de Chapuy	29,3	40,8	9,9	22,0
Similar a C4 de Chapuy	2,6	3,1	1,3	3,4
Similar a C5 de Chapuy	9,0	2,1	22,5	8,5
Chapuy múltiple	17,0	12,5	27,8	11,9
Dependiente de BCR (CD79B+)	9,4	1,7	24,5	8,5
Doble mutación	1,4	2,1	0,0	1,7
TP53 TN	78,8	77,5	86,1	66,1
Alteración de BCL2	23,2	33,6	9,9	6,8
Alteración de EZH2	10,0	15,9	1,3	3,4
Seleccionable como diana	85,0	86,2	88,7	69,5
Alteración de B2M	15,8	19,0	11,3	11,9
Amplificación de CD274	4,4	4,8	4,0	3,4
Alteración de CD58	8,2	8,3	9,9	3,4
Alteración de CD70	8,8	8,3	9,3	10,2
Alteración de CIITA	4,8	6,2	2,6	3,4
Evasión inmunitaria	33,9	35,6	32,5	28,8

Tabla 5: COO más FM-Clinical

Subtipo por COODC	TODAS	GCB	ABC	NO CLASIFICADO
Similar a MCD de Schmitz	13,0	1,1	38,4	9,8
Similar a BN2 de Schmitz	12,8	9,9	13,4	27,9
Similar a EZB de Schmitz	29,7	47,4	2,3	4,9
Similar a N1 de Schmitz	2,6	0,9	5,8	3,3
Schmitz_múltiple	7,0	7,1	8,1	3,3
Similar a C1 de Chapuy	2,2	1,7	2,3	4,9
Similar a C2 de Chapuy	4,1	2,6	7,0	4,9
Similar a C3 de Chapuy	29,9	42,9	9,9	11,5
Similar a C4 de Chapuy	1,9	1,4	2,9	1,6
Similar a C5 de Chapuy	10,3	2,8	26,7	6,6
Chapuy múltiple	23,8	22,7	27,3	19,7
Dependiente de BCR (CD79B+)	8,0	1,4	22,7	4,9
Doble mutación	1,7	2,8	0,0	0,0
TP53 TN	62,7	61,1	68,0	57,4
Alteración de BCL2	32,8	50,3	5,8	8,2
Alteración de EZH2	13,0	20,2	2,3	1,6
Seleccionable como diana	80,7	86,4	75,0	63,9
Alteración de B2M	15,9	21,0	7,0	11,5
Amplificación de CD274	2,7	2,0	4,7	1,6
Alteración de CD58	9,4	8,2	13,4	4,9
Alteración de CD70	6,8	8,0	6,4	1,6
Alteración de CIITA	5,3	6,8	3,5	1,6
Evasión inmunitaria	31,1	34,1	29,1	19,7

REIVINDICACIONES

1. Un método para determinar la célula de origen (COO) de un linfoma difuso de linfocitos B grandes (DLBCL) de un paciente diagnosticado con DLBCL, comprendiendo el método:

a. realizar la secuenciación de ADN en ADN de una muestra obtenida del paciente para obtener una lista de características genómicas asociadas a la muestra,
b. aplicar, mediante un sistema informático, un modelo de clasificación del ADN de la célula de origen predefinido a la lista de características genómicas para calcular una puntuación de probabilidad, en donde el modelo de clasificación del ADN de la célula de origen se describe mediante un conjunto de características ponderadas basadas en el ADN y se entrena para identificar muestras de COO de linfocitos B activados (ABC) y COO de linfocitos B del centro germinal (GCB) utilizando características basadas en el ADN, en donde el conjunto de características ponderadas basadas en el ADN incluye métricas del número de copias a nivel de brazo cromosómico y de cigosidad, características a nivel de genes y una o más características de clase de frecuencia de alteración, y
c. clasificar, mediante el sistema informático, al paciente como que tiene una COO de linfocitos B del centro germinal (GCB) cuando la puntuación de probabilidad es inferior a un primer punto de corte predefinido, y clasificar, mediante el sistema informático, al paciente como que tiene una COO de linfocitos B activados (ABC) cuando la puntuación de probabilidad es superior o igual a un segundo punto de corte predefinido, en donde el primer corte predefinido está en o por debajo del segundo corte predefinido.

2. El método de la reivindicación 1, en donde el primer corte predefinido está entre 0,2-0,3, y el segundo corte predefinido está entre 0,4-0,6.

3. El método de la reivindicación 1 o la reivindicación 2, en donde no se realiza ningún análisis de ARN, y/o en donde el análisis de ARN no contribuye a la determinación de la COO, o en donde el ADN comprende ADN obtenido sometiendo la muestra a transcripción inversa de ARNm.

4. El método de una cualquiera de las reivindicaciones 1-3, en donde no se realiza ningún análisis de inmunohistoquímica (IHC), y/o en donde el análisis de IHC no contribuye a la determinación de la COO.

5. El método de cualquiera de las reivindicaciones 1-4, en donde se detectan alteraciones en los genes BCL2, EZH2 y TNSFRSF14 y/o en donde se detectan alteraciones en el número de copias del cromosoma 3q, los genes MYD88 y CD79B y/o en donde se detectan alteraciones en los genes NOTCH1, NOTCH2 y BCL6 y/o en donde se detectan un reordenamiento de IGH:BCL2, una alteración de CREBBP y una alteración de TNFRSF14 y/o en donde se detectan una deleción de CDKN2A/B, una alteración de CD79B en el aminoácido 196 y una alteración de MYD88 en el aminoácido 265; opcionalmente en donde la detección de alteraciones en los genes BCL2, EZH2 y TNSFRSF14 utilizando la COODC predice que la COO del DLBCL es un linfocito B del centro germinal (GCB) y/o en donde la detección de alteraciones en el número de copias del cromosoma 3q, los genes MYD88 y CD79B utilizando la COODC predice que la COO del DLBCL es un linfocito B activado (ABC) y/o en donde la detección de alteraciones en los genes NOTCH1, NOTCH2 y BCL6 utilizando la COODC predice que la COO del DLBCL es un linfocito B activado (ABC) y/o en donde la detección de un reordenamiento de IGH:BCL2, una alteración de CREBBP y una alteración de TNFRSF14 predice que la COO del DLBCL es un GCB y/o en donde la detección de una deleción de CDKN2A/B, una alteración de CD79B en el aminoácido 196 y una alteración de MYD88 en el aminoácido 265 predice que la COO del DLBCL es un ABC.

6. El método de cualquier reivindicación anterior, comprendiendo además clasificar a un paciente que tiene una puntuación predictiva superior o igual al primer punto de corte predefinido e inferior al segundo punto de corte predefinido como no clasificado.

7. El método de cualquiera de las reivindicaciones 1-6, en donde el paciente se identifica como que tiene una COO de GCB y se selecciona para la administración de una terapia que es eficaz para tratar COO de GCB, o en donde el paciente se identifica como que tiene una COO de ABC y se selecciona para la administración de una terapia que es eficaz para COO de ABC.

8. El método de cualquiera de las reivindicaciones 1-7, en donde la muestra es una muestra clínica.

9. El método de la reivindicación 8, en donde la muestra clínica se selecciona del grupo que consiste en una biopsia de tumor, sangre, aspirado de médula ósea, ADN extracelular circulante, ADN tumoral circulante y células tumorales circulantes.

10. El método de la reivindicación 9, en donde la muestra clínica comprende ADN extracelular circulante.

11. El método de la reivindicación 9, en donde la muestra clínica es una biopsia de tumor conservada como preparación fijada en formol e incluida en parafina (FFPE).

12. El método de cualquier reivindicación anterior, en donde el modelo de clasificación del ADN de la célula de origen se desarrolló usando regresión logística penalizada.

13. El método de cualquier reivindicación anterior, en donde el conjunto de características incluye una o más características seleccionadas de la Tabla 1.

14. Un sistema informático para determinar la célula de origen (COO) de un linfoma difuso de linfocitos B grandes (DLBCL) de un paciente diagnosticado con DLBCL, el sistema informático configurado para realizar un método que comprende:

a. obtener una lista de características genómicas asociadas a una muestra del paciente adquirida mediante la secuenciación de ADN en el ADN de la muestra,
b. aplicar un modelo de clasificación del ADN de la célula de origen predefinido a la lista de características genómicas para calcular una puntuación de probabilidad, en donde el modelo de clasificación del ADN de la célula de origen se describe mediante un conjunto de características ponderadas basadas en el ADN y se entrena para identificar muestras de COO de linfocitos B activados (ABC) y COO de linfocitos B del centro germinal (GCB) utilizando el conjunto de características ponderadas basadas en el ADN, en donde el conjunto de características ponderadas basadas en el ADN incluye métricas del número de copias a nivel de brazo cromosómico y de cigosidad, características a nivel de genes y una o más características de clase de frecuencia de alteración, y
c. clasificar al paciente como que tiene una COO de linfocitos B del centro germinal (GCB) cuando la puntuación de probabilidad es inferior a un primer punto de corte predefinido, y clasificar, mediante el sistema informático, al paciente como que tiene una COO de linfocitos B activados (ABC) cuando la puntuación de probabilidad es superior o igual a un segundo punto de corte predefinido, en donde el primer corte predefinido está en o por debajo del segundo corte predefinido.

15. El sistema informático de la reivindicación 14, en donde el sistema informático está configurado para realizar un método que tiene las características de cualquiera de las reivindicaciones 2 a 13.

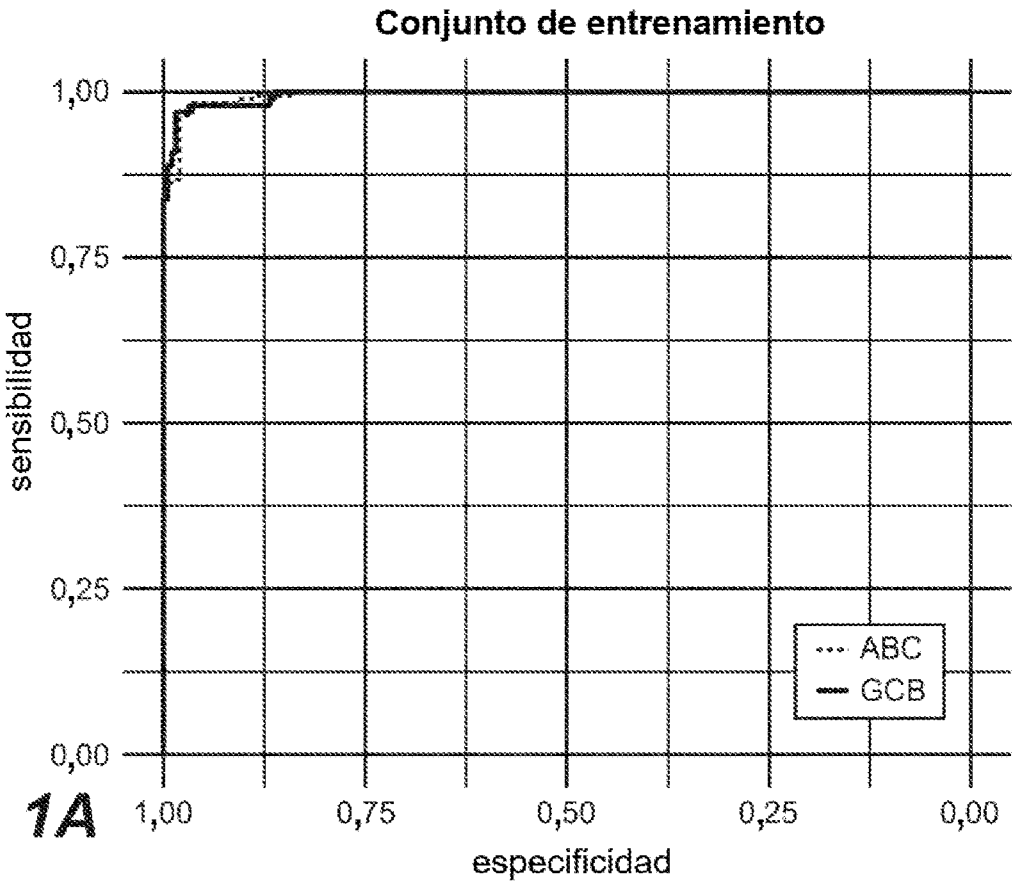

Conjunto de entrenamiento
1,00
0,75
0,50
0,25
0,00
sensibilidad
ABC
GCB
1,00
0,75
0,50
0,25
0,00
especificidad

FIG. 1A

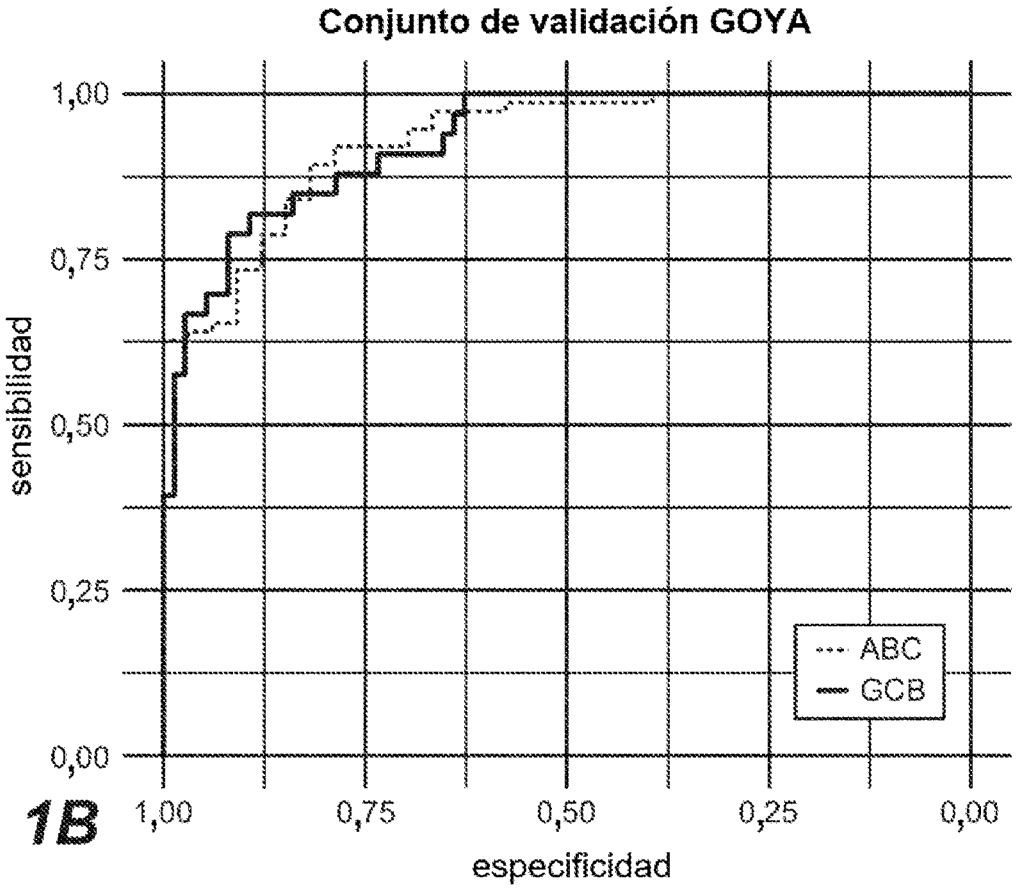
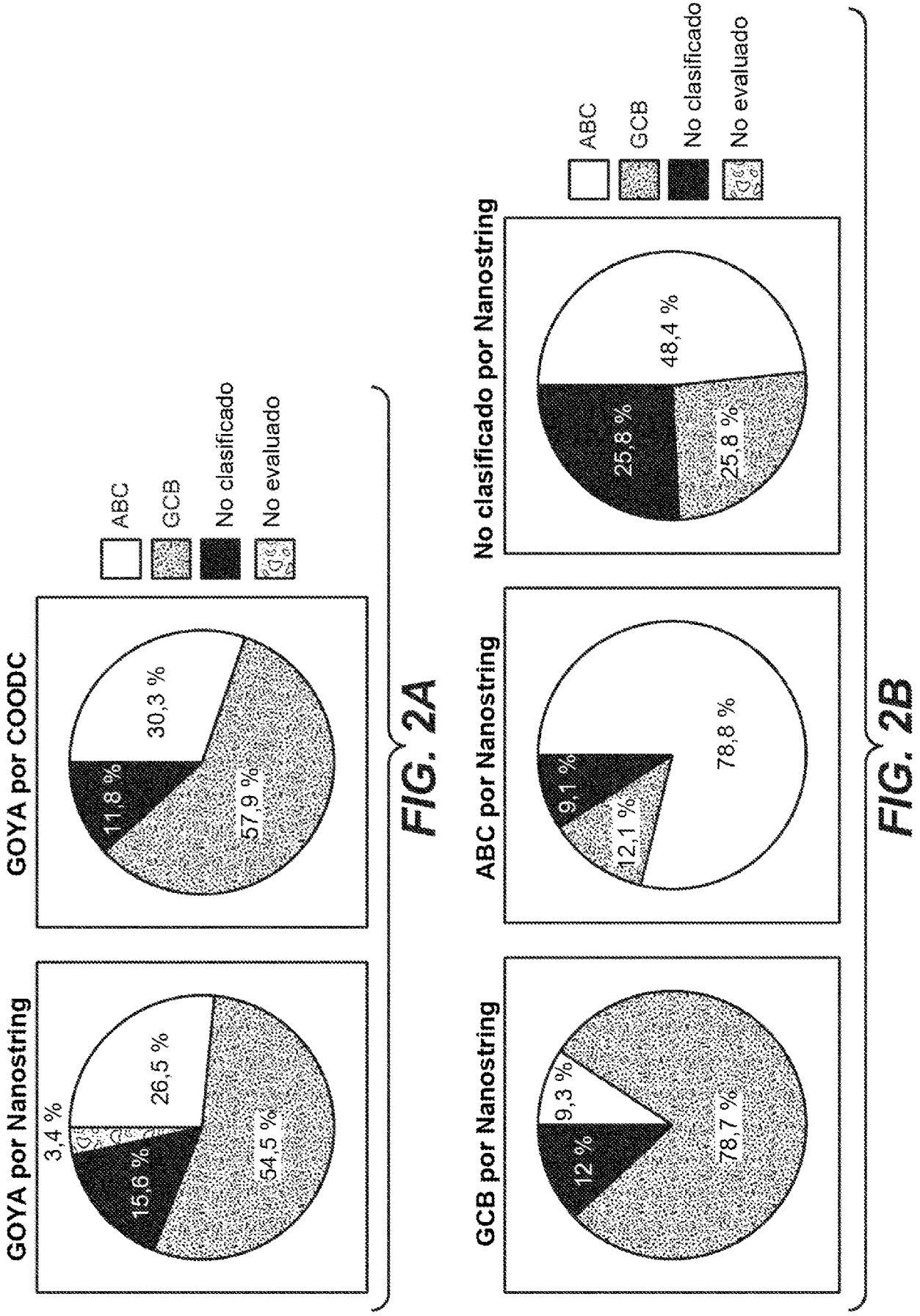

Conjunto de validación GOYA
1,00
0,75
0,50
0,25
0,00
sensibilidad
ABC
GCB
1,00
0,75
0,50
0,25
0,00
especificidad

FIG. 1B


GOYA por Nanostring
3,4 %
26,5 %
15,6 %
54,5 %
GOYA por COODC
11,8 %
30,3 %
57,9 %
ABC
GCB
No clasificado
No evaluado
FIG. 2A
GCB por Nanostring
9,3 %
12 %
78,7 %
ABC por Nanostring
9,1 %
12,1 %
78,8 %
No clasificado por Nanostring
25,8 %
48,4 %
25,8 %
ABC
GCB
No clasificado
No evaluado
FIG. 2B

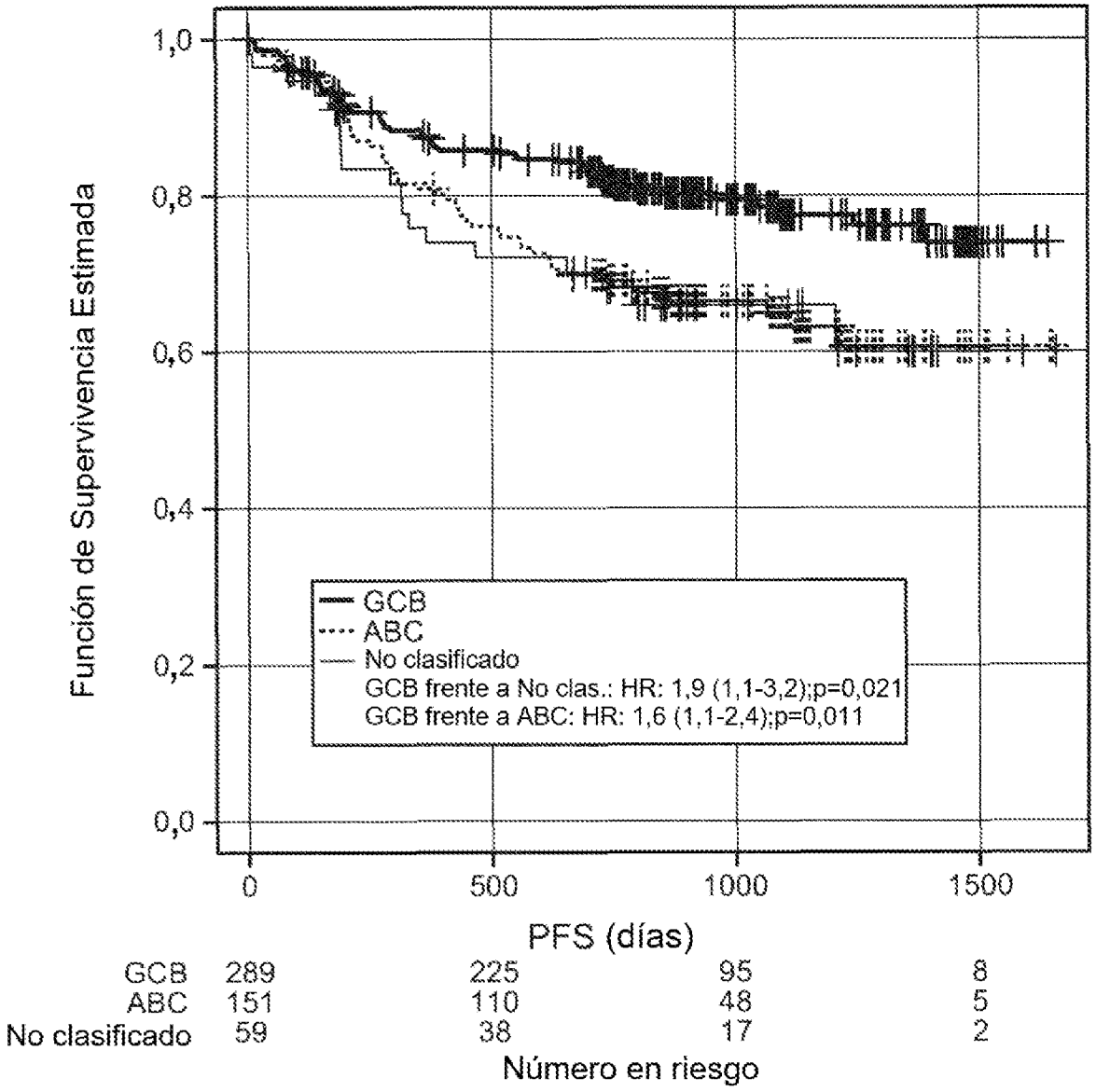

Función de Supervivencia Estimada
1,0
0,8
0,6
0,4
0,2
0,0
GCB
ABC
No clasificado
GCB frente a No clas.: HR: 1,9 (1,1-3,2);p=0,021
GCB frente a ABC: HR: 1,6 (1,1-2,4);p=0,011
0
500
1000
1500
PFS (días)
GCB 289 225 95 8
ABC 151 110 48 5
No clasificado 59 38 17 2
Número en riesgo

FIG. 2C

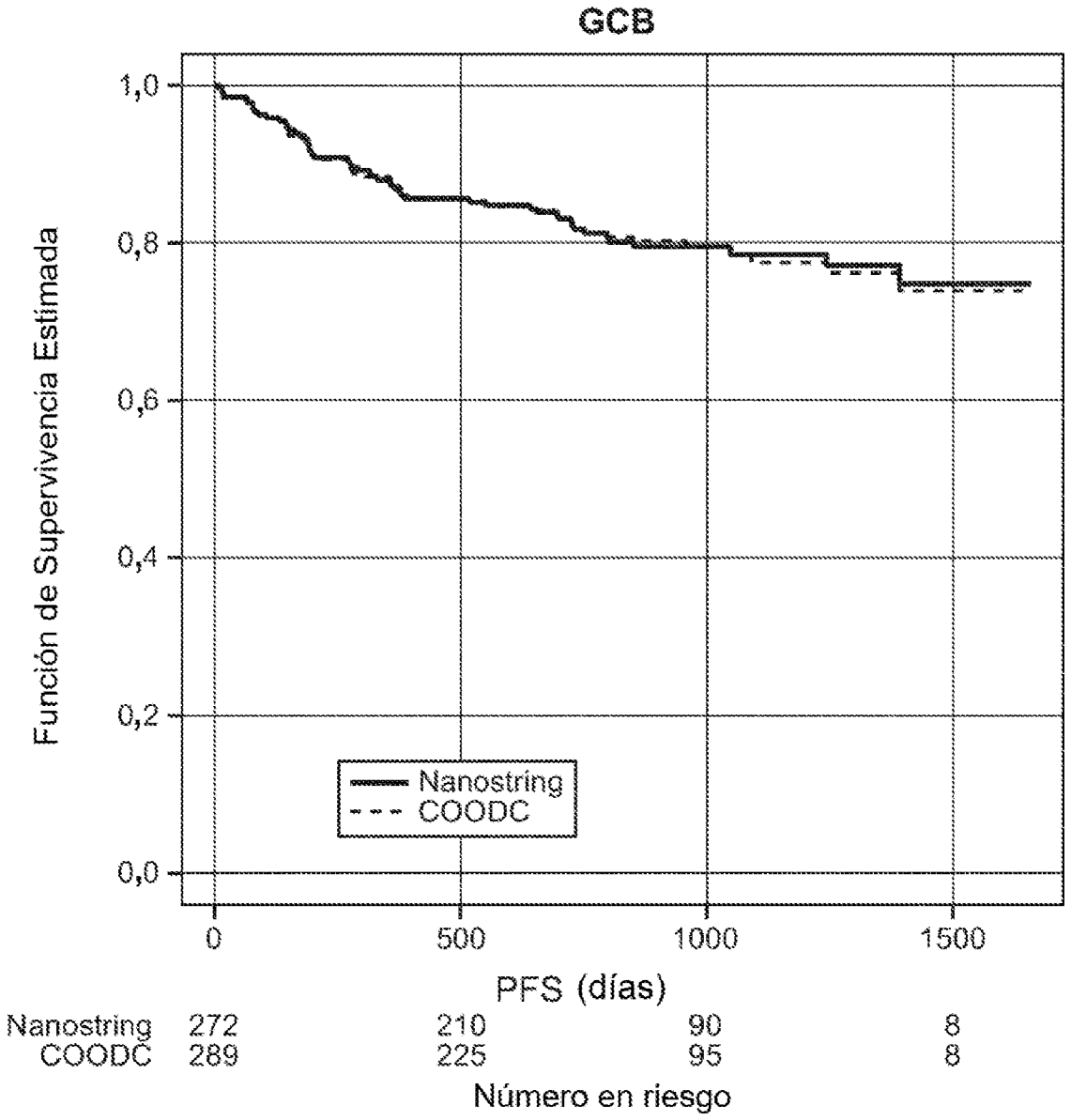

GCB
Función de Supervivencia Estimada
1,0
0,8
0,6
0,4
0,2
0,0
Nanostring
COODC
0
500
1000
1500
PFS (días)
Nanostring 272 210 90 8
COODC 289 225 95 8
Número en riesgo

FIG. 2D

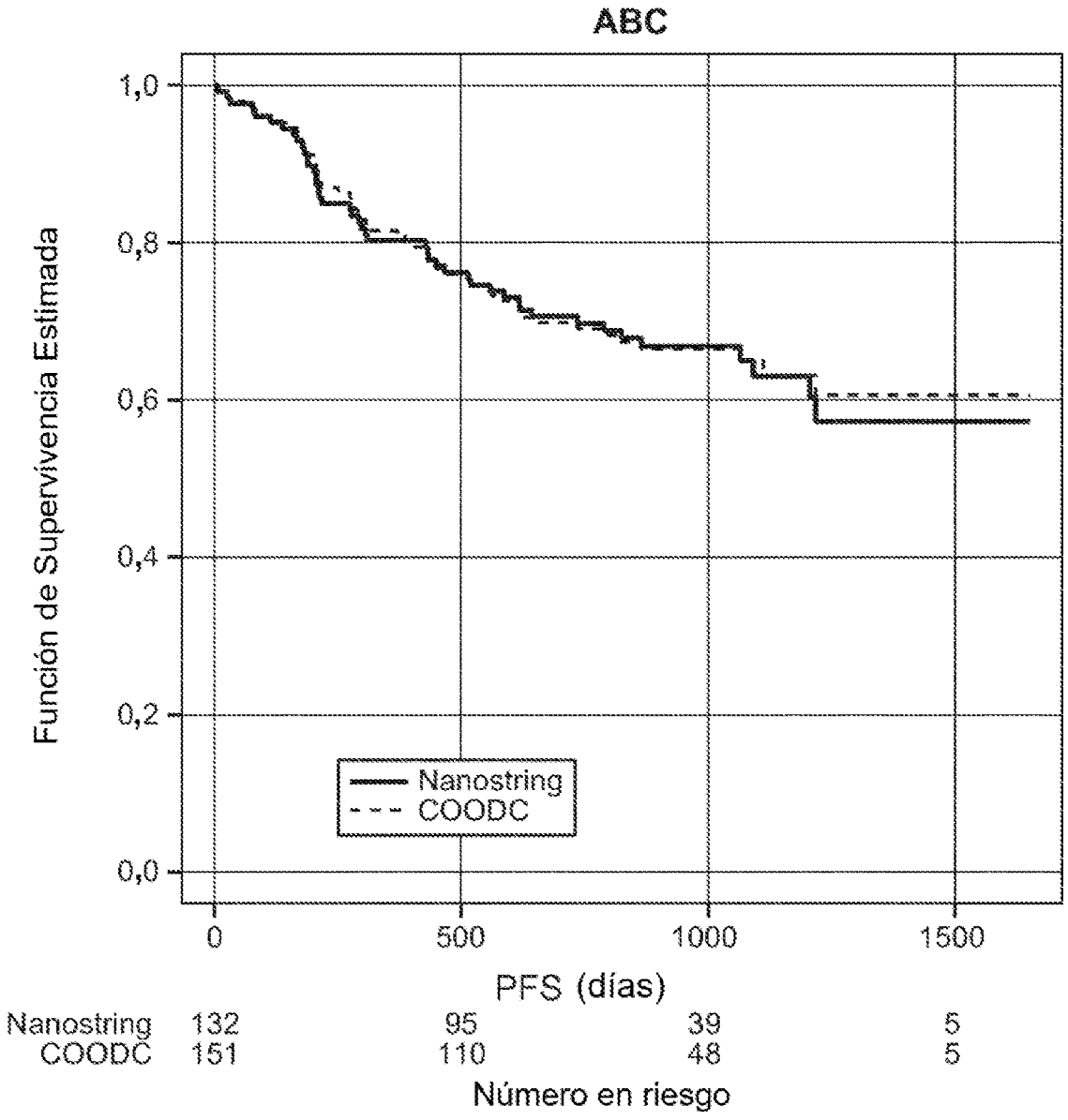

ABC
Función de Supervivencia Estimada
1,0
0,8
0,6
0,4
0,2
0,0
Nanostring
COODC
0
500
1000
1500
PFS (días)
Nanostring 132 95 39 5
COODC 151 110 48 5
Número en riesgo

FIG. 2E

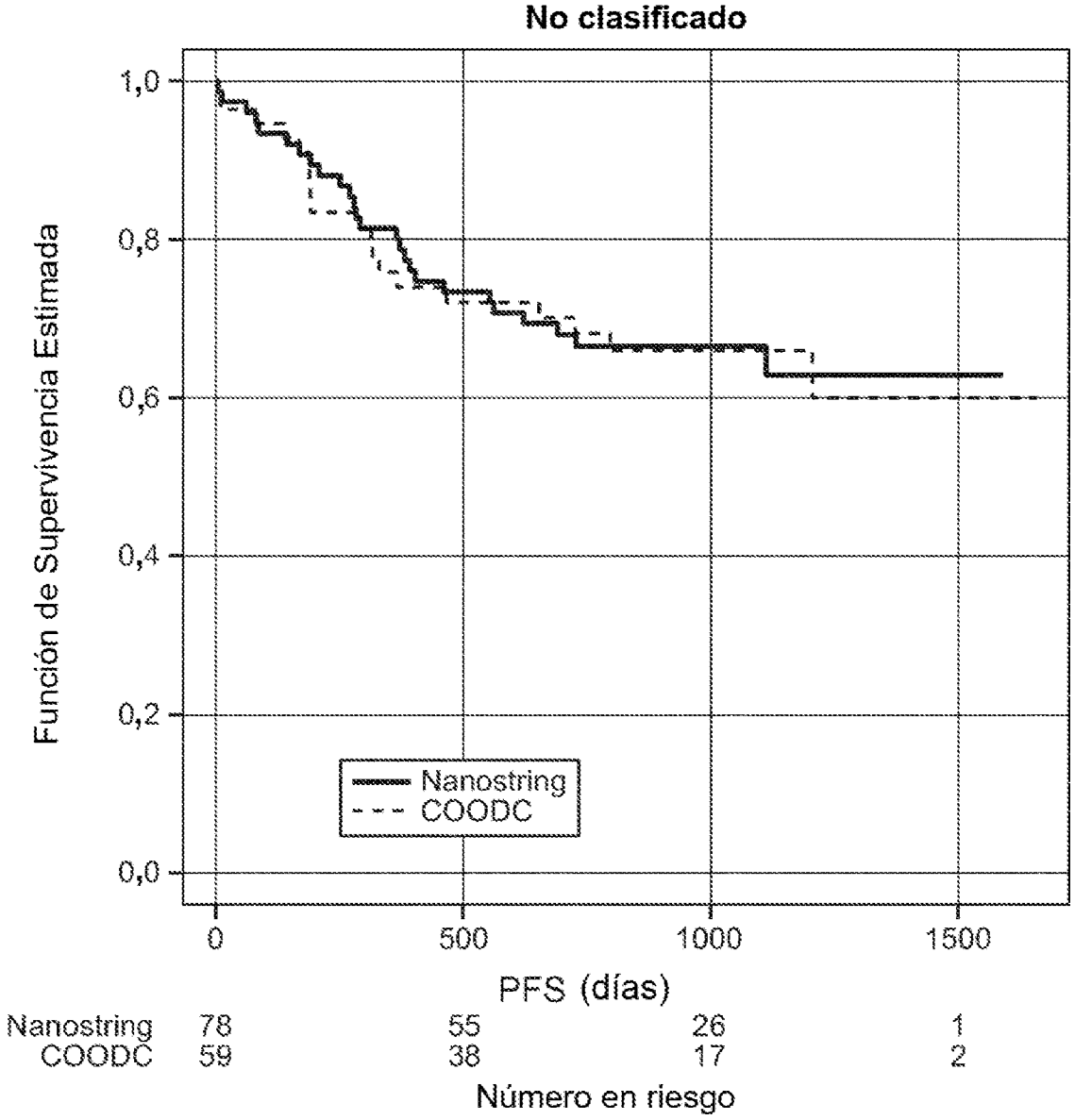
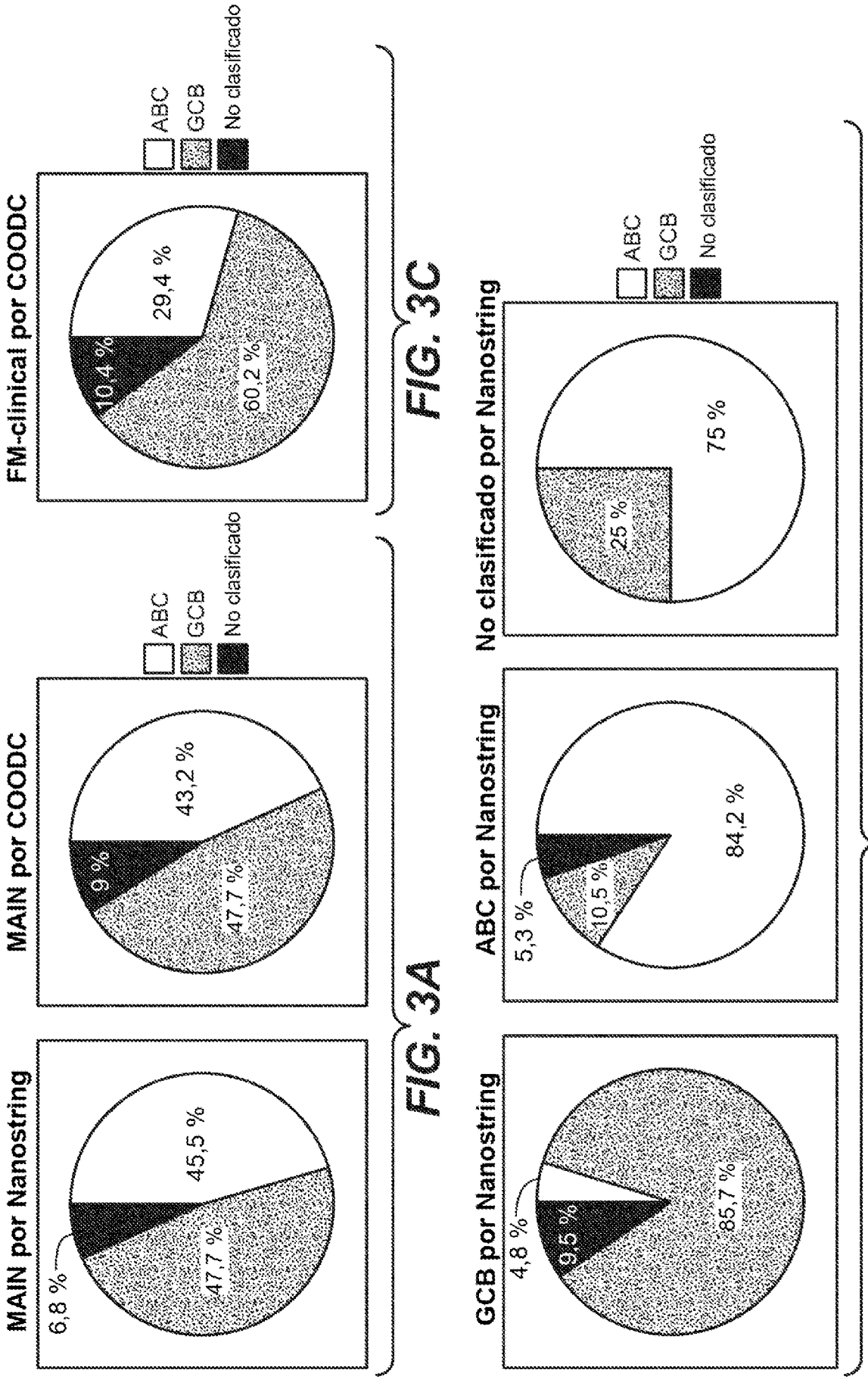

No clasificado
Función de Supervivencia Estimada
1,0
0,8
0,6
0,4
0,2
0,0
Nanostring
COODC
0
500
1000
1500
PFS (días)
Nanostring 78 55 26 1
COODC 59 38 17 2
Número en riesgo

FIG. 2F


MAIN por Nanostring
6,8 %
45,5 %
47,7 %
MAIN por COODC
9 %
43,2 %
47,7 %
ABC
GCB
No clasificado
FIG. 3A
FM-clinical por COODC
10,4 %
29,4 %
60,2 %
ABC
GCB
No clasificado
FIG. 3C
GCB por Nanostring
4,8 %
9,5 %
85,7 %
ABC por Nanostring
5,3 %
10,5 %
84,2 %
No clasificado por Nanostring
25 %
75 %
ABC
GCB
No clasificado
FIG. 3B

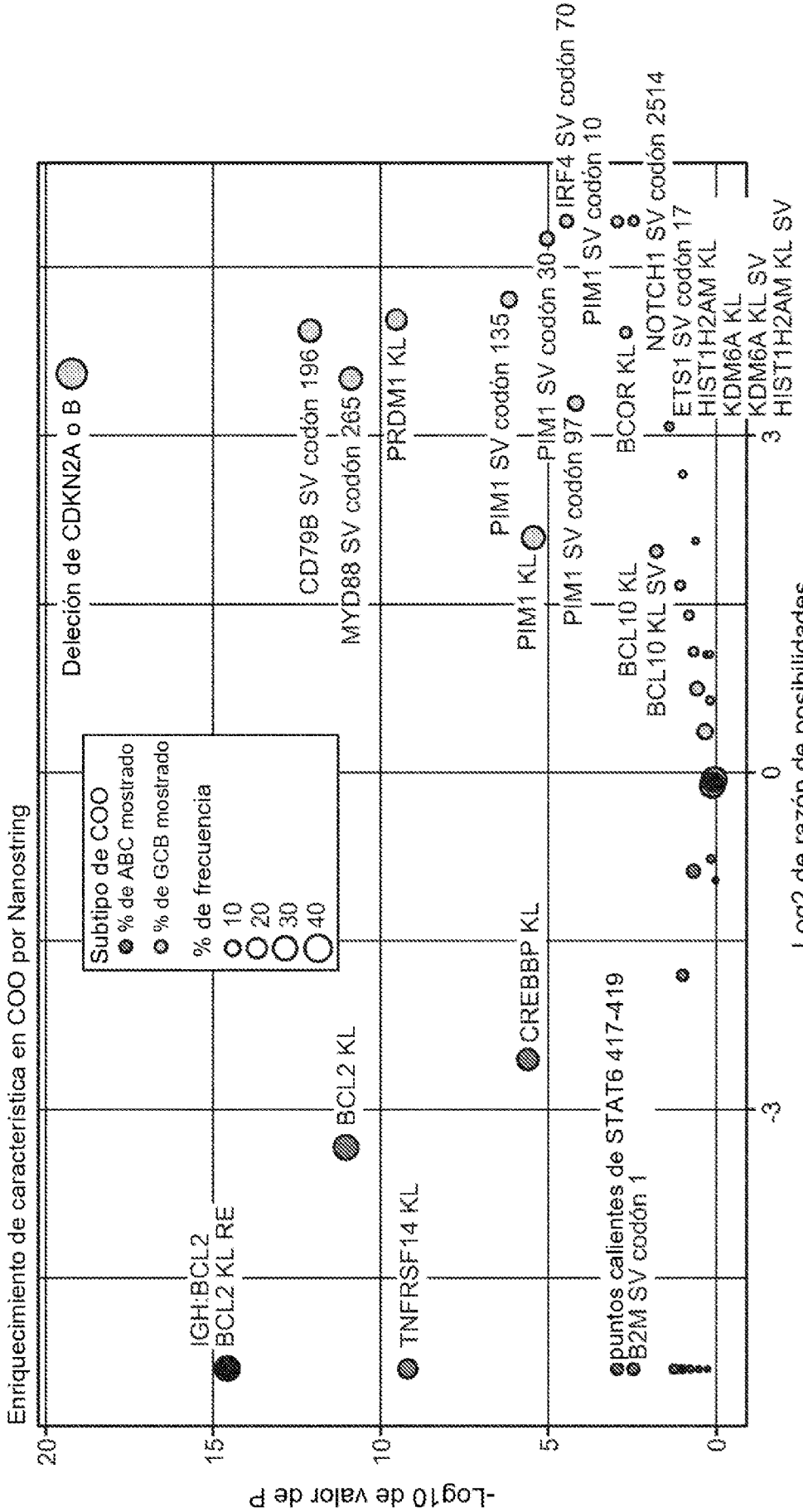

Enriquecimiento de característica en COO por Nanostring
-Log10 de valor de P
Log2 de razón de posibilidades
Subtipo de COO
% de ABC mostrado
% de GCB mostrado
% de frecuencia
10
20
30
40
Deleción de CDKN2A o B
IGH:BCL2
BCL2 KL RE
BCL2 KL
TNFRSF14 KL
puntos calientes de STAT6 417-419
B2M SV codón 1
CREBBP KL
CD79B SV codón 196
MYD88 SV codón 265
PRDM1 KL
PIM1 KL
PIM1 SV codón 135
PIM1 SV codón 97
PIM1 SV codón 30
IRF4 SV codón 70
PIM1 SV codón 10
BCL10 KL
BCL10 KL SV
BCOR KL
NOTCH1 SV codón 2514
ETS1 SV codón 17
HIST1H2AM KL
KDM6A KL
KDM6A KL SV
HIST1H2AM KL SV

FIG. 4A

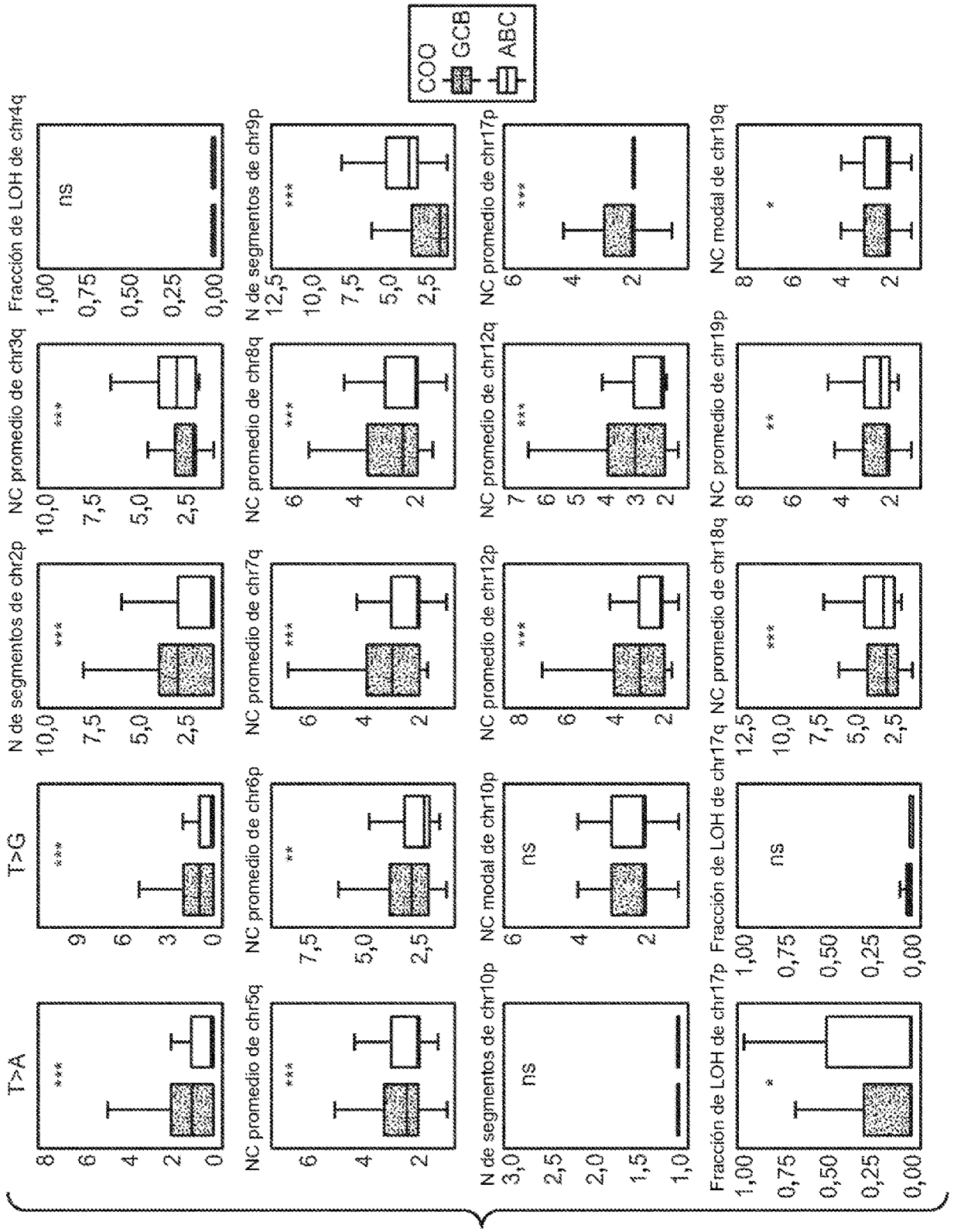

T>A
T>G
N de segmentos de chr2p
NC promedio de chr3q
Fracción de LOH de chr4q
NC promedio de chr5q
NC promedio de chr6p
NC promedio de chr7q
NC promedio de chr8q
N de segmentos de chr9p
N de segmentos de chr10p
NC modal de chr10p
NC promedio de chr12p
NC promedio de chr12q
NC promedio de chr17p
Fracción de LOH de chr17p
Fracción de LOH de chr17q
NC promedio de chr18q
NC promedio de chr19p
NC modal de chr19q
COO
GCB
ABC

FIG. 4B

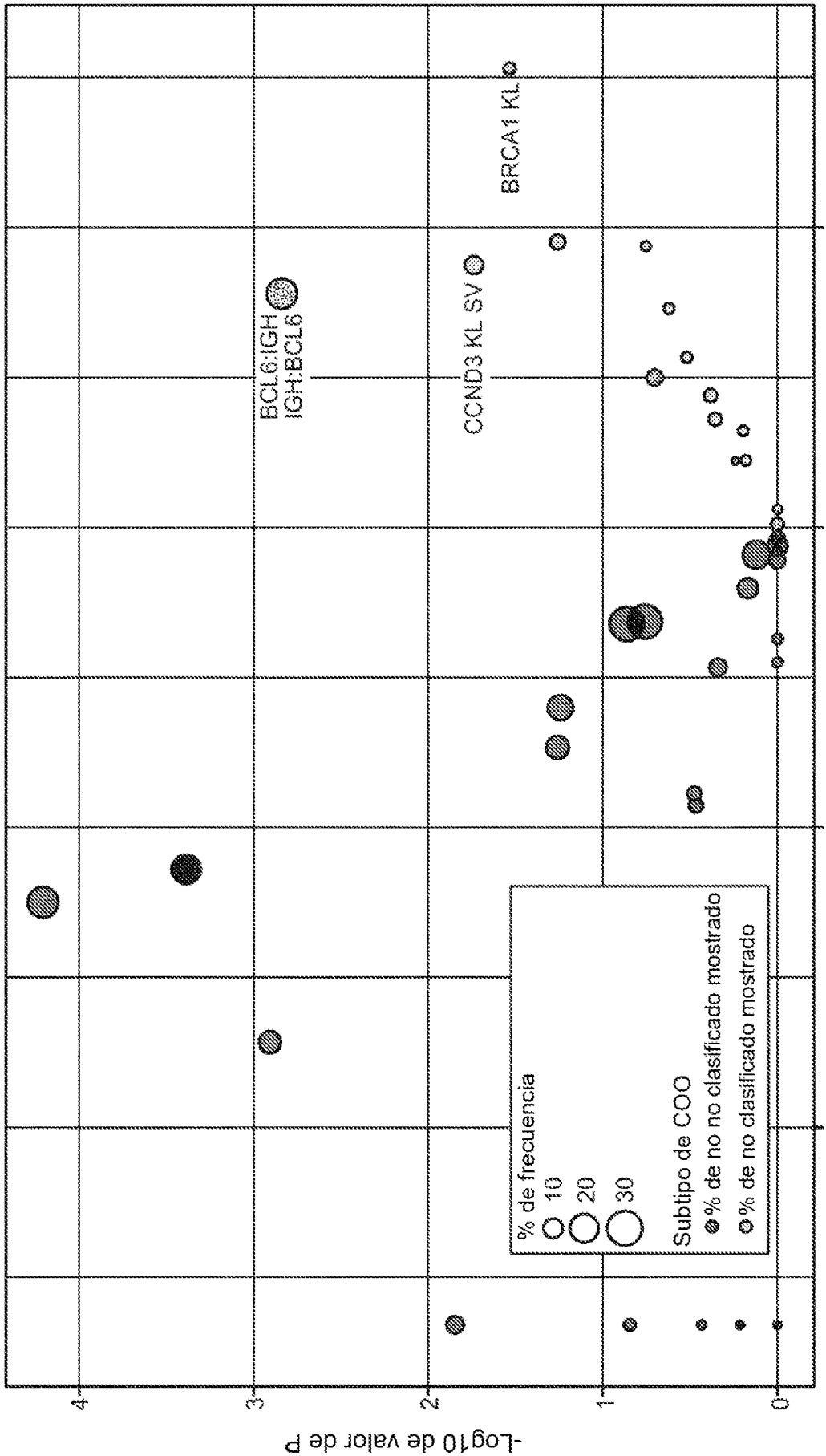

BCL6:IGH
IGH:BCL6
CCND3 KL SV
BRCA1 KL
% de frecuencia
10
20
30
Subtipo de COO
% de no no clasificado mostrado
% de no clasificado mostrado
-Log10 de valor de P
Log2 de razón de posibilidades
0
1
2
3
4
-4
-2
0
2

FIG. 4C

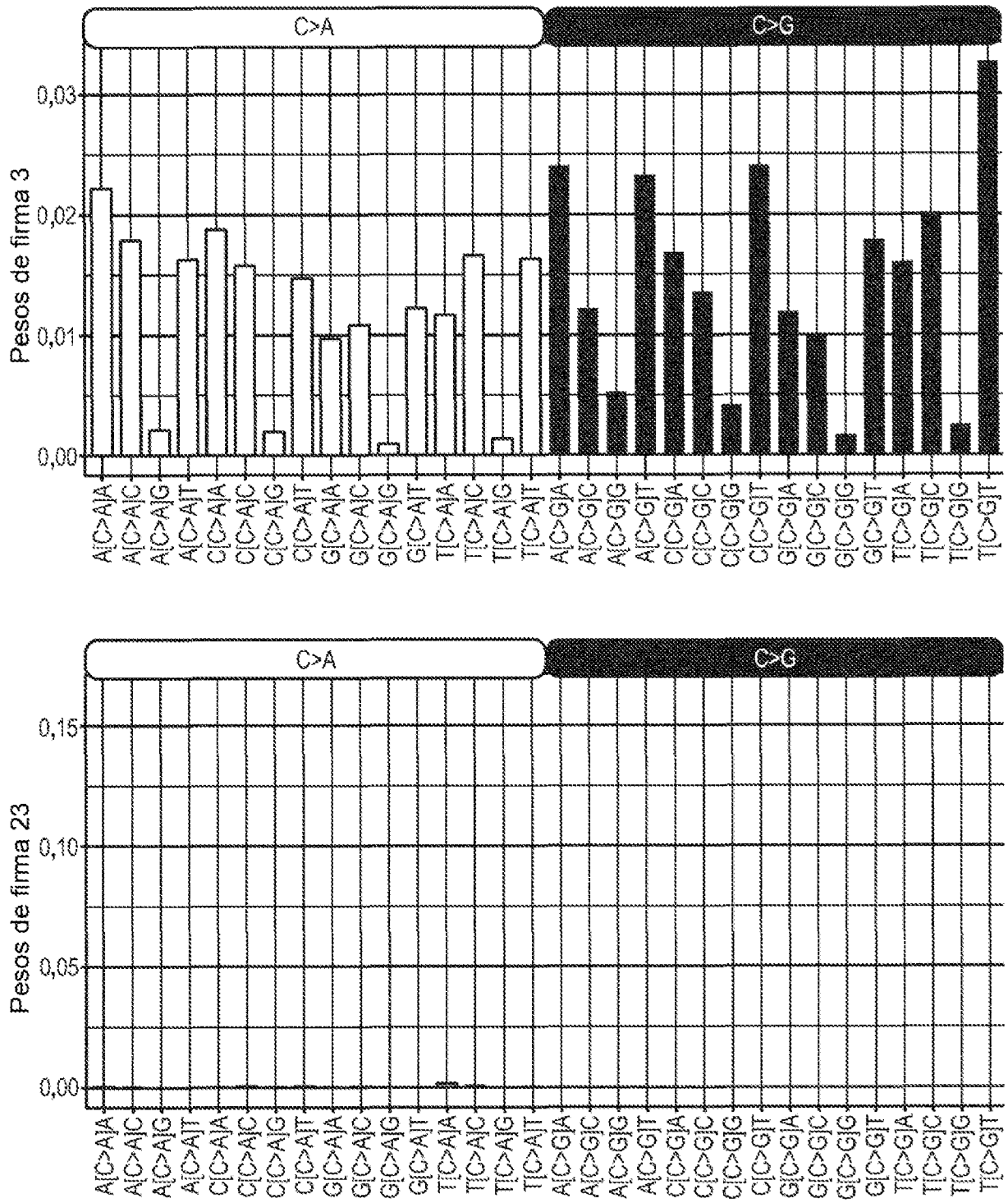


FIG. 5A-1

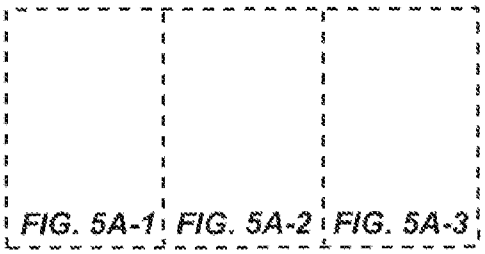


FIG. 5A

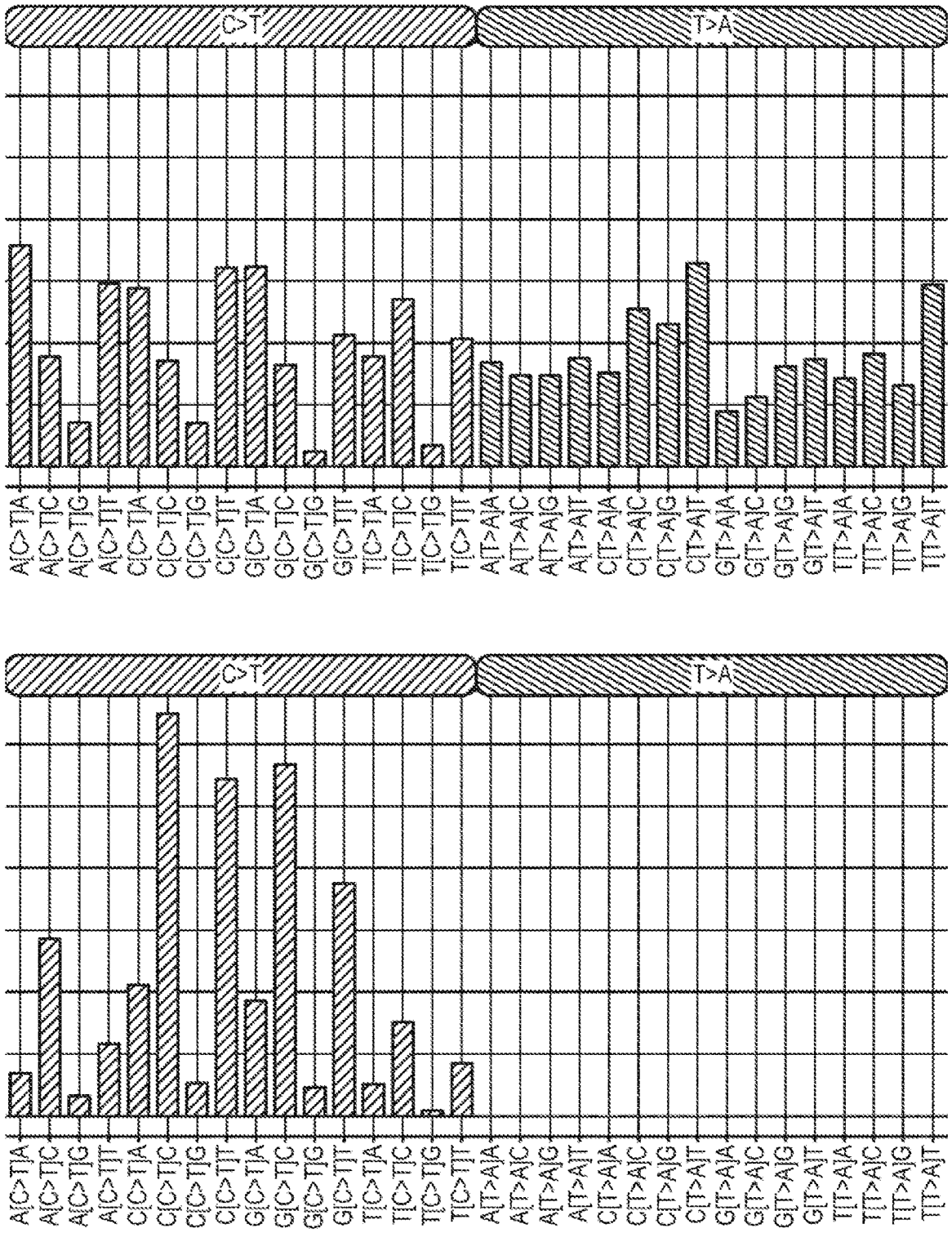

C>T
T>A
A[C>T]A
A[C>T]C
A[C>T]G
A[C>T]T
C[C>T]A
C[C>T]C
C[C>T]G
C[C>T]T
G[C>T]A
G[C>T]C
G[C>T]G
G[C>T]T
T[C>T]A
T[C>T]C
T[C>T]G
T[C>T]T
A[T>A]A
A[T>A]C
A[T>A]G
A[T>A]T
C[T>A]A
C[T>A]C
C[T>A]G
C[T>A]T
G[T>A]A
G[T>A]C
G[T>A]G
G[T>A]T
T[T>A]A
T[T>A]C
T[T>A]G
T[T>A]T

FIG. 5A-2

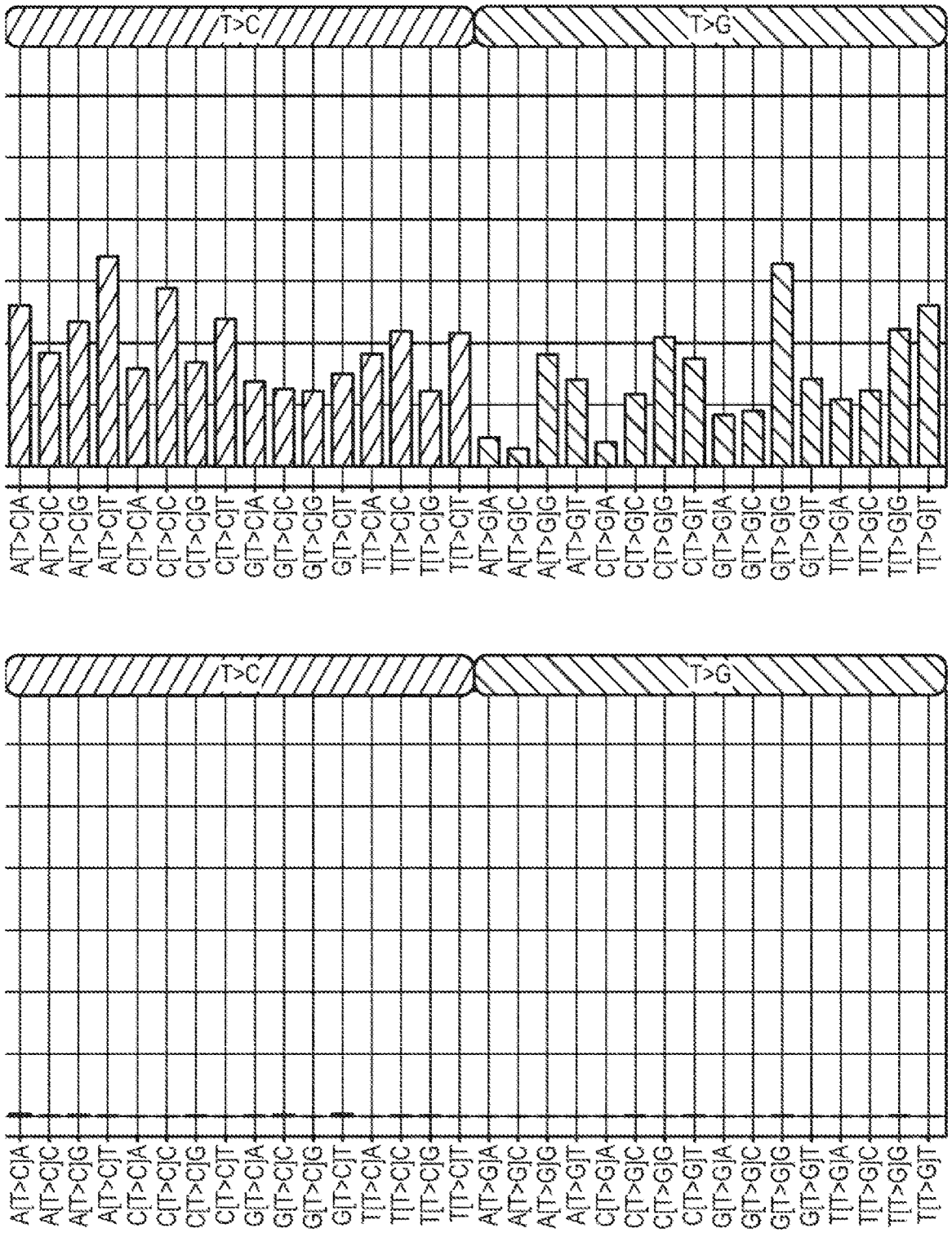

T>C
T>G
A[T>C]A
A[T>C]C
A[T>C]G
A[T>C]T
C[T>C]A
C[T>C]C
C[T>C]G
C[T>C]T
G[T>C]A
G[T>C]C
G[T>C]G
G[T>C]T
T[T>C]A
T[T>C]C
T[T>C]G
T[T>C]T
A[T>G]A
A[T>G]C
A[T>G]G
A[T>G]T
C[T>G]A
C[T>G]C
C[T>G]G
C[T>G]T
G[T>G]A
G[T>G]C
G[T>G]G
G[T>G]T
T[T>G]A
T[T>G]C
T[T>G]G
T[T>G]T

FIG. 5A-3

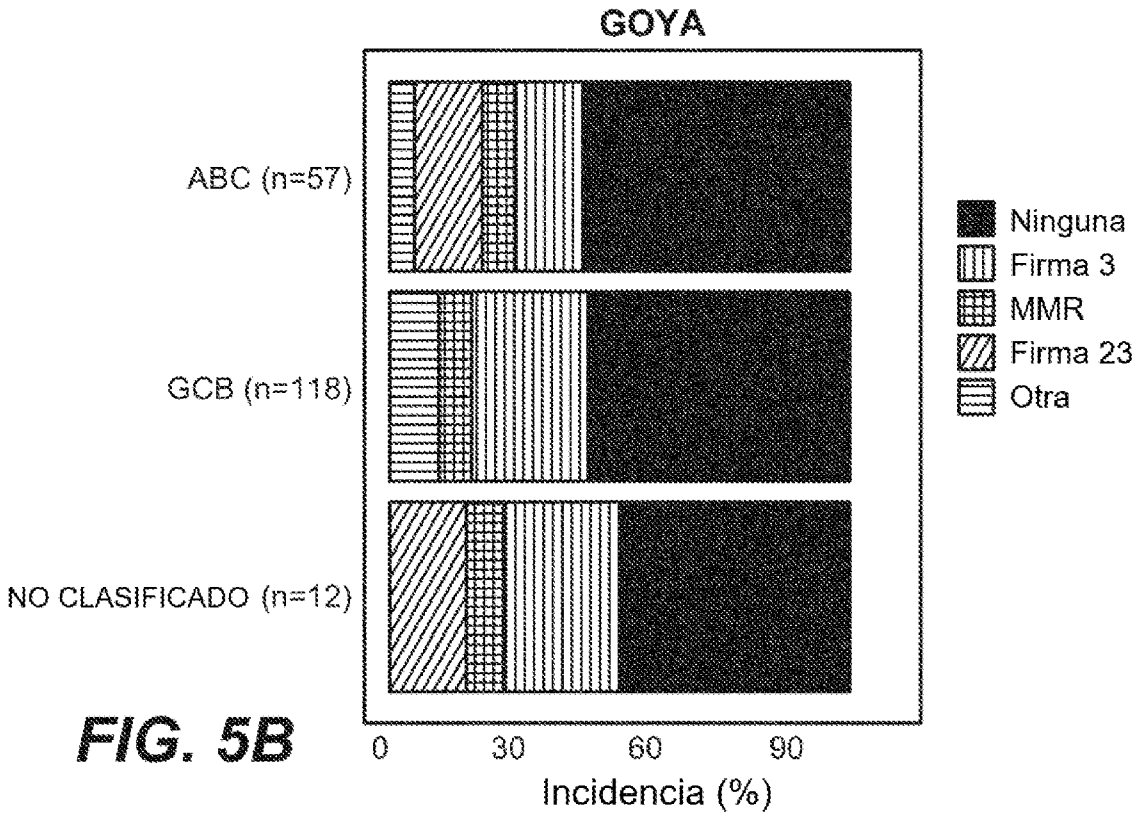

GOYA
ABC (n=57)
GCB (n=118)
NO CLASIFICADO (n=12)
Ninguna
Firma 3
MMR
Firma 23
Otra
0
30
60
90
Incidencia (%)

FIG. 5B

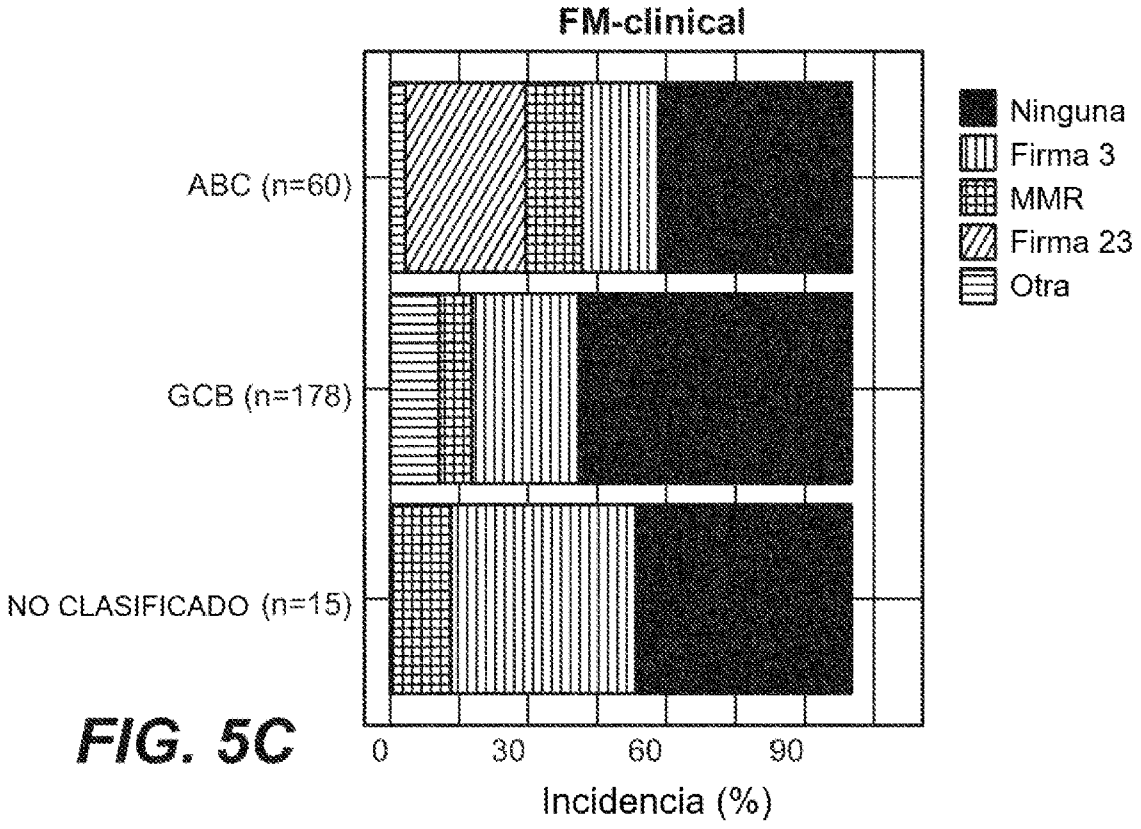
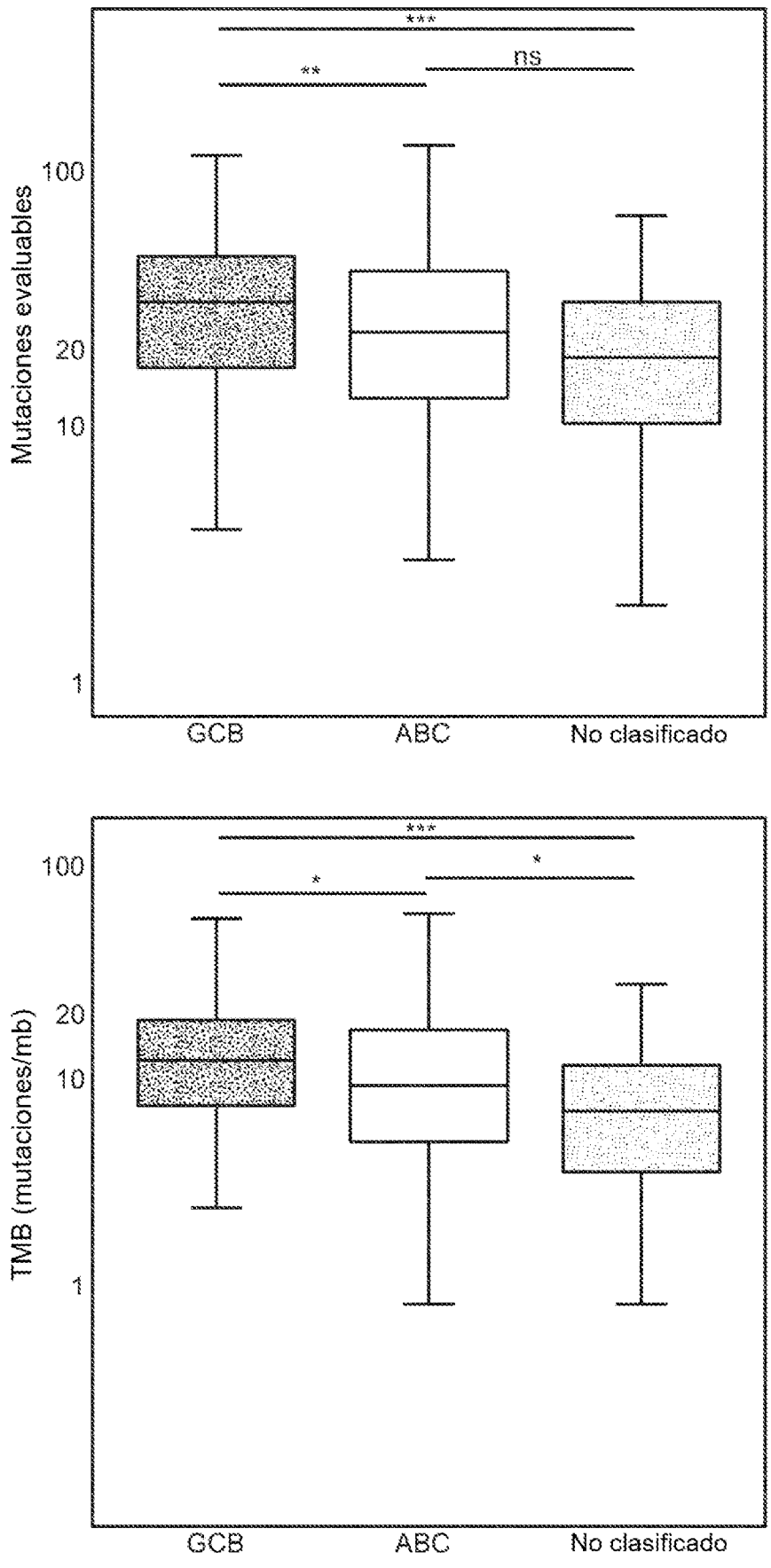

FM-clinical
ABC (n=60)
GCB (n=178)
NO CLASIFICADO (n=15)
Ninguna
Firma 3
MMR
Firma 23
Otra
0
30
60
90
Incidencia (%)

FIG. 5C

FIG. 5D


***
**
ns
Mutaciones evaluables
100
20
10
1
GCB
ABC
No clasificado
***
*
*
TMB (mutaciones/mb)
100
20
10
1
GCB
ABC
No clasificado

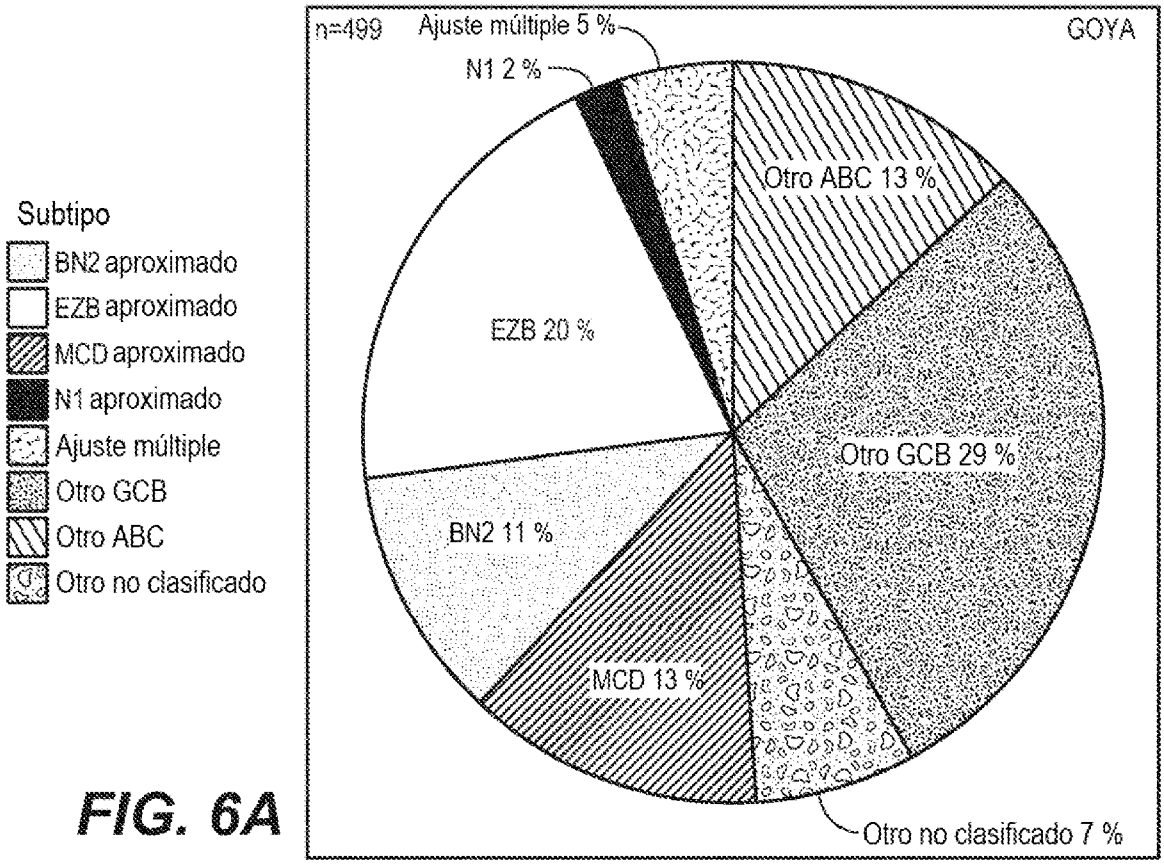

n=499
Ajuste múltiple 5 %
GOYA
N1 2 %
Otro ABC 13 %
EZB 20 %
Otro GCB 29 %
BN2 11 %
MCD 13 %
Otro no clasificado 7 %
Subtipo
BN2 aproximado
EZB aproximado
MCD aproximado
N1 aproximado
Ajuste múltiple
Otro GCB
Otro ABC
Otro no clasificado

FIG. 6A

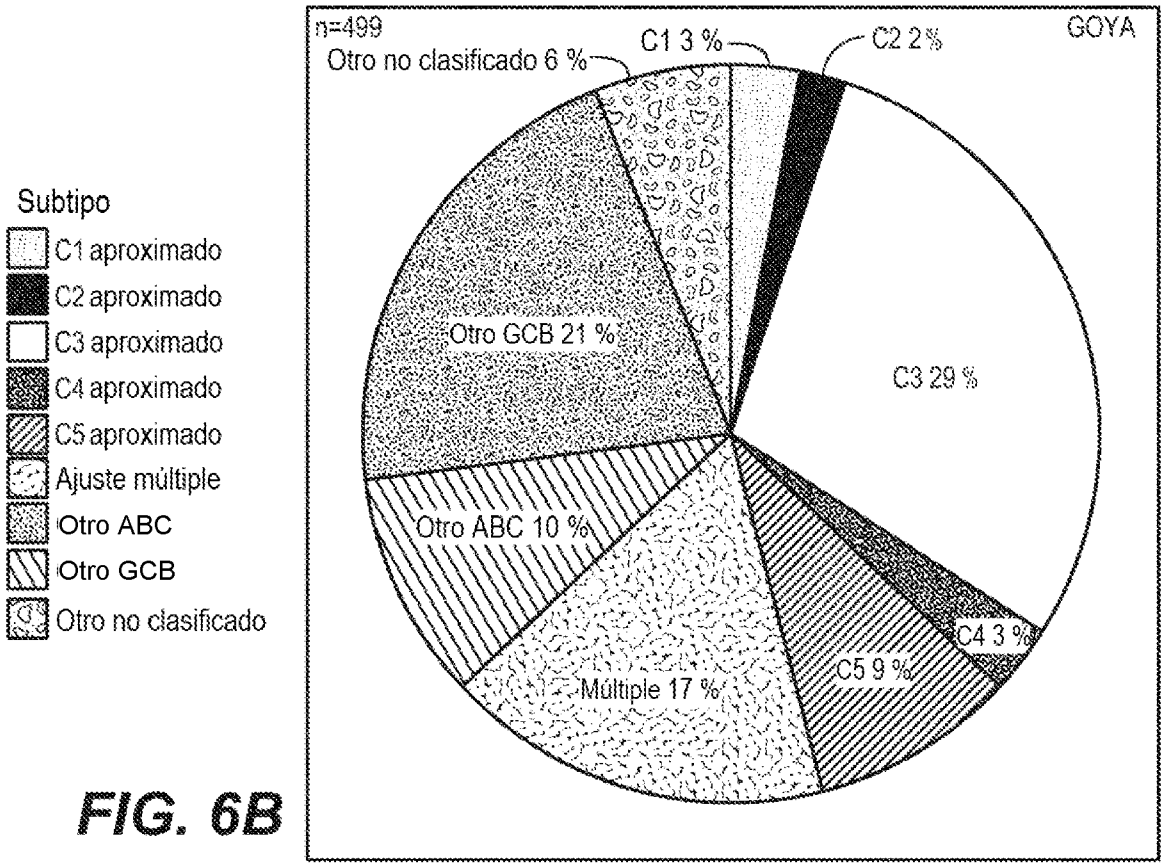

n=499
GOYA
Otro no clasificado 6 %
C1 3 %
C2 2 %
Otro GCB 21 %
C3 29 %
Otro ABC 10 %
C4 3 %
C5 9 %
Múltiple 17 %
Subtipo
C1 aproximado
C2 aproximado
C3 aproximado
C4 aproximado
C5 aproximado
Ajuste múltiple
Otro ABC
Otro GCB
Otro no clasificado

FIG. 6B

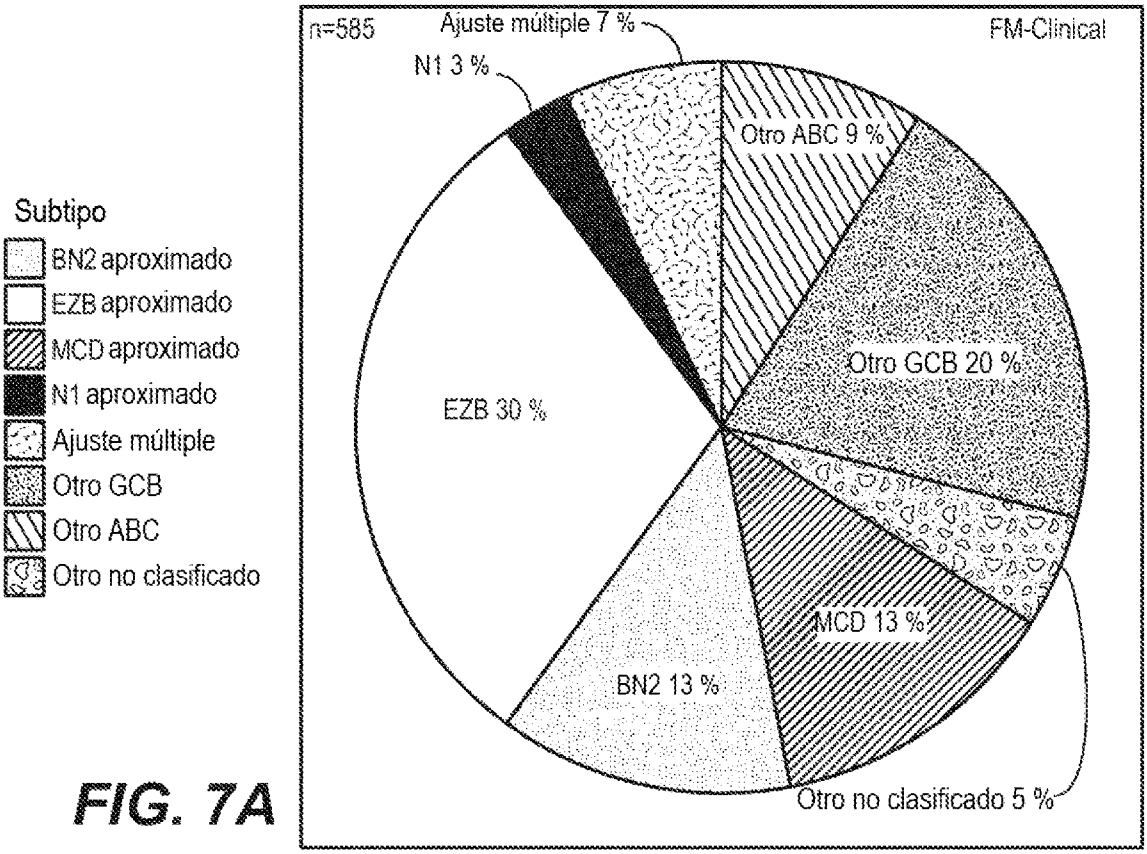

Subtipo
BN2 aproximado
EZB aproximado
MCD aproximado
N1 aproximado
Ajuste múltiple
Otro GCB
Otro ABC
Otro no clasificado
n=585
Ajuste múltiple 7 %
FM-Clinical
N1 3 %
Otro ABC 9 %
Otro GCB 20 %
EZB 30 %
MCD 13 %
BN2 13 %
Otro no clasificado 5 %

FIG. 7A

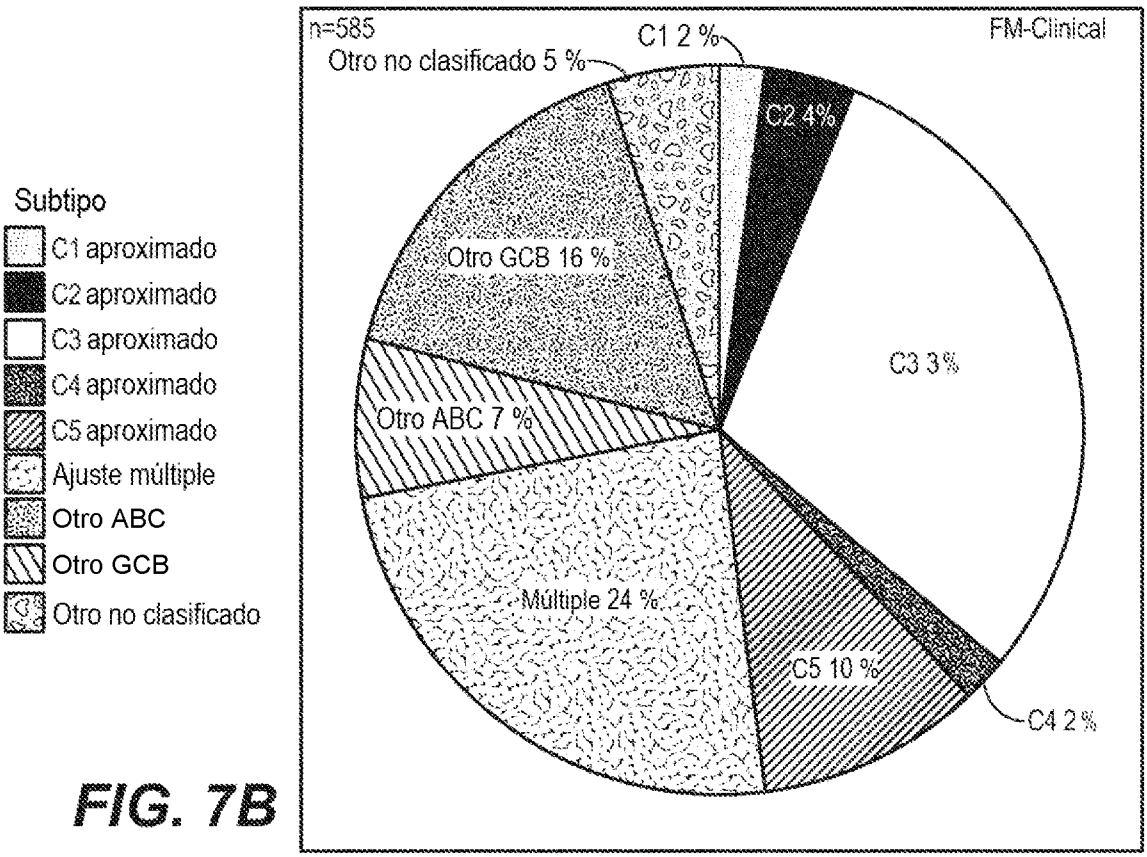

Subtipo
C1 aproximado
C2 aproximado
C3 aproximado
C4 aproximado
C5 aproximado
Ajuste múltiple
Otro ABC
Otro GCB
Otro no clasificado
n=585
C1 2 %
FM-Clinical
Otro no clasificado 5 %
C2 4%
Otro GCB 16 %
C3 3 %
Otro ABC 7 %
Múltiple 24 %
C5 10 %
C4 2 %

FIG. 7B