

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 792 227**

51 Int. Cl.:

C12Q 1/6883 (2008.01)

G01N 33/68 (2006.01)

G16H 50/20 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **03.11.2014 PCT/US2014/063714**

87 Fecha y número de publicación internacional: **31.03.2016 WO16048388**

96 Fecha de presentación y número de la solicitud europea: **03.11.2014 E 14802991 (1)**

97 Fecha y número de publicación de la concesión europea: **22.04.2020 EP 3198023**

54 Título: **Predicción de evento de riesgo cardiovascular y usos de la misma**

30 Prioridad:

26.09.2014 US 201462055984 P

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

10.11.2020

73 Titular/es:

**SOMALOGIC, INC. (100.0%)
2945 Wilderness Place
Boulder, CO 80301, US**

72 Inventor/es:

**STERLING, DAVID;
KATO, SHINTARO;
BRODY, EDWARD N. y
WILLIAMS, STEPHEN A.**

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 792 227 T3

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

DESCRIPCIÓN

Predicción de evento de riesgo cardiovascular y usos de la misma

5 **Campo de la invención**

La presente solicitud se refiere, generalmente, a la detección de biomarcadores y a un método de evaluación del riesgo de un evento cardiovascular futuro en un individuo y, más específicamente, a métodos usados para evaluar a un individuo para realizar predicción del riesgo de desarrollar un evento cardiovascular (CV) en un período de 1 a 5 años. Tales eventos incluyen, pero sin limitaciones, infarto de miocardio, ictus, insuficiencia cardíaca congestiva o muerte.

Antecedentes

15 La siguiente descripción proporciona un resumen de información relevante para la presente solicitud y no es una admisión que cualquier información proporcionada o publicaciones a las que se hace referencia en el presente documento sea técnica anterior de la presente solicitud.

20 La enfermedad cardiovascular es la principal causa de muerte en EE.UU. Hay numerosos e importantes predictores del riesgo de eventos primarios (D'Agostino, R et al., "General Cardiovascular Risk Profile for Use in Primary Care: The Framingham Heart Study" *Circulation* 117:743-53 (2008); y Ridker, P. et al., "Development and Validation of Improved Algorithms for the Assessment of Global Cardiovascular Risk in Women" *JAMA* 297(6):611-619 (2007)) y eventos secundarios (Shlipak, M. et al. "Biomarkers to Predict Recurrent Cardiovascular Disease: The Heart & Soul Study" *Am. J. Med.* 121:50-57 (2008)) que se usan ampliamente en la práctica clínica y ensayos terapéuticos. Desafortunadamente, las curvas características de funcionamiento del receptor (ROC), los cocientes de riesgos instantáneos y la concordancia muestran que el rendimiento de los factores de riesgo existentes y los biomarcadores es modesto (AUC de ~0,75 significan que estos factores están únicamente a medio camino entre echar una moneda al aire y la perfección). Además de la necesidad de mejorar el rendimiento diagnóstico, existe la necesidad de un producto de riesgo que responda tanto a corto plazo como personalmente a en los individuos a las intervenciones beneficiosas (y destructivas) y a los cambios en el estilo de vida. La ecuación de Framingham usada habitualmente tiene tres problemas principales. En primer lugar, es a demasiado largo plazo: da cálculos de riesgo a 10 años, pero los seres humanos no tienen en cuenta los riesgos futuros y son reacios a realizar modificaciones de comportamiento y estilo de vida basadas en ellos. En segundo lugar, no responde muy bien a las intervenciones: depende mucho de la edad cronológica, que no puede descender; y del sexo, que no se puede cambiar. En tercer lugar, dentro de la población de alto riesgo prevista en el presente documento, los factores de Framingham no logran discriminar bien entre alto y bajo riesgo: el cociente de riesgos instantáneos entre los cuartiles alto y bajo es solo 2, y cuando se intenta usar las puntuaciones de Framingham para personalizar el riesgo estratificando a los sujetos en capas más finas (deciles, por ejemplo), las tasas de eventos observadas son similares para muchos de los deciles.

40 Los factores de riesgo para la enfermedad cardiovascular se utilizan ampliamente para impulsar la intensidad y la naturaleza del tratamiento médico y su uso ha contribuido sin duda a la reducción de la morbilidad y mortalidad cardiovascular que se ha observado en las últimas dos décadas. Estos factores se han combinado rutinariamente en algoritmos, pero desafortunadamente no capturan todo el riesgo (la presentación inicial más común para la enfermedad cardíaca sigue siendo la muerte). De hecho, probablemente solo capturan la mitad del riesgo. Un área bajo la curva ROC de ~0,76 es típica para tales factores de riesgo en prevención primaria, con un rendimiento mucho peor en la prevención secundaria (0,62 es típico), solo representa de aproximadamente a un cuarto a la mitad del rendimiento entre un lanzamiento de moneda a 0,5 y la perfección a 1,0.

50 La adición de nuevos biomarcadores a las puntuaciones de riesgo clínico ha sido decepcionante. Por ejemplo, en el estudio de Framingham (Wang et al., "Multiple Biomarkers for the Prediction of First Major Cardiovascular Events and Death" *N. Eng. J. Med.* 355:2631-2637 (2006)) en 3.209 personas, la adición de 10 biomarcadores (CRP, BNP, NT-proBNP, aldosterona, renina, fibrinógeno, dímero D, Inhibidor del activador de plasminógeno de tipo 1, homocisteína y la proporción entre albúmina urinaria a creatinina), no mejoró significativamente la AUC cuando se añadió a los factores de riesgo existentes: el AUC para eventos de 0-5 años fue de 0,76 con la edad, el sexo y los factores de riesgo convencionales y 0,77 con la mejor combinación de biomarcadores añadidos a la mezcla, y para la prevención secundaria, la situación es peor.

60 La identificación temprana de pacientes con mayor riesgo de un evento cardiovascular dentro de un período de 1-5 años es importante porque un tratamiento más agresivo de los individuos con riesgo elevado puede mejorar el resultado. Por lo tanto, el manejo óptimo requiere una intervención agresiva para reducir el riesgo de un evento cardiovascular en los pacientes que se considera que tienen un riesgo mayor, mientras que los pacientes con menor riesgo de sufrir un evento cardiovascular pueden evitar tratamientos costosos y potencialmente invasivos, que probablemente no tengan ningún efecto beneficioso para el paciente.

65 La selección de biomarcadores para la predicción del riesgo de tener un estado de enfermedad o afección específica dentro de un período de tiempo definido implica primero la identificación de marcadores que tienen una relación

5 mensurable y estadísticamente significativa con la probabilidad y/o el momento de un evento para una aplicación médica específica. Los biomarcadores pueden incluir moléculas secretadas o desprendidas que están en la ruta causal a la afección de interés, o que están aguas abajo o paralelas al desarrollo o progresión de la enfermedad o afección, o ambas. Se liberan en el torrente sanguíneo desde el tejido cardiovascular o desde otros órganos y tejidos circundantes y las células circulantes en respuesta a los procesos biológicos que predisponen a un evento cardiovascular o pueden reflejar los efectos posteriores de la fisiopatología, tal como una disminución en la función renal. Los biomarcadores pueden incluir moléculas pequeñas, péptidos, proteínas y ácidos nucleicos. Algunos de los problemas clave que afectan la identificación de biomarcadores incluyen el exceso de ajuste de los datos disponibles y el sesgo en los datos.

10 Se han utilizado diversos procedimientos en un intento de identificar biomarcadores y diagnosticar o predecir el riesgo de sufrir una enfermedad o afección. Para los marcadores basados en proteínas, estos incluyen electroforesis bidimensional, espectrometría de masas y métodos de inmunoensayo. Para los marcadores de ácido nucleico, estos incluyen perfiles de expresión de ARNm, perfiles de microARN, FISH, análisis en serie de la expresión génica (SAGE), matrices de expresión génica a gran escala, secuenciación génica y genotipado (SNP o análisis de variantes pequeñas).

20 La utilidad de la electroforesis bidimensional está limitada por la baja sensibilidad de detección; los problemas con la solubilidad, la carga y la hidrofobicidad de las proteínas; reproducibilidad en gel; y la posibilidad de que una sola mancha represente varias proteínas. Para la espectrometría de masas, dependiendo del formato utilizado, las limitaciones giran en torno al procesamiento y la separación de la muestra, la sensibilidad a las proteínas poco abundantes, las consideraciones de señal a ruido y la incapacidad de identificar inmediatamente la proteína detectada. Las limitaciones en los enfoques de inmunoensayo para el descubrimiento de biomarcadores se centran en la incapacidad de los ensayos multiplexados basados en anticuerpos para medir una gran cantidad de analitos. SE podría simplemente imprimir una matriz de anticuerpos de alta calidad y, sin sándwiches, medir los analitos unidos a esos anticuerpos. (Este sería el equivalente formal del uso de un genoma completo de secuencias de ácido nucleico para medir por hibridación todas las secuencias de ADN o ARN en un organismo o una célula. El experimento de hibridación funciona porque la hibridación puede ser una prueba estricta de identidad). Sin embargo, incluso los anticuerpos muy buenos no son lo suficientemente rigurosos al seleccionar sus compañeros de unión para funcionar en el contexto de sangre o incluso extractos celulares porque el conjunto de proteínas en esas matrices tiene abundancias ampliamente variables, lo que puede conducir a una mala relación señal/ruido. Por lo tanto, se debe usar un enfoque diferente con los enfoques basados en inmunoensayos para el descubrimiento de biomarcadores, sería necesario usar ensayos ELISA multiplexados (es decir, sándwiches) para obtener la suficiente rigurosidad para medir muchos analitos simultáneamente para decidir qué analitos son realmente biomarcadores. Los inmunoensayos de tipo sándwich no se adaptan a un alto contenido y, por lo tanto, no es posible el descubrimiento de biomarcadores utilizando inmunoensayos de tipo sándwich rigurosos utilizando formatos de matriz estándar. Por último, los reactivos de anticuerpos están sujetos a una gran variabilidad del lote e inestabilidad del reactivo. La presente plataforma para el descubrimiento de biomarcadores de proteínas supera este problema.

40 Muchos de estos procedimientos dependen o requieren algún tipo de fraccionamiento de la muestra antes del análisis. Por lo tanto, la preparación de la muestra requerida para ejecutar un estudio suficientemente potenciado diseñado para identificar y descubrir biomarcadores estadísticamente relevantes en una serie de poblaciones de muestra bien definidas es extremadamente difícil, costoso y lento. Durante el fraccionamiento, se puede introducir un amplio rango de variabilidad en las diversas muestras. Por ejemplo, un marcador potencial podría ser inestable para el proceso, la concentración del marcador podría cambiarse, podría producirse una agregación o desagregación inapropiada y podría ocurrir una contaminación involuntaria de la muestra que enmascararía los cambios sutiles anticipados en el estadio inicial de la enfermedad.

50 Está ampliamente aceptado que los procedimientos de descubrimiento y detección de biomarcadores que usan estas tecnologías tienen serias limitaciones para la identificación de biomarcadores diagnósticos o predictivos. Estas limitaciones incluyen la incapacidad de detectar biomarcadores poco abundantes, la incapacidad de cubrir sistemáticamente todo el rango dinámico del proteoma, la irreproducibilidad en el procesamiento y fraccionamiento de muestras, y la irreproducibilidad general y la falta de solidez del procedimiento. Además, estos estudios han introducido sesgos en los datos y no han abordado adecuadamente la complejidad de las poblaciones de muestra, incluidos los controles apropiados, en términos de distribución y aleatorización requeridas para identificar y validar biomarcadores dentro de una población de enfermedad objetivo.

60 Aunque durante varias décadas han continuado los esfuerzos destinados al descubrimiento de biomarcadores nuevos y efectivos, los esfuerzos han sido en gran medida infructuosos. Los biomarcadores para diversas enfermedades generalmente se han identificado en laboratorios académicos, normalmente a través de un descubrimiento accidental mientras se hacía investigación básica sobre algún proceso de enfermedad. En función del descubrimiento y con pequeñas cantidades de datos clínicos, se publicaron documentos que sugerían la identificación de un nuevo biomarcador. La mayoría de estos biomarcadores propuestos, sin embargo, no se han confirmado como biomarcadores reales o útiles, principalmente porque el pequeño número de muestras clínicas analizadas proporciona solo pruebas estadísticas débiles de que se ha encontrado un biomarcador efectivo. Esto es, la identificación inicial no fue rigurosa con respecto a los elementos básicos de las estadísticas. En cada uno de los

años 1994 a 2003, una búsqueda de la literatura científica muestra que se publicaron miles de referencias en relación con biomarcadores. Durante ese mismo período de tiempo, sin embargo, la FDA aprobó para el uso diagnóstico, como máximo, tres nuevos biomarcadores de proteínas al año y, en varios años, no se aprobaron nuevos biomarcadores de proteínas.

5 Conforme a la historia de los esfuerzos fallidos de descubrimiento de biomarcadores, Se han propuesto teorías que promueven aún más la comprensión general de que los biomarcadores para el diagnóstico, el pronóstico o la predicción del riesgo de desarrollar enfermedades y afecciones son poco frecuentes y difíciles de encontrar. La investigación de biomarcadores basada en geles 2D o espectrometría de masas respalda estas ideas. A través de
10 estos enfoques se han identificado muy pocos biomarcadores útiles. Sin embargo, generalmente se pasa por alto que el gel 2D y la espectrometría de masas miden las proteínas que están presentes en la sangre a concentraciones de aproximadamente 1 nM y superiores, y que este conjunto de proteínas puede ser el menos propenso a cambiar con la enfermedad o el desarrollo de una afección concreta. Además de la presente plataforma de descubrimiento de biomarcadores, no existen plataformas de descubrimiento de biomarcadores proteómicos que puedan medir con
15 precisión los niveles de expresión de proteínas a concentraciones mucho más bajas.

Se sabe mucho sobre las rutas bioquímicas para la compleja biología humana. Muchas rutas bioquímicas culminan o son iniciadas por proteínas secretadas que funcionan localmente dentro de la patología; por ejemplo, los factores de crecimiento se secretan para estimular la replicación de otras células en la patología y se secretan otros factores
20 para proteger al sistema inmunitario, y así sucesivamente. Si bien muchas de estas proteínas secretadas funcionan de manera paracrina, algunas operan distalmente en el cuerpo. Un experto en la técnica con una comprensión básica de las rutas bioquímicas comprendería que muchas proteínas específicas de la patología deberían existir en la sangre en concentraciones inferiores (incluso muy por debajo) a los límites de detección de los geles 2D y la espectrometría de masas. Lo que debe preceder a la identificación de este número relativamente abundante de
25 biomarcadores de enfermedades es una plataforma proteómica que puede analizar proteínas en concentraciones inferiores a las detectables mediante geles 2D o espectrometría de masas.

Como se ha analizado anteriormente, los eventos cardiovasculares pueden prevenirse mediante un tratamiento agresivo si la propensión a tales eventos se puede determinar con precisión y dirigiendo tales intervenciones a las
30 personas que más las necesitan y/o lejos de las personas que menos las necesitan, la eficiencia de los recursos médicos puede mejorarse y los costes pueden reducirse al mismo tiempo. Además, cuando el paciente tiene el conocimiento de información precisa y a corto plazo sobre su probabilidad personal de eventos cardiovasculares, esto es menos negable que la información basada en la población a largo plazo y conducirá a mejores opciones de estilo de vida y un mejor cumplimiento de la medicación, lo que se sumará a los beneficios. Las pruebas de
35 marcadores múltiples existentes requieren la obtención de varias muestras de un individuo o requieren que una muestra se reparta entre múltiples ensayos. De manera óptima, una prueba mejorada requeriría solo una sola muestra de sangre, orina o de otro tipo, y un solo ensayo. Por consiguiente, existe la necesidad de biomarcadores, métodos, dispositivos, reactivos, sistemas y kits que permitan la predicción de eventos cardiovasculares dentro de un período de 5 años. El documento WO2013/049674 describe biomarcadores, métodos, dispositivos, reactivos,
40 sistemas y kits para la evaluación del riesgo de un evento cardiovascular (CV) en un plazo de 5 años.

Sumario de la invención

La presente invención está definida por las reivindicaciones adjuntas. La solicitud incluye biomarcadores, métodos,
45 reactivos, dispositivos, sistemas y kits para la predicción del riesgo de sufrir un evento cardiovascular (CV) en un período de 1 año, un período de 2 años, un período de 3 años o un período de 4 años. Los biomarcadores de la presente solicitud se identificaron utilizando un ensayo multiplexado basado en aptámeros de velocidad de ralentización (SOMAmer) que se describe con detalle en el presente documento. Mediante el uso del procedimiento de identificación de biomarcadores basado en SOMAmer descrito en el presente documento, Esta aplicación
50 describe un conjunto de biomarcadores que son útiles para predecir la probabilidad de un evento CV e el plazo de 1 año, 2 años, 3 años o 4 años.

Los eventos cardiovasculares pueden evitarse mediante un tratamiento agresivo si la propensión a tales eventos se puede determinar con precisión. Las pruebas de marcadores múltiples de la técnica anterior requieren la obtención
55 de varias muestras de un individuo o requieren que una muestra se reparta entre múltiples ensayos. Sería preferente proporcionar un ensayo de pronóstico que requeriría solo una muestra biológica, medido en un ensayo particular, en lugar de múltiples muestras para diferentes tipos de analitos (lípidos, proteínas, metabolitos) o paneles de analitos. El beneficio central de una prueba de muestra única es la simplicidad en el punto de uso, desde una prueba con múltiples colecciones de muestras y/o múltiples tipos de tecnología (tal como la integración de resultados de sangre
60 con una o más fuentes complementarias de información, tales como datos demográficos, ecocardiografía, formación de imágenes, análisis de orina, presión arterial o distensibilidad vascular) es más complejo de administrar y esto constituye una barrera para la adopción. Una ventaja adicional se obtiene al analizar esa única muestra en un solo ensayo para múltiples proteínas. Un solo ensayo debería mitigar la variación no deseada debido a la calibración de múltiples resultados de análisis o formatos tecnológicos juntos. La prueba que forma la base de esta aplicación es una prueba de "una muestra, un ensayo". Esta combinación de una muestra única y un ensayo es una característica
65 novedosa de esta prueba del riesgo de evento cardiovascular que aborda la complejidad logística de obtener

múltiples muestras y usar múltiples modalidades de medición y los problemas y riesgos biológicos implicados en la división de muestras en múltiples alícuotas para múltiples procedimientos analíticos independientes.

5 Se sabe que la enfermedad cardiovascular implica múltiples procesos biológicos y tejidos. Ejemplos bien conocidos de sistemas y procesos biológicos asociados con enfermedades cardiovasculares son inflamación, trombosis, angiogénesis asociada a la enfermedad, activación plaquetaria, activación de macrófagos, respuesta hepática aguda, remodelación de la matriz extracelular y función renal. Estos procesos pueden observarse en función del sexo, el estado menopáusico y la edad, y según el estado de coagulación y la función vascular. Dado que estos sistemas se comunican parcialmente a través de sistemas de señalización basados en proteínas y se pueden medir múltiples proteínas en una sola muestra de sangre, la invención proporciona una única muestra, la prueba basada en múltiples proteínas de ensayo único enfocada a proteínas de sistemas biológicos específicos y procesos implicados en enfermedades cardiovasculares.

15 Como se trata en el presente documento, una de las funciones centrales de medir el riesgo de un evento cardiovascular es permitir la evaluación del progreso en respuesta al tratamiento y los cambios de comportamiento, tales como la dieta y el ejercicio. Los métodos actuales de predicción de riesgos, tal como la ecuación de Framingham, incluyen información clínica covariable claramente no sensible, los factores clave son la edad y el sexo del sujeto. Esto hace que la ecuación de Framingham sea menos útil para monitorizar el cambio en el riesgo de un individuo, aunque puede ser precisa para una población. Una característica novedosa de esta prueba del riesgo de evento CV es que no requiere la edad como parte del modelo de pronóstico. La invención sujeto se basa en la premisa de que, dentro de la biología del envejecimiento, existen factores biológicos subyacentes que están más directamente asociados con el riesgo, pero que son variables entre individuos y, por lo tanto, se utilizan mejor para evaluar el riesgo que la edad cronológica. La invención se basa en la creencia de que la edad en sí misma no es un factor causal en la enfermedad y que la edad está actuando como un sustituto o representante de la biología subyacente. Si bien la edad es, de hecho, pronóstica de eventos CV, no puede usarse para evaluar la mejora individual y, presumiblemente, el efecto de la edad está mediado por la función biológica. Este efecto puede determinarse mejor mediante la medición de la biología relevante. En la presente invención, las proteínas dirigidas están implicadas en la biología de la enfermedad. Por lo tanto, la invención captura la información biológica que se refleja en la correlación entre la edad y el riesgo de un evento CV.

20 La estrategia para identificar proteínas de múltiples procesos implicados en la enfermedad cardiovascular requería elegir parámetros que proporcionaran una amplia gama/diversidad de pacientes con enfermedad CV que presentaban diversos eventos o síntomas. Los eventos debidos a enfermedad cardiovascular son heterogéneos, que implica muerte súbita de causa desconocida y dos clases principales de eventos conocidos: trombótico (ictus, ataques isquémicos transitorios, infarto de miocardio) y eventos relacionados con ICC. Algunos eventos de presentación pueden carecer de información de diagnóstico específica (por ejemplo, muerte en el hogar). En vista de estas características de la enfermedad CV, la prueba de la invención se desarrolló midiendo las proteínas implicadas en los procesos biológicos asociados con la enfermedad CV, en muestras de sangre de una amplia gama de eventos. Esta estrategia dio como resultado la inclusión de información de múltiples procesos implicados en la enfermedad (por ejemplo, angiogénesis, activación plaquetaria, activación de macrófagos, respuesta hepática aguda, otra inflamación linfocitaria, remodelación de la matriz extracelular y función renal). Con el fin de desarrollar una prueba de una sola muestra pronóstica basada en múltiples proteínas para la enfermedad CV, la población de estudio elegida fue un estudio de cohortes de un grupo de sujetos de riesgo alto con arteriopatía coronaria aparentemente estable: el estudio "Heart & Soul". Al elegir este grupo de sujetos con una tasa alta de eventos CV, fue posible determinar el riesgo asociado con las mediciones de proteínas con mayor exactitud de lo que habría sido posible en la población general (en qué eventos son más raros). El desarrollo de la prueba del sujeto en este grupo de alto riesgo permitió la identificación de combinaciones de biomarcadores proteicos que podrían generalizarse debido a la biología común. Como resultado, es probable que la prueba de la invención objeto y los biomarcadores sean efectivos más allá de la predicción de eventos en una población más grande que los individuos que coinciden con los criterios de entrada en el estudio "Heart & Soul".

En algunas realizaciones de la presente divulgación, se proporcionan métodos para la detección selectiva de un sujeto del riesgo de un evento cardiovascular (CV). En algunas realizaciones, un método comprende

- 55 (a) formar un panel de biomarcadores que comprende N biomarcadores seleccionados de entre MMP12, angiopoyetina-2, complemento C7, troponina cardíaca I, proteína 4 relacionada con angiopoyetina, CCL18/PARC, complejo alfa-1-antiquimotripsina, GDF11 y alfa-2-antiplasmina, en el que N es un número entero de 2 a 9; y
- 60 (b) detectar el nivel de cada uno de los N biomarcadores del panel en una muestra del sujeto.

En algunas realizaciones de la presente divulgación, se proporcionan métodos para predecir la probabilidad de que un sujeto tenga un evento CV. En algunas realizaciones, un método comprende

- 65 (a) formar un panel de biomarcadores que comprende N biomarcadores seleccionados de entre MMP12, angiopoyetina-2, complemento C7, troponina cardíaca I, proteína 4 relacionada con angiopoyetina, CCL18/PARC, complejo alfa-1-antiquimotripsina, GDF11 y alfa-2-antiplasmina, en el que N es un número entero

de 2 a 9; y

(b) detectar el nivel de cada uno de los N biomarcadores del panel en una muestra del sujeto.

- 5 En algunas realizaciones, se proporcionan métodos para la detección selectiva de un sujeto del riesgo de un evento cardiovascular (CV), que comprende detectar el nivel de al menos cinco, al menos seis, al menos siete, al menos ocho o los nueve biomarcadores seleccionados de entre MMP12, angiopoyetina-2, complemento C7, troponina cardíaca I, proteína 4 relacionada con angiopoyetina, CCL18/PARC, complejo alfa-1-antiquimotripsina, GDF11 y alfa-2-antiplasmina en una muestra del sujeto.
- 10 En algunas realizaciones, se proporcionan métodos para predecir la probabilidad de que un sujeto sufra un evento CV, que comprende detectar el nivel de al menos cinco, al menos seis, al menos siete, al menos ocho o los nueve biomarcadores seleccionados de entre MMP12, angiopoyetina-2, complemento C7, troponina cardíaca I, proteína 4 relacionada con angiopoyetina, CCL18/PARC, complejo alfa-1-antiquimotripsina, GDF11 y alfa-2-antiplasmina en una muestra del sujeto.
- 15 En alguna realización, la probabilidad de que el sujeto sufra un evento CV en un plazo de 4 años es alta si el nivel de al menos cinco, al menos seis o los siete biomarcadores seleccionados de entre MMP12, angiopoyetina-2, complemento C7, troponina cardíaca I, proteína 4 relacionada con angiopoyetina, el complejo CCL18/PARC y alfa1-antiquimotripsina es más alto que un nivel de control de la proteína respectiva y si el nivel de al menos un biomarcador o ambos biomarcadores seleccionados de entre GDF11 y alfa2-antiplasmina es más bajo que un nivel de control de la proteína respectiva.
- 20 En algunas realizaciones, se proporcionan métodos para la detección selectiva de un sujeto del riesgo de un evento cardiovascular (CV), que comprenden detectar el nivel de GDF11 y FSTL3 en una muestra del sujeto.
- 25 En algunas realizaciones, se proporcionan métodos para predecir la probabilidad de que un sujeto sufra un evento CV, que comprenden detectar el nivel de GDF11 y FSTL3 en una muestra del sujeto. En algunas realizaciones, se proporcionan métodos para predecir la probabilidad de que un sujeto sufra un evento trombótico, que comprenden detectar el nivel de GDF11 y FSTL3 en una muestra del sujeto. En algunas realizaciones, el evento trombótico se selecciona incluso de entre infarto de miocardio, ictus y ataques isquémico transitorio.
- 30 En algunas realizaciones, la probabilidad de que el sujeto sufra un evento CV (tal como un evento trombótico) en un plazo de 4 años es alta si el nivel de GDF11 es más bajo que un nivel de control de GDF11 y/o el nivel de FSTL3 es más alto que un nivel de control de FSTL3.
- 35 En algunas realizaciones, el método comprende detectar el nivel de MMP12. En algunas realizaciones, El método comprende detectar el nivel de angiopoyetina-2. En algunas realizaciones, el método comprende detectar el nivel de complemento C7. En algunas realizaciones, el método comprende detectar el nivel de troponina I cardíaca. En algunas realizaciones, el método comprende detectar el nivel de proteína 4 relacionada con la angiopoyetina. En algunas realizaciones, el método comprende detectar el nivel de CCL18/PARC. En algunas realizaciones, el método comprende detectar el nivel de complejo alfa-1-anti-quimotripsina. En algunas realizaciones, el método comprende detectar el nivel de GDF11. En algunas realizaciones, el método comprende detectar el nivel de alfa-2-antiplasmina. En algunas realizaciones, el método comprende detectar el nivel de MMP12, angiopoyetina-2, complemento C7, troponina cardíaca I, proteína 4 relacionada con angiopoyetina, CCL18/PARC, complejo alfa-1-antiquimotripsina, GDF11 y alfa-2-antiplasmina.
- 40 En algunas realizaciones, el sujeto padece arteriopatía coronaria. En algunas realizaciones, el sujeto no tiene antecedentes de eventos CV. En algunas realizaciones, el sujeto tiene una puntuación elevada en el American College of Cardiology (ACC). En algunas realizaciones, el sujeto tiene una puntuación de riesgo intermedia en el ACC. En algunas realizaciones, El sujeto tiene una puntuación del riesgo baja en el ACC. En algunas realizaciones, el sujeto ha sufrido al menos un evento CV. En algunas realizaciones, el evento CV se selecciona de entre infarto de miocardio, ictus, insuficiencia cardíaca congestiva, ataque isquémico transitorio y muerte.
- 45 En algunas realizaciones, la muestra se selecciona de una muestra de sangre, una muestra de suero, una muestra de plasma y una muestra de orina. En algunas realizaciones, la muestra es una muestra de plasma. En algunas realizaciones, el método se realiza *in vitro*.
- 50 En algunas realizaciones, cada biomarcador es un biomarcador proteico. En algunas realizaciones, el método comprende poner en contacto biomarcadores de la muestra del sujeto con un conjunto de reactivos de captura de biomarcadores, en el que cada reactivo de captura de biomarcadores del conjunto de reactivos de captura de biomarcadores se une específicamente a un biomarcador diferente que se esté detectando. En algunas realizaciones, cada reactivo de captura de biomarcadores es un anticuerpo o un aptámero. En algunas realizaciones, cada reactivo de captura de biomarcadores es un aptámero. En algunas realizaciones, al menos un aptámero es un aptámero de tasa de disociación lenta. En algunas realizaciones, al menos un aptámero de tasa de disociación lenta comprende al menos uno, al menos dos, al menos tres, al menos cuatro, al menos cinco, al menos seis, al menos siete, al menos ocho, al menos nueve o al menos 10 nucleótidos con modificaciones. En algunas realizaciones, cada
- 55
- 60
- 65

aptámero de tasa de disociación lenta se une a su proteína objetivo con una tasa de disociación ($t_{1/2}$) de ≥ 30 minutos, ≥ 60 minutos, ≥ 90 minutos, ≥ 120 minutos, ≥ 150 minutos, ≥ 180 minutos, ≥ 210 minutos, o ≥ 240 minutos.

5 En algunas realizaciones, la probabilidad de un evento CV se basa en los niveles de biomarcadores y al menos un elemento de información biomédica adicional seleccionada de

- 10 a) información correspondiente a la presencia de factores de riesgo cardiovascular seleccionados del grupo que consiste en infarto de miocardio previo, evidencia angiográfica de más del 50 % de estenosis en uno o más vasos coronarios, isquemia inducida por el ejercicio mediante tapiz rodante o pruebas nucleares o revascularización coronaria previa,
- b) información correspondiente a los descriptores físicos de dicho individuo,
- c) información correspondiente a un cambio en el peso de dicho individuo,
- d) información correspondiente a la etnia de dicho individuo,
- 15 e) información correspondiente al sexo de dicho individuo,
- f) información correspondiente al historial de tabaquismo de dicho individuo,
- g) información correspondiente al historial de alcoholismo de dicho individuo,
- h) información correspondiente al historial ocupacional de dicho individuo,
- i) información correspondiente a los antecedentes familiares de enfermedad cardiovascular u otras afecciones del sistema circulatorio de dicho individuo,
- 20 j) información correspondiente a la presencia o ausencia en dicho individuo de al menos un marcador genético que se correlaciona con un mayor riesgo de enfermedad cardiovascular en dicho individuo o un miembro de la familia de dicho individuo,
- k) información correspondiente a los síntomas clínicos de dicho individuo,
- l) información correspondiente a otras pruebas de laboratorio,
- 25 m) información correspondiente a los valores de expresión génica de dicho individuo, y
- n) información correspondiente al consumo de factores de riesgo cardiovascular conocidos de dicho individuo, tales como una dieta rica en grasas saturadas, rica en sales, rica en colesterol,
- o) información correspondiente a los resultados de imagen del individuo obtenidos por técnicas seleccionadas del grupo que consisten en electrocardiograma, ecocardiografía, ultrasonido carotídeo para determinar el grosor de la íntima-media, dilatación mediada por flujo, velocidad de la onda de pulso, índice tobillo-brazo, ecocardiografía de esfuerzo, imagen de perfusión miocárdica, calcio coronario por TC, angiografía por TC de alta resolución, imagen de resonancia magnética y otras modalidades de imagen,
- 30 p) información sobre los medicamentos del individuo, e
- q) información sobre la función renal del individuo.

35 En algunas realizaciones, el método comprende determinar la probabilidad de un evento CV con el propósito de determinar una prima de seguro médico o una prima de seguro de vida. En algunas realizaciones, el método comprende además determinar la cobertura o prima para seguro médico o seguro de vida. En algunas realizaciones, el método comprende además usar la información resultante del método para predecir y/o gestionar la utilización de recursos médicos. En algunas realizaciones, el método comprende además usar la información resultante del método para permitir una decisión de adquirir o comprar una consulta médica, hospital o empresa.

40 En algunas realizaciones, se proporciona un método implementado por ordenador para evaluar el riesgo de un evento cardiovascular (CV). En algunas realizaciones, el procedimiento comprende recuperar en un ordenador información de biomarcadores para un sujeto, en el que la información del biomarcador comprende los niveles de al menos cinco, al menos seis, al menos siete, al menos ocho o los nueve biomarcadores seleccionados de entre MMP12, angiopoyetina-2, complemento C7, troponina cardíaca I, proteína 4 relacionada con angiopoyetina, CCL18/PARC, complejo alfa-1-antiquimotripsina, GDF11 y alfa-2-antiplasmina en una muestra del sujeto; realizar con el ordenador una clasificación de cada uno de dichos valores de biomarcadores; indicar un resultado de la evaluación del riesgo para un evento CV para dicho individuo basado en una pluralidad de clasificaciones. En algunas realizaciones, indicar el resultado de la evaluación del riesgo de un evento CV para el sujeto comprende mostrar el resultado en una pantalla de ordenador.

55 Breve descripción de los dibujos

La **figura 1** muestra gráficos de cajas de la distribución del factor a escala de normalización para proteínas medidas en conjuntos de descubrimiento y validación en cada dilución de muestra. En los gráficos de cajas, la línea roja indica la mediana del valor, la extensión de la caja muestra el rango intercuartílico que contiene el 50 % de los datos y los bigotes se extienden 1,5 veces el rango intercuartílico fuera de la caja. Las muestras con factores de escala de normalización extrema están marcadas con el signo "+" en rojo. La normalización aumenta (disminuye) los niveles medianos de señal en el conjunto de descubrimiento (validación) para compensar el sesgo de intensidad sistemático evidente en la señal de proteína medida en las muestras de validación.

La **figura 2** muestra los gráficos de cocientes de riesgos instantáneos en el modelo Cox univariado por desviación estándar de RFU (arriba) o entre los cuartiles de RFU externos (abajo). La línea discontinua horizontal indica un nivel de significación corregido de Bonferroni $p = 0,05$. Los nombres de genes del Centro Nacional de Información Biotecnológica (NCBI) se utilizan como etiquetas sucintas para proteínas con cocientes de riesgos

instantáneos extremos. Las proteínas marcadas en rojo se incluyen en el modelo CVD9: ANGPT2 = "Angiopoyetina-2"; C7="Complemento C7"; SERPINF2 = "inhibidor de la serina proteasa F2" o "α2-antiplasmina"; CCL18 = "Ligando 18 de quimiocina (motivo C-C)" también conocido como "quimiocina pulmonar y regulada por activación (PARC)"; ANGL4 = "proteína 4 relacionada con la angiopoyetina"; KLK3.SERPINA3 = "complejo de α1-antiquimotripsina"; y TNNT3 = "Troponina-I, cardíaca".

La **figura 3** muestra los niveles medios de señal de las proteínas CVD9 en conjuntos de descubrimiento y validación y un modelo de regresión lineal robusto utilizado para estimar el sesgo de intensidad residual resultante del procedimiento de normalización.

La **figura 4** muestra una comparación del riesgo predicho y real generado por el modelo de Framingham en el conjunto de descubrimiento antes (izquierda) y después (derecha) de la recalibración con el modelo de calibración de Cox.

La **figura 5** muestra una comparación del riesgo predicho y real generado por el modelo de Framingham en el conjunto de validación HUNT3 antes (izquierda) y después (derecha) de la recalibración con el modelo de calibración de Cox.

La **figura 6** muestra un diagrama de flujo de muestra y del proceso estadístico aplicado a los conjuntos de muestras de descubrimiento (izquierda, gris) y validación (derecha, rosa).

La **figura 7** muestra la proporción de los cocientes de riesgos instantáneos univariadas del cuarto al primer cuartil (con intervalos de confianza del 95 %) para un grupo complementario de 16 proteínas seleccionadas por el procedimiento LASSO multivariado en el conjunto de descubrimiento (símbolos negros, línea superior de cada par de líneas) y las mismas proteínas del conjunto de validación (símbolos rojos, línea inferior de cada par de líneas). Las proteínas marcadas con asteriscos se incluyen en el modelo paramétrico final (CVD9) después de la retroeliminación gradual de las proteínas menos importantes. Para las propiedades biológicas relevantes de estas 16 proteínas, véanse los Ejemplos. Leyenda: MMP-7= metaloproteínasa de la matriz 7; MMP12= metaloproteínasa de la matriz 12; TIM3 = inmunoglobulina de linfocitos T y proteína 3 que contiene el dominio de mucina; CCL18= ligando 18 de quimiocina (motivo C-C), anteriormente conocido como PARC = quimiocina pulmonar y regulada por activación; GDF11= factor de diferenciación de crecimiento 11; CDO = oncogén regulado asociado a la adhesión celular; EGF= factor de crecimiento epidérmico.

La **figura 8** muestra el rendimiento de calibración por decil del riesgo predicho en el conjunto de validación HUNT-3 para CVD9 (izquierda) y Framingham (derecha).

La **figura 9** muestra el riesgo predicho de CVD9 (rosa) y Framingham (gris) frente al percentil de riesgo de CVD9. Los puntos sólidos indican la frecuencia de eventos observados para los pacientes en cada decil de riesgo predicho generado por los modelos CVD9 (rosa) y Framingham (gris). La línea horizontal indica la incidencia del evento a los 4 años.

La **figura 10** muestra las curvas ROC para el modelo aplicado al conjunto de descubrimiento (negro, indicado por la flecha) y el conjunto de validación independiente (rojo, indicado por la flecha) a 1 año y a 4 años, el tiempo máximo válido para la puntuación de Framingham en esta población. También se incluyen las curvas ROC para la puntuación de Framingham en la cohorte de descubrimiento (verde) y de validación (azul).

La **figura 11** muestra curvas de supervivencia de Kaplan-Meier para cada cuartil de riesgo predicho por CVD9 en las cohortes de descubrimiento (izquierda) y validación (derecha). Las marcas de verificación muestran el tiempo de censura (última observación) para sujetos individuales y los intervalos sombreados indican intervalos de confianza del 95 %.

La **figura 12** ilustra un sistema informático a modo de ejemplo no limitante para su uso con diversos procedimientos implementados por ordenador descritos en el presente documento.

La **figura 13** ilustra un ensayo de aptámero a modo de ejemplo no limitante que puede usarse para detectar uno o más biomarcadores en una muestra biológica.

La **figura 14** muestra ciertas pirimidinas modificadas de ejemplo que pueden incorporarse en aptámeros, tales como los aptámeros de tasa de disociación lenta.

La **figura 15** muestra la correlación entre GDF11 y FSTL3.

La **figura 16** muestra las curvas de supervivencia para cada cuartil para cada modelo. Los cuartiles primero a cuarto se describen con color negro (línea superior), rojo (segunda línea hacia abajo), verde (tercera línea hacia abajo) y azul (línea inferior). El sombreado muestra los intervalos de confianza del 95 %. El carácter "+" significa muestras censuradas.

La **figura 17** muestra una comparación de las curvas de supervivencia entre GDF11 y GDF11.FSTL3 para el grupo de bajo riesgo y el grupo de alto riesgo. En el panel de la izquierda, la línea superior representa el modelo GDF11.FSTL3 y la línea inferior representa el modelo GDF11. En el panel de la derecha, la línea superior representa el modelo GDF11 y la línea inferior representa el modelo GDF11.FSTL3.

La **figura 18** muestra una comparación de la probabilidad a 4 años entre GDF11 y GDF11.FSTL3 (izquierda) y entre FSTL3 y GDF11.FSTL3 (derecha).

La **figura 19** muestra la curva ROC a 4 años para los tres modelos.

La **figura 20** muestra curvas de supervivencia para cada cuartil del predictor lineal de cada grupo (todos, CHF-Muerte y evento trombótico) del modelo GDF11.FSTL3. Los cuartiles primero a cuarto se describen con color negro (línea superior), rojo (segunda línea hacia abajo), verde (tercera línea hacia abajo) y azul (línea inferior). El sombreado muestra los intervalos de confianza del 95 %.

La **figura 21** muestra las curvas de supervivencia para cada cuartil para los modelos GDF11, GDF11.WFIKKN1, GDF11.WFIKKN2 y GDF11.WFIKKN1.WFIKKN2.

La **figura 22** muestra la probabilidad del riesgo entre el modelo GDF11 y los modelos GDF11.WFIKKN1,

GDF11.WFIKKN2 y GDF11.WFIKKN1.WFIKKN2.

La **figura 23** muestra las curvas ROC para cada modelo: GDF11, GDF11.WFIKKN1, GDF11.WFIKKN2 y GDF11.WFIKKN1.WFIKKN2.

5 Descripción detallada

Aunque la invención se describirá junto con ciertas realizaciones representativas, debe entenderse que la invención está definida por las reivindicaciones y no está limitada a dichas realizaciones.

10 Un experto en la técnica reconocerá muchos métodos y en la práctica de la presente invención se pueden usar materiales similares o equivalentes a los descritos en el presente documento. La presente invención no está limitada de ningún modo a los procedimientos y materiales descritos.

15 Salvo que se defina de otro modo, todos los términos técnicos y científicos usados en el presente documento tienen el significado comprendido habitualmente por un experto en la materia a la que pertenece la presente invención. Aunque cualquier método, dispositivos y materiales similares o equivalentes a los descritos en el presente documento se pueden utilizar en la práctica de la presente invención, determinados métodos, dispositivos y materiales se describen en el presente documento.

20 Tal como se usa en la presente solicitud, incluyendo las reivindicaciones adjuntas, las formas en singular "un", "una", y "el/la" incluye el plural, a menos que el contenido dicte claramente otra cosa, y se puede usar indistintamente con "al menos uno" y "uno o más". Por lo tanto, la referencia a "un aptámero" incluye mezclas de aptámeros, la referencia a "una sonda" incluye mezclas de sondas y similares.

25 Tal como se usa en el presente documento, las expresiones "comprende", "que comprende", "incluye", "que incluye", "contiene", "que contiene", y cualquier variación de los mismos, están destinados a cubrir una inclusión no exclusiva, tal como un proceso, método, producto definido por su proceso o composición de la materia que comprende, incluye o contiene un elemento o una lista de elementos, puede incluir otros elementos que no están expresamente enumerados.

30 La presente solicitud incluye biomarcadores, métodos, dispositivos, reactivos, sistemas y kits para la predicción del riesgo de eventos CV a corto plazo dentro de un período de tiempo definido, tal como en 1 año, en 2 años, en 3 años o en 4 años.

35 "Evento cardiovascular" significa un fallo o mal funcionamiento de cualquier parte del sistema circulatorio. En una realización, "evento cardiovascular" significa ictus, ataque isquémico transitorio (TIA), infarto de miocardio (IM), muerte súbita atribuible al mal funcionamiento del sistema circulatorio y/o insuficiencia cardíaca, o muerte súbita de causa desconocida en una población en la que la causa más probable es cardiovascular. En otra realización, "evento cardiovascular" significa cualquiera de las disfunciones anteriores y/o angina inestable, necesidad de endoprótesis vascular o angioplastia, o similares.

40 Los eventos cardiovasculares incluyen "insuficiencia cardíaca congestiva" o "ICC" y "eventos trombóticos". Los eventos trombóticos incluyen IM, ataques isquémicos transitorios (AIT), ictus, síndrome coronario agudo y necesidad de revascularización coronaria.

45 En determinadas realizaciones, los biomarcadores se proporcionan para su uso solo o en varias combinaciones para evaluar el riesgo de muerte súbita o un evento CV futuro en un período de tiempo de 4 años con eventos CV definidos como infarto de miocardio, ictus, muerte e insuficiencia cardíaca congestiva. Los eventos trombóticos consisten en infarto de miocardio e ictus combinados. Como se describe con detalle en lo sucesivo, las realizaciones
50 a modo de ejemplo incluyen los biomarcadores proporcionados en la Tabla 3, que se identificaron usando un ensayo multiplexado basado en SOMAmer que se describe generalmente en los Ejemplos.

Si bien algunos de los biomarcadores de eventos CV descritos pueden ser útiles solos para evaluar el riesgo de un evento CV, en el presente documento también se describen procedimientos para la agrupación de múltiples
55 subconjuntos de biomarcadores de evento CV, en los que cada selección de grupos o subconjuntos es útil como panel de tres o más biomarcadores, a los que en el presente documento se hace referencia de forma indistinta como "panel de biomarcador" y un panel. Por lo tanto, varias realizaciones de la presente solicitud proporcionan combinaciones que comprenden al menos cinco, al menos seis, al menos siete, al menos ocho o los nueve biomarcadores en la Tabla 3.

60 En una realización, el número de biomarcadores útiles para un subconjunto o panel de biomarcadores se basa en el valor de la sensibilidad y la especificidad para la combinación particular de valores de biomarcadores. Los términos "sensibilidad" y "especificidad" se usan en el presente documento con respecto a la capacidad para clasificar correctamente a un individuo, basándose en uno o más valores de biomarcadores detectados en su muestra
65 biológica, como tener un mayor riesgo de sufrir un evento CV en un plazo de 4 años o no tener un mayor riesgo de sufrir un evento CV en el mismo período de tiempo. "Sensibilidad" indica el rendimiento del biomarcador o

biomarcadores con respecto a la clasificación correcta de los individuos que tienen mayor riesgo de evento CV. "Especificidad" indica el rendimiento del biomarcador o biomarcadores con respecto a la clasificación correcta de los individuos que no tienen un riesgo mayor de evento CV. Por ejemplo, una especificidad del 85 % y una sensibilidad del 90 % para un panel de marcadores utilizados para analizar un conjunto de muestras negativas para el evento y muestras positivas para el evento indica que el 85 % de las muestras de control se clasificaron correctamente como muestras negativas para el evento en el panel y el 90 % de las muestras positivas para el evento se clasificaron correctamente como muestras positivas para el evento en el panel.

En un método alternativo, las puntuaciones pueden notificarse en un intervalo continuo, con un umbral de riesgo alto, intermedio o bajo de un evento CV dentro de una unidad de tiempo definida, con umbrales determinados en función de los hallazgos clínicos; una expresión alternativa de los mismos datos es fijar el umbral de probabilidad (tal como el 50 %) y predecir el momento en que esta proporción de sujetos sufriría su evento (por ejemplo, análogo a la semivida en la desintegración radiactiva, el momento en que la mitad del isótopo se habría desintegrado).

un factor que puede afectar a la cantidad de biomarcadores que se utilizarán en un subconjunto o panel de biomarcadores es el procedimiento utilizado para obtener muestras biológicas de individuos a los que se está evaluando por un riesgo de un evento CV. En un entorno de obtención de muestras cuidadosamente controlado, la cantidad de biomarcadores necesarios para alcanzar los valores deseados de sensibilidad y especificidad y umbral será menor que en una situación en la que puede haber más variación en la recolección, manipulación y almacenamiento de muestras. Como alternativa, se puede obtener una mayor sensibilidad y especificidad utilizando más marcadores que sean menos robustos para la obtención de muestras (por ejemplo, que no sobreviven en una situación de recolección variable) junto con marcadores de manipulación de muestras que permiten el rechazo de muestras mal recolectadas o la eliminación de marcadores sensibles del algoritmo de predicción de riesgos.

"Muestra biológica", "muestra" y "muestra de ensayo" se usan indistintamente en el presente documento para hacer referencia a cualquier material, fluido biológico, tejido o célula obtenido, o de otro modo, derivado de un individuo. Esta incluye sangre (incluida sangre entera, leucocitos, células mononucleares de sangre periférica, capa leucocitaria, plasma y suero), esputo, lágrimas, moco, lavados nasales, aspirados nasales, orina, saliva, lavados peritoneales, ascitis, líquido cístico, fluido glandular, fluido linfático, aspirado bronquial, líquido sinovial, aspirado articular, secreciones orgánicas, células, un extracto celular y un fluido cerebroespinal. Esto también incluye fracciones experimentalmente separadas de todos los anteriores. Por ejemplo, una muestra de sangre puede fraccionarse en suero, plasma o en fracciones que contengan tipos concretos de células sanguíneas, tales como glóbulos rojos o glóbulos blancos (leucocitos). En algunas realizaciones, una muestra de sangre es una mancha de sangre seca. En algunas realizaciones, una muestra de plasma es un mancha de plasma seca. En algunas realizaciones, una muestra puede ser una combinación de muestras de un individuo, tal como una combinación de una muestra de tejido y de fluido. La expresión "muestra biológica" también incluye materiales que contienen material sólido homogeneizado, tal como a partir de una muestra de heces, una muestra tisular o una biopsia tisular, por ejemplo. La expresión "muestra biológica" también incluye materiales derivados de un cultivo tisular o un cultivo celular. Puede emplearse cualquier método adecuado para obtener una muestra biológica; métodos a modo de ejemplo incluyen, por ejemplo, flebotomía, hisopado (por ejemplo, hisopado bucal) y un procedimiento de biopsia de aspirado con agua fija. Tejidos a modo de ejemplo susceptibles de aspiración con agua fina incluyen ganglio linfático, pulmón, tiroides, mama, páncreas e hígado. También se pueden recoger muestras, por ejemplo, mediante microdissección (por ejemplo, microdissección de captura láser (MCL) o microdissección láser (MDL)), lavado de vejiga, frotis (por ejemplo, un frotis de PAP) o lavado ductal. Una "muestra biológica" obtenida o derivada de un individuo incluye cualquier tal muestra que ha sido procesada de cualquier modo adecuado después de haber sido obtenida a partir del individuo. En algunas realizaciones, una muestra biológica es una muestra de plasma.

Además, en algunas realizaciones, una muestra biológica se puede obtener tomando muestras biológicas de un número de individuos y combinándolas o combinando una alícuota de la muestra biológica de cada individuo. La muestra agrupada puede tratarse como se describe en el presente documento para una muestra de un solo individuo y, por ejemplo, si se establece un mal pronóstico en la muestra agrupada, cada muestra biológica individual puede volver a analizarse para determinar qué individuo o individuos tienen un mayor o menor riesgo de evento CV.

Para los fines de la presente memoria descriptiva, se pretende que la frase "datos atribuidos a una muestra biológica de un individuo" signifique que los datos en alguna forma derivaron o se generaron utilizando, la muestra biológica del individuo. Los datos pueden haberse reformateado, revisado o alterado matemáticamente hasta cierto punto después de haber sido generados, tal como mediante la conversión desde unas unidades en un sistema de medición a unidades en otro sistema de medición; pero, se entiende que los datos se han obtenido a partir de la muestra biológica o se han generado usando la muestra biológica.

"Diana", "molécula diana" y "analito" se usan de forma intercambiable en el presente documento para hacer referencia a cualquier molécula de interés que pueda estar presente en una muestra biológica. Una "molécula de interés" incluye cualquier variación menor de una molécula en particular, tal como, en el caso de una proteína, por ejemplo, variaciones menores en la secuencia de aminoácidos, formación de enlaces disulfuro, glucosilación, lipidación, acetilación, fosforilación o cualquier otra manipulación o modificación, tal como conjugación con un componente de etiquetado, que no altera sustancialmente la identidad de la molécula. Una "molécula diana", "diana"

- o "analito" se refiere a un conjunto de copias de un tipo o especie de molécula o estructura multimolecular. "Moléculas diana", "diana" y "analitos" se refiere a un conjunto de especie de molécula o estructura multimolecular. Las moléculas diana de ejemplo incluyen proteínas, polipéptidos, ácidos nucleicos, hidratos de carbono, lípidos, polisacáridos, glucoproteínas, hormonas, receptores, antígenos, anticuerpos, aficuerpos, miméticos de anticuerpos,
- 5 virus, patógenos, sustancias tóxicas, sustratos, metabolitos, análogos del estado de transición, cofactores, inhibidores, fármacos, colorantes, nutrientes, factores de crecimiento, células, tejidos y cualquier fragmento o porción de cualquiera de los anteriores. En algunas realizaciones, una molécula diana es una proteína, en cuyo caso, la molécula diana puede denominarse "proteína diana".
- 10 Tal como se usa en el presente documento, un "agente de captura" o "reactivo de captura" se refiere a una molécula que es capaz de unirse específicamente a un biomarcador. un "reactivo de captura de la proteína diana" se refiere a una molécula que es capaz de unirse específicamente a una proteína diana. Los reactivos de captura de ejemplo no limitantes incluyen aptámeros, anticuerpos, adnectinas, anquirina, otros miméticos de anticuerpos y otros armazones de proteínas, autoanticuerpos, quimeras, moléculas pequeñas, ácidos nucleicos, lectinas, receptores de unión a
- 15 ligando, polímeros impresos, avímeros, peptidomiméticos, receptores hormonales, receptores de citocina, receptores sintéticos, y modificaciones y fragmentos de cualquiera de los reactivos de captura mencionados anteriormente. En algunas realizaciones, se selecciona un reactivo de captura de un aptámero y un anticuerpo.
- 20 El término "anticuerpo" se refiere a anticuerpos de longitud completa de cualquier especie y fragmentos y derivados de tales anticuerpos, incluidos fragmentos Fab, fragmentos F(ab')₂, anticuerpos monocatenarios, fragmentos Fv y fragmentos Fv monocatenarios. El término "anticuerpo" también se refiere a anticuerpos derivados de forma sintética, tales como anticuerpos y fragmentos derivados de la presentación en fagos, aficuerpos, nanocuerpos, etc.
- 25 Tal como se usa en el presente documento, "marcador" y "biomarcador" se usan de forma intercambiable para hacer referencia a una molécula diana que indica o es un signo de un proceso normal o anormal en un individuo o de una enfermedad u otra afección en un individuo. De manera más específica, un "marcador" o "biomarcador" es un parámetro anatómico, fisiológica, bioquímico o molecular asociado con la presencia de un estado o proceso fisiológico específico, ya sea normal o anormal y, si es anormal, ya sea crónico o agudo. Los biomarcadores son detectables y se pueden medir mediante diversos procedimientos que incluyen análisis de laboratorio e imágenes
- 30 médicas. En algunas realizaciones, un biomarcador es una proteína diana.
- 35 Tal como se usa en el presente documento, "nivel de biomarcador" y "nivel" se refieren a una medición que se realiza usando cualquier procedimiento analítico para detectar el biomarcador en una muestra biológica y que indica la presencia, ausencia, cantidad o concentración absoluta, cantidad o concentración relativa, título, un nivel, un nivel de expresión, una relación de niveles medidos, o similares, de, para, o correspondientes al biomarcador en la muestra biológica. La naturaleza exacta del "nivel" depende del diseño específico y los componentes del procedimiento analítico en particular empleado para detectar el biomarcador.
- 40 Cuando un biomarcador indica o es un signo de un proceso anormal o una enfermedad u otra afección en un individuo, dicho biomarcador generalmente se describe como sobreexpresado o subexpresado en comparación con un nivel de expresión o valor del biomarcador que indica o es un signo de un proceso normal o una ausencia de una enfermedad u otra afección en un individuo. "Regulación por aumento", "regulado por aumento", "sobreexpresión", "sobreexpresado" y cualquier variación de los mismos se utilizan indistintamente para hacer referencia a un valor o nivel de un biomarcador en una muestra biológica que es mayor que un valor o nivel (o intervalo de valores o niveles) del biomarcador que normalmente se detecta en muestras biológicas similares de individuos sanos o normales. Los términos también pueden hacer referencia a un valor o nivel de un biomarcador en una muestra biológica que es mayor que un valor o nivel (o intervalo de valores o niveles) del biomarcador que puede detectarse en un estadio diferente de una enfermedad en particular.
- 45
- 50 "Regulación por disminución", "regulado por disminución", "subexpresión", "subexpresado" y cualquier variación de los mismos se utilizan indistintamente para hacer referencia a un valor o nivel de un biomarcador en una muestra biológica que es menor que un valor o nivel (o intervalo de valores o niveles) del biomarcador que normalmente se detecta en muestras biológicas similares de individuos sanos o normales. Los términos también pueden hacer referencia a un valor o nivel de un biomarcador en una muestra biológica que es menor que un valor o nivel (o intervalo de valores o niveles) del biomarcador que puede detectarse en un estadio diferente de una enfermedad en particular.
- 55
- 60 Además, un biomarcador que está sobreexpresado o subexpresado también se puede decir que "se expresa de forma diferencial" o que tiene un "nivel diferencial" o "valor diferencial" en comparación con un nivel de expresión o valor del biomarcador "normal" que indica o es un signo de un proceso normal o una ausencia de una enfermedad u otra afección en un individuo. Por lo tanto, la "expresión diferencial" de un biomarcador también se puede denominar como una variación desde un nivel de expresión "normal" del biomarcador.
- 65 Un "nivel de control" de una molécula diana se refiere al nivel de la molécula diana en el mismo tipo de muestra de un individuo que no tiene la enfermedad o afección o de un individuo que no se sospecha o no está en riesgo de tener la enfermedad o afección, o de un individuo que ha sufrido un primer evento cardiovascular o primario pero no

un evento cardiovascular secundario o de un individuo que tiene una enfermedad cardiovascular estable. El nivel de control puede referirse al nivel promedio de la molécula diana en muestras de una población de individuos que no tiene la enfermedad o afección, o que no se sospecha o en riesgo de sufrir la enfermedad o afección, o que ha tenido una enfermedad primaria o primer evento cardiovascular pero no un evento cardiovascular secundario o que

5 tiene una enfermedad cardiovascular estable o una combinación de las mismas.

Tal como se usa en el presente documento, "individuo", "sujeto", y "paciente" se usan indistintamente para referirse a un mamífero. Un individuo mamífero puede ser un ser humano o no humano. En diversas realizaciones, el individuo es un ser humano. Un individuo sano o normal es un individuo en el que la enfermedad o afección de

10 interés (que incluye, por ejemplo, Los eventos cardiovasculares, tales como infarto de miocardio, ictus e insuficiencia cardíaca congestiva) no son detectable por los métodos de diagnóstico convencionales.

"Diagnosticar", "diagnosticar", "diagnóstico" y sus variaciones se refieren a la detección, determinación o reconocimiento de un estado o condición de salud de un individuo sobre la base de uno o más signos, los síntomas, datos u otra información pertinente a ese individuo. El estado de salud de un individuo puede diagnosticarse como sano/normal (es decir, un diagnóstico de la ausencia de una enfermedad o afección) o diagnosticarse como enfermo/anormal (es decir, un diagnóstico de la presencia, o una evaluación de las características, de una enfermedad o afección). Los términos "diagnosticar", "diagnosticar", "diagnóstico", etc., abarcan, con respecto a una enfermedad o afección en particular, la detección inicial de la enfermedad; la caracterización o clasificación de la enfermedad; la detección de la progresión, remisión o recurrencia de la enfermedad; y la detección de la respuesta de la enfermedad después de la administración al individuo de un tratamiento o terapia. La predicción del riesgo de un evento CV incluye distinguir a las personas que tienen un mayor riesgo de sufrir un evento CV de los individuos que no lo tienen.

15 "Pronóstico", "pronosticar", "pronóstico" y variaciones de los mismos se refieren a la predicción de un curso futuro de una enfermedad o afección en un individuo que tiene la enfermedad o afección (por ejemplo, predecir la supervivencia del paciente) y tales términos abarcan la evaluación de la respuesta de la enfermedad o afección después de la administración al individuo de un tratamiento o terapia.

20 "Evaluar", "que evalúa", "evaluación" y sus variaciones abarcan tanto "diagnosticar" como "pronosticar" y también abarcar determinaciones o predicciones sobre el curso futuro de una enfermedad o afección en un individuo que no padece la enfermedad, así como determinaciones o predicciones con respecto al riesgo de una enfermedad o afección recurra en un individuo que aparentemente se ha curado de la enfermedad o se ha resuelto la afección. El término "evaluar" también abarca la evaluación de la respuesta de un individuo a una terapia, tal como, por ejemplo, predecir si es probable que un individuo responda favorablemente a un agente terapéutico o si es poco probable que responda a un agente terapéutico (o experimentará toxicidad u otros efectos secundarios indeseables, por ejemplo), seleccionar un agente terapéutico para la administración a un individuo o controlar o determinar la respuesta de un individuo a una terapia que se ha administrado al individuo. Por lo tanto, "evaluar" el riesgo de un evento CV puede incluir, por ejemplo, cualquiera de los siguientes: predecir el riesgo futuro de un evento CV en un individuo; predecir el riesgo de un evento CV en un individuo que aparentemente no tiene problemas CV; predecir un tipo particular de evento CV; predecir el tiempo hasta que se produzca un evento CV; o determinar o predecir la respuesta de un individuo a un tratamiento de CV o seleccionar un tratamiento CV para administrar a un individuo basándose en una determinación de los valores de biomarcador derivados de la muestra biológica del individuo. La evaluación del riesgo de un evento CV puede incluir realizaciones tales como la evaluación del riesgo de un evento CV en una escala continua o la clasificación del riesgo de un evento CV en clasificaciones de escalas. La clasificación del riesgo incluye, por ejemplo, clasificación en dos o más clasificaciones, tales como "Sin riesgo elevado de un evento CV"; "Riesgo elevado de un evento CV"; y/o "Riesgo inferior al promedio de evento CV". En algunas realizaciones, la evaluación del riesgo de un evento CV es durante un periodo definido. Los periodos definidos de ejemplo no limitantes incluyen 1 año, 2 años, 3 años, 4 años, 5 años y más de 5 años.

25 Tal como se usa en el presente documento, "información biomédica adicional" se refiere a una o más evaluaciones de un individuo, distintas de usar cualquiera de los biomarcadores descritos en el presente documento, que están asociadas con el riesgo CV o, más específicamente, riesgo de evento CV. "Información biomédica adicional" incluye cualquiera de los siguientes: descriptores físicos de un individuo, incluyendo la altura y/o el peso de un individuo; la edad de un individuo; el sexo de un individuo; variaciones en el peso; la etnia de un individuo; los antecedentes ocupacionales; antecedentes familiares de enfermedad cardiovascular (u otros trastornos del sistema circulatorio); la presencia de uno o más marcadores genéticos que se correlacionan con un mayor riesgo de enfermedad cardiovascular (u otros trastornos del sistema circulatorio) en el individuo o alteraciones en un miembro de la familia en el grosor de la íntima de la carótida; síntomas clínicos, tal como dolor en el pecho, valores de expresión génica de ganancia o pérdida de peso; descriptores físicos de un individuo, incluidos los descriptores físicos observados mediante imágenes radiológicas; estado de tabaquismo; trastornos por abuso de alcohol; los antecedentes ocupacionales; hábitos alimenticios - ingesta de sal, grasas saturadas y colesterol; consumo de cafeína; e información de técnicas de imagen, tal como electrocardiograma, ecocardiografía, ultrasonido carotídeo para determinar el grosor de la íntima-media, dilatación mediada por flujo, velocidad de la onda de pulso, índice tobillo-brazo, ecocardiografía de esfuerzo, imagen de perfusión miocárdica, calcio coronario por TC, angiografía por TC de alta resolución, resonancia magnética y otras modalidades de imagen; y los medicamentos del individuo. El análisis

de los niveles de biomarcadores en combinación con una evaluación de cualquier información biomédica adicional, incluyendo otras pruebas de laboratorio (por ejemplo, análisis de HDL, LDL, niveles de CRP, análisis de Nt-proBNP, análisis de BNP, prueba de troponina de alta sensibilidad, prueba de glectina-3, prueba de albúmina sérica, prueba de creatina), pueden, por ejemplo, mejorar la sensibilidad, la especificidad y/o el AUC para la predicción de eventos CV en comparación con pruebas de biomarcadores de forma individual o la evaluación de cualquier artículo concreto de información biomédica adicional solo (por ejemplo, imágenes del espesor de la íntima de la carótida). Se puede obtener información biomédica adicional de un individuo usando técnicas rutinarias conocidas en la materia, tales como los propios individuos mediante el uso de un cuestionario de rutina para el paciente o un cuestionario del historial de salud, etc., o de un médico, etc. El análisis de los niveles de biomarcadores en combinación con una evaluación de cualquier información biomédica adicional puede, por ejemplo, mejorar la sensibilidad, la especificidad y/o los umbrales para la predicción de eventos CV (u otros usos relacionados con eventos cardiovasculares) en comparación con la prueba de biomarcadores de forma individual o la evaluación de cualquier artículo concreto de información biomédica adicional solo (por ejemplo, imágenes de TC de forma individual).

Tal como se usa en el presente documento, "detectar" o "determinar" con respecto a un valor de biomarcador incluye el uso tanto del instrumento usado para observar y registrar una señal correspondiente a un nivel de biomarcador como el material o materiales requeridos para generar esa señal. En diversas realizaciones, el nivel del biomarcador se detecta usando cualquier procedimiento adecuado, incluyendo fluorescencia, quimioluminiscencia, resonancia del plasmón superficial, ondas acústicas superficiales, espectrometría de masas, espectroscopía infrarroja, espectroscopia Raman, microscopio de fuerza atómica, microscopia de túnel de detección, procedimientos de detección electroquímica, resonancia magnética nuclear, puntos cuánticos y similares.

Tal como se usa en el presente documento, se determina una "puntuación del riesgo del American College of Cardiology (ACC)" de acuerdo con Goff et al., "2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines," publicado online en Circulation on November 12, 2013 (Print ISSN: 0009-7322, ISSN en línea: 1524-4539). Tal como se usa en el presente documento, una puntuación del riesgo "alta" es un pronóstico del riesgo a 10 años de 20,0 % o más para un evento de enfermedad aterosclerótica/cardiovascular (ASCVD) dura (definido como la primera aparición de infarto de miocardio no mortal o muerte por enfermedad coronaria (CHD), o ictus fatal o no fatal); una puntuación del riesgo "intermedia" es un pronóstico del riesgo a 10 años de 10,0-19,9 % para un evento de ASCVD dura; y una puntuación del riesgo "baja" es un pronóstico del riesgo a 10 años <10,0 % para un evento ASCVD dura. Véase Goff en la página 16, Tabla 5.

"Soporte sólido" se refiere en el presente documento a cualquier sustrato que tenga una superficie a la que las moléculas puedan unirse, directa o indirectamente, a través de o bien enlaces covalentes o no covalentes. Un "soporte sólido" puede tener diversos formatos físicos, que pueden incluir, por ejemplo, una membrana; un chip (un chip de proteína); un portaobjetos (por ejemplo, un portaobjetos o un cubreobjetos de vidrio); una columna; una partícula hueca, sólida, semisólida, que contiene poros o cavidades, tal como, por ejemplo, una perla; un gel; una fibra, incluyendo un material de fibra óptica; una matriz; y un receptáculo de muestras. Los receptáculos de muestras de ejemplo incluyen pocillos para muestras, tubos, capilares, viales y cualquier otro recipiente, ranura o hendidura que puede contener una muestra. Un receptáculo de muestras puede estar contenido en una plataforma de múltiples muestras, tal como una placa de microtitulación, un portaobjetos, un dispositivo microfluídico y similares. Un soporte puede estar compuesto por un material natural o sintético, un material orgánico o inorgánico. La composición del soporte sólido al cual se fijan los reactivos de captura generalmente depende del procedimiento de fijación (por ejemplo, unión covalente). Otros receptáculos a modo de ejemplo incluyen microgotas y emulsiones acuosas/oleosas controladas en microfluidos o en volumen dentro de las cuales pueden producirse ensayos y manipulaciones relacionadas. Los soportes sólidos adecuados incluyen, por ejemplo, plásticos, resinas, polisacáridos, sílice o materiales basados en sílice, vidrio funcionalizado, silicona modificada, carbono, metales, vidrios inorgánicos, membranas, nylon, fibras naturales (tales como, por ejemplo, seda, lana y algodón), polímeros y similares. El material que compone el soporte sólido puede incluir grupos reactivos tales como, por ejemplo, carboxi, grupos amino o hidroxilo, que se usan para la unión de los reactivos de captura. Los soportes sólidos poliméricos pueden incluir, por ejemplo, poliestireno, tetraftalato de polietilenglicol, acetato de polivinilo, cloruro de polivinilo, polivinilpirrolidona, poliacrilonitrilo, metacrilato de polimetilo, politetrafluoroetileno, caucho de butilo, caucho de estirenobutadieno, caucho natural, polietileno, polipropileno, (poli)tetrafluoroetileno, fluoruro de (poli)vinilideno, policarbonato y polimetilpenteno. Las partículas de soporte sólido adecuadas que se pueden usar incluyen, por ejemplo, partículas codificadas, tales como partículas codificadas de tipo Luminex®, partículas magnéticas y partículas de vidrio.

Ejemplos de usos de biomarcadores

En diversas realizaciones de ejemplo, se proporcionan procedimientos para evaluar el riesgo de un evento CV en un individuo mediante la detección de uno o más valores de biomarcadores correspondientes a uno o más biomarcadores que están presentes en la circulación de un individuo, tal como en suero o plasma, mediante cualquier número de procedimientos analíticos, incluido cualquiera de los procedimientos analíticos descritos en el presente documento. Estos biomarcadores, por ejemplo, se expresan diferencialmente en individuos con un mayor riesgo de evento CV en comparación con

los individuos sin un mayor riesgo de evento CV. La detección de la expresión diferencial de un biomarcador en un individuo puede usarse, por ejemplo, para permitir la predicción del riesgo de un evento CV en un año, 2 años, 3 años, 4 años o 5 años.

5 Además de evaluar los niveles de biomarcadores como una prueba de diagnóstico independiente, los niveles de biomarcadores también pueden evaluarse junto con la determinación de polimorfismos de un solo nucleótido (SNP) u otras lesiones genéticas o variabilidad que son indicativas de un mayor riesgo de susceptibilidad de la enfermedad o afección. (Véase, por ejemplo, Amos et al., Nature Genetics 40, 616-622 (2009)).

10 Además de evaluar los niveles de biomarcadores como una prueba de diagnóstico independiente, los niveles de biomarcadores también se pueden usar junto con exploración radiológica. Los niveles de biomarcadores también se pueden usar junto con síntomas relevantes o pruebas genéticas. La detección de cualquiera de los biomarcadores descritos en el presente documento puede ser útil después de que se haya evaluado el riesgo de evento CV para guiar la atención clínica apropiada del individuo, incluido el aumento a niveles de atención más agresivos en individuos de alto riesgo después de que se haya determinado el riesgo de evento CV. Además de evaluar los niveles de biomarcadores junto con síntomas relevantes o factores de riesgo, la información con respecto a los biomarcadores también se puede evaluar junto con otros tipos de datos, particularmente datos que indican el riesgo de un individuo de sufrir eventos cardiovasculares (por ejemplo, historial clínico del paciente, los síntomas, antecedentes familiares de enfermedad cardiovascular, antecedentes de tabaquismo o consumo de alcohol, factores de riesgo tales como la presencia de uno o más marcadores genéticos y/o el estado de otros biomarcadores, etc.). Estos diversos datos pueden evaluarse mediante procedimientos automáticos, tal como un programa/software informático, que pueden incorporarse en un ordenador u otro aparato/dispositivo.

25 Además de evaluar los niveles de biomarcadores junto con la exploración radiológica en individuos de alto riesgo (por ejemplo, evaluar los niveles de biomarcadores junto con el bloqueo detectado en un angiograma coronario), la información con respecto a los biomarcadores también se puede evaluar junto con otros tipos de datos, particularmente datos que indican el riesgo de un individuo de sufrir un evento CV (por ejemplo, historial clínico del paciente, los síntomas, antecedentes familiares de enfermedad cardiovascular, factores de riesgo, tales como si el individuo es fumador o no, consumidor de alcohol y/o estado de otros biomarcadores, etc.). Estos diversos datos pueden evaluarse mediante procedimientos automáticos, tal como un programa/software informático, que pueden incorporarse en un ordenador u otro aparato/dispositivo.

35 Las pruebas de biomarcadores también pueden asociarse con pautas y algoritmos de riesgo cardiovascular actualmente en uso en la práctica clínica. Por ejemplo, la puntuación del riesgo de Framingham utiliza factores de riesgo para proporcionar una puntuación de riesgo, factores de riesgo tales como niveles de colesterol LDL y colesterol HDL, niveles alterados de glucosa, tabaquismo, presión arterial sistólica y diabetes. La frecuencia de los pacientes de alto riesgo aumenta con la edad y los varones comprenden una mayor proporción de pacientes de alto riesgo que las mujeres.

40 Cualquiera de los biomarcadores descritos también se puede usar en pruebas de imagen. Por ejemplo, un agente de formación de imágenes puede acoplarse a cualquiera de los biomarcadores descritos, que pueden usarse para ayudar a predecir el riesgo de un evento cardiovascular, para monitorizar la respuesta a las intervenciones terapéuticas, para seleccionar para poblaciones diana en un ensayo clínico entre otros usos.

45 **Detección y determinación de biomarcadores y niveles de biomarcadores**

Un nivel de biomarcador para los biomarcadores descritos en el presente documento puede detectarse usando cualquiera de diversos procedimientos analíticos conocidos. En una realización, se detecta un valor de biomarcador usando un reactivo de captura. En diversas realizaciones, el reactivo de captura puede exponerse al biomarcador en solución o puede exponerse al biomarcador mientras el reactivo de captura está inmovilizado en un soporte sólido. En otras realizaciones, el reactivo de captura contiene una característica que reacciona con una característica secundaria en un soporte sólido. En estas realizaciones, el reactivo de captura puede exponerse al biomarcador en solución y, después, la característica del reactivo de captura puede usarse junto con la característica secundaria sobre el soporte sólido para inmovilizar el biomarcador sobre el soporte sólido. El reactivo de captura se selecciona según el tipo de análisis que se va a realizar. Los reactivos de captura incluyen, pero sin limitación, aptámeros, anticuerpos, adnectinas, anquirina, otros miméticos de anticuerpos y otros armazones de proteínas, autoanticuerpos, quimeras, moléculas pequeñas, fragmentos $F(ab')_2$, fragmentos de anticuerpo monocatenarios, fragmentos Fv, fragmentos Fv monocatenarios, ácidos nucleicos, lectinas, receptores de unión a ligando, anticuerpos, nanocuerpos, polímeros impresos, avímeros, peptidomiméticos, receptores hormonales, receptores de citocinas y receptores sintéticos, y modificaciones y fragmentos de estos.

En algunas realizaciones, se detecta un nivel de biomarcador usando un biomarcador/reactivo de captura.

65 En algunas realizaciones, el nivel del biomarcador se obtiene del complejo biomarcador/reactivo de captura y se detecta indirectamente, tal como, por ejemplo, como resultado de una reacción posterior a la interacción del

biomarcador/reactivo de captura, pero depende de la formación. del complejo biomarcador/reactivo de captura.

En algunas realizaciones, el nivel del biomarcador se detecta directamente a partir del biomarcador en una muestra biológica.

5 En algunas realizaciones, los biomarcadores se detectan usando un formato multiplexado que permite la detección simultánea de dos o más biomarcadores en una muestra biológica. En algunas realizaciones del formato multiplexado, los reactivos de captura se inmovilizan, directa o indirectamente, de forma covalente o no covalente, en ubicaciones discretas sobre un soporte sólido. En algunas realizaciones, un formato multiplexado usa soportes sólidos discretos en el que cada soporte sólido tiene un reactivo de captura único asociado con ese soporte sólido, tal como, por ejemplo, puntos cuánticos. En algunas realizaciones, se usa un dispositivo individual para la detección de cada uno de múltiples biomarcadores para detectar en una muestra biológica. Los dispositivos individuales se pueden configurar para permitir que cada biomarcador en la muestra biológica se procese simultáneamente. Por ejemplo, se puede usar una placa de microtitulación de manera que cada pocillo en la placa se use para analizar de manera única uno o más biomarcadores que se detectarán en una muestra biológica.

20 En una o más de las realizaciones anteriores, se puede usar un marcador fluorescente para marcar un componente del complejo biomarcador/reactivo de captura para permitir la detección del nivel del biomarcador. En diversas realizaciones, el marcador fluorescente puede conjugarse con un reactivo de captura específico de cualquiera de los biomarcadores descritos en el presente documento usando técnicas conocidas y el marcador fluorescente puede usarse después para detectar el nivel del biomarcador correspondiente. Los marcadores fluorescentes adecuados incluyen quelatos de tierras raras, fluoresceína y sus derivados, rodamina y sus derivados, dansilo, alofocianina, PBXL-3, Qdot 605, lisamina, ficoeritrina, rojo Texas y otros compuestos similares.

25 En algunas realizaciones, el marcador fluorescente es una molécula de colorante fluorescente. En algunas realizaciones, la molécula de colorante fluorescente incluye al menos un sistema de anillo indolio sustituido, en el que el sustituyente en el carbono 3 del anillo indolio contiene un grupo químicamente reactivo o una sustancia conjugada. En algunas realizaciones, la molécula de colorante incluye una molécula de AlexaFluor, tal como, por ejemplo, AlexaFluor 488, AlexaFluor 532, AlexaFluor 647, AlexaFluor 680 o AlexaFluor 700. En otras realizaciones, la molécula de colorante incluye un primer tipo y un segundo tipo de molécula de colorante, tal como, por ejemplo, dos moléculas de AlexaFluor. En algunas realizaciones, la molécula de colorante incluye un primer tipo y un segundo tipo de molécula de colorante y las dos moléculas de colorante tienen diferentes espectros de emisión.

35 La fluorescencia se puede medir con diversos instrumentos compatibles con una amplia gama de formatos de ensayo. Por ejemplo, los espectrofluorímetros se han diseñado para analizar placas de microtitulación, portaobjetos de microscopio, matrices impresas, cubetas, etc. Véase Principles of Fluorescence Spectroscopy, de J.R. Lakowicz, Springer Science + Business Media, Inc., 2004. Véase Bioluminescence & Chemiluminescence: Progress & Current Applications; Philip E. Stanley y Larry J. Kricka editors, World Scientific Publishing Company, enero de 2002.

40 En una o más realizaciones, opcionalmente, se puede usar un marcador de quimioluminiscencia para marcar un componente del complejo biomarcador/reactivo de captura para permitir la detección del nivel del biomarcador. Los materiales quimioluminiscentes adecuados incluyen cualquiera de cloruro de oxalilo, rodamina 6G, Ru(bipi)₃²⁺, TMAE (tetrakis(dimetilamino)etileno), pirogalol (1,2,3-trihidroxibenceno), lucigenina, peroxioxalatos, oxalatos de arilo, ésteres de acridinio, dioxetanos y otros.

45 En algunas realizaciones, el procedimiento de detección incluye una combinación de enzima/sustrato que genera una señal detectable que corresponde al nivel del biomarcador. Generalmente, la enzima cataliza una alteración química del sustrato cromogénico que se puede medir usando diversas técnicas, incluyendo espectrofotometría, fluorescencia y quimioluminiscencia. Las enzimas adecuadas incluyen, por ejemplo, luciferasas, luciferina, malato deshidrogenasa, ureasa, peroxidasa de rábano picante (HRPO), fosfatasa alcalina, beta-galactosidasa, glucoamilasa, la lisozima, glucosa oxidasa, galactosa oxidasa y glucosa-6-fosfato deshidrogenasa, uricasa, xantina oxidasa, lactoperoxidasa, microperoxidasa y similares.

50 En algunas realizaciones, el procedimiento de detección puede ser una combinación de fluorescencia, quimioluminiscencia, radionúclido o combinaciones de enzima/sustrato que generan una señal mensurable. En algunas realizaciones, la señalización multimodal podría tener características únicas y ventajosas en los formatos de ensayo de para biomarcadores.

60 En algunas realizaciones, los niveles de biomarcadores para los biomarcadores descritos en el presente documento pueden detectarse utilizando cualquier procedimiento analítico, incluidos ensayos de aptámeros unplexados, ensayos de aptámeros multiplexados, inmunoensayos multiplexados o unplexados, perfiles de expresión de ARNm, perfiles de expresión de miARN, análisis de espectrometría de masas, procedimientos histológicos/citológicos, etc. como se analiza más adelante.

65 **Determinación de niveles de biomarcadores utilizando ensayos basados en aptámeros**

Los ensayos dirigidos a la detección y cuantificación de moléculas fisiológicamente significativas en muestras biológicas y otras muestras son herramientas importantes en la investigación científica y en el campo del cuidado de la salud. Una clase de tales ensayos implica el uso de una micromatriz que incluye uno o más aptámeros inmovilizados sobre un soporte sólido. Los aptámeros son capaces de unirse a una molécula diana de una manera altamente específica y con una afinidad muy alta. Véase, por ejemplo, la patente de Estados Unidos n.º 5,475,096 titulada "Nucleic Acid Ligands"; véanse también, por ejemplo, la patente de Estados Unidos n.º 6,242,246, la patente de Estados Unidos n.º 6,458,543 y la patente de Estados Unidos n.º 6,503,715, cada una de las cuales se titula "Nucleic Acid Ligand Diagnostic Biochip". Una vez que la micromatriz se pone en contacto con una muestra, los aptámeros se unen a sus moléculas diana respectivas presentes en la muestra y, de ese modo, permiten la determinación de un nivel de biomarcador correspondiente a un biomarcador.

Tal como se usa en el presente documento, un "aptámero" se refiere a un ácido nucleico que tiene una afinidad de unión específica por una molécula diana. Se reconoce que las interacciones de afinidad son una cuestión de grado; sin embargo, en este contexto, la "afinidad de unión específica" de un aptámero por su diana significa que el aptámero se une a su diana generalmente con un grado de afinidad mucho más alto que con el que se une a otros componentes en una muestra de prueba. Un "aptámero" es un conjunto de copias de un tipo o especie de molécula de ácido nucleico que tiene una secuencia de nucleótidos en particular. Un aptámero puede incluir cualquier número adecuado de nucleótidos, incluyendo cualquier número de nucleótidos químicamente modificados. "Aptámeros" se refiere a más de uno de dichos conjuntos de moléculas. Diferentes aptámeros pueden tener el mismo o diferente número de nucleótidos. Los aptámeros pueden ser ADN o ARN o ácidos nucleicos modificados químicamente y pueden ser monocatenarios, bicatenarios o contener regiones bicatenarias, y pueden incluir estructuras superiores ordenadas. Un aptámero también puede ser un fotoaptámero, en el que un grupo funcional fotorreactivo o químicamente reactivo se incluye en el aptámero para permitir que se una covalentemente a su diana correspondiente. Cualquiera de los procedimientos del aptámero divulgados en el presente documento puede incluir el uso de dos o más aptámeros que se unen específicamente a la misma molécula diana. Como se describe con más detalle a continuación, un aptámero puede incluir un marcador. Si un aptámero incluye un marcador, no es necesario que todas las copias del aptámero tengan el mismo marcador. Por otra parte, si cada uno de los diferentes aptámeros incluye un marcador, estos diferentes aptámeros pueden tener el mismo marcador o un marcador diferente.

Un aptámero puede identificarse utilizando cualquier procedimiento conocido, incluido el proceso SELEX. Una vez identificado, un aptámero puede prepararse o sintetizarse de acuerdo con cualquier procedimiento conocido, incluidos los procedimientos de síntesis química y los procedimientos de síntesis enzimática.

Los términos "SELEX" y "proceso SELEX" se usan indistintamente en el presente documento para hacer referencia en general a una combinación de (1) la selección de aptámeros que interaccionan con una molécula diana de una manera deseable, por ejemplo, unión con alta afinidad a una proteína, con (2) la amplificación de esos ácidos nucleicos seleccionados. El proceso SELEX se puede usar para identificar los aptámeros con alta afinidad a una diana o biomarcador específico.

Normalmente, SELEX incluye preparar una mezcla candidata de ácidos nucleicos, unir la mezcla candidata a la molécula diana deseada para formar un complejo de afinidad, separar los complejos de afinidad de los ácidos nucleicos candidatos no unidos, separar y aislar el ácido nucleico del complejo de afinidad, purificar el ácido nucleico e identificar una secuencia de aptámero específica. El proceso puede incluir múltiples ciclos para perfeccionar adicionalmente la afinidad del aptámero seleccionado. El proceso puede incluir fases de amplificación en uno o más puntos en el proceso. Véase, por ejemplo, la patente de Estados Unidos n.º 5,475,096, titulada "Nucleic Acid Ligands". El proceso SELEX se puede utilizar para generar un aptámero que se une covalentemente a su diana así como un aptámero que se une no covalentemente a su diana. Véase, por ejemplo, la patente de Estados Unidos n.º 5,705,337 titulada "Systematic Evolution of Nucleic Acid Ligands by Exponential Enrichment: Chemi-SELEX."

El proceso SELEX se puede utilizar para identificar aptámeros de alta afinidad que contienen nucleótidos modificados que confieren características mejoradas en el aptámero, tal como, por ejemplo, características de administración mejoradas o estabilidad *in vivo* mejorada. Los ejemplos de dichas modificaciones incluyen sustituciones químicas en las posiciones de base y/o ribosa y/o fosfato. Se describen aptámeros identificados con el proceso SELEX que contienen nucleótidos modificados en la patente de Estados Unidos n.º 5,660,985, titulada "High Affinity Nucleic Acid Ligands Containing Modified Nucleotides", que describe oligonucleótidos que contienen derivados de nucleótidos químicamente modificados en las posiciones 5' y 2' de las pirimidinas. La patente de Estados Unidos n.º 5,580,737, véase más arriba, describe aptámeros altamente específicos que contienen uno o más nucleótidos modificados con 2'-amino (2'-NH₂), 2'-fluoro (2'-F) y/o 2'-O-metilo (2'-OMe). Véase también, la publicación de solicitud de patente de Estados Unidos n.º 20090098549, titulada "SELEX and PHOTOSELEX", que describe bibliotecas de ácidos nucleicos que tienen propiedades químicas y físicas expandidas y su uso en SELEX y photoSELEX.

También puede utilizarse SELEX para identificar aptámeros que tienen características de tasa de disociación deseables. Véase la publicación de Estados Unidos n.º 20090004667, titulada "Method for Generating Aptamers with Improved Off-Rates", que describe métodos SELEX mejorados para generar aptámeros que pueden unirse a

moléculas diana. Se describen métodos para producir aptámeros y fotoaptámeros que tienen tasas más lentas de disociación de sus respectivas moléculas diana. Los métodos implican poner en contacto la mezcla candidata con la molécula diana, permitir que ocurra la formación de complejos ácido nucleico-diana y realizar un proceso de enriquecimiento de tasa de disociación lenta en donde los complejos ácido nucleico-diana con tasas de disociación rápidas se disociarán y no se reformarán, mientras que los complejos con tasa de disociación lenta permanecerán intactos. Además, los procedimientos incluyen el uso de nucleótidos modificados en la producción de mezclas de ácidos nucleicos candidatas para generar aptámeros con un rendimiento de velocidad de disociación mejorado. Los nucleótidos modificados de ejemplo no limitantes incluyen, por ejemplo, en la figura 14 se muestran las pirimidinas modificadas. En algunas realizaciones, un aptámero comprende al menos un nucleótido con una modificación, tal como una modificación de una base. En algunas realizaciones, un aptámero comprende al menos un nucleótido con una modificación hidrofóbica, tal como una modificación de base hidrofóbica, lo que permite contactos hidrofóbicos con una proteína diana. Tales contactos hidrofóbicos, en algunas realizaciones, contribuyen a una mayor afinidad y/o una unión con tasa de disociación más lenta por parte del aptámero. Los ejemplos de nucleótidos no limitantes con modificaciones hidrofóbicas se muestran en la figura 14. En algunas realizaciones, un aptámero comprende al menos dos, al menos tres, al menos cuatro, al menos cinco, al menos seis, al menos siete, al menos ocho, al menos nueve o al menos 10 nucleótidos con modificaciones hidrofóbicas, en la que cada modificación hidrofóbica puede ser la misma o diferente de las demás. En algunas realizaciones, al menos una, al menos dos, al menos tres, al menos cuatro, al menos cinco, al menos seis, al menos siete, al menos ocho, al menos nueve o al menos 10 modificaciones hidrofóbicas en un aptámero pueden seleccionarse independientemente de las modificaciones hidrofóbicas que se muestran en la figura 14.

En algunas realizaciones, un aptámero de tasa de disociación lenta (incluidos aptámeros que comprenden al menos un nucleótido con una modificación hidrofóbica) tiene una tasa de disociación ($t_{1/2}$) de ≥ 30 minutos, ≥ 60 minutos, ≥ 90 minutos, ≥ 120 minutos, ≥ 150 minutos, ≥ 180 minutos, ≥ 210 minutos, o ≥ 240 minutos.

En algunas realizaciones, un ensayo emplea aptámeros que incluyen grupo funcionales fotorreactivos que permiten que los aptámeros se unan covalentemente o "fotorreticulen" sus moléculas diana. Véase, por ejemplo, la patente de Estados Unidos n.º 6,544,776 titulada "Nucleic Acid Ligand Diagnostic Biochip". Estos aptámeros fotorreactivos también se denominan como fotoaptámeros. Véase, por ejemplo, la patente de Estados Unidos n.º 5,763,177, la patente de Estados Unidos n.º 6,001,577 y la patente de Estados Unidos n.º 6,291,184, cada una de las cuales se titula "Systematic Evolution of Nucleic Acid Ligands by Exponential Enrichment: Photoselection of Nucleic Acid Ligands and Solution SELEX"; véanse también, por ejemplo, la patente de Estados Unidos n.º 6,458,539, titulada "Photoselection of Nucleic Acid Ligands". Después de poner la micromatriz en contacto con la muestra y de que los fotoaptámeros hayan tenido la oportunidad de unirse a sus moléculas diana, los fotoaptámeros se fotoactivan y el soporte sólido se lava para retirar cualquier molécula no específicamente unida. Pueden usarse condiciones de lavado duro, puesto que las moléculas diana que están unidas a los fotoaptámeros no se retiran generalmente, debido a los enlaces covalentes creados por el/los grupo(s) funcional(es) fotoactivado(s) sobre los fotoaptámeros. De esta manera, el ensayo permite la detección de un nivel de biomarcador correspondiente a un biomarcador en la muestra de prueba.

En estos formatos de ensayo, los aptámeros se inmovilizan en el soporte sólido antes de ponerse en contacto con la muestra. En determinadas circunstancias, sin embargo, la inmovilización de los aptámeros antes del contacto con la muestra puede no proporcionar un ensayo óptimo. Por ejemplo, la inmovilización previa de los aptámeros puede dar como resultado una mezcla ineficiente de los aptámeros con las moléculas diana en la superficie del soporte sólido, conduciendo, puede ser, a tiempo de reacción largos y, por lo tanto, períodos de incubación prolongados para permitir la unión eficaz de los aptámeros a sus moléculas diana. Además, cuando se emplean fotoaptámeros en el ensayo y dependiendo del material utilizado como soporte sólido, el soporte sólido puede tender a dispersar o absorber la luz utilizada para efectuar la formación de enlaces covalentes entre los fotoaptámeros y sus moléculas diana. Por otra parte, dependiendo del método empleado, la detección de moléculas diana unidas a sus aptámeros puede estar sujeta a imprecisión, puesto que la superficie del soporte sólido puede exponerse también y estar afectada por cualquier agente de etiquetado que se use. Por último, la inmovilización de los aptámeros en el soporte sólido generalmente implica una etapa de preparación del aptámero (es decir, la inmovilización) antes de la exposición de los aptámeros a la muestra y esta etapa de preparación puede afectar a la actividad o funcionalidad de los aptámeros.

También se han descrito ensayos de aptámeros que permiten a un aptámero capturar su diana en solución y a continuación emplear etapas de separación que están designadas para retirar componentes específicos de la mezcla de aptámero-diana antes de su detección (véase la Publicación de Estados Unidos n.º 20090042206, titulada "Multiplexed Analyses of Test Samples"). Los procedimientos de ensayo de aptámeros descritos permiten la detección y cuantificación de una diana de ácido no nucleico (por ejemplo, una diana proteica) en una muestra de prueba detectando y cuantificando un ácido nucleico (es decir, un aptámero). Los procedimientos descritos crean un sustituto de ácido nucleico (es decir, el aptámero) para detectar y cuantificar una diana de ácido no nucleico, permitiendo, de este modo, la amplia variedad de tecnologías de ácido nucleico, incluyendo la amplificación, a aplicar a un intervalo más amplio de dianas deseadas, incluidas dianas de proteína.

Los aptámeros pueden construirse para facilitar la separación de los componentes del ensayo de un complejo de

biomarcador aptámero (o complejo covalente de biomarcador fotoaptámero) y permiten el aislamiento del aptámero para su detección y/o cuantificación. En una realización, estas construcciones pueden incluir un elemento escindible o liberable dentro de la secuencia del aptámero. En otras realizaciones, se puede introducir una funcionalidad adicional en el aptámero, por ejemplo, un componente marcado o detectable, un componente espaciador o un
 5 marcador de unión específico o elemento de inmovilización. Por ejemplo, el aptámero puede incluir un marcador conectado al aptámero a través de un resto escindible, un marcador, un componente espaciador que separa el marcador y el resto escindible. En una realización, un elemento escindible es un enlazador fotoescindible. El enlazador fotoescindible puede unirse a un resto de biotina y una sección espaciadora, puede incluir un grupo NHS para la derivatización de aminas y puede usarse para introducir un grupo biotina en un aptámero, permitiendo de ese
 10 modo la liberación del aptámero más adelante en un procedimiento de ensayo.

Los ensayos homogéneos, realizados con todos los componentes del ensayo en solución, no requieren la separación de la muestra y los reactivos antes de la detección de la señal. Estos procedimientos son rápidos y fáciles de usar. Estos procedimientos generan señales basadas en un reactivo de unión o de captura molecular que
 15 reacciona con su diana específica. En algunas realizaciones de los métodos descritos en el presente documento, los reactivos de captura molecular comprenden un aptámero o un anticuerpo o similar y la diana específica sería un biomarcador que se muestra en la Tabla 3.

En algunas realizaciones, un procedimiento para la generación de señal aprovecha el cambio de señal de anisotropía debido a la interacción de un reactivo de captura marcado con fluoróforo con su diana de biomarcador específica. Cuando la captura marcada reacciona con su diana, el aumento del peso molecular hace que el movimiento de rotación del fluoróforo unido al complejo sea mucho más lento cambiando el valor de anisotropía. Al
 20 monitorizar el cambio de anisotropía, los acontecimientos de unión se pueden usar para medir cuantitativamente los biomarcadores en soluciones. Otros procedimientos incluyen ensayos de polarización de fluorescencia, procedimientos de baliza molecular, inactivación de fluorescencia resuelta en el tiempo, quimioluminiscencia, transferencia de energía de resonancia de fluorescencia y similares.

Un ejemplo de ensayo de aptámero basado en solución que puede usarse para detectar un nivel de biomarcador en una muestra biológica incluye lo siguiente: (a) preparar una mezcla poniendo en contacto la muestra biológica con
 30 un aptámero que incluye un primer marcador y tiene un afinidad específica por el biomarcador, en el que se forma un complejo de afinidad de aptámero cuando el biomarcador está presente en la muestra; (b) exponer la mezcla a un primer soporte sólido, que incluye un primer elemento de captura, y permitir que el primer marcador se asocie con el primer elemento de captura; (c) eliminar cualquier componente de la mezcla no asociado con el primer soporte sólido; (d) unir un segundo marcador al componente biomarcador del complejo de afinidad de aptámero; (e) liberar el complejo de afinidad del aptámero del primer soporte sólido; (f) exponer el complejo de afinidad de aptámero liberado a un segundo soporte sólido que incluye un segundo elemento de captura y permitir que el segundo
 35 marcador se asocie con el segundo elemento de captura; (g) eliminar cualquier aptámero que no está en complejo de la mezcla mediante reparto del aptámero que no está en complejo del complejo de afinidad de aptámero; (h) eluir el aptámero del soporte sólido; y (i) detectar el biomarcador detectando el componente aptámero del complejo de afinidad del aptámero.

Se puede usar cualquier medio conocido en la técnica para detectar un valor de biomarcador detectando el componente aptámero de un complejo de afinidad de aptámero. Se pueden utilizar varios procedimientos de detección diferentes para detectar el componente aptámero de un complejo de afinidad, tal como, por ejemplo,
 45 ensayos de hibridación, espectroscopia de masas o QPCR. En algunas realizaciones, los procedimientos de secuenciación de ácido nucleico pueden usarse para detectar el componente aptámero de un complejo de afinidad de aptámero y detectar de ese modo un valor de biomarcador. Brevemente, una muestra de prueba puede someterse a cualquier tipo de procedimiento de secuenciación de ácido nucleico para identificar y cuantificar la secuencia o secuencias de uno o más aptámeros presentes en la muestra de prueba. En algunas realizaciones, la secuencia incluye la molécula de aptámero completa o cualquier porción de la molécula que puede usarse para
 50 identificar de manera única la molécula. En otras realizaciones, la secuenciación de identificación es una secuencia específica añadida al aptámero; dichas secuencias a menudo se denominan "marcadores", "códigos de barras", o "códigos postales". En algunas realizaciones, el procedimiento de secuenciación incluye etapas enzimáticas para amplificar la secuencia de aptámero o para convertir cualquier tipo de ácido nucleico, incluyendo ARN y ADN que contienen modificaciones químicas en cualquier posición, en cualquier otro tipo de ácido nucleico apropiado para la
 55 secuenciación.

En algunas realizaciones, el procedimiento de secuenciación incluye una o más etapas de clonación. En otras realizaciones, el procedimiento de secuenciación incluye un procedimiento de secuenciación directa sin clonación.
 60

En algunas realizaciones, el procedimiento de secuenciación incluye un enfoque dirigido con cebadores específicos que se dirigen a uno o más aptámeros en la muestra de prueba. En otras realizaciones, el procedimiento de secuenciación incluye un enfoque aleatorio dirigido a todos los aptámeros en la muestra de prueba.

En algunas realizaciones, el procedimiento de secuenciación incluye etapas enzimáticas para amplificar la molécula dirigida para la secuenciación. En otras realizaciones, el procedimiento de secuenciación secuencia directamente
 65

moléculas individuales. Un procedimiento de ejemplo basado en secuenciación de ácido nucleico que puede usarse para detectar un valor de biomarcador correspondiente a un biomarcador en una muestra biológica incluye lo siguiente: (a) convertir una mezcla de aptámeros que contienen nucleótidos modificados químicamente en ácidos nucleicos no modificados con una etapa enzimática; (b) secuenciar aleatoriamente los ácidos nucleicos no modificados resultantes con una plataforma de secuenciación masivamente paralela, tal como, por ejemplo, el sistema de secuenciación 454 (454 Life Sciences/Roche), el sistema de secuenciación Illumina (Illumina), el sistema de secuenciación ABI SOLiD (Applied Biosystems), el secuenciador HeliScope Single Molecule Sequencer (Helicos Biosciences) o el sistema de secuenciación de molécula sencilla en tiempo real Pacific Biosciences (Pacific BioSciences) o el sistema de secuenciación Polonator G (Dover Systems); y (c) identificar y cuantificar los aptámeros presentes en la mezcla mediante secuencia específica y recuento de secuencias.

En el Ejemplo 1 se describe un método de ejemplo no limitante para detectar biomarcadores en una muestra biológica usando aptámeros. Véase también Kraemer et al., 2011, PLoS One 6(10): e26332.

15 **Determinación de los niveles de biomarcadores usando inmunoensayos**

Los procedimientos de inmunoensayo se basan en la reacción de un anticuerpo contra su diana o analito correspondiente y pueden detectar el analito en una muestra dependiendo del formato del ensayo específico. Para mejorar la especificidad y la sensibilidad de un procedimiento de ensayo basado en inmunorreactividad, a menudo se usan anticuerpos monoclonales y fragmentos de los mismos debido a su reconocimiento específico de epítomos. Los anticuerpos policlonales también se han utilizado con éxito en diversos inmunoensayos debido a su mayor afinidad por la diana en comparación con los anticuerpos monoclonales. Los inmunoensayos se han diseñado para su uso con una amplia gama de matrices de muestras biológicas. Los formatos de inmunoensayo se han diseñado para proporcionar resultados cualitativos, semicuantitativos y cuantitativos.

Los resultados cuantitativos se generan mediante el uso de una curva estándar creada con concentraciones conocidas del analito específico a detectar. La respuesta o señal de una muestra desconocida se representa en la curva estándar y se establece una cantidad o nivel correspondiente a la diana en la muestra desconocida.

Se han diseñado numerosos formatos de inmunoensayo. Los ensayos ELISA o EIA pueden ser cuantitativos para la detección de un analito. Este procedimiento se basa en la unión de un marcador al analito o al anticuerpo y el componente marcador incluye, tanto directa como indirectamente, una enzima. Las pruebas ELISA pueden ser formateadas para una detección directa, indirecta, competitiva o de tipo sándwich del analito. Otros procedimientos dependen de marcadores, tales como, por ejemplo, radioisótopos (I^{125}) o fluorescencia. Otras técnicas incluyen, por ejemplo, la aglutinación, la nefelometría, turbidimetría, transferencia de tipo Western, inmunoprecipitación, inmunocitoquímica, inmunohistoquímica, citometría de flujo, ensayo Luminex y otros (véase *ImmunoAssay: A Practical Guide*, editado por Brian Law, publicado por Taylor & Francis, Ltd., edición de 2005).

Entre los formatos de ensayo de ejemplo se incluyen ensayo de inmunoabsorción ligada a enzima (ELISA), radioinmunoensayo, fluorescentes, quimioluminiscencia y transferencia de energía de resonancia de fluorescencia (FRET) o inmunoensayos FRET resueltos en el tiempo (TR-FRET). Los ejemplos de procedimientos para detectar biomarcadores incluyen inmunoprecipitación de biomarcadores, seguida de procedimientos cuantitativos que permiten la discriminación por tamaño y nivel de péptido, tales como electroforesis en gel, electroforesis capilar, electrocromatografía plana y similares.

Los procedimientos para detectar y/o cuantificar un marcador detectable o material generador de señal dependen de la naturaleza del marcador. Los productos de reacciones catalizadas por enzimas apropiadas (en las que el marcador detectable es una enzima; véase anteriormente) pueden ser, sin limitación, fluorescentes, luminiscentes o radiactivos o pueden absorber luz visible o ultravioleta. Los ejemplos de detectores adecuados para detectar tales marcadores detectables incluyen, sin limitación, película de rayos X, contadores de radiactividad, contadores de centelleo, espectrofotómetros, colorímetros, fluorómetros, luminómetros y densitómetros.

Cualquiera de los procedimientos para la detección se puede realizar en cualquier formato que permita cualquier preparación, procesamiento y análisis adecuados de las reacciones. Esto puede ser, por ejemplo, en placas de ensayo de múltiples pocillos (por ejemplo, 96 pocillos o 386 pocillos) o usando cualquier matriz o micromatriz adecuada. Las soluciones madre para varios agentes se pueden realizar de forma manual o robótica, y el pipeteo, dilución, mezcla, distribución, lavado, incubación, lectura de muestras, recopilación y análisis de datos posteriores se pueden realizar de forma robotizada utilizando software de análisis, robótica e instrumentos de detección disponibles comercialmente capaces de detectar un marcador detectable.

60 **Determinación de niveles de biomarcadores utilizando perfiles de expresión génica**

La medición de ARNm en una muestra biológica puede, en algunas realizaciones, usarse como un sustituto para la detección del nivel de la proteína correspondiente en la muestra biológica. Por lo tanto, en algunas realizaciones, un biomarcador o panel de biomarcadores descritos en el presente documento también puede detectarse detectando el ARN apropiado.

En algunas realizaciones, Los niveles de expresión de ARNm se miden mediante reacción en cadena de la polimerasa cuantitativa con transcripción inversa (RT-PCR, seguida de qPCR). La RT-PCR se usa para crear un ADNc a partir del ARNm. El ADNc puede usarse en un ensayo de qPCR para producir fluorescencia a medida que progresa el proceso de amplificación del ADN. En comparación con una curva patrón, la qPCR puede producir una medición absoluta, tal como el número de copias de ARNm por célula. Se han usado transferencias de tipo Northern, micromatrices, ensayos Invader y RT-PCR combinados con electroforesis capilar para medir los niveles de expresión de ARNm en una muestra. Véase Gene Expression Profiling: Methods and Protocols, Richard A. Shimkets, editor, Humana Press, 2004.

10 **Detección de biomarcadores utilizando tecnologías de imagen molecular *in vivo***

En algunas realizaciones, un biomarcador descrito en el presente documento puede usarse en pruebas de imagen molecular. Por ejemplo, un agente de formación de imágenes puede acoplarse a un reactivo de captura, que puede usarse para detectar el biomarcador *in vivo*.

Las tecnologías de imagen *in vivo* proporcionan procedimientos no invasivos para determinar el estado de una enfermedad en particular en el cuerpo de un individuo. Por ejemplo, partes enteras del cuerpo, o incluso todo el cuerpo, pueden verse como una imagen tridimensional, proporcionando así información valiosa sobre la morfología y las estructuras del cuerpo. Dichas tecnologías se pueden combinar con la detección de los biomarcadores descritos en el presente documento para proporcionar información concerniente al biomarcador *in vivo*.

El uso de tecnologías de imagen molecular *in vivo* se está expandiendo debido a diversos avances en la tecnología. Estos avances incluyen el desarrollo de nuevos agentes de contraste o marcadores, tales como marcadores radiactivos y/o marcadores fluorescentes, que pueden proporcionar señales fuertes dentro del cuerpo; y el desarrollo de una potente y nueva tecnología de obtención de imagen, que puede detectar y analizar estas señales desde el exterior del cuerpo, con suficiente sensibilidad y precisión para proporcionar información útil. El agente de contraste se puede visualizar en un sistema de formación de imágenes apropiado, proporcionando de ese modo una imagen de la porción o porciones del cuerpo en el que se encuentra el agente de contraste. El agente de contraste puede estar unido a o asociado con un reactivo de captura, tal como un aptámero o un anticuerpo, por ejemplo, y/o con un péptido o proteína, o un oligonucleótido (por ejemplo, para la detección de la expresión génica), o un complejo que contiene cualquiera de estos con una o más macromoléculas y/u otras formas particuladas.

El agente de contraste también puede presentar un átomo radiactivo que es útil en la obtención de imágenes. Los átomos radiactivos adecuados incluyen tecnecio-99m o yodo-123 para estudios gammagráficos. Otros restos fácilmente detectables incluyen, por ejemplo, marcadores de espín para imágenes de resonancia magnética (IRM) tales como, por ejemplo, yodo-123 de nuevo, yodo-131, indio-111, flúor-19, carbono-13, nitrógeno-15, oxígeno-17, gadolinio, manganeso o hierro. Tales marcadores son bien conocidos en la técnica y r un experto en la técnica podría seleccionarlos fácilmente.

Las técnicas de obtención de imágenes estándar incluyen, pero sin limitaciones, imágenes por resonancia magnética, tomografía computarizada, tomografía de emisión de positrones (PET), tomografía computarizada por emisión de fotón único (SPECT) y similares. Para la obtención de imágenes de diagnóstico *in vivo*, el tipo de instrumento de detección disponible es un factor principal en la selección de un agente de contraste dado, tal como un radionúclido dado y el biomarcador en particular que se usa como diana (proteína, ARNm y similares). El radionúclido elegido normalmente tiene un tipo de descomposición que es detectable por un tipo dado de instrumento. Asimismo, cuando se selecciona un radionúclido para el diagnóstico *in vivo*, su semivida debería ser lo suficientemente larga como para permitir la detección en el momento de la captación máxima por el tejido diana, pero lo suficientemente corto como para minimizar la radiación perjudicial del huésped.

Las técnicas de formación de imágenes a modo de ejemplo incluyen, pero sin limitaciones, PET y SPECT, que son técnicas de formación de imágenes en las que un radionúclido se administra de forma sintética o local a un individuo. La captación posterior del radiotrazador se mide a lo largo del tiempo y se utiliza para obtener información sobre el tejido diana y el biomarcador. Debido a las emisiones de alta energía (rayos gamma) de los isótopos específicos empleados y a la sensibilidad y sofisticación de los instrumentos utilizados para detectarlos, la distribución bidimensional de la radiactividad puede deducirse desde fuera del cuerpo.

Los nucleidos emisores de positrones de uso habitual en PET incluyen, por ejemplo, carbono-11, nitrógeno-13, oxígeno-15 y flúor-18. Los isótopos que se descomponen por captura de electrones y/o emisión gamma se usan en SPECT e incluyen, por ejemplo, yodo-123 y tecnecio-99m. Un procedimiento de ejemplo para marcar aminoácidos con tecnecio-99m es la reducción del ion pertechnetato en presencia de un precursor quelante para formar el complejo de tecnecio 99m precursor lábil, que, a su vez, reacciona con el grupo de unión a un metal de un péptido quimiotáctico bifuncionalmente modificado para formar un conjugado peptídico de tecnecio-99m quimiotáctico.

Los anticuerpos se usan con frecuencia para tales procedimientos de diagnóstico por imagen *in vivo*. La preparación y uso de anticuerpos para el diagnóstico *in vivo* se conocen bien en la técnica. De manera similar, los aptámeros

pueden usarse para tales procedimientos de diagnóstico por imagen *in vivo*. Por ejemplo, un aptámero que se usó para identificar un biomarcador concreto descrito en el presente documento puede marcarse e inyectarse apropiadamente en un individuo para detectar el biomarcador *in vivo*. El marcador utilizado se seleccionará de acuerdo con la modalidad de imágenes que se utilizará, como se ha descrito anteriormente. Los agentes de formación de imágenes dirigidos por el aptámero podrían tener características únicas y ventajosas relacionadas con la penetración del tejido, la distribución tisular, la cinética, eliminación, la potencia y la selectividad en comparación con otros agentes de formación de imágenes.

Dichas técnicas también se pueden realizar opcionalmente con oligonucleótidos marcados, por ejemplo, para la detección de la expresión génica a través de imágenes con oligonucleótidos antisentido. Estos procedimientos se usan para la hibridación *in situ*, por ejemplo, con moléculas fluorescentes o radionúclidos como marcador. Otros procedimientos para la detección de la expresión génica incluyen, por ejemplo, la detección de la actividad de un gen indicador.

Otro tipo general de tecnología de imagen es la imagen óptica, en la que las señales fluorescentes dentro del sujeto son detectadas por un dispositivo óptico que es externo al sujeto. Estas señales pueden deberse a la fluorescencia real y/o a la bioluminiscencia. Las mejoras en la sensibilidad de los dispositivos de detección óptica han aumentado la utilidad de las imágenes ópticas para los ensayos de diagnóstico *in vivo*.

Para una revisión de otras técnicas, véase N. Blow, *Nature Methods*, 6, 465-469, 2009.

Determinación de niveles de biomarcadores utilizando procedimientos de espectrometría de masas

Se pueden usar diversas configuraciones de espectrómetros de masas para detectar niveles de biomarcadores. Se dispone de varios tipos de espectrómetros de masas o pueden producirse con diversas configuraciones. En general, un espectrómetro de masas tiene los siguientes componentes principales: una entrada para muestras, una fuente de iones, un analizador de masas, un detector, un sistema de vacío y un sistema de control de instrumentos. La diferencia en la entrada para muestras, la fuente de iones y el analizador de masas generalmente define el tipo de instrumento y sus capacidades. Por ejemplo, una entrada puede ser una fuente de cromatografía líquida de columna capilar o puede ser una sonda o directa o paso tal como se utiliza en la desorción láser asistida por matriz. Las fuentes de iones comunes son, por ejemplo, electropulverización, incluyendo nanopulverización y micropulverización o desorción láser asistida por matriz. Los analizadores de masas comunes incluyen un filtro de masa cuadrupolar, un analizador de masas de trampa iónica y un analizador de masas de tiempo de vuelo. Los procedimientos de espectrometría de masas adicionales son bien conocidos en la técnica (véase Burlingame et al. *Anal. Chem.* 70:647 R-716R (1998); Kinter y Sherman, Nueva York (2000)).

Los biomarcadores de proteínas y los niveles de biomarcadores se pueden detectar y medir mediante cualquiera de los siguientes: espectrometría de masas por ionización por electropulverización (ESI-MS), ESI-MS/MS, ESI-MS/(MS)_n, ionización por desorción láser asistida por matriz espectrometría de masas de tiempo de vuelo (MALDI-TOF-MS), espectrometría de masas de tiempo de vuelo de desorción/ionización láser mejorada en superficie (SELDI-TOF-MS), desorción/ionización en silicio (DIOS), espectrometría de masas de iones secundarios (SIMS), tiempo de vuelo cuadrupolar (Q-TOF), tiempo de vuelo en tándem (TOF/TOF), llamado ultraflex III TOF/TOF, espectrometría de masas por ionización química a presión atmosférica (APCI-MS), APCI-MS/MS, APCI-(MS)_n, espectrometría de masas de fotoionización a presión atmosférica (APPI-MS), APPI-MS/MS y APPI-(MS)_n, espectrometría de masas cuadrupolar, espectrometría de masas por transformada de Fourier (FTMS), espectrometría de masas cuantitativa y espectrometría de masas de trampa iónica.

Las estrategias de preparación de muestras se utilizan para marcar y enriquecer muestras antes de la caracterización espectroscópica de masas de los biomarcadores de proteínas y la determinación de los niveles de biomarcadores. Los procedimientos de marcaje incluyen, pero sin limitaciones, el marcador isobárico para la cuantificación relativa y absoluta (iTRAQ) y el marcaje de isótopos estables con aminoácidos en cultivo celular (SILAC). Los reactivos de captura utilizados para enriquecer selectivamente muestras de proteínas biomarcadoras candidatas antes del análisis espectroscópico de masas incluyen, pero sin limitaciones, anticuerpos, sondas de ácido nucleico, quimeras, moléculas pequeñas, un fragmento F(ab')₂, un fragmento de anticuerpo monocatenario, un fragmento Fv, un fragmento Fv monocatenario, un ácido nucleico, una lectina, un receptor de unión a ligando, anticuerpos, nanocuerpos, anquirina, anticuerpos de dominio, armazones de anticuerpos alternativos (por ejemplo, diacuerpos, etc.), polímeros impresos, avimeros, peptidomiméticos, peptoides, ácidos peptidonucleicos, ácido nucleico treosa, un receptor hormonal, un receptor de citocinas y receptores sintéticos, y modificaciones y fragmentos de estos.

Determinación de niveles de biomarcadores usando un ensayo de ligadura de proximidad

Se puede usar un ensayo de ligadura de proximidad para determinar los valores de biomarcadores. Brevemente, una muestra de prueba se pone en contacto con un par de sondas de afinidad que pueden ser un par de anticuerpos o un par de aptámeros, con cada miembro del par extendido con un oligonucleótido. Las dianas para el par de sondas de afinidad pueden ser dos determinantes distintos en una proteína o un determinante en cada una de las

dos proteínas diferentes, que pueden existir como complejos homopoliméricos o heteromultiméricos. Cuando las sondas se unen a las dianas determinantes, los extremos libres de las extensiones de oligonucleótidos se ponen lo suficientemente cerca para hibridar juntos. La hibridación de las extensiones de oligonucleótidos se ve facilitada por un oligonucleótido conector común que sirve para unir entre sí las extensiones de los oligonucleótidos cuando se colocan en una proximidad suficiente. Una vez que las extensiones de oligonucleótidos de las sondas se hibridan, los extremos de las extensiones se unen mediante ligamiento enzimático de ADN.

Cada extensión de oligonucleótido comprende un sitio de cebador para la amplificación por PCR. Una vez que las extensiones de oligonucleótidos se ligan, los oligonucleótidos forman una secuencia continua de ADN que, a través de amplificación por PCR, revela información sobre la identidad y la cantidad de la proteína diana, así como, información sobre las interacciones proteína-proteína en las que los determinantes de la diana están en dos proteínas diferentes. La ligación de proximidad puede proporcionar un ensayo altamente sensible y específico para la información sobre la interacción y concentración de la proteína en tiempo real mediante el uso de PCR en tiempo real. Las sondas que no se unen a los determinantes de interés no tienen las correspondientes extensiones de oligonucleótidos puestas en proximidad y no puede procederse a la ligación ni a la amplificación por PCR, lo que da como resultado que no se produzca ninguna señal.

Los ensayos anteriores permiten la detección de valores de biomarcadores que son útiles en los procedimientos para la predicción del riesgo de eventos CV, en los que los procedimientos comprenden detectar, en una muestra biológica de un individuo, al menos cinco, al menos seis, al menos siete, al menos ocho o los nueve biomarcadores seleccionados de entre MMP12, angiopoyetina-2, complemento C7, troponina cardíaca I, proteína 4 relacionada con angiopoyetina, CCL18/PARC, complejo alfa-1-antiquimotripsina, GDF11 y alfa-2-antiplasmina, en el que una clasificación, tal como se describe más adelante, el uso de los valores de biomarcadores indica si e individuo tiene un riesgo elevado de sufrir un evento CV en el plazo de 1 año, 2 años, 3 años o 4 años. De acuerdo con cualquiera de los procedimientos descritos en el presente documento, los valores de los biomarcadores pueden detectarse y clasificarse individualmente o pueden detectarse y clasificarse colectivamente, como por ejemplo en un formato de ensayo múltiple.

Clasificación de biomarcadores y cálculo de puntuaciones de la enfermedad

En algunas realizaciones, una "firma" de biomarcador para una prueba de diagnóstico dada contiene un conjunto de biomarcadores, teniendo cada marcador niveles característicos en las poblaciones de interés. Los niveles característicos, en algunas realizaciones, pueden hacer referencia a la media o promedio de los niveles de biomarcador para los individuos en un grupo concreto. En algunas realizaciones, un método diagnóstico descrito en el presente documento puede usarse para asignar una muestra desconocida de un individuo a uno de dos grupos, bien a un riesgo mayor de evento CV no.

La asignación de una muestra a uno de dos grupos se conoce como clasificación y el procedimiento utilizado para realizar esta asignación se conoce como clasificador o procedimiento de clasificación. Los procedimientos de clasificación también pueden denominarse procedimientos de puntuación. Hay muchos procedimientos de clasificación que se pueden usar para construir un clasificador de diagnóstico a partir de un conjunto de niveles de biomarcadores. En algunos casos, los procedimientos de clasificación se realizan utilizando técnicas de aprendizaje supervisado en las que se recopila un conjunto de datos utilizando muestras obtenidas de individuos dentro de dos (o más, para múltiples estados de clasificación) grupos distintos que se desea distinguir. Dado que la clase (grupo o población) a la que pertenece cada muestra se conoce de antemano para cada muestra, se puede entrenar al procedimiento de clasificación para dar la respuesta de clasificación deseada. También es posible usar técnicas de aprendizaje no supervisado para producir un clasificador de diagnóstico.

Los enfoques comunes para desarrollar clasificadores de diagnóstico incluyen árboles de decisión; empaquetado + potenciación + bosques; aprendizaje basado en inferencia de reglas; ventanas de Parzen; modelos lineales; logística; procedimientos de redes neuronales; agrupamiento no supervisado; K-medias; ascendente/descendente jerárquico; aprendizaje semisupervisado; procedimientos prototipo; vecino más próximo; estimación de la densidad de Kernel; máquinas de vectores de soporte; modelos ocultos de Markov; aprendizaje Boltzmann; y los clasificadores pueden combinarse de manera simple o de maneras que minimicen funciones objetivo particulares. Para una revisión, véase, por ejemplo, Pattern Classification, R. O. Duda, et al., editores, John Wiley & Sons, 2ª edición, 2001; véanse también, The Elements of Statistical Learning - Data Mining, Inference, and Prediction, T. Hastie, et al., editores, Springer Science+Business Media, LLC, 2ª edición, 2009.

Para producir un clasificador usando técnicas de aprendizaje supervisado, se obtiene un conjunto de muestras llamadas datos de entrenamiento. En el contexto de las pruebas de diagnóstico, los datos de entrenamiento incluyen muestras de los distintos grupos (clases) a los que posteriormente se asignarán muestras desconocidas. Por ejemplo, las muestras recolectadas de individuos en una población de control y de individuos en una población de enfermedad en particular pueden constituir datos de entrenamiento para desarrollar un clasificador que pueda clasificar las muestras desconocidas (o, más particularmente, los individuos de los que se obtuvieron las muestras) como que tienen la enfermedad o están libres de la enfermedad. El desarrollo del clasificador a partir de los datos de entrenamiento se conoce como entrenamiento del clasificador. Los detalles específicos sobre el entrenamiento del

clasificador dependen de la naturaleza de la técnica de aprendizaje supervisado. El entrenamiento de un clasificador bayesiano ingenuo es un ejemplo de dicha técnica de aprendizaje supervisado (véase, por ejemplo, *Pattern Classification*, R. O. Duda, et al., editores, John Wiley & Sons, 2ª edición, 2001; véanse también, *The Elements of Statistical Learning - Data Mining, Inference, and Prediction*, T. Hastie, et al., editores, Springer Science+Business Media, LLC, 2ª edición, 2009). El entrenamiento de un clasificador bayesiano ingenuo se describe, por ejemplo, en las publicaciones de Estados Unidos n.º: 2012/0101002 y 2012/0077695.

Dado que, por lo general, hay muchos más niveles de biomarcadores potenciales que muestras en un conjunto de entrenamiento, se debe tener cuidado para evitar el exceso de ajuste. El exceso de ajuste ocurre cuando un modelo estadístico describe un error o ruido aleatorio en lugar de la relación subyacente. El exceso de ajuste puede evitarse de varias maneras, incluyendo, por ejemplo, limitando el número de biomarcadores utilizados para desarrollar el clasificador, suponiendo que las respuestas del biomarcador son independientes unas de otras, limitando la complejidad del modelo estadístico subyacente y asegurando que el modelo estadístico subyacente se ajuste a los datos.

Un ejemplo ilustrativo del desarrollo de una prueba de diagnóstico usando un conjunto de biomarcadores incluye la aplicación de un clasificador ingenuo de Bayes, un clasificador probabilístico simple basado en el teorema de Bayes con un tratamiento estrictamente independiente de los biomarcadores. Cada biomarcador se describe mediante una función de densidad de probabilidad dependiente de clase (pdf) para los valores de RFU medidos o los valores de log de RFU (unidades de fluorescencia relativas) en cada clase. Se supone que los pdf conjuntos para el conjunto de biomarcadores en una clase son el producto de los pdf individuales dependientes de clase para cada biomarcador. Entrenamiento de un clasificador ingenuo de Bayes en este contexto equivale a la asignación de parámetros ("parametrización") para caracterizar los pdf dependientes de la clase. Se puede usar cualquier modelo subyacente para los pdf dependientes de la clase, pero el modelo generalmente debe ajustarse a los datos observados en el conjunto de entrenamiento.

El rendimiento del clasificador ingenuo de Bayes depende del número y la calidad de los biomarcadores utilizados para construir y entrenar al clasificador. Un único biomarcador funcionará de acuerdo con su distancia KS (Kolmogorov-Smirnov). La adición de biomarcadores posteriores con buenas distancias KS ($> 0,3$, por ejemplo, en general, mejorará el rendimiento de la clasificación si los biomarcadores añadidos posteriormente son independientes del primer biomarcador. Usando la sensibilidad más la especificidad como una puntuación del clasificador, se pueden generar muchos clasificadores de puntuación con una variación de un algoritmo ambicioso. (Un algoritmo ambicioso es cualquier algoritmo que sigue la metaheurística de resolución de problemas de realizar la elección localmente óptima en cada etapa con la esperanza de encontrar el óptimo global).

Otra forma de representar el rendimiento del clasificador es a través de una característica operativa del receptor (ROC), o simplemente una curva ROC o un gráfico ROC. El ROC es una representación gráfica de la sensibilidad, o tasa positiva verdadera, frente a una tasa falsa positiva (1-especificidad o 1-tasa de verdadero negativo), para un sistema clasificador binario ya que su umbral de discriminación varía. El ROC también se puede representar de manera equivalente al trazar la fracción de positivos verdaderos de los positivos (TPR = tasa positiva de verdaderos) frente a la fracción de falsos positivos de los negativos (FPR = tasa de falsos positivos). También conocida como curva de características operativas relativas, porque es una comparación de dos características operativas (TPR y FPR) a medida que cambia el criterio. El área bajo la curva ROC (AUC) se usa habitualmente como una medida resumida de la precisión diagnóstica. Puede tomar valores de 0,0 a 1,0. Las AUC tienen una propiedad estadística importante: la AUC de un clasificador es equivalente a la probabilidad de que el clasificador clasifique un caso positivo elegido aleatoriamente por encima de un caso negativo elegido aleatoriamente (Fawcett T, 2006. An introduction to ROC analysis. *Pattern Recognition Letters*.27: 861-874). Esto es equivalente a la prueba de rangos de Wilcoxon (Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143, 29-36.). Otra forma de describir el rendimiento de una prueba de diagnóstico en relación con un patrón de referencia conocido es el índice de reclasificación neta: la capacidad de la nueva prueba para actualizar o reducir correctamente el riesgo en comparación con la prueba estándar de referencia. Véase, por ejemplo, Pencina et al., 2011, *Stat. Med.* 30: 11-21. Si bien la AUC bajo la curva ROC es óptima para evaluar el rendimiento de un clasificador de 2 clases, la medicina estratificada y personalizada se basa en la inferencia de que la población contiene más clases que 2. Para tales comparaciones, el cociente de riesgos instantáneos de los cuartiles superior frente a inferior (u otras estratificaciones, tales como los deciles) se puede usar de manera más apropiada.

Las predicciones del riesgo habilitadas a través de la presente invención pueden aplicarse a individuos en atención primaria o en centros cardiovasculares especializados, o incluso directamente al consumidor. En algunas realizaciones, los clasificadores utilizados para predecir eventos pueden implicar alguna calibración a la población a la que se aplican; por ejemplo, puede haber variaciones debido a la etnia o la geografía. Tales calibraciones, en algunas realizaciones, pueden establecerse de antemano a partir de grandes estudios de población, por lo que cuando se aplican a un paciente individual, estos se incorporan antes de realizar una predicción de riesgo. Se toma una muestra de sangre venosa, se procesa adecuadamente y se analiza como se describe en el presente documento. Una vez que el análisis se ha completado, las predicciones de riesgo pueden efectuarse matemáticamente, con o sin incorporación de otros metadatos de los registros médicos descritos en el presente

documento, tal como genéticos o demográficos. Son posibles diversas formas de salida de información dependiendo del nivel de experiencia del consumidor. Para los consumidores que buscan el tipo más simple de salida, la información puede ser, en algunas realizaciones, "es probable que esta persona sufra un evento en los próximos x años (donde x es 1-4), sí/no" o, como alternativa, similar a un "semáforo" rojo/naranja/verde o su equivalente verbal o escrito, como riesgo alto/medio/bajo. Para los consumidores que buscan mayor detalle, en algunas realizaciones, el riesgo puede aparecer como un número o un gráfico que ilustra la probabilidad de un evento por unidad de tiempo como una puntuación continua, o un mayor número de estratos (tales como deciles) y/o el tiempo promedio hasta el evento y/o el tipo más probable de evento. En algunas realizaciones, el resultado puede incluir recomendaciones terapéuticas. La monitorización longitudinal del mismo paciente a lo largo del tiempo permitirá gráficos que muestren la respuesta a intervenciones o cambios en el estilo de vida. En algunas realizaciones, se puede proporcionar más de un tipo de resultado al mismo tiempo para satisfacer las necesidades del paciente y de los miembros individuales del equipo de atención con diferentes niveles de experiencia.

En algunas realizaciones de la invención, los nueve biomarcadores que se muestran en la Tabla 3 (los "biomarcadores CVD9") se detectan en una muestra de sangre (tal como una muestra de plasma o una muestra de suero) de un sujeto, por ejemplo, usando aptámeros, tales como los aptámeros de tasa de disociación lenta. Los valores de log RFU se utilizan para calcular un índice de pronóstico (PI). A continuación se muestra una fórmula de PI de ejemplo no limitante:

$$PI = -16,61 + 1,55 \times \text{ANGPT2} - 1,22 \times \text{GDF11} + 2,12 \times \text{C7} - 2,64 \times \text{SERPINF2} + 0,57 \times \text{CCL18} + 1,02 \times \text{ANGPTL4} + 1,43 \times \text{KLK3} \cdot \text{SERPINA3} + 0,72 \times \text{MMP12} + 0,59 \times \text{TNNI3}, s = 0,85,$$

donde los niveles de proteína se toman en log₁₀ RFU. Un experto en la técnica apreciará que la fórmula de PI puede recalibrarse de acuerdo con la población de la que se toma el sujeto. Tal recalibración puede llevarse a cabo de acuerdo con los métodos descritos en el presente documento y/o métodos conocidos en la técnica.

Dado el PI, la probabilidad de que el sujeto sufra un evento cardiovascular (evento CV) en los próximos "t" años viene dada por la fórmula:

$$Pr[T \leq t] = 1 - e^{-e^{\frac{\text{Log}(t) - PI}{s}}},$$

donde PI es el *índice pronóstico* (o predictor lineal) y s es el parámetro de escala asociado para la distribución de valores extremos. En diversas realizaciones, "t" es de 5 años o menos, 4 años o menos, 3 años o menos o 2 años o menos.

Kits

Se puede detectar cualquier combinación de los biomarcadores describen en el presente documento utilizando un kit adecuado, tal como para usar en la realización de los procedimientos descritos en el presente documento. Además, cualquier kit puede contener uno o más marcadores detectables como se describe en el presente documento, tales como un resto fluorescente, etc.

En algunas realizaciones de la divulgación, un kit incluye (a) uno o más reactivos de captura (tales como, por ejemplo, al menos un aptámero o anticuerpo) para detectar uno o más biomarcadores en una muestra biológica, en el que los biomarcadores incluyen al menos cinco, al menos seis, al menos siete, al menos ocho o los nueve biomarcadores seleccionados de entre MMP12, angiopoyetina-2, complemento C7, troponina cardíaca I, proteína 4 relacionada con angiopoyetina, CCL18/PARC, complejo alfa-1-antiquimotripsina, GDF11 y alfa-2-antiplasmina, y, opcionalmente, (b) uno o más productos de software o programas de ordenador para clasificar al individuo de quien se obtuvo la muestra biológica como que tiene o no mayor riesgo de sufrir un evento CV o para determinar la probabilidad de que el individuo tiene un mayor riesgo de un evento CV, tal como se describe adicionalmente en el presente documento. Como alternativa, en lugar de uno o más productos de programas informáticos, se puede proporcionar una o más instrucciones para realizar manualmente las etapas anteriores por parte de un ser humano.

En algunas realizaciones, un kit comprende un soporte sólido, un reactivo de captura y un material generador de señal. El kit también puede incluir instrucciones para usar los dispositivos y reactivos, manipular la muestra y analizar los datos. Además, el kit puede usarse con un sistema informático o software para analizar e informar el resultado del análisis de la muestra biológica.

Los kits también pueden contener uno o más reactivos (por ejemplo, tampones de solubilización, detergentes, lavados o tampones) para procesar una muestra biológica. Cualquiera de los kits descritos en el presente documento también puede incluir, por ejemplo, tampones, agentes de bloqueo, materiales de matriz de espectrometría de masas, agentes de captura de anticuerpos, muestras de control positivo, muestras de control negativo, software e información, tales como protocolos, guías y datos de referencia.

En algunas realizaciones, se proporcionan kits para el análisis del estado de riesgo de evento CV, en el que los kits

comprenden cebadores de PCR para uno o más aptámeros específicos para biomarcadores descritos en el presente documento. En algunas realizaciones, un kit puede incluir además instrucciones de uso y correlación de los biomarcadores con predicción de riesgo de un evento CV. En algunas realizaciones, un kit también puede incluir una matriz de ADN que contiene el complementario de uno o más de los aptámeros específicos de los biomarcadores descritos en el presente documento, reactivos y/o enzimas para amplificar o aislar el ADN de la muestra. En algunas realizaciones, los kits pueden incluir reactivos para PCR en tiempo real, por ejemplo, sondas y/o cebadores TaqMan y enzimas.

Por ejemplo, un kit puede comprender (a) reactivos que comprenden al menos un reactivo de captura para determinar el nivel de uno o más biomarcadores en una muestra de prueba, y, opcionalmente, (b) uno o más algoritmos o programas de ordenador para realizar las etapas de comparar la cantidad de cada biomarcador cuantificado en la muestra de prueba a uno o más puntos de corte predeterminados. En algunas realizaciones, un algoritmo o programa de ordenador asigna una puntuación para cada biomarcador cuantificado basándose en dicha comparación y, en algunas realizaciones, combina las puntuaciones asignadas para cada biomarcador cuantificado para obtener una puntuación total. Además, en algunas realizaciones, un algoritmo o programa de ordenador compara la puntuación total con una puntuación predeterminada y usa la comparación para determinar si un individuo tiene un mayor riesgo de sufrir un evento CV. Como alternativa, en lugar de uno o más algoritmos o programas informáticos, se puede proporcionar una o más instrucciones para realizar manualmente las etapas anteriores por parte de un ser humano.

Procedimientos y software informáticos

Una vez que se selecciona un biomarcador o panel de biomarcadores, un procedimiento para diagnosticar a un individuo puede comprender lo siguiente: 1) obtener una muestra biológica; 2) realizar un procedimiento analítico para detectar y medir el biomarcador o biomarcadores en el panel en la muestra biológica; 3) opcionalmente realizar cualquier normalización o estandarización de datos; 4) determinar cada nivel de biomarcador; y 5) informar de los resultados. En algunas realizaciones, los resultados se calibran para la población/etnia del sujeto. En algunas realizaciones, los niveles de biomarcadores se combinan de alguna manera y se informa un valor único para los niveles de biomarcadores combinados. En este enfoque, en algunas realizaciones, la puntuación puede ser un número único determinado a partir de la integración de todos los biomarcadores que se compara con un valor umbral preestablecido que es una indicación de la presencia o ausencia de enfermedad. O la puntuación de diagnóstico o predictiva puede ser una serie de barras que representan un valor de biomarcador y el patrón de las respuestas se puede comparar con un patrón preestablecido para la determinación de la presencia o ausencia de enfermedad, afección o el mayor riesgo (o no) de sufrir un evento.

Al menos algunas realizaciones de los procedimientos descritos en el presente documento pueden implementarse con el uso de un ordenador. En la figura 12 se muestra un ejemplo de un sistema informático 100. Con referencia a la figura 12, el sistema 100 se muestra compuesto por elementos de hardware que están acoplados eléctricamente a través del bus 108, que incluye un procesador 101, un dispositivo de entrada 102, un dispositivo de salida 103, un dispositivo de almacenamiento 104, un lector de medios de almacenamiento legible por ordenador 105a, sistema de comunicaciones 106, aceleración de procesamiento (por ejemplo, DSP o procesadores de propósito especial) 107 y memoria 109. El lector de medios de almacenamiento legible por ordenador 105a está acoplado además a medios de almacenamiento legibles por ordenador 105b, representando la combinación de manera integral dispositivos de almacenamiento remotos, locales, fijos y/o extraíbles, además de medios de almacenamiento, memoria, etc., para información que puede leerse de forma temporal y/o permanente, que puede incluir el dispositivo de almacenamiento 104, memoria 109 y/o cualquier otro recurso de sistema 100 accesible. El sistema 100 también comprende elementos de software (que se muestran localizados actualmente dentro de la memoria de trabajo 191) que incluyen un sistema operativo 192 y otro código 193, tales como programas, datos y similares.

Con respecto a la Figura 12, el sistema 100 tiene una amplia flexibilidad y configurabilidad. Por lo tanto, por ejemplo, se puede utilizar una única arquitectura para implementar uno o más servidores que pueden configurarse adicionalmente de acuerdo con protocolos deseables, variaciones de protocolo, extensiones, etc. Sin embargo, será evidente para los expertos en la técnica que las realizaciones bien puede utilizarse de acuerdo con requisitos de aplicación más específicos. Por ejemplo, uno o más elementos del sistema podrían implementarse como subelementos dentro de un componente del sistema 100 (por ejemplo, dentro del sistema de comunicaciones 106). También se puede utilizar hardware personalizado y/o se pueden implementar elementos particulares en el hardware, el software o ambos. Además, aunque puede emplearse la conexión a otros dispositivos informáticos, tales como dispositivos de entrada/salida de red (no mostrados), debe entenderse que también se pueden utilizar conexiones cableadas, inalámbricas, módem y/o de otros tipos a otros dispositivos informáticos.

En un aspecto, el sistema puede comprender una base de datos que contiene características de biomarcadores característicos de predicción del riesgo de un evento CV. Los datos de biomarcadores (o información de biomarcadores) se pueden utilizar como una entrada al ordenador para usar como parte de un procedimiento implementado por ordenador. Los datos de biomarcadores pueden incluir los datos como se describe en el presente documento.

En un aspecto, el sistema comprende adicionalmente uno o más dispositivos para proporcionar datos de entrada al uno o más procesadores.

5 El sistema comprende además una memoria para almacenar un conjunto de datos de elementos de datos clasificados.

En otro aspecto, el dispositivo para proporcionar datos de entrada comprende un detector para detectar la característica del elemento de datos, por ejemplo, tal como un espectrómetro de masas o un lector de chips de genes.

10 El sistema adicionalmente puede comprender un sistema de gestión de base de datos. Las solicitudes o consultas de los usuarios pueden formatearse en un lenguaje apropiado que entienda el sistema de gestión de la base de datos que procesa la consulta para extraer la información relevante de la base de datos de los conjuntos de entrenamiento.

15 El sistema puede conectarse a una red a la que están conectados un servidor de red y uno o más clientes. La red puede ser una red de área local (LAN) o una red de área amplia (WAN), como se conoce en la técnica. Preferentemente, el servidor incluye el hardware necesario para ejecutar los productos del programa informático (por ejemplo, el software) para acceder a los datos de la base de datos para procesar las solicitudes del usuario.

20 El sistema puede incluir un sistema operativo (por ejemplo, UNIX o Linux) para ejecutar instrucciones desde un sistema de gestión de bases de datos. En un aspecto, el sistema operativo puede operar en una red de comunicaciones global, tal como internet y utilizar un servidor de red de comunicaciones global para conectarse a dicha red.

25 El sistema puede incluir uno o más dispositivos que comprenden una interfaz de visualización gráfica que comprende elementos de interfaz tales como botones, menús desplegables, barras de desplazamiento, campos para introducir texto y similares, como se encuentran rutinariamente en interfaces gráficas de usuario conocidas en la técnica. Las solicitudes introducidas en una interfaz de usuario pueden transmitirse a un programa de aplicación en el sistema para formatear para buscar información relevante en una o más de las bases de datos del sistema. Las solicitudes o consultas introducidas por un usuario se pueden construir en cualquier lenguaje de base de datos adecuado.

30 La interfaz gráfica de usuario puede generarse mediante un código de interfaz gráfica de usuario como parte del sistema operativo y puede usarse para ingresar datos y/o para visualizar los datos introducidos. El resultado de los datos procesados puede mostrarse en la interfaz, imprimirse en una impresora en comunicación con el sistema, guardarse en un dispositivo de memoria y/o transmitirse a través de la red o puede proporcionarse en forma de un medio legible por ordenador.

35 El sistema puede estar en comunicación con un dispositivo de entrada para proporcionar datos con respecto a elementos de datos al sistema (por ejemplo, valores de expresión). En un aspecto, el dispositivo de entrada puede incluir un sistema de perfil de expresión génica que incluye, por ejemplo, un espectrómetro de masas, un chip de gen o un lector de matriz, y similares.

40 Los procedimientos y aparatos para analizar la información del biomarcador de predicción del riesgo de evento CV de acuerdo con diversas realizaciones pueden implementarse de cualquier manera adecuada, por ejemplo, usando un programa informático que opera en un sistema informático. Se puede usar un sistema informático convencional que comprende un procesador y una memoria de acceso aleatorio, tal como un servidor de aplicaciones accesible remotamente, un servidor de red, un ordenador personal o una estación de trabajo. Los componentes adicionales del sistema informático pueden incluir dispositivos de memoria o sistemas de almacenamiento de información, tales como un sistema de almacenamiento masivo y una interfaz de usuario, por ejemplo un monitor, un teclado y dispositivo de seguimiento convencionales. El sistema informático puede ser un sistema autónomo o parte de una red de ordenadores que incluye un servidor y una o más bases de datos.

45 El sistema de análisis de biomarcador de predicción del riesgo de evento CV puede proporcionar funciones y operaciones para completar el análisis de datos, tal como recopilación de datos, procesamiento, análisis, notificación y/o diagnóstico. Por ejemplo, en una realización, el sistema informático puede ejecutar el programa informático que puede recibir, almacenar, buscar, analizar y notificar información relacionada con los biomarcadores de predicción de riesgo de evento CV. El programa informático puede comprender múltiples módulos que realizan diversas funciones u operaciones, tales como un módulo de procesamiento para procesar datos brutos y generar datos complementarios y un módulo de análisis para analizar datos brutos y datos complementarios para generar un estado y/o diagnóstico de predicción de evento CV o cálculo del riesgo. El cálculo del estado de riesgo para un evento CV puede comprender, opcionalmente, generar o recopilar cualquier otra información, incluida información biomédica adicional, con respecto a la afección del individuo en relación con la enfermedad, afección o evento, la identificación de si pueden ser deseables pruebas adicionales o la evaluación del estado de salud del individuo.

65

Algunas realizaciones descritas en el presente documento pueden implementarse para incluir un producto de programa informático. Un producto de programa informático puede incluir un medio legible por ordenador que tiene un código de programa legible por ordenador incorporado en el medio para hacer que un programa de aplicación se ejecute en un ordenador con una base de datos.

5 Tal como se usa en el presente documento, un "producto de programa informático" se refiere a un conjunto organizado de instrucciones en forma de enunciados de lenguaje natural o de programación que están contenidas en un medio físico de cualquier naturaleza (por ejemplo, escrito, electrónica, magnético, óptico u otro) y que se puede usar con un ordenador u otro sistema automatizado de procesamiento de datos. Tales enunciados de lenguaje de programación, cuando son ejecutadas por una computadora o sistema de procesamiento de datos, hacen que el ordenador o el sistema de procesamiento de datos actúen de acuerdo con el contenido particular de las enunciados. Los productos del programa informático incluyen sin limitación: programas en código fuente y objeto y/o bibliotecas de prueba o de datos integradas en un medio legible por ordenador. Además, el producto de programa informático que permite que un sistema informático o un dispositivo de procesamiento de datos actúe de formas preseleccionadas puede proporcionarse de varias formas, incluyendo, pero sin limitación, código fuente original, código de ensamblaje, código de objeto, lenguaje de la máquina, versiones encriptadas o comprimidas de lo anterior y todos los equivalentes.

20 En un aspecto, se proporciona un producto de programa informático para la evaluación del riesgo de evento CV. El producto de programa informático incluye un medio legible por ordenador que incorpora código de programa ejecutable por un procesador de un dispositivo o sistema informático, comprendiendo el código de programa: un código que recupera datos atribuidos a una muestra biológica de un individuo, en el que los datos comprenden niveles de biomarcadores que corresponden cada uno a uno de los biomarcadores de la Tabla 3; y un código que ejecuta un procedimiento de clasificación que indica un estado de riesgo de evento CV del individuo en función de los valores de biomarcadores.

30 En otro aspecto más, se proporciona un producto de programa informático para indicar una probabilidad de un evento CV. El producto de programa informático incluye un medio legible por ordenador que incorpora código de programa ejecutable por un procesador de un dispositivo o sistema informático, comprendiendo el código de programa: un código que recupera datos atribuidos a una muestra biológica de un individuo, en el que los datos comprenden un valor de biomarcador correspondiente a un biomarcador en la muestra biológica seleccionado entre los biomarcadores proporcionado en la Tabla 3; y un código que ejecuta un procedimiento de clasificación que indica un estado de riesgo de evento CV del individuo en función del valor de biomarcador.

35 Aunque se han descrito varias realizaciones como procedimientos o aparatos, debe entenderse que las realizaciones pueden implementarse a través del código acoplado a un ordenador, por ejemplo, un código residente en una computadora o accesible por el ordenador. Por ejemplo, el software y las bases de datos podrían utilizarse para implementar muchos de los procedimientos mencionados anteriormente. Por lo tanto, además de las realizaciones realizadas por hardware, también se observa que estas realizaciones se pueden realizar mediante el uso de un artículo de fabricación compuesto por un medio utilizable por ordenador que tiene un código de programa legible por ordenador incorporado en el mismo, lo que provoca la habilitación de las funciones divulgadas en esta descripción. Por lo tanto, se desea que las realizaciones también se consideren protegidas por esta patente en sus medios de código de programa también. Además, las realizaciones pueden incorporarse como código almacenado en una memoria legible por ordenador de prácticamente cualquier tipo que incluye, sin limitación, RAM, ROM, medios magnéticos, medios ópticos o medios magneto-ópticos. Incluso más generalmente, las realizaciones podrían implementarse en el software o en el hardware, o cualquier combinación de los mismos, incluyendo, pero sin limitación, software que se ejecuta en un procesador de propósito general, microcódigo, matrices lógicas programables (PLA) o circuitos integrados específicos de la aplicación (ASIC).

50 También se prevé que las realizaciones podrían realizarse como señales de ordenador incorporadas en una onda portadora, así como señales (por ejemplo, eléctricas y ópticas) propagadas a través de un medio de transmisión. Por lo tanto, los diversos tipos de información discutidos anteriormente podrían formatearse en una estructura, tales como una estructura de datos y transmitirse como una señal eléctrica a través de un medio de transmisión o almacenarse en un medio legible por ordenador.

55 También se observa que muchas de las estructuras, materiales y actos enumerados en el presente documento se pueden citar como medios para realizar una función o paso para realizar una función. Por lo tanto, debe entenderse que dicho lenguaje tiene derecho a cubrir todas las estructuras, materiales o actos desvelados en esta especificación y sus equivalentes,

60 El proceso de identificación de biomarcadores, la utilización de los biomarcadores divulgados en el presente documento y los diversos procedimientos para determinar los valores de biomarcadores se describen con detalle anteriormente con respecto a la evaluación del riesgo de un evento CV. Sin embargo, la aplicación del proceso, el uso de biomarcadores identificados y los procedimientos para determinar los valores de biomarcadores son totalmente aplicables a otros tipos específicos de afecciones cardiovasculares, o a cualquier otra enfermedad o afección médica o a la identificación de individuos que pueden o no ser beneficiados por un tratamiento médico

auxiliar.

Otros métodos

5 En algunas realizaciones, los biomarcadores y los métodos descritos en el presente documento se utilizan para determinar una prima de seguro médico o una decisión de cobertura y/o una prima de seguro de vida o una decisión de cobertura. En algunas realizaciones, los resultados de los métodos descritos en el presente documento se utilizan para determinar una prima de seguro médico y/o una prima de seguro de vida. En algunos de dichos casos, una organización que proporciona solicitudes de seguro médico o de seguro de vida u obtiene información sobre el riesgo de un sujeto de sufrir un evento de CV y usa esa información para determinar un seguro médico apropiado o una prima de seguro de vida para el sujeto. En algunas realizaciones, la prueba es solicitada y pagada por la organización que proporciona seguro médico o seguro de vida. En algunas realizaciones, el comprador potencial de una consulta o sistema de salud o compañía utiliza la prueba para predecir futuras responsabilidades o costes en caso de que la adquisición continúe.

15 En algunas realizaciones, los biomarcadores y los métodos descritos en el presente documento se utilizan para predecir y/o gestionar la utilización de los recursos médicos. En algunas de dichas realizaciones, los métodos no se llevan a cabo con el propósito de tal predicción, pero la información obtenida del método se utiliza en dicha predicción y/o gestión de la utilización de recursos médicos. Por ejemplo, un centro de pruebas u hospital puede recopilar información de los presentes métodos para muchos sujetos con el fin de predecir y/o administrar la utilización de recursos médicos en un centro particular o en un área geográfica particular.

Ejemplos

25 Los siguientes ejemplos se proporcionan solo con fines ilustrativos y no pretenden limitar el alcance de la aplicación tal como se define en las reivindicaciones adjuntas. Las técnicas de biología molecular rutinarias descritas en los siguientes ejemplos pueden llevarse a cabo tal como se describe en los manuales de laboratorio convencionales, tal como Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 3ª ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., (2001).

30

Ejemplo 1: Detección de biomarcadores de ejemplo usando aptámeros

Se describe un método de ejemplo para detectar uno o más biomarcadores en una muestra, por ejemplo, en Kraemer et al., *PLoS One* 6(10): e26332, y se describe a continuación. Se describen tres métodos diferentes de cuantificación: hibridación basada en micromatrices, un método basado en cuentas Luminex y qPCR.

35

Reactivos

40 HEPES, NaCl, KCl, EDTA, EGTA, MgCl₂ y Tween-20 se pueden adquirir, por ejemplo, de Fisher Biosciences. La sal sódica de sulfato de dextrano (DxSO₄), nominalmente 8.000 de peso molecular, puede adquirirse, por ejemplo, en AIC y se dializa contra agua desionizada durante al menos 20 horas con un intercambio. La ADN polimerasa KOD EX se puede adquirir, por ejemplo, en VWR. Se pueden adquirir cloruro de tetrametilamonio y CAPSO, por ejemplo, en Sigma-Aldrich y estreptavidina-ficoeritrina (SAPE), por ejemplo, en Moss Inc. El clorhidrato de 4-(2-aminoetil)-bencenosulfonilfluoruro (AEBSF) se puede adquirir, por ejemplo, en Gold Biotechnology. Las placas de 96 pocillos recubiertas con estreptavidina se pueden adquirir en, por ejemplo, Thermo Scientific (placas HBC transparentes de 96 pocillos recubiertas de estreptavidina Pierce, número de producto 15500 o 15501). Se puede comprar NHS-PEO4-biotina, por ejemplo, en Thermo Scientific (EZ-Link NHS-PEO4-Biotina, número de producto 21329), disuelta en DMSO anhidro y puede almacenarse congelada en alícuotas de un solo uso. IL-8, MIP-4, lipocalina-2, RANTES, MMP-7 y MMP-9 se pueden adquirir en, por ejemplo, R&D Systems. Resistina y MCP-1 se pueden adquirir, por ejemplo, en PeproTech, y se puede comprar tPA, por ejemplo, en VWR.

50

Ácidos nucleicos

55 Se pueden comprar oligodesoxinucleótidos convencionales (incluidos los sustituidos con amina y biotina), por ejemplo, en Integrated DNA Technologies (IDT). Bloque Z es un oligodesoxinucleótido monocatenario de secuencia 5'-(AC-BnBn) 7-AC-3', donde Bn indica un resto de desoxiuridina sustituido con bencilo. El bloque Z puede sintetizarse usando química convencional de fosforamidita. Los reactivos de captura de aptámeros también pueden sintetizarse mediante la química convencional de fosforamidita y pueden purificarse, por ejemplo, en una columna PRP-3 de 21,5x75 mm, funcionando a 80 °C en un sistema Waters Autopurification 2767 (o sistema semiautomático Waters serie 600), usando, por ejemplo, un calentador timberline TL-600 o TL-150 y un gradiente de bicarbonato de trietilamonio (TEAB)/ACN para eluir el producto. La detección se realiza a 260 nm y las fracciones se recogen a través del pico principal antes de agrupar las mejores fracciones.

60

Tampones

65

El tampón SB18 se compone de HEPES 40 mM, NaCl 101 mM, KCl 5 mM, MgCl₂ 5 mM y Tween 20 al 0,05 % (v/v)

ajustado a pH 7,5 con NaOH. El tampón SB17 es SB18 suplementado con EDTA trisódico 1 mM. El tampón PB1 se compone de HEPES 10 mM, NaCl 101 mM, KCl 5 mM, MgCl₂ 5 mM, EDTA trisódico 1 mM y Tween-20 al 0,05 % (v/v) ajustado a pH 7,5 con NaOH. El tampón de elución CAPSO consiste en CAPSO 100 mM a pH 10,0 y NaCl 1 M. El tampón de incubación contiene HEPES 500 mM, HCl 500 mM y 0,05 % (v/v) de Tween-20. El tampón de hibridación Agilent es una formulación patentada que se suministra como parte de un kit (kit de hibridación Oligo aCGH/ChIP-on-chip). El tampón de lavado 1 Agilent es una formulación patentada (kit de tampón de lavado 1 Oligo aCGH/ChIP-on-chip, Agilent). El tampón de lavado 2 Agilent es una formulación patentada (kit de tampón de lavado 2 Oligo aCGH/ChIP-on-chip, Agilent). La solución de hibridación TMAC consiste en cloruro de tetrametilamonio 4,5 M, EDTA trisódico 6 mM, Tris-HCl 75 mM (pH 8,0) y 0,15 % (v/v) de Sarkosyl. El tampón KOD (concentrado 10 veces) consiste en Tris-HCl 1200 mM, MgSO₄ 15 mM, KCl 100 mM, (NH₄)₂SO₄ 60 mM, 1 % v/v de Triton-X 100 y 1 mg/ml de BSA.

Preparación de la muestra

15 El suero (almacenado a -80 °C en alícuotas de 100 µl) se descongela en un baño de agua a 25 °C durante 10 minutos, luego se almacena en hielo antes de la dilución de la muestra. Las muestras se mezclan mediante centrifugación vorticial suave durante 8 segundos. Se prepara una solución de muestra de suero al 6 % por dilución en 0,94x SB17 suplementado con MgCl₂ 0,6 mM, EGTA trisódico 1 mM, AEBSF 0,8 mM y bloque Z 2 µM. Una porción de la solución madre de suero al 6 % se diluye 10 veces en SB17 para crear una solución madre de suero al 0,6 %. Se utilizan soluciones madre del 6 % y del 0,6 %, en algunas realizaciones, para detectar analitos de alta y baja abundancia, respectivamente.

Reactivo de captura (aptámero) y preparación de placas de estreptavidina

25 Los aptámeros se agrupan en 2 mezclas de acuerdo con la abundancia relativa de sus analitos afines (o biomarcadores). Las concentraciones de las soluciones madre son 4 nM para cada aptámero y la concentración final de cada aptámero es 0,5 nM. Las mezclas madre de aptámero se diluyen 4 veces en tampón SB17, se calientan a 95 °C durante 5 minutos y se enfrían a 37 °C durante un período de 15 minutos antes de su uso. Este ciclo de desnaturalización-renaturalización está destinado a normalizar las distribuciones de conformadores de aptámeros y, por lo tanto, a garantizar una actividad de aptámero reproducible a pesar de las historias variables. Las placas de estreptavidina se lavan dos veces con 150 µl de tampón PB1 antes de su uso.

Equilibrio y captura de placa

35 Las mezclas de 2 veces el aptámero (55 µl) sometidas a calor-enfriamiento se combinan con un volumen igual de diluciones de suero al 6 % o 0,6 %, produciendo mezclas de equilibrio que contienen 3 % y 0,3 % de suero. Las placas se sellan con una estera de sellado de silicona (estera de sellado de silicona Axymat, VWR) y se incuban durante 1,5 h a 37 °C. Las mezclas de equilibrio se transfieren después a los pocillos de una placa de estreptavidina de 96 pocillos lavada y se incuban adicionalmente en un equipo Eppendorf Thermomixer a 37 °C, con agitación a 800 rpm, durante dos horas.

Ensayo manual

45 A menos que se especifique otra cosa, el líquido se elimina mediante volcado, seguido de dos golpecitos sobre toallas de papel en capas. Los volúmenes de lavado son de 150 µl y todas las incubaciones con agitación se realizan en un Eppendorf Thermomixer a 25 °C, 800 rpm. Las mezclas de equilibrio se eliminan mediante pipeteo y las placas se lavan dos veces durante 1 minuto con tampón PB1 suplementado con sulfato de dextrano 1 mM y biotina 500 µM, después 4 veces durante 15 segundos con tampón PB1. Se añade una solución recién preparada de NHS-PEO4-biotina 1 mM en el tampón PB1 (150 µl/pocillo) y las placas se incuban durante 5 minutos con agitación. 50 La solución de NHS-biotina se retira y las placas se lavan 3 veces con tampón PB1 suplementado con glicina 20 mM, y 3 veces con tampón PB1. A continuación se añaden ochenta y cinco µl de tampón PB1 suplementado con DxSO₄ 1 mM a cada pocillo y las placas se irradian bajo una lámpara UV BlackRay (longitud de onda nominal 365 nm) a una distancia de 5 cm durante 20 minutos con agitación. Las muestras se transfieren a una placa nueva, lavada y recubierta con estreptavidina o a un pocillo no utilizado de la placa de estreptavidina lavada existente, combinando mezclas de dilución de muestra alta y baja en un solo pocillo. Las muestras se incuban a temperatura ambiente con agitación durante 10 minutos. El material sin adsorber se elimina y las placas se lavan 8 veces durante 15 segundos cada una con tampón PB1 suplementado con 30 % de glicerol. A continuación, las placas se lavan una vez con tampón PB1. Los aptámeros se eluyen durante 5 minutos a temperatura ambiente con 100 µl de tampón de elución CAPSO. 90 µl de la fracción eluida se transfieren a una placa de 96 pocillos HybAid y se añaden 10 µl de tampón de neutralización.

Ensayo semiautomatizado

65 Las placas con estreptavidina que portan las mezclas en equilibrio adsorbidas se colocan en la plataforma de una lavadora de placas BioTek EL406, que se programa para realizar en las siguientes etapas: el material no absorbido se elimina por aspiración y los pocillos se lavan 4 veces con 300 µl de tampón PB1 suplementado con sulfato de

dextrano 1 mM y biotina 500 μ M. A continuación, los pocillos se lavan tres veces con 300 μ l de tampón PB1. Se añaden ciento cincuenta μ l de una solución recién preparada (de una reserva madre de 100 mM en DMSO) de NHS-PEO4-biotina 1 mM en el tampón PB1. Las placas se incuban durante 5 minutos con agitación. Se aspira el líquido y los pocillos se lavan 8 veces con 300 μ l de tampón PB1 suplementado con glicina 10 mM. Se añaden cien μ l de
 5 tampón PB1 suplementado con sulfato de dextrano 1 mM. Después de estas etapas automatizadas, las placas se retiran del lavador de placas y se colocan en un termo-agitador montado bajo una fuente de luz UV (BlackRay, longitud de onda nominal de 365 nm) a una distancia de 5 cm durante 20 minutos. El agitador térmico se ajusta a 800 rpm y 25 °C. Después de 20 minutos de irradiación, las muestras se transfieren manualmente a una placa de estreptavidina recién preparada y lavada (o a un pocillo no utilizado de la placa lavada existente). Las mezclas de
 10 reacción de alta abundancia (mezcla de 3 % de suero + 3 % de aptámero) y de baja abundancia (mezcla de 0,3 % de suero + 0,3 % de aptámero) se combinan en un solo pocillo en este punto. Esta placa "Catch-2" se coloca en la plataforma de la lavadora de placas BioTek EL406, que se programa para realizar en las siguientes etapas: la placa se incuba durante 10 minutos con agitación. Se aspira el líquido y los pocillos se lavan 21 veces con 300 μ l de tampón PB1 suplementado con glicerol al 30 %. Los pocillos se lavan 5 veces con 300 μ l de tampón PB1, y se aspira el lavado final. Se añaden cien μ l de tampón de elución CAPSO, y los aptámeros se eluyen durante 5 minutos
 15 con agitación. Tras estas etapas automatizadas, se retira la placa de la plataforma del lavador de placas y se transfieren manualmente alícuotas de 90 μ l de las muestras a los pocillos de una placa HybAid de 96 pocillos que contiene 10 μ l de tampón de neutralización.

20 Hibridación a micromatrices Agilent 8x15k personalizados

Se transfieren 24 μ l de la fracción eluida neutralizada a una nueva placa de 96 pocillos y a cada pocillo se añaden 6 μ l de 10x Agilent Block (kit de hibridación Oligo aCGH/ChIP-on-chip, Large Volume, Agilent 5188-5380), que
 25 contiene un conjunto de controles de hibridación compuestos por 10 aptámeros Cy3. A cada pocillo se añaden 30 μ l de tampón de hibridación Agilent 2x a cada muestra y se mezclan. Cuarenta μ l de la solución de hibridación resultante se pipetea manualmente en cada "pocillo" del portaobjetos de tipo junta de hibridación (portaobjetos de tipo junta de hibridación, 8 micromatrices por formato de portaobjetos, Agilent). Los portaobjetos de micromatric Custom Agilent, con 10 sondas por matriz complementarias a la región aleatoria de 40 nucleótidos de cada aptámero con un enlazador dx 20x, se colocan sobre los portaobjetos de tipo junta de acuerdo con el protocolo del fabricante.
 30 El conjunto (kit de cámara de hibridación - habilitado para SureHyb, Agilent) se sujeta e incuba durante 19 horas a 60 °C mientras gira a 20 rpm.

Lavado posterior a la hibridación

35 Se colocaron aproximadamente 400 ml de tampón de lavado Agilent 1 en cada uno de dos placas de tinción de vidrio separadas. Los portaobjetos (no más de dos a la vez) se desmontan y se separan sumergidos en el tampón de lavado 1, a continuación se transfieren a una rejilla de portaobjetos en una segunda placa de tinción que también contiene tampón de lavado 1. Los portaobjetos se incuban durante 5 minutos adicionales en el tampón de lavado 1 con agitación. Los portaobjetos se transfieren al tampón de lavado 2 preequilibrado a 37 °C y se incuban durante 5
 40 minutos con agitación. Los portaobjetos se transfieren a una cuarta placa de tinción que contiene acetonitrilo y se incuban durante 5 minutos con agitación.

Formación de imágenes por micromatrices

45 Se toman imágenes de los portaobjetos de micromatrices con un sistema de escáner de micromatrices Agilent G2565CA, usando el canal Cy3 a una resolución de 5 μ m en el ajuste 100 % de PMT y la opción XRD habilitada en 0,05. Las imágenes TIFF resultantes se procesan con el software de extracción de características Agilent versión 10.5.1.1 con el protocolo GE1_105_Dec08.

50 Diseño de sonda Luminex

Las sondas inmovilizadas en perlas tienen 40 desoxinucleótidos complementarios al extremo 3' de la región aleatoria de 40 nucleótidos del aptámero diana. La región complementaria del aptámero se acopla a las microesferas Luminex a través de un enlazador de hexaetilenglicol (HEG) que lleva un terminal amino 5'. Los desoxioligonucleótidos de
 55 detección biotinilados comprenden 17-21 desoxinucleótidos complementarios a la región del cebador 5' de los aptámeros diana. Los restos de biotina se añaden a los extremos 3' de los oligos de detección.

Acoplamiento de sondas a microesferas Luminex

60 Las sondas se acoplan a las microesferas Luminex Microplex esencialmente según las instrucciones del fabricante, pero con las siguientes modificaciones: las cantidades de oligonucleótidos amino-terminales son 0,08 nMol por $2,5 \times 10^6$ microesferas, y la segunda adición de EDC es de 5 μ l a 10 mg/ml. Las reacciones de acoplamiento se realizan en un equipo Eppendorf ThermoShaker a 25 °C y 600 rpm.

65 Hibridación de microesferas

Las soluciones madre de microesferas (aproximadamente 40000 microesferas/ μ l) se agitan en vórtex y se sonicán en un limpiador ultrasónico Health Sonics (Modelo: T1.9C) durante 60 segundos para suspender las microesferas. Las microesferas suspendidas se diluyen a 2000 microesferas por reacción en soluciones de hibridación de 1,5x TMAC y se mezclan con agitación en vortex y sonicación. Treinta y tres μ l por reacción de la mezcla de perlas se transfieren a una placa HybAid de 96 pocillos. Se añaden a cada reacción siete μ l de reserva madre de oligonucleótidos de detección biotinilado 15 nM en 1x tampón TE y se mezclan. Se añaden diez μ l de muestra de ensayo neutralizada y la placa se sella con un sello de estera de con tapa de silicona. La placa se incuba primero a 96 °C durante 5 minutos y se incuba a 50 °C sin agitación durante la noche en un horno de hibridación convencional. Una placa de filtro (poro Dura, Millipore, número de pieza Millipore MSBVN1250, tamaño de poro de 1,2 μ m) se humedece previamente con 75 μ l de solución de hibridación TMAC 1x suplementada con 0,5 % (p/v) de BSA. Todo el volumen de muestra de la reacción de hibridación se transfiere a la placa de filtro. La placa de hibridación se enjuaga con 75 μ l de solución de hibridación 1 x TMAC que contiene BSA al 0,5 % y cualquier material restante se transfiere a la placa de filtro. Las muestras se filtran al vacío lento, con 150 μ l de tampón evacuado durante aproximadamente 8 segundos. La placa de filtro se lava una vez con 75 μ l de solución de hibridación TMAC 1x que contiene 0,5 % de BSA y las microesferas en la placa de filtro se resuspenden en 75 μ l de solución de hibridación TMAC 1x que contiene 0,5 % de BSA. La placa del filtro se protege de la luz y se incuba en un Eppendorf Thermalmixer R durante 5 minutos a 1000 rpm. La placa del filtro se lava una vez con 75 μ l de solución de hibridación TMAC 1x que contiene BSA al 0,5 %. Se añaden 75 μ l de 10 μ g/ml de estreptavidina ficoeritrina (SAPE-100, MOSS, Inc.) en una solución de hibridación TMAC 1x a cada reacción y se incuban en Eppendorf Thermalmixer R a 25 °C a 1000 rpm durante 60 minutos. La placa de filtro se lava dos veces con 75 μ l de solución de hibridación 1 x TMAC que contiene BSA al 0,5 % y las microesferas en la placa de filtro se resuspenden en 75 μ l de solución de hibridación TMAC 1x que contiene BSA al 0,5 %. La placa de filtro se incuba después protegida de la luz en un Eppendorf Thermalmixer R durante 5 minutos, 1000 rpm. La placa del filtro se lava una vez con 75 μ l de solución de hibridación TMAC 1x que contiene BSA al 0,5 %. Las microesferas se resuspenden en 75 μ l de solución de hibridación TMAC 1x suplementada con BSA al 0,5 % y se analizan en un instrumento Luminex 100 que ejecuta el software Xponent 3.0. Se cuentan al menos 100 microesferas por tipo de perla, bajo calibración PMT alta y una configuración de discriminador de doblete de 7500 a 18000.

Lectura QPCR

Las curvas estándar para qPCR se preparan en agua que van de 108 a 102 copias con diluciones de 10 veces y un control sin molde. Las muestras de ensayo neutralizadas se diluyen 40 veces en H₂O desionizada. La mezcla maestra de qPCR se prepara a una concentración final 2x (2x tampón KOD, mezcla de dNTP 400 μ M, mezcla 400 nM de cebadores directos e inversos, 2x verde SYBR Green I y 0,5 U de KOD EX). Se añaden diez μ l de mezcla maestra 2x qPCR a 10 μ l de muestra de ensayo diluida. La qPCR se ejecuta en un BioRad MyIQ iCycler con 2 minutos a 96 °C, seguido de 40 ciclos de 96 °C durante 5 segundos y 72 °C durante 30 segundos.

Ejemplo 2. Procedimientos

40 Diseño del estudio y recolección de muestras

Las muestras de plasma archivadas de sujetos con CHD estable se obtuvieron de dos estudios de cohortes independientes bien conocidos. Las características de la población del estudio se muestran en la Tabla 1. Los inventores descubrieron los biomarcadores de proteínas y entrenamiento modelo en 938 muestras de plasma del estudio Heart and Soul, con seguimiento posterior de 10 años. Véase, por ejemplo, Shlipak et al., Am J Med. 2008;121:50-57; Whooley et al., JAMA. 2008;300:2379-2388. Los inventores validaron el modelo en 971 muestras de HUNT3, un estudio noruego de cohortes prospectivo con seguimiento de 5 años. Véase Krokstad et al., Int J Epidemiol. 2013;42:968-977. Los inventores usaron los criterios de inclusión y exclusión de Heart and Soul para seleccionar a todos los participantes con CHD estable de la cohorte HUNT3 más grande para este análisis. Las muestras de plasma del descubrimiento fueron representativas de un estudio prospectivo académico bien controlado: los sujetos estaban en ayunas, las muestras se recogieron a la misma hora del día y se centrifugaron y se congelaron a -80 °C en una hora de la recolección. Por el contrario, la recolección de muestras en el conjunto de validación HUNT3 era representativa de las condiciones probables del "mundo real"; los sujetos no estaban en ayunas, se observaron en diferentes momentos del día y el plasma no se separó de las células durante hasta 24 horas, mientras las muestras permanecieron a 4 °C. Evaluar el rendimiento del modelo de esta manera permite determinar la solidez del modelo a los factores asociados con la recolección práctica de muestras clínicas, una consideración importante para la validación de biomarcadores. Véase McShane et al., Nature. 2013;502:317-320. Ambos estudios fueron aprobados por las juntas de revisión institucional relevantes.

Tabla 1: Características de la población de estudio

	Descubrimiento (Heart and Soul)	Validación (HUNT3)
Origen de la muestra	Estudio de cohortes prospectivo basado en UCSF en 12 clínicas ambulatorias en el área de la Bahía de San Francisco	Cohorte anidada de 1017 pacientes del estudio de cohorte prospectivo noruego en 50.807 participantes
Criterios de entrada	Diagnóstico de arteriopatía coronaria estable mediante IM previo, > 50% de estenosis en la angiografía, isquemia inducida por ejercicio, revascularización previa	Todos los sujetos tenían arteriopatía coronaria estable, seleccionados por los mismos criterios que para la cohorte Heart and Soul, excepto que los datos del ejercicio no estaban disponibles.
Procesamiento de muestras	Fechas de obtención: 2000-2002 En ayunas Hora fija del día Plasma EDTA Tiempo hasta la separación de las células normalmente <1 hora Almacenamiento a -80 °C	Fechas de obtención: 2006-2008 No en ayunas Hora aleatoria del día Plasma EDTA Tiempo hasta la separación de las células hasta 24 h, muestra mantenida a 4 °C Almacenamiento a -80 °C
Evento y adjudicación de eventos	Criterio de valoración de evento compuesto definido como el primero de: muerte por cualquier causa; hospitalización por infarto de miocardio; ictus o ataque isquémico transitorio; hospitalización por signos y síntomas de insuficiencia cardíaca. Cada evento fue juzgado por 2 revisores independientes y enmascarados. En caso de desacuerdo, los jueces conferidos, reconsideraron su clasificación y, si es necesario, solicitaron la consulta de un tercer juez enmascarado.	Las definiciones de eventos son las mismas que para descubrimiento, adjudicadas a partir de la revisión de registros médicos por un cardiólogo experimentado
Tiempo de seguimiento	Fecha del último seguimiento: 11,09 mediana de años (IQR) del tiempo de seguimiento: 7,9 (5,5) Años	Fecha del último seguimiento: 5,57 mediana de años (IQR) del tiempo de seguimiento: 4,3(1,0) Años
Enmascaramiento	Técnicos de laboratorio enmascarados para las características clínicas y los resultados. Adjudicación de resultados con enmascaramiento de los resultados proteómicos.	Técnicos de laboratorio enmascarados para las características clínicas y los resultados. Adjudicación de resultados con enmascaramiento de los resultados proteómicos.
Aplicación del modelo	Biomarcadores identificados, modelos entrenados, modelos aplicados	No se identificaron biomarcadores, no hay modelos entrenados, solo se aplicaron modelos entrenados para el Heart and Soul

Ensayo proteómico SOMAscan

- 5 Los reactivos de afinidad individual utilizados en el ensayo de proteínas son aptámeros de ADN modificados (SOMAmer) de tasa de disociación baja con una afinidad muy baja por sus dianas proteicas. Véase Vaught et al., J Am Chem Soc. 2010;132:4141-4151. Las modificaciones químicas de las bases de ADN en los aptámeros mejoran sus características de unión. Véase Davies et al., Proc Natl Acad Sci USA. 2012;109:19971-19976. Los inventores usaron 1130 de estos reactivos en el ensayo multiplex SOMAscan™ 8-10. En resumen, se incubaba una muestra de plasma en cada pocillo de una placa de 96 pocillos con una mezcla de SOMAmer que se unen a sus proteínas diana. Dos etapas de inmovilización basadas en perlas permiten la eliminación de las proteínas no unidas o unidas de forma inespecífica y la eliminación de los SOMAmer no unidos. Solo los reactivos unidos a la proteínas diana sobreviven al ensayo, siendo el número de cada uno proporcional cuantitativamente a la concentración de proteína en la muestra original. El ADN en cada reactivo se cuantifica en una matriz de hibridación Agilent y las muestras se normalizan y calibran de manera que el grado de fluorescencia en el lugar de la matriz se relacione con la concentración de una proteína específica. Las 1054 proteínas que pasaron el control de calidad tenían una mediana de coeficiente de variación intraensayo e interensayos <5 %. Véase Gold et al., PLoS One. 2010;5:e15004.

Ensayo y datos de SOMAscan

- 20 Las muestras de plasma se analizaron durante un período de 3 semanas hábiles en 32 ciclos de ensayos separados. Las muestras de estudio se asignaron aleatoriamente a los ciclos de ensayo junto con un conjunto de

muestras de calibración y control. No hubo información de identificación disponible para los técnicos de laboratorio que realizan el ensayo.

5 La normalización intraciclos y la calibración entre ciclos se realizaron de acuerdo con los procedimientos de control de calidad (QC) de datos del ensayo SOMAscan versión 3 como se define en el sistema de calidad de buenas prácticas de laboratorio (BPL) de SomaLogic. La calibración entre ciclos está diseñada para eliminar los "efectos por lotes" entre los ciclos sucesivos del ensayo, mientras que la normalización intraciclos elimina los cambios masivos en la concentración de proteínas (y, por lo tanto, la intensidad de la señal) entre las muestras dentro de cada ciclo.

10 Brevemente, la calibración entre ciclos escala el nivel de señal para cada proteína de modo que el nivel observado en el patrón de calibración del ciclo coincida con el nivel esperado representado por la referencia de calibración externa. Las tolerancias del QC se definen en términos de la magnitud de la escalada multiplicativa requerida para hacer coincidir el nivel mediano de la señal en los patrones de calibración replicados con los niveles de señal generados por referencia externa.

15 La normalización intraciclos controla los sesgos de intensidad de señal "masiva" que pueden ser el resultado de la eficiencia de hibridación diferencial o la dilución de muestras diferencial (u otros artefactos de protocolo de recolección) que cambian la concentración de proteína total en la muestra. El primer efecto es capturado por un conjunto de controles usados para monitorizar la reacción de hibridación para cada muestra y el segundo usa la mediana de la relación de los niveles medianos de señal en cada muestra con los niveles medianos de señal en todas las muestras dentro del ciclo. No es raro que las diferencias en el protocolo de recolección de muestras generen un sesgo sistemático de intensidad en los niveles de señal para una gran cantidad de proteínas. La Figura 1 muestra gráficos de cajas de los factores de escala multiplicativos en las dos cohortes cuando las 1130 proteínas se agrupan por dilución de muestra. Las proteínas medidas en las diluciones de muestra de 40 % y 1 % tenían niveles de señal sistemáticamente *más altas* (más bajas) en el conjunto de validación (descubrimiento), lo que da como resultado factores de escala de normalización correspondientes más pequeños (más grandes) que uno. Después del procedimiento de normalización, el nivel mediano de señal para las proteínas en cada una de las tres diluciones es el mismo en los conjuntos de descubrimiento y validación.

30 Los niveles de proteína se informan en unidades de fluorescencia relativa (RFU) y se transformaron logarítmicamente antes del análisis posterior.

Muestras y proteínas excluidas del análisis

35 Las proteínas se excluyeron del análisis si la tolerancia al control de calidad (QC) de la calibración entre ciclos asociada se excedía en al menos uno de los 32 ciclos de ensayo independientes. Esto sucedió para 76 proteínas; en muchos casos, la *mayoría* de los ciclos estaban dentro de la tolerancia requerida, pero, por simplicidad, los inventores eligieron excluir las 76 proteínas en el análisis de descubrimiento de biomarcadores presentado en el presente documento.

40 Las muestras se excluyeron del análisis de descubrimiento de biomarcadores por las siguientes razones: 1) incumplimiento de la tolerancia al QC de la normalización intraciclos, 2) un número inusualmente alto de valores atípicos o 3) evidencia de hemólisis como lo indican los niveles extremos de hemoglobina o los técnicos de ensayo que notan un color plasmático aberrante (rojo). Los valores atípicos de proteínas individuales se definieron como proteínas con niveles de señal fuera del intervalo dado por la mediana $\pm 6^*$ desviaciones medias absolutas (MADN)¹: las muestras de pacientes con valores atípicos en más del 5 % de las proteínas medidas se excluyeron del análisis. La Tabla 2 resume el número de muestras excluidas en función de cada criterio.

50 ¹ Sea $\Phi^{-1}(z)$ la inversa de la función de distribución acumulativa normal. Luego, para datos distribuidos normalmente, la estimación robusta MADN (x) = $\sigma * \Phi^{-1}(3/4)$, por lo que $3^*MADN \approx 2\sigma$ y el intervalo establecido es $\pm 4\sigma$ para las mediciones gaussianas.

Tabla 2: Muestras excluidas por criterios

	Criterios de normalización del QC fallidos	Hemólisis potencial	> 5 % de valores atípicos de proteínas	Total
Conjunto de descubrimiento	18	22	15	55
Conjunto de validación	10	27	10	47
total	28	49	25	

Métodos estadísticos

55 El resultado en este estudio se definió como el primer evento entre muerte, infarto de miocardio (IM), ictus, ataque isquémico transitorio (AIT) u hospitalización por insuficiencia cardíaca. Los inventores usaron modelos de riesgos proporcionales de Cox para estimar las asociaciones univariadas entre los niveles de proteínas y el riesgo de eventos cardiovasculares, como se expone a continuación.

Selección de proteínas predictivas del riesgo cardiovascular

Se usaron modelos de riesgos proporcionales de Cox de variable única para identificar un conjunto de proteínas asociadas individualmente con un mayor riesgo de eventos cardiovasculares secundarios. Con un nivel de significancia corregido de Bonferroni del 5 %, exactamente 200 proteínas se asociaron con un mayor riesgo de resultados cardiovasculares. Las gráficas de "volcán" en la Figura 2 muestran el logaritmo negativo del valor p estadístico de Wald en función del cociente de riesgos instantáneos, ya sea por desviación estándar de las unidades de fluorescencia relativa (RFU) (arriba) o entre los niveles extremos del indicador categórico para pertenencia al cuartil RFU (abajo). En el último caso, el cociente de riesgos instantáneos informado da el aumento en el riesgo experimentado por un sujeto en el cuartil de riesgo más alto (4^o) en comparación con un sujeto en el cuartil del riesgo más bajo (1^o).

Algunas de estas 200 proteínas están asociadas con tamaños de efecto relativamente pequeños, pero las 117 enumeradas en la Tabla 14 tienen cociente de riesgos instantáneos fuera del intervalo [0,75 - 1,25]. El examen de la estructura de correlación correspondiente (datos no mostrados) entre estas 200 proteínas revela varios grupos de proteínas con correlaciones pareadas similares. Una discusión exhaustiva de la función biológica de estos grupos de proteínas está más allá del alcance de este manuscrito y se analizará en otra parte.

El LASSO (Tibshirani, Stat Med 1997;16:385-95) se usó como procedimiento de detección selectiva variable para identificar un subconjunto de proteínas asociadas en conjunto con un mayor riesgo CV. La validación cruzada generalizada usando *coxnet* (Simon et al., Journal of Statistical Software 2011;39:1-13) en el paquete R *glmnet* (Friedman et al., Journal of Statistical Software 2010;33:1-22) para establecer el parámetro de regularización LASSO. Los inventores usaron la heurística de "un error estándar" (Hastie et al., Elements of Statistical Learning, Segunda ed. 2^a Ed: Springer; 2009) para establecer el nivel de regularización. La perturbación de la etapa de validación cruzada se usó como una simple verificación de la "estabilidad" del conjunto resultante de proteínas seleccionadas. Este análisis dio la confianza de que las proteínas incluidas en CVD9 son "estables" en la medida en que se seleccionarían la mayoría de las veces que se aplicó el procedimiento "LASSO seguido de eliminación hacia atrás". Para generar resultados reproducibles para el análisis subsiguiente, los inventores fijaron la semilla de número aleatorio en 1 antes de la validación cruzada de LASSO. La inicialización de este valor y el establecimiento del parámetro de regularización LASSO en el error estándar de valor 1 por encima del cual minimiza la desviación de probabilidad parcial validada cruzada da como resultado un modelo LASSO que contiene las 16 proteínas tratadas en el presente documento.

Los presentes inventores usaron LASSO solo para la selección de variables, prefiriendo el modelo de supervivencia totalmente paramétrico (Weibull) como modelo de pronóstico final. Este último tiene una representación simple y una forma matemática aceptable (Royston et al., Metodología de investigación médica BMC 2013; 13: 33; van Houwelingen, Stat Med 2000; 19: 3401-15) a la calibración para su uso en estudios de validación externos. La retroeliminación progresiva hacia atrás comenzada desde el modelo LASSO completo se usó para eliminar proteínas que no eran predictores significativos en *ausencia* de la restricción impuesta por la penalización LASSO. Cuando se utiliza el criterio de detención del criterio de información bayesiano (BIC) para equilibrar el rendimiento y la complejidad del modelo, la retroeliminación descartó 7 proteínas: catepsina H, receptor del EGF, receptor de la hormona de crecimiento, proteína TIM-3 de la membrana de linfocitos T, MMP-7, CDO relacionado con el oncogén de adhesión celular y la trombospondina-2 dan como resultado el modelo CVD9 de 9 proteínas que se muestra en la Tabla 3.

Tabla 3: Características de rendimiento analítico de los biomarcadores CVD9

Diana	% de dilución de la muestra	Calibrador % CV	Límites de cuantificación (pg/ml)		Intervalo (Log)	Precisión entre ensayos (n = 3) (% CV)		
			Inferior	superior		Baja	Media	Alta
Angiopoyetina-2	0,01	5,66	8.4x10 ⁰	1.8x10 ⁴	3,3	1,6	2,0	0,6
Complemento C7	0,00	8,66	1.7x10 ¹	3.0x10 ⁴	3,2	7,0	7,2	6,3
MMP12	0,40	4,89	2.8x10 ⁻¹	1.4x10 ⁴	4,7	1,1	2,0	2,5
troponina I, cardíaca	0,40	7,62	-	-	-	-	-	-
proteína 4 relacionada con angiopoyetina	0,40	6,46	1.4x10 ¹	1.4x10 ⁴	3,0	9,6	4,9	7,6
Complejo α1-antiquimotripsina	0,00	5,97	1.8x10 ⁰	1.9x10 ⁴	4,0	4,5	5,8	5,3
GDF11	0,01	6,03	6.5x10 ⁰	4.2x10 ³	2,8	1,5	1,2	3,6
CCL18/PARC	0,00	6,83	1.6x10 ⁻¹	2.6x10 ³	4,2	5,1	1,6	5,6
α 2-Anti-plasmina	0,00	7,46	9.8x10 ⁰	1.7x10 ⁴	3,2	3,7	2,7	1,8

Modelo CVD9

50

El modelo final (CVD9) contiene las 9 proteínas. Si bien el ajuste de este modelo para las variables clínicas mejoró ligeramente el ajuste (véase a continuación), estos ajustes no lograron una mejora significativa en el rendimiento de discriminación ni de la calibración sobre el logrado con el modelo de "solo proteínas" en el conjunto de descubrimiento. Esto nos llevó a designar CVD9 como el modelo "primario" de los inventores para evaluar el rendimiento de validación y un modelo que incluye la edad, el sexo, el estado de diabetes y la tasa de filtración glomerular estimada (TFGe) como modelo secundario.

Para un modelo de tiempo de fallo acelerado, la probabilidad de que ocurra un evento en el intervalo [0, t] viene dada por

$$\Pr[T \leq t] = 1 - e^{-e^{\frac{\text{Log}(t) - PI}{s}}}$$

donde PI es el *índice pronóstico* (o predictor lineal) y s es el parámetro de escala asociado para la distribución de valores extremos. Al ajustar el modelo, los inventores trabajaron con variables estandarizadas, en el presente documento, se ha absorbido la media de la población y la desviación estándar en el término de intersección para poder informar del índice pronóstico y el factor de escala como,

$$PI = -16,61 + 1,55 \times \text{ANGPT2} - 1,22 \times \text{GDF11} + 2,12 \times \text{C7} - 2,64 \times \text{SERPINF2} + 0,57 \times \text{CCL18} + 1,02 \times \text{ANGPTL4} + 1,43 \times \text{KLK3.SERPINA3} + 0,72 \times \text{MMP12} + 0,59 \times \text{TNNI3}, s = 0,85,$$

donde los niveles de proteína se toman en log₁₀ RFU.

Incorporación de variables clínicas

El estudio HUNT3 no se diseñó específicamente como un estudio de enfermedad cardiovascular, por lo que algunos parámetros del historial médico y las mediciones de laboratorio clínico que estaban disponibles en el conjunto de descubrimiento no estaban disponibles en el conjunto de validación (por ejemplo, fracción de eyección ventricular izquierda ecocardiográfica, hipertrofia ventricular izquierda, función diastólica). Con esto en mente, los inventores solo consideraron ajustar las variables clínicas que estaban disponibles en ambas colecciones y que diferían entre pacientes con eventos y aquellos sin ellos.

Cuando se añadieron a CVD9, las variables clínicas sexo (masculino), edad, diabetes (sí), inhibidores de la ECA (sí) y la tasa de filtración glomerular estimada (TFGe) de forma individual (y conjunta) aumentaron el ajuste del modelo combinado resultante (p < 0,001). El uso de inhibidores de la ECA o ARB no se incluyó en el modelo final, porque la medicación no estaba disponible en la cohorte HUNT3.

Además de las 9 proteínas utilizadas en CVD9, los inventores añadieron primero la edad y el sexo, y, a continuación, añadieron el estado de diabetes y la TFGe para dar a modelos adicionales que combinan proteínas y variables clínicas disponibles habitualmente que predicen el resultado. Las estimaciones puntuales de los coeficientes del predictor lineal del modelo de tiempo de fallo acelerado (AFT) y el parámetro de escala estimado para la distribución de valores extremos se enumeran en la Tabla 4. En la Tabla 4, las abreviaturas son ANGPT2 = "Angiopoyetina-2"; C7="Complemento C7"; SERPINF2 = "α2-antiplasmina"; CCL18 = "Ligando 18 de quimiocina (motivo C-C)" también conocido como "quimiocina pulmonar y regulada por activación (PARC)", ANGL4 = "proteína 4 relacionada con la angiopoyetina; KLK3.SERPINA3 = "complejo α1-antiquimotripsina"; TNNI3= "Troponina I, cardíaca"; y TFGe = "tasa de filtración glomerular estimada".

Tabla 4: Coeficientes estimados para 3 modelos candidatos

Término modelo	CVD9	CVD9 + Edad + Sexo	CVD9 + Edad + Sexo + Diabetes + TFGe
Escala	0,848	0,849	0,845
Intersección	-16,612	-18,614	-17,478
Diabetes = Sí	0	0	0,277
TFGe	0	0	-0,005
Edad	0	0,012	0,012
Sexo= masculino	0	0,358	0,391
ANGPT2	1,547	1,712	1,510
GDF11	-1,224	-1,320	-1,347
C7	2,115	2,092	2,025
SERPINF2	-2,643	-2,057	-1,663
CCL18	0,574	0,554	0,375
ANGL4	1,022	0,902	0,848
KLK3.SERPINA3	1,433	1,409	1,361
MMP12	0,718	0,522	0,436
TNNI3	0,588	0,587	0,610

Habitualmente se informan varias medidas diferentes de rendimiento de discriminación, los inventores informaron de una "estadística c", el índice de discriminación integrada (IDI) y el índice de reclasificación neta sin categoría (NRI).

- 5 La Tabla 5 enumera estas medidas de discriminación junto con el cociente de riesgos instantáneos Q4/Q1 y la estadística de Hosmer-Lemeshow para resumir el rendimiento de calibración para los 3 modelos. Los intervalos de confianza informados son IC del 95 % empíricos generados utilizando 100 muestras de bootstrap. La primera columna enumera el valor p para la prueba de razón de verosimilitud que compara los modelos ampliados con el modelo de referencia (solo proteína). La primera columna proporciona el valor p para la prueba de razón de verosimilitud (LR) que compara el modelo ampliado con el modelo de solo proteína. Las medidas posteriores de discriminación son el área ponderada bajo la curva ROC incidente/dinámica (C_T), el índice de discriminación integrado (IDI), el índice de reclasificación neta (NRI) y el cociente de riesgos instantáneos del cuarto al primer cuartil (Q4/Q1). Rendimiento de calibración evaluado con la estadística de Hosmer-Lemeshow.
- 10
- 15 La adición de variables clínicas cuyos valores basales distinguen los grupos de eventos y sin eventos proporciona una ligera mejora en las estimaciones puntuales de IDI, NRI (> 0) y cociente de riesgos instantáneos Q4/Q1, aunque la "estadística C" integrada de la AUC permanece esencialmente sin cambios.

Tabla 5: Medidas del rendimiento de discriminación y calibración en el conjunto de descubrimiento para el modelo CVD9

Modelo	Prueba LR	C ^T		IDI	NRI			Cociente de riesgos instantáneos Q4/Q1	Calibración	
		Año 1	Año 4		NRI (>0)	Evento	Ningún evento		Hosmer-Lemeshow X ²	valor p
Solo proteína (CVD9)	NA	0,76 (0,72-0,79)	0,74 (0,71-0,77)	0,15 (0,11-0,18)	0,57 (0,42-0,70)	0,16 (0,01-0,30)	0,41 (0,35-0,47)	8,2	8,14	0,42
CVD9 + Edad + Varón	0,0002	0,76 (0,73-0,79)	0,75 (0,72-0,77)	0,16 (0,12-0,20)	0,64 (0,50-0,76)	0,19 (0,08-0,30)	0,45 (0,37-0,50)	10,0	2,69	0,95
CVD9 + Edad + Varón + Diabetes + TFGe ²	6,3e-6	0,77 (0,73-0,79)	0,75 (0,72-0,77)	0,17 (0,13-0,20)	0,68 (0,54-0,80)	0,22 (0,11-0,32)	0,46 (0,39-0,52)	10,3	3,81	0,87

² Se usó la fórmula CKD-EPI 2009 TFGe porque estaba disponible en el conjunto tanto de descubrimiento como de validación.

Recalibración de CVD9 para validación

Antes de comparar el rendimiento de CVD9 con la puntuación de Framingham, ambos modelos fueron recalibrados para su uso en el conjunto de validación. Como en van Houwelingen ((Stat Med 2000;19:3401-15), los inventores usaron un modelo de calibración de tiempo de fallo acelerado de Weibull para recalibrar los coeficientes del modelo para su uso en la población de validación. Si se deja que PI sea el índice pronóstico y $H(t|PI)$ denote la función de riesgo acumulativo, entonces el modelo de calibración es

$$\log(H(t|PI)) = \gamma_0 + \gamma_1 PI + \gamma_2 e,$$

donde el término de error e , tiene una distribución de valores extremos. Denotando el riesgo acumulado basal por $H_0(t)$ y usando $H(t|PI) = H_0(t)e^{PI}$ da,

$$\log(H_0(t)) = \gamma_0 + (\gamma_1 - 1)PI + \gamma_2 e. \quad (1)$$

Una evaluación de calibración formal (llamada "validación por calibración" por Van Houwelingen) implica probar la hipótesis de calibración perfecta, $H_0: \gamma_0 = 0, \gamma_1 = 0, \gamma_2 = 1$. El ajuste del modelo (1) utilizando *survreg* del paquete R supervivencia (Therneau, A Package for Survival Analysis in S. R package versión 237-7 2014) proporciona los coeficientes de calibración enumerados en la Tabla 6.

Tabla 6: Coeficientes para el modelo de calibración de Weibull aplicado a CVD9 para su uso en el conjunto de validación.

CVD9				
	Estimación	IC del 95 %		valor p
$\hat{\gamma}_0$	-0,230	-0,4189	-0,0412	0,02
$\hat{\gamma}_1 - 1$	-0,998	-1,1752	-0,8212	0,98
$\log(\hat{\gamma}_2)$	0,149	0,0384	0,2598	0,008

La intersección ($\hat{\gamma}_0$) y el término de escala ($\hat{\gamma}_2$) indican que CVD9 necesita calibración antes de aplicarse a la cohorte de validación, aunque, como se trata a continuación, el sesgo sistemático de intensidad en el conjunto de validación es responsable de la mayor parte de la contribución al término de intersección.

Las muestras de sangre en el conjunto de validación HUNT3 se recolectaron utilizando un protocolo de recolección más indulgente que en el conjunto de descubrimiento y, como resultado, los inventores observaron un sesgo de intensidad sistemático en la mayoría de las 1054 proteínas medidas en las muestras de validación. Tal como se trata en el presente documento, este sesgo fue eliminado en gran medida por las etapas de normalización, aunque, como se muestra a continuación, queda un pequeño sesgo *residual* en los niveles de señal para las 9 proteínas utilizadas en el modelo CVD9. Este sesgo es un artefacto del proceso de normalización (las muestras de validación tienen niveles de señal *más altos* que las muestras de descubrimiento antes de la normalización, pero niveles de señal más bajos después) y, como se muestra a continuación, es en gran parte responsable del valor estimado del coeficiente, $\hat{\gamma}_0$.

La intersección de la línea de regresión robusta en la Figura 3 da una estimación del sesgo de intensidad común a las 9 proteínas en CVD9. Si dejamos que Δ denote el sesgo estimado y β_j y σ_j sean el coeficiente del modelo y la desviación estándar de la población para la proteína $j^{\text{ésima}}$, luego, aplicar CVD9 a los datos de validación da como resultado la adición del factor constante,

$$\Delta \sum_{j=1}^9 \frac{\beta_j}{\sigma_j} \quad (2),$$

sobre lo que de otro modo sería la contribución de la intersección del modelo. De esta manera, el sesgo de intensidad en las señales de proteínas aparece como una discrepancia en la escala de *tiempo* de la función de supervivencia basal en los conjuntos de descubrimiento y validación, precisamente el término asociado con el parámetro γ_0 en el modelo de calibración (1). Usando la estimación $\Delta = -0,056$ generada por la intersección de la regresión lineal robusta en (2) junto con los coeficientes del modelo CVD9 y las desviaciones estándar de la población, resta 0,23145 del predictor lineal, casi exactamente el valor ($\hat{\gamma}_0$) estimado para la intersección en el modelo de calibración. Por lo tanto, el sesgo de intensidad "residual" que queda después del procedimiento de normalización es en gran parte responsable de la magnitud de ($\hat{\gamma}_0$) en lugar de una discrepancia real entre las funciones de supervivencia basales en las cohortes de descubrimiento y validación.

El sesgo de la intensidad de la señal en la recolección de muestras HUNT3 es un aspecto de este conjunto de validación particular que no se espera generalizar a las muestras recolectadas bajo protocolos de recolección más estrictos. Teniendo esto en cuenta, los inventores evaluaron el rendimiento en el conjunto de validación utilizando el

modelo CVD9 *recalibrado* que se describe a continuación.

Cuando la distribución de tiempo del evento es Weibull con escala a y forma b, la función de supervivencia basal correspondiente es

5

$$\Pr_{\text{cal}}[T \geq t] = e^{-\left(\frac{t}{a}\right)^b},$$

que los inventores escriben en términos del riesgo basal acumulativo (H_0) como $\log(H_0) = b \log(t/a)$. Sustituyendo esto en el lado izquierdo de la ecuación (1), y usando los coeficientes de calibración en **¡Error! Fuente de referencia no encontrada.**, la expresión resultante para generar las puntuaciones del riesgo puede expresarse en forma del modelo de tiempo de fallo acelerado,

10

$$\log(t|z) = \beta_0^{\text{cal}} + \beta^{\text{cal}T} z + s^{\text{cal}} e,$$

15 con coeficientes modelo "calibrados":

$$\beta_0^{\text{cal}} = \log\left(ae^{\frac{y_0}{b}}\right), \beta_j^{\text{cal}} = -\frac{(y_1-1)}{b} \beta_j^{\text{cox}}, s^{\text{cal}} = \frac{y_2}{b}.$$

Usando estos coeficientes modelo, la puntuación del riesgo *calibrado* asociado se genera usando

20

$$\Pr_{\text{cal}}[T \leq t | z] = 1 - e^{-e^{u^{\text{cal}}(z)}},$$

donde

25

$$u^{\text{cal}}(z) = \frac{(\log t - [\beta_0^{\text{cal}} + \beta^{\text{cal}T} z])}{(s^{\text{cal}})}.$$

El índice de pronóstico (PI) resultante y el factor de escala de valores extremos para el modelo CVD9 *recalibrado* utilizado en el conjunto de validación son:

30

$$\left\{ \begin{array}{l} \text{PI} = -16,39 + 1,55 \times \text{ANGPT2} - 1,22 \times \text{GDF11} + 2,11 \times \text{C7} \\ \quad - 2,64 \times \text{SERPINF2} + 0,57 \times \text{CCL18} \\ + 1,02 \times \text{ANGPTL4} + 1,43 \times \text{KLL3.SERPINA3} + 0,72 \times \text{MMP12} + 0,59 \times \text{TNNI3} \\ s = 0,98 \end{array} \right.$$

Se construyeron modelos de calibración similares para las variantes de CVD9 que incluyen variables clínicas. Los coeficientes del modelo de calibración resultantes se indican en la Tabla 7. Como fue el caso de CVD9, los modelos que incluyen variables clínicas tenían el mismo sesgo sistemático de intensidad en las 9 proteínas y, juntos, este sesgo generó una contribución de -0,254 y -0,245 a las estimaciones (\hat{y}_0) en los respectivos modelos de calibración.

35

Tabla 7: Coeficientes para los modelos de calibración de Weibull aplicados a variantes de CVD9 que incluyen variables clínicas.

CVD9 + Edad + Sexo = "Varón"				
	Estimación	IC del 95 %		valor p
\hat{y}_0	-0,275	-0,4633	-0,08623	0,004
$\hat{y}_1 - 1$	-1,042	-1,2388	-0,8460	0,67
$\log(\hat{y}_2)$	0,1577			0,005
CVD9 + Edad + Sexo = "Varón" + Diabetes = "Sí" + TFGe				
	Estimación	IC del 95 %		valor p
\hat{y}_0	-0,228	-0,4228	-0,03383	0,03
$\hat{y}_1 - 1$	-1,079	-1,3041	-0,8532	0,49
$\log(\hat{y}_2)$	0,1737			0,002

40 Después de la identificación de proteínas significativamente asociadas con eventos cardiovasculares (después de la corrección de Bonferroni a un nivel de significancia del 5 %), los inventores usaron L1 penalizado (LASSO; véase Tibshirani, Stat Med. 1997;16:385-395) regresión de Cox con fines de selección de variables (proteína). En virtud de seleccionar simultáneamente variables y reducir los coeficientes concomitantes de mermas, el LASSO produce buenos modelos predictivos, como ha sido ampliamente demostrado. Véase Hastie et al., Elements of statistical learning, segunda ed. Springer; 2009. Tales enfoques de penalización de L1 son especialmente efectivos en

45

configuraciones de predictores de alta dimensión ejemplificadas por las 1054 proteínas de los inventores. Para obtener un modelo totalmente paramétrico, los inventores aplicaron la retroeliminación escalonada a un tiempo de fallo acelerado de Weibull usando el conjunto completo de proteínas seleccionadas por LASSO. Esto eliminó a los 7 contribuyentes menos importantes y dio como resultado el modelo parsimonioso de 9 proteínas (CVD9), un modelo de pronóstico completamente paramétrico en el espíritu de Framingham.

Como un patrón de referencia comparativo respetado, las predicciones del riesgo se generaron a partir del modelo de riesgo de evento secundario de Framingham (D'Agostino, et al., Am Heart J. 2000; 139: 272-281) recalibrado para su uso en los conjuntos de datos de descubrimiento y validación, como se expone a continuación.

D'Agostino presenta el siguiente modelo de tiempo de fallo acelerado para la predicción de eventos cardiovasculares secundarios:

$$\Pr[T \leq t] = 1 - e^{-e^{\frac{\text{Log}(t) - PI_{FR}}{s}}}$$

donde el índice pronóstico y el parámetro de escala para varones es

$$\text{Varones: } \begin{cases} PI_{FR} = 4,995 - 0,0145 \times \text{Edad} - 0,6738 \times \text{Ln} \left(\frac{\text{Col. total}}{\text{HDL}} \right) - 0,3042 \times \text{Estado de diabetes,} \\ s = 0,9994 \end{cases}$$

$$\text{Mujeres: } \begin{cases} PI_{FR} = 13,537 - 0,0225 \times \text{Edad} - 0,834 \times \text{Ln} \left(\frac{\text{Col. total}}{\text{HDL}} \right) - 0,7829 \times \text{Estado de diabetes} - 1,3713 \times \text{Ln}(\text{PAS}) - 0,3669 \times \text{Fumador} \\ s = 1,031 \end{cases}$$

Antes de comparar el modelo de Framingham con CVD9, los inventores recalibraron el modelo para su uso en los conjuntos de descubrimiento y validación.

Recalibrado de Framingham para conjuntos de descubrimiento y validación

Para recalibrar la puntuación de riesgo secundario de Framingham para su uso en el conjunto de descubrimiento y conjunto de validación, se utilizó un modelo de calibración de riesgo proporcional de Cox de variable única (an Houwelingen, Stat Med 2000;19:3401-15; Steyerberg. Clinical Prediction Models: Springer; 2010). Denotando la función de supervivencia basal por $S_0(t)$, la puntuación del riesgo de Framingham calibrado a 4 años es

$$\Pr_{\text{cal}}[T \leq t] = 1 - \hat{S}_0(4)^{e^{\hat{\beta} \times PI_{FR}}}$$

donde $\hat{\beta}$ *es la estimación del modelo de calibración ajustada a los valores del índice de pronóstico de Framingham en el conjunto de muestra particular y $\hat{S}_0(t)$ es la estimación de Kaplan-Meier de la función de supervivencia en esa población. La Tabla 8 enumera los coeficientes del modelo de calibración resultante.

Tabla 8: Coeficientes de calibración para el modelo de calibración de riesgo proporcional de Cox de variable única utilizado para recalibrar el riesgo secundario de Framingham para el conjunto de descubrimiento o validación respectivo

	Coeficiente estimado de calibración de Cox ($\hat{\beta}$ *)	Error típico	valor p
Discovery	0,472	0,066	<0,001
Validación	0,396	0,067	<0,001

El rendimiento de la calibración se evaluó evaluando el acuerdo entre la frecuencia de los eventos observados y predichos. La figura 4 muestra la frecuencia de los eventos predichos y observados para cada decil de riesgo para el modelo de Framingham en el descubrimiento y la Figura 5 en los conjuntos de validación. En cada caso, el cuadro de la izquierda muestra la puntuación original de Framingham y el cuadro de la derecha muestra la puntuación recalibrada utilizando el modelo con los coeficientes enumerados en la Tabla 8.

Si bien el rendimiento de calibración fue aceptable en la cohorte de descubrimiento, no se logró un nivel similar de acuerdo entre las frecuencias de eventos predichas y observadas en la cohorte de validación como se puede ver en la Figura 5. Los inventores informaron sobre la estadística χ^2 (Hosmer-Lemeshow) para resumir el rendimiento de calibración que se muestra gráficamente en la Figura 4 y la Figura 5, esta estadística y el valor p asociado se calcularon con la función plotCalibration en el paquete R predecible. Kundu et al., PredictABEL: Assessment of risk prediction models. Paquete R versión 12-1 2012. Los valores p de 0,70 y 0,02 para la prueba de Hosmer-Lemeshow son consistentes con una buena calibración del modelo de Framingham en el descubrimiento y una mala calibración

en la cohorte de validación.

Las entradas en la Tabla 9 resumen el rendimiento de discriminación y calibración de las puntuaciones de Framingham recalibradas en los conjuntos de descubrimiento y validación. Tal como se trata con mayor detalle en el presente documento, los inventores informaron de dos medidas de discriminación en el rendimiento, el cociente de riesgos instantáneos entre el cuarto y el primer cuartil y la "estadística C". Para el último índice de concordancia, los inventores notificaron el área ponderada bajo la curva ROC incidente/dinámica, C^T con para $\tau = 4$ años. Las estadísticas c son consistentes con una discriminación relativamente pobre del modelo de Framingham en las cohortes de descubrimiento y validación.

Tabla 9: Rendimiento de discriminación y calibración de modelos Framingham recalibrados en los conjuntos de descubrimiento y validación.

Conjunto de datos	Discriminación			Calibración	
	C^1 (Año 1)	C^4 (Año 4)	HR Q4/Q1	Hosmer-Lemeshow χ^2	valor p
Discovery	0,620	0,615	2,8	5,54	0,70
Validación	0,616	0,609	2,3	18,75	0,02

Como esta puntuación se validó para predicciones de hasta 4 años inclusive, los inventores usaron el intervalo de tiempo de cuatro años para las comparaciones de rendimiento con el modelo de proteína CVD9. También se calculó el índice de reclasificación neta sin categoría (NRI; Pencina et al., Stat Med. 2011; 30: 11-21) para el modelo de proteína CVD9 frente a Framingham como se trata más adelante. La puntuación de riesgo de Framingham se validó previamente solo para predicciones de IM y muerte, pero también se predicen eventos de ictus e insuficiencia cardíaca. Los inventores conservaron la puntuación de riesgo de evento secundario de Framingham como comparador porque en este estudio su rendimiento es similar en todos los tipos de eventos y porque se considera la puntuación de interés más probable para la comunidad científica para esta población. El proceso que generó el modelo de predicción de riesgo cardiovascular de múltiples proteínas y las métricas que lo comparan con la puntuación del riesgo de evento secundario de Framingham (D'Agostino et al., Am Heart J. 2000; 139: 272-281) se resumen en la Figura 6 y se analizan a continuación. El impacto de la adición de parámetros clínicos comúnmente disponibles (seleccionados de las variables que estaban disponibles en ambas cohortes y diferían entre pacientes con eventos y aquellos sin ellos) a CVD9 también se evaluó en modelos secundarios (véase anteriormente). Toda la computación estadística se realizó utilizando el lenguaje R para computación estadística. Véase R Core Team RfSC, Viena, Austria R: A language and environment for statistical computing. Manual. 2013.

Rendimiento de validación

El gráfico de bosques que se muestra en la Figura 7 muestra una comparación de los cociente de riesgos instantáneos para las 16 proteínas LASSO en los conjuntos de descubrimiento y validación. Con la excepción de la proteína 4 relacionada con la angiopoyetina y el complemento C7, los cociente de riesgos instantáneos para las proteínas individuales en el modelo CVD9 son similares en los conjuntos de descubrimiento y validación. Esta es una medida del rendimiento de validación de las proteínas individuales, el resto de esta sección analiza el rendimiento de validación de la combinación específica de esas proteínas que da como resultado el modelo CVD9.

El rendimiento de la calibración es particularmente importante cuando las predicciones del modelo se utilizan para notificar las decisiones clínicas. En primer lugar, los inventores evaluaron las estimaciones CVD9 de riesgo *absoluto* en la población de validación como en Steyerberg (Epidemiology 2010;21:128-38) y, a continuación, evaluaron el rendimiento de discriminación en términos de cambio en la estadística C y la reclasificación del riesgo con respecto al modelo de Framingham.

Calibración

El rendimiento de la calibración se evaluó evaluando la concordancia entre la frecuencia de eventos observados y predichos en el intervalo de cuatro años después de la muestra de sangre basal. La Figura 8 muestra la frecuencia de eventos predichos y observados para cada decil de riesgo en el conjunto de validación para CVD9 (izquierda) y el modelo de Framingham (derecha) recalibrado para su uso en el conjunto de validación.

En todo el intervalo, la frecuencia de eventos prevista en un decil de riesgo dado generado por CVD9 está dentro del 8 (y normalmente dentro del 3) por ciento de la frecuencia de eventos observada. Cada barra derecha de cada par de barras representa aproximadamente 100 pacientes y, como indican las barras de error, las puntuaciones de riesgo para los pacientes en cada decil son más similares entre sí que las de los pacientes en los deciles de riesgo vecinos. Es en este sentido que se habla de evaluación del riesgo "individualizada" cuando se considera la información proporcionada por las proteínas en CVD9.

En general, la concordancia entre las frecuencias de eventos predichas y observadas es más débil en el modelo de

Framingham (particularmente para los pacientes en los percentiles de riesgo 10-20). La Figura 9 muestra las curvas de predicción (Pepe et al., *Stat Med* 2013;32:1467-82) para CVD9 y la puntuación de Framingham se recalibró para su uso en el conjunto de validación.

- 5 Con las puntuaciones del riesgo de los dos modelos en la misma escala en la Figura 9, se ve que el modelo CVD9 genera una representación más precisa del riesgo absoluto que Framingham en ambos extremos del espectro de riesgo al predecir correctamente el riesgo (bajo) de los sujetos por debajo del percentil 10 y el riesgo catastróficamente alto (65 %) para los sujetos por encima el percentil 90 de riesgo. Además, la pendiente de la curva de predicción para CVD9 es más pronunciada en la mitad superior de los percentiles de riesgo, lo que indica que
- 10 CVD9 proporciona una estimación de resolución más fina del riesgo absoluto para los pacientes en cada decil de riesgo que el modelo tradicional de Framingham.

Discriminación

- 15 Las entradas en la Tabla 10 resumen el rendimiento de discriminación de los modelos CVD9 y Framingham recalibrados para la cohorte de validación. Como índice de concordancia, los inventores notificaron el área ponderada bajo la curva ROC incidente/dinámica, C^T , asociado con el intervalo de seguimiento fijo $[0, \tau]$ (Heagerty et al., *Biometrics* 2005;61:92-105), que es equivalente a Harrell, Pencina and D'Agostino's "C-statistic". Véase, por ejemplo, Harrell et al., *Stat Med* 1996;15:361-87; y Pencina et al., *Stat Med* 2012;31:1543-53. Se calculó C^T para $\tau =$
- 20 1 y 4 años usando la función `risksetAUC` en el paquete R `riskSetROC`. Véase Heagerty et al., `risksetROC`: Riskset ROC curve estimation from censored survival data. R Package versión 104 2012.

Tabla 10: Resumen de rendimiento de discriminación y calibración para modelos de CVD y Framingham recalibrados para su uso en cohorte de validación.

Modelo	C _t		NRI		Evento		Ningún evento	IDI	Cociente de riesgos instantáneos Q4/Q1	Calibración	
	Año 1	Año 4	NRI(>0)		Evento	Hosmer-Lemeshow X ²				valor p	
CVD9	0,71 (0,68-0,74)	0,70 (0,67-0,73)	52 % (37-67 %)	18 % (5-31 %)	34 % (26-41 %)	0,10 (0,07-0,13)	6,0	7,90	0,44		
CVD9 + Edad + Varón	0,69 (0,66,0,73)	0,68 (0,65,0,71)	41 % (27-55 %)	16 % (3-29 %)	26 % (19-33 %)	0,08 (0,05-0,12)	4,9	1,51	0,99		
CVD9 + Edad + Varón + Diabetes + TFGe	0,67 (0,64-0,71)	0,65 (0,63-0,69)	35 % (22-51 %)	11 % (1-25 %)	24 % (18-30 %)	0,07 (0,04-0,09)	4,6	9,99	0,266		
Framingham	0,616 (0,58-0,66)	0,609 (0,58,0,64)	-	-	-	0,02	2,3	18,75	0,02		

Curvas ROC

La Figura 10 muestra las curvas ROC para los modelos de CVD9 y Framingham generados con el paquete riskSetROC para los conjuntos de descubrimiento y validación. Se generaron curvas ROC en los años uno y cuatro para cada modelo.

Reclasificación de riesgos

Las probabilidades de eventos a cuatro años se generaron con CVD9 y el modelo secundario de Framingham con este último recalibrado para su uso en el conjunto de descubrimiento. El índice de reclasificación neta libre de categoría¹⁹ NRI (> 0) se calculó utilizando el paquete R *nricsens*. Véase Eisuke. NRI for risk prediction models with time to event and binary response data. R Package versión 12 2013.

La Tabla 11 enumera los términos en NRI (> 0) y las probabilidades de reclasificación que comparan CVD9 con Framingham en los conjuntos de descubrimiento y validación. Los intervalos de confianza notificados son intervalos empíricos del 95 % calculados con 100 muestras de bootstrap. Como se trata en la Sección ¡Error! Fuente de referencia no encontrada., tanto el modelo de CVD9 como el de Framingham se recalibraron para su uso en el conjunto de validación antes de este cálculo.

Tabla 11: Índices de reclasificación neta y probabilidades de reclasificación para CVD9 en comparación con el modelo de Framingham en el conjunto de descubrimiento.

	Descubrimiento			Validación		
	Estimación puntual	Límite inferior del IC del 95 %	Límite superior del IC del 95 %	Estimación puntual	Límite inferior del IC del 95 %	Límite superior del IC del 95 %
NRI(>0)	0,57	0,43	0,71	0,52	0,38	0,67
Evento NRI	0,16	0,04	0,29	0,18	0,06	0,31
Ningún evento NRI	0,41	0,34	0,47	0,34	0,27	0,41
Pr (Riesgo ascendente Evento)	0,58	0,52	0,64	0,59	0,53	0,65
Pr (Riesgo descendente Evento)	0,42	0,36	0,48	0,41	0,35	0,47
Pr (Riesgo descendente sin evento)	0,71	0,67	0,74	0,67	0,63	0,70
Pr (Riesgo ascendente sin evento)	0,29	0,26	0,33	0,33	0,30	0,37

Ejemplo 3. Resultados

Características basales

Las características clínicas de las dos poblaciones de estudio al inicio se resumen en la Tabla 12. Como cabía esperar, los factores de riesgo conocidos son significativamente más prevalentes en los grupos con eventos. Hubo menos eventos generales en HUNT3 que en Heart and Soul, debido a un seguimiento más corto; no obstante, las poblaciones fueron generalmente comparables en las tasas de eventos por unidad de tiempo y la distribución de los tipos de eventos. En la Tabla 12, los valores P están asociados con la prueba exacta de Fisher para covariables categóricas y la prueba U de Mann-Whitney para covariables continuas. Valores continuos resumidos con mediana y rango intercuartílico (IQR). El conjunto de validación HUNT3 no se diseñó como un estudio de CHD y, como resultado, parte de la información clínica no estaba disponible y está marcada como N/A. Leyendas: IMC = índice de masa corporal; ECA = enzima convertidora de angiotensina; ARB= bloqueador del receptor de angiotensina; LDL-C = colesterol de lipoproteínas de baja densidad; HDL-C = colesterol de lipoproteínas de alta densidad; TG = triglicéridos; TFGe = tasa de filtración glomerular estimada.

Tabla 12: Características basales de la población de estudio

Resumen del evento	Conjunto de descubrimiento (Heart and Soul)			Conjunto de validación (HUNT3)		
	Ningún evento (10 años)	Evento (10 años)	Valor p	Ningún evento (5 años)	Evento (5 años)	Valor p
N.º de sujetos	473	465		699	272	
Edad (años)	64,0 (57,0-71,0)	71,0 (63,0-78,0)	<0,001	67,6 (60,0-75,3)	75,9 (67,9-81,0)	<0,001

(continuación)

Resumen del evento	Conjunto de descubrimiento (Heart and Soul)			Conjunto de validación (HUNT3)		
	Ningún evento (10 años)	Evento (10 años)	Valor p	Ningún evento (5 años)	Evento (5 años)	Valor p
Varón (%)	361 (76,3)	412 (88,6)	<0,001	508(72,7)	192(70,6)	<0,001
Caucásico (%)	275 (58,1)	290 (62,5)	0,18	NA	NA	NA
Diabetes (%)	96 (20,3)	151 (32,6)	<0,001	84 (12,0)	49 (18,0)	0,02
Tabaquismo, actual (%)	88 (18,6)	96 (20,7)	0,46	142 (20,3)	56 (20,6)	0,93
IMC (kg/m ²)	28,2 (25,2-31,6)	27,1 (24,4-30,5)	<0,001	27,9 (25,8-30,7)	28,0 (25,3-30,8)	0,89
Medicamentos cardiovasculares						
Estatina (%)	275 (58,1)	290 (62,5)	0,17	NA	NA	NA
ECA/ARB (%)	89 (23,5)	89 (39,9)	<0,001	NA	NA	NA
Beta-bloqueante (%)	88 (18,6)	96 (20,7)	0,69	NA	NA	NA
Aspirina (%)	28,2(25,2-31,6)	27,1 (24,4-30,5)	0,26	NA	NA	NA
Análisis de laboratorio						
LDL-C (mg/dl)	98,0 (83,0-121)	100 (81,2-124,0)	0,76	NA	NA	NA
HDL-C (mg/dl)	44 (36-54)	42(35-52)	0,044	46,4 (38,7-54,1)	42,5 (34,8-50,3)	0,002
Colesterol total (mg/dl)	174 (152-196)	169 (146-201)	0,29	174 (151-197)	178 (147-209)	0,32
TG (mg/dl)	110,0 (74,8-169)	111,0 (74,0-163,0)	0,78	142,0 (106,9-195,0)	142,0 (106,0, 195,0)	0,22
TFGe ³ (ml/min)	78,2 (65,3-91,9)	65,8 (52,3-82,7)	<0,001	70,9 (58,5-82,9)	60,3 (46,7-73,6)	<0,001
Creatinina (mg/dl)	1,0 (0,8-1,1)	1,1 (0,9-1,3)	<0,001	1,0 (0,9-1,1)	1,1 (0,9-1,3)	<0,001

¹ CKD-EPI 2009 en el conjunto de validación y CKD-EPI 2012 en el conjunto de descubrimiento cuando sea posible con CKD-EPI 2009 cuando los valores faltantes impidieron el cálculo de la fórmula de 2012.
³ CKD-EPI 2009 en el conjunto de validación y CKD-EPI 2012 en el conjunto de descubrimiento cuando sea posible con CKD-EPI 2009 cuando los valores faltantes impidieron el cálculo de la fórmula de 2012.

Proteínas relacionadas con el riesgo cardiovascular (CV)

5 Con un nivel de significancia corregido de Bonferroni del 5 %, el análisis de regresión de Cox univariado reveló que 117 de las 1054 proteínas que pasaron el control de calidad estaban asociadas con un mayor riesgo de eventos cardiovasculares y también tenían cocientes de riesgos instantáneos del cuarto al primer cuartil univariados de > 1,25 o <0,75 (estas 117 proteínas se enumeran en la Tabla 14 a continuación). Algunas de estas proteínas estaban correlacionadas, lo que sugiere la presencia de mucho menos de 117 procesos biológicos distintos; la biología de estas proteínas será la diana de un análisis posterior. Los cociente de riesgos instantáneos para las 16 proteínas seleccionadas de esta lista por el proceso LASSO y el subconjunto de 9 proteínas elegidas para el modelo CVD9 final se muestran en la Figura 7. Las propiedades biológicas relevantes de las 16 proteínas seleccionadas con LASSO se resumen a continuación.

15 Los biomarcadores identificados en este análisis no solo sirven para derivar un potente modelo de predicción de riesgo cardiovascular, sino que también informan de la comprensión de la biología de la enfermedad cardiovascular (CVD) e identifica posibles dianas farmacológicas y opciones de tratamiento. A continuación, los inventores dan una breve descripción de la función o funciones conocidas de las 16 proteínas seleccionadas por LASSO en el modelo de predicción de riesgo CV.

20 Crecimiento y remodelación

25 El factor de diferenciación del crecimiento 11 (GDF11) es un ejemplo de descubrimiento biológico utilizando una herramienta de ensayo de proteómica sin sesgos con hallazgos de potencial importancia clínica. Usando SOMAScan, Lee et al. identificaron la pérdida de GDF11 relacionada con la edad como la causa de la hipertrofia cardíaca relacionada con la edad en ratones. Véase Loffredo et al., Cell 2013;153:828-39. GDF11 está ahora bajo investigación activa por su papel en la supresión de la hipertrofia cardíaca y la insuficiencia cardíaca diastólica en seres humanos. Véase, por ejemplo, Olsen et al., Journal of the American College of Cardiology 2014;63:A780. De forma interesante, mientras que las concentraciones de GDF-11 se reducen con el aumento del riesgo de eventos cardiovasculares en el presente estudio, un inhibidor de la actividad GDF11, La similar a folistatina 3 se asocia positivamente con el aumento del riesgo cardiovascular (véase la Tabla 14).

El receptor del factor de crecimiento epidérmico (EGF) (EGFR) se expresa en monocitos y macrófagos en lesiones ateroscleróticas. La activación por unión del ligando estimula la proliferación celular y la quimiotaxis. Dreux et al., *Atherosclerosis* 2006;186:38-53. La evidencia de estudios en animales muestra que el receptor de EGF protege

5 contra la hipertrofia cardíaca y apoya la arquitectura apropiada de la pared vascular y la reactividad de los vasos. Schreier et al, *Hypertension* 2013;61:333-40.

Las formas solubles del receptor de la hormona del crecimiento (GHR) y los factores de crecimiento epidérmico pueden servir como reservorios e inhibidores de los factores circulantes implicados en la mitogénesis, la función

10 celular, y tienen funciones bien conocidas en el cáncer. De forma interesante, señalización del receptor de la hormona del crecimiento, a través de la estimulación de su mediador anabólico el factor I de crecimiento similar a la insulina, Ya se ha demostrado que tiene una correlación negativa con el riesgo de desarrollar arteriopatía coronaria. Juul et al., *Circulation* 2002;106:939-44.

La angiopoyetina-2 (ANGPT2), que antagoniza la actividad de la angiopoyetina-1 en el receptor Tie-2 del receptor de

15 tirosina-proteína quinasa y actúa en concierto con la angiopoyetina-1 durante la angiogénesis, promueve la relajación de los contactos de la matriz celular y puede inducir la apoptosis de las células endoteliales y la rotura de los vasos durante la angiogénesis²⁶. Maisonpierre et al., *Science* 1997;277:55-60. Un miembro de la misma familia de genes, La proteína 4 relacionada con la angiopoyetina (ANGPTL4) es inducida por hipoxia y no solo afecta a la

20 función vascular y la interacción de las células endoteliales de la matriz, pero también el metabolismo de los lípidos como un potente inhibidor de la lipoproteína lipasa²⁷. Li et al., *Current opinion in lipidology* 2006;17:152-6.

Las interacciones controladas de la matriz extracelular y las células son vitales para la fisiología normal de los

25 órganos, durante el desarrollo normal, en respuesta a lesiones vasculares y miocárdicas, y durante metástasis de cáncer. Las metaloproteinasas de la matriz y sus inhibidores tienen varias dianas en la matriz extracelular vascular y se han asociado con la estabilidad de la placa aterosclerótica, la formación de aneurismas y otras enfermedades cardiovasculares. Dollery et al., *Cardiovascular research* 2006;69:625-35. La metaloproteinasa de matriz (MMP)-7 y MMP12 están representadas en nuestro modelo predictivo, mientras que TIMP1 también tiene una asociación significativa con el riesgo cardiovascular (véase la Tabla 14). La trombospondina-2 (THBS2) media las interacciones

30 célula-célula y matriz-célula vasculares y cardíacas y se ha implicado en la regulación de la angiogénesis, la trombosis y la inflamación. El aumento de la concentración sérica de trombospondina-2 se asocia con el riesgo de mortalidad cardíaca en varones mayores. Golledge et al., *The American journal of cardiology* 2013;111:1800-4. El CDO relacionado con el oncogén de la adhesión celular CDON) es un miembro de la proteína de la superficie celular de la superfamilia de Ig/fibronectina implicada en la miogénesis y la adhesión celular muscular. Tenzen et al.,

35 *Developmental cell* 2006;10:647-56. Su papel en la interacción célula-célula se ha observado en la invasividad tumoral, pero se sabe poco sobre su relación con el sistema cardiovascular.

Inflamación

Representando los complejos papeles de la inflamación y la inmunidad en la enfermedad cardiovascular, este

40 modelo incorpora el ligando 18 de la quimiocina inflamatoria Quimiocina (motivo C-C), anteriormente conocida como quimiocina pulmonar y regulada por activación CCL18/PARC, que es una quimiocina elaborada por monocitos/macrófagos que parece estar implicada en el reclutamiento de linfocitos T. Chenivresse et al., *J Immunol* 2012;189:128-37. Los niveles plasmáticos de CCL18/PARC están elevados durante los episodios de angina

45 inestable y también se ha descubierto que predicen eventos CV en pacientes con angina estable. De Sutter et al., *Journal of molecular and cellular cardiology* 2010;49:894-6. La inmunoglobulina de linfocitos T y la proteína 3 que contiene el dominio de mucina (TIM-3) está implicada en la activación de macrófagos y otras actividades del sistema inmunitario. Anderson, *Expert opinion on therapeutic targets* 2007;11:1005-9.

El complemento C7 (C7) es uno de los 5 componentes que forman el complejo del complemento terminal bioactivo

50 (TCC). El TCC depositada en las células endoteliales da como resultado la proliferación celular, la liberación de factores de crecimiento y citocinas inflamatorias, y aumento de la expresión del factor tisular. El TCC también estimula la proliferación de células musculares lisas en placas ateroscleróticas. Speidl et al., *JTH* 2011;9:428-40. En pacientes con insuficiencia cardíaca sintomática, el TCC soluble en suero elevado predice resultados adversos

55 (muerte, trasplante de corazón urgente u hospitalización con insuficiencia cardíaca agravada). Clark et al., *Am Heart J* 2001;141:684-90. Complemento C9, otro miembro del TCC, también está elevado en el estudio de los inventores (tabla 14).

Proteasas

El complejo α 1-antiquimotripsina (SERPINA3) representa la forma unida del inhibidor de la proteasa α 1-

60 antiquimotripsina que tiene varios sustratos biológicos. Puede modular múltiples procesos de enfermedades agudas y crónicas, incluida la presión arterial. Tang et al., *Clin Exp Hypertens* 2008;30:648-61. La α 2-antiplasmina (SERPINF2) es un inhibidor de la serina proteasa (SERPIN) que inactiva la plasmina y, por lo tanto, reduce la fibrinólisis. Matsuno et al., *Journal of thrombosis and haemostasis*: *JTH* 2003;1:1734-9; Mutch et al., *JTH*

65 2007;5:812-7. La catepsina H (CTSH) es una cisteína proteinasa lisosómica importante en la degradación de las

proteínas lisosómicas³⁹. Cheng et al., *Circulation* 2012;125:1551-62. Sin embargo, su relación con la enfermedad CV hasta el presente estudio ha sido incierta. Lutgens et al., *FASEB J.* 2007;21:3029-41.

Marcador de necrosis miocárdica

5 A diferencia de muchas de las proteínas mencionadas que están potencialmente implicadas en las vías causales de las enfermedades cardiovasculares, la troponina I es un marcador bien establecido de necrosis de cardiomiocitos y de riesgo cardiovascular.

10 Aplicación del modelo de riesgo CVD9

Se calculó el riesgo de CVD9 para cada sujeto, se dividió en cuartiles y las curvas de supervivencia sin eventos a 5 años resultantes se muestran en la Figura 11. Los cociente de riesgos instantáneos Q4/Q1 para CVD9 son 8,2 en el conjunto de descubrimiento y 6,0 en el conjunto de validación; para la puntuación del riesgo secundario de Framingham (recalibrado para su uso en estas poblaciones), el cociente de riesgos instantáneos Q4/Q1 es 2,8 en la cohorte de descubrimiento y 2,3 para la cohorte de validación.

Los inventores también evaluaron el rendimiento comparativo de los modelos CVD9 frente a Framingham utilizando el índice de reclasificación neta y la estadística C a 1 año (un punto de tiempo recomendado por el National Heart, Lung and Blood Institute Working Group; véase Eagle et al., *Circulation.* 2010; 121: 1447-1454) y en el horizonte temporal máximo validado de 4 años para el modelo de Framingham. Véase D'Agostino et al., *Am Heart J.* 2000;139:272-281. Tal como se muestra en la Tabla 13, el modelo de predicción de riesgo CVD9 ofrece mejoras sustanciales en la discriminación, evidenciado por aumentos en la estadística C de 0,14/0,09 y un NRI sin categoría de 0,57/0,54 en cohortes de descubrimiento/validación respectivamente. Existe una buena concordancia para el modelo CVD9 entre las frecuencias de eventos observadas y predichas (calibración) en la cohorte de validación. La adición de parámetros clínicos y demográficos comúnmente disponibles (edad, sexo, diabetes y la tasa de filtración glomerular estimada) no mejoraron significativamente el modelo de CVD9 (Tabla 13). Los datos comparativos de rendimiento para todos los modelos se muestran en la Tabla 13. En la Tabla 2, NRI (> 0) = índice de reclasificación neta libre de categoría, TFGe = tasa de filtración glomerular estimada.

Tabla 13: Rendimiento comparativo del modelo CVD9 y el modelo Framingham

	Secundario de Framingham	Solo proteína (CVD9)	CVD9 + Edad + Sexo	CVD9 + Edad + Sexo + Diabetes + TFGe
Todas las métricas se muestran en el formato: descubrimiento/validación				
Estadística C A 1 año	0,62/0,62	0,76/0,71	0,76/0,69	0,77/0,67
Estadística C a 4 año	0,62/0,61	0,74/0,70	0,75/0,68	0,75/0,65
4 años NRI (> 0) frente a Framingham		0,57/0,52	0,64/0,41	0,67/0,35
Evento NRI frente a Framingham		0,16/0,18	0,19/0,16	0,20/0,11
Ningún evento NRI frente a Framingham		0,41/0,34	0,45/0,26	0,47/0,24
Cociente de riesgos instantáneos Q4/Q1	2,8/2,3	8,2/6,0	10,0/4,9	10,3/4,6

Ejemplo 4: Análisis

35 En este estudio, los inventores buscaban mejorar la predicción de los resultados cardiovasculares, especialmente a corto plazo, utilizando el descubrimiento de biomarcadores en el análisis proteómico más grande realizado hasta la fecha. Usaron la tecnología de aptámeros modificada para analizar 1130 proteínas plasmáticas en la cohorte de descubrimiento de 938 pacientes con CHD estable y validaron los hallazgos en una cohorte independiente de 971 pacientes. En la cohorte de descubrimiento, los inventores descubrieron 117 proteínas pronósticas del criterio de valoración cardiovascular compuesto con cociente de riesgos instantáneos superiores al 25 % de la unidad (Tabla 14). A partir de estos, los inventores construyeron un modelo multivariable que consiste en 9 proteínas (CVD9) cuyo rendimiento fue superior al de los factores de riesgo tradicionales o biomarcadores sanguíneos descritos en la literatura (véase, por ejemplo, Eagle et al., *Circulation.* 2010;121:1447-1454; D'Agostino et al., *Am Heart J.* 2000;139:272-281; y Pearson et al., *Circulation.* 2003; 107: 499-511), que muestra las ventajas potenciales de la proteómica de base amplia en comparación con los enfoques basados en candidatos. Las proteínas biomarcadoras individuales y el modelo CVD9 se replicaron bien en la cohorte de validación, a pesar de la baja calidad de la muestra de sangre consistente con la práctica clínica típica.

50 La aplicación adecuada de estrategias preventivas y terapéuticas se basa en el sistema de clasificación de riesgos que permite a los profesionales de la salud orientar los tratamientos más intensivos a las personas de mayor riesgo. Véase Eagle et al., *Circulation.* 2010;121:1447-1454. Los enfoques de uso habitual se basan en evaluaciones del riesgo basadas en factores de riesgo tradicionales y tienen limitaciones. Muchos de estos factores de riesgo son

condiciones no modificables crónicas o incluso fijas, tales como el sexo, la raza, la edad avanzada o los antecedentes familiares. No es sorprendente, que sean mucho más adecuados para predecir el riesgo a largo plazo (10 años) o de por vida que el riesgo a corto plazo. Los factores de riesgo tradicionales predicen eventos secundarios particularmente pobres en sujetos con CHD prevalente. La identificación de pacientes con riesgo alto de eventos cardiovasculares a corto plazo representa una importante necesidad no satisfecha, ya que identificaría a las personas con mayor necesidad urgente de prevención cardiovascular, intervención y cumplimiento de los tratamientos prescritos.

Se han propuesto varias tecnologías "ómicas" para complementar los factores de riesgo tradicionales en la evaluación del riesgo cardiovascular. Véase, por ejemplo, McShane et al., *Nature*. 2013;502:317-320. Entre ellos, las puntuaciones de riesgo genómico se han investigado más ampliamente. Los enfoques genómicos basados en polimorfismos comunes de un solo nucleótido no han logrado mejorar la discriminación de riesgo o la reclasificación sobre los factores de riesgo tradicionales, juzgado por las mismas métricas que se vieron afectadas favorablemente por la puntuación proteómica CVD9 en el presente estudio (estadística *c* y reclasificación neta). Véase, por ejemplo, Paynter et al., *JAMA*. 2010;303:631-637; Ripatti et al., *Lancet*. 2010;376:1393-1400. Incluso si los enfoques genómicos son finalmente exitosos, será en la predicción del riesgo a largo plazo en lugar del riesgo a corto plazo ya que los factores de riesgo genéticos no cambian con el tiempo y ejercen su efecto a través de exposiciones de por vida. En comparación con la genómica, la proteómica ofrece varias ventajas potenciales. La proteómica integra influencias ambientales y genéticas, los niveles de proteínas pueden cambiar con el tiempo, que reflejan los beneficios o daños de los tratamientos o los cambios en el estilo de vida y las proteínas a menudo se encuentran en las vías causales de las enfermedades y, por lo tanto, en dianas potenciales de las terapias. Véase, por ejemplo, Nissen et al., *N Engl J Med*. 2005;352:29-38; Ridker et al., *Lancet*. 2009;373:1175-1182; y Stein et al., *N Engl J Med*. 2012;366:1108-1118.

Se usó una nueva plataforma proteómica que consta de aptámeros modificados para medir 1130 proteínas en un pequeño volumen (<100 µl) de plasma. Se descubrieron 117 biomarcadores proteicos candidatos de riesgo cardiovascular (Tabla 14). Notablemente, muchas de estas proteínas no se han notificado previamente como biomarcadores de riesgo cardiovascular. A partir de estas proteínas, se construyó un modelo totalmente paramétrico parsimonioso utilizando un enfoque estadístico (LASSO junto con la retroeliminación) en lugar del enfoque biológico. En este proceso, se omiten algunas proteínas con cocientes de riesgos instantáneos razonables (CRP, por ejemplo) ya que transmiten información capturada por proteínas que ya están en el modelo, mientras que otras proteínas con relaciones de riesgo univariadas más bajas se retienen debido a la información única que proporcionan. Las funciones biológicas de las proteínas seleccionadas por LASSO se tratan en el presente documento.

La puntuación del riesgo de la proteína CVD9 funcionó mejor que la puntuación del riesgo secundario de Framingham (D'Agostino et al., *Am Heart J*. 2000; 139: 272-281), que se basa en factores de riesgo tradicionales. Incluyendo variables clínicas que fueron significativamente diferentes en la población de eventos como, edad, sexo, diabetes o la tasa de filtración glomerular estimada (TFGe) en modelos secundarios solo proporcionaron mejoras modestas en CVD9 en las cocientes de riesgos instantáneos intercuartílicos y los índices de reclasificación neta en la cohorte de descubrimiento (Tabla 13). Es posible que CVD9 ya encapsule la biología subyacente al riesgo asociado con los factores de riesgo tradicionales, aunque los inventores no están proponiendo que la evaluación de proteínas con CVD9 o modelos similares las reemplace, ya que este último podría ser un mejor indicador del riesgo a largo plazo y una diana específica de los tratamientos. No obstante, los niveles de proteínas CVD9 proporcionan una evaluación individualizada superior del riesgo cardiovascular a corto plazo que Framingham, particularmente para pacientes en riesgo extremo (Figuras 6 y 7), presumiblemente porque indican si las vías asociadas con complicaciones cardiovasculares se han activado y si el daño del órgano terminal ha ocurrido (por ejemplo, troponina; véase Beatty et al., *JAMA Intern Med*. 2013;173:763-769).

El estudio de los presentes inventores es el primer análisis proteómico a gran escala del riesgo cardiovascular, utilizando una plataforma proteómica a gran escala de alto rendimiento. Este enfoque resultó en el descubrimiento de numerosos biomarcadores de proteínas individuales novedosos y condujo a la construcción de un modelo robusto de predicción de riesgo multivariable con un rendimiento superior para predecir el riesgo a corto plazo de eventos cardiovasculares secundarios. El estudio se realizó en dos cohortes bien caracterizadas grandes con excelente adjudicación de eventos de resultado, a través de dos continentes. The US National Cancer Institute, en colaboración con un panel de científicos expertos, ha desarrollado una lista de verificación de criterios que se pueden usar para determinar la disponibilidad de tecnologías basadas en ómicas para guiar la atención al paciente en ensayos clínicos. La calidad de la muestra se observó como una razón importante por la cual los hallazgos ómicos informados en un laboratorio pueden no replicarse en otros. Por consiguiente, los inventores han realizado el análisis proteómico a través de una gama de calidades de muestras, representante de instituciones académicas (Heart and Soul) y patrones de práctica clínica (HUNT3) y sus hallazgos son sólidos en este rango de calidad de muestra.

Los inventores han enfocado su investigación inicial a propósito en una población de sujetos de alto riesgo con enfermedad coronaria establecida (CHD). Existe la necesidad adicional de una predicción precisa del riesgo cardiovascular en la población general de menor riesgo o en personas con CHD de riesgo aún mayor. Estos estudios se están realizando actualmente con otras cohortes. Otra limitación es que hay muchas más proteínas en la sangre

que las 1130 que se cuantificó. Todavía no se sabe si su evaluación mejoraría la evaluación del riesgo cardiovascular, ya que podrían estar en las mismas vías y, por lo tanto, ser redundantes con las proteínas que ya se evaluaron. También se están realizando estudios que evalúan una cantidad aún mayor de proteínas que la informada en el presente estudio.

5 En resumen, se ha realizado con éxito el estudio proteómico más grande del riesgo cardiovascular hasta la fecha, con más de 2 millones de mediciones individuales de proteínas, identificó numerosos biomarcadores nuevos de riesgo y demostró un modelo de predicción de riesgo con un rendimiento superior y robusto.

10 Tabla 14: Tabla de proteínas individuales asociadas al riesgo cardiovascular. Los biomarcadores en el panel de CVD9 están en negrita. Si el cociente de riesgos instantáneos (HR) es superior a 1, los niveles aumentados del biomarcador se asocian con un mayor riesgo; si el HR es inferior a 1, los niveles disminuidos del biomarcador se asocian con un mayor riesgo.

Diana	ID UniProt	Q4/Q1 HR	HR por desviación estándar	Valor P para HR continua
Angiopoyetina-2	O15123	3,13	1,67	<1e-16
MMP12	P39900	3,52	1,65	<1e-16
proteína TIM-3 de la membrana de linfocitos T	Q8TDQ0	2,98	1,61	<1e-16
Proteína 2 de unión al factor de crecimiento de tipo insulina	P18065	2,93	1,58	<1e-16
TNF R-II	P20333	3,33	1,56	<1e-16
Folistatina-tipo 3	O95633	3,52	1,56	<1e-16
Hemofiltrado CC quimiocina 1	Q16627	2,63	1,55	<1e-16
β2-microglobulina	P61769	3,58	1,54	<1e-16
Trombospondina-2	P35442	3,19	1,54	<1e-16
MMP-7	P09237	3,54	1,53	<1e-16
Endostatina	P39060	2,45	1,52	<1e-16
catepsina H	P09668	4,06	1,52	<1e-16
receptor B2 de EPH	P29323	2,20	1,50	<1e-16
proteína de unión a la interleucina 18	O95998	2,55	1,49	<1e-16
Cordina-tipo 1	Q9BU40	2,87	1,49	<1e-16
Cistatina C	P01034	3,54	1,49	<1e-16
Complemento C9	P02748	2,81	1,48	8,80E-14
CCL18/PARC	P55774	2,55	1,47	1,11E-16
Complemento C7	P10643	3,09	1,47	<1e-16
RELT receptor de factor de necrosis tumoral	Q969Z4	3,23	1,46	<1e-16
Dentado-	P78504	2,17	1,45	3,66E-15
Receptor de netrina UNC5H3	095185	3,15	1,44	<1e-16
Efrina-A4	P52798	3,37	1,44	<1e-16
Serina proteasa 4 específica del cerebro Neuroblastoma supresor de tumorigenicidad 1 DAN	Q9GZN4	2,54	1,44	<1e-16
	P41271	3,24	1,43	<1e-16
receptor 2 de efrina de tipo A	P29317	2,86	1,43	<1e-16
Espondina-1	Q9HCB6	2,99	1,42	<1e-16
periostina	Q15063	2,11	1,40	6,08E-12
factor A de crecimiento endotelial vascular	P15692	2,43	1,40	1,02E-12
Receptor secuestrante del miembro 1 de clase F	Q14162	2,90	1,39	<1e-16
Complejo α1-antiquimotripsina	P07288, P01011	2,53	1,39	1,95E-14
proteína adaptadora Crk-i	P46108	2,71	1,39	<1e-16
Efrina-A5	P52803	2,35	1,37	2,89E-15
Molécula de adhesión selectiva de células endoteliales	Q96AP7	2,13	1,37	1,02E-14
Glutación S-transferasa Pi 1	P09211	2,37	1,37	1,27E-12
receptor 6 de muerte DR6	O75509	1,98	1,36	1,35E-10
Proteína protectora de macrófagos	P40121	3,01	1,36	<1e-16
Proteína 80 que contiene el dominio de hélice superenrollada	Q76M96	2,07	1,35	1,36E-10

(continuación)

Diana	ID UniProt	Q4/Q1 HR	HR por desviación estándar	Valor P para HR continua
gen 3 de activación de linfocitos	P18627	2,06	1,35	2,47E-10
Ck-β-8-1 Variante de empalme de proteína 3 inflamatoria de macrófagos	P55773	1,84	1,35	1,26E-08
Elafina	P19957	2,18	1,35	9,09E-12
TIMP-1	P01033	3,94	1,34	<1e-16
HSP 70	P08107	2,08	1,34	1,24E-10
Estanniocalcina-1	P52823	2,29	1,34	2,62E-11
Receptor III-B región Fc de la inmunoglobulina G	075015	1,91	1,34	2,53E-10
Inhibidor de la proteasa leucocitaria secretora	P03973	2,37	1,34	7,91E-11
TRAIL R4	Q9UBN6	2,02	1,33	1,29E-09
MMP-3	P08254	2,40	1,33	1,36E-09
Hormona pancreática PH	P01298	2,19	1,33	1,18E-10
Factor neurotrófico de dopamina conservado	Q49AH0	1,82	1,32	5,46E-10
Cistatina D	P28325	2,11	1,32	9,36E-10
GPVI	Q9HCN6	1,79	1,32	3,09E-09
Catepsina Z/X/P	Q9UBR2	2,07	1,32	9,36E-10
proteína 1 de tipo delta	O00548	3,28	1,31	<1e-16
MPIF-1	P55773	1,88	1,31	2,73E-09
Kalikeína 11	Q9UBX7	1,86	1,31	1,22E-11
1 similar al receptor de interleucina-1 ST2	Q01638	2,02	1,30	1,91E-09
Molécula 5 de señalización de la activación linfocitaria	Q9UIB8	2,49	1,30	9,99E-16
TFF3	Q07654	3,62	1,30	3,50E-13
Subunidad β de PAFAH	P68402	2,41	1,29	7,94E-14
Proteína 1 de unión al factor de crecimiento de tipo insulina	P08833	1,87	1,29	1,11E-08
CD48	P09326	2,07	1,29	4,61E-10
Renina	P00797	1,70	1,29	1,67E-08
Neuroigina 4, ligado a X	Q8N0W4	2,24	1,29	1,91E-12
sustancia quimioatrayente de linfocitos B	O43927	2,34	1,29	1,24E-11
Proteína A plasmática asociada al embarazo	Q13219	1,69	1,29	6,82E-08
uPAR	Q03405	3,00	1,28	3,77E-13
resistina	Q9HD89	1,77	1,28	2,26E-08
Fucosiltransferasa 5	Q11128	1,55	1,28	8,30E-07
factor 1 derivado de células estromales	P48061	1,79	1,28	2,17E-07
Nidogen	P14543	1,67	1,28	1,11E-07
Ligando 1A tipo TNF	O95150	2,66	1,28	6,61E-13
Requisito de alta temperatura serina peptidasa A2	O43464	2,01	1,28	6,35E-09
Proteína 7 de unión al factor de crecimiento de tipo insulina	Q16270	1,78	1,28	1,55E-07
receptor 1 de interleucina 1	P14778	1,83	1,27	1,12E-06
Fosfolipasa secretora no pancreática A2	P14555	1,98	1,27	5,23E-09
Proteína 4 relacionada con angiopoyetina	Q9BY76	2,84	1,27	4,93E-11
Proteína de unión a ácidos grasos, de tipo cardíaco	P05413	2,58	1,27	1,29E-10
Proteína de unión a lipopolisacáridos	P18428	1,99	1,27	1,80E-06
receptor factor I de crecimiento similar a la insulina IGF-I sR	P08069	1,75	1,27	9,87E-07
tenascina C	P24821	1,89	1,27	3,14E-07
Receptor de ectodisplasina-A2 ligado al cromosoma X XEDAR	Q9HAV5	3,64	1,27	<1e-16
troponina I, cardíaca	P19429	2,94	1,27	1,01E-12
Sialoproteína ósea 2	P21815	1,79	1,27	7,52E-08
Proteína 6 de unión al factor de crecimiento de tipo insulina	P24592	2,29	1,26	3,04E-11

(continuación)

Diana	ID UniProt	Q4/Q1 HR	HR por desviación estándar	Valor P para HR continua
Matrilina-2	O00339	1,89	1,26	4,44E-07
antígeno Ly-9 de superficie de linfocitos T	Q9HBG7	1,56	1,26	1,79E-06
Laillina	Q6UX15	2,50	1,26	1,42E-09
dCTP pirofosfatasa 1	Q9H773	1,59	1,26	7,35E-06
Dímero de cadena γ de fibrinógeno	P02679	1,98	1,25	3,29E-06
receptor B6 de EPH	O15197	1,79	1,25	3,26E-09
Anhidrasa carbónica III	P07451	1,88	1,25	5,33E-07
receptor 1 de lipoproteínas de baja densidad oxidadas	P78380	1,89	1,25	2,37E-07
Cistatina SA	P09228	1,59	1,25	2,82E-06
Factor 7 de crecimiento de fibroblastos	P21781	2,08	1,25	2,90E-12
Neurexofilina-1	P58417	0,58	0,75	1,11E-08
Soggy-1	Q9UK85	0,54	0,75	2,57E-08
15-hidroxi prostaglandina deshidrogenasa	P15428	0,51	0,74	1,93E-08
Proteína C	P04070	0,38	0,74	1,29E-13
proteína α de activación de fibroblastos	Q12884	0,49	0,74	1,47E-06
TWEAK	O43508	0,44	0,74	4,06E-06
receptor 2 del factor de crecimiento endotelial vascular	P35968	0,47	0,74	2,21E-10
Complemento de la proteína de unión C1q	Q07021	0,49	0,74	9,01E-05
Angiostatina	P00747	0,48	0,73	5,89E-11
ErbB3	P21860	0,39	0,72	2,03E-10
GDF11	O95390	0,41	0,72	8,75E-09
BMP-1	P13497	0,39	0,71	2,54E-13
CDO regulada por el oncogén de adhesión celular	O4KMG0	0,39	0,70	9,06E-14
CK-MM	P06732	0,45	0,70	3,60E-11
Dipeptidasa de carnosina 1	Q96KN2	0,36	0,69	<1e-16
AMPc y GMPc fosfodiesterasa 11A	Q9HCR9	0,39	0,69	5,18E-12
CK-MB	P12277 P06732	0,41	0,69	5,06E-13
Cadherina-3	P22223	0,41	0,67	1,11E-16
Protooncogén del receptor de proteína tirosina cinasa Ret	P07949	0,42	0,66	4,89E-13
α2-antiplasmina	P08697	0,37	0,64	<1e-16
Hormona del crecimiento	P10912	0,29	0,63	<1e-16
Receptor del EGF	P00533	0,29	0,60	<1e-16

Ejemplo 5: Modelo GDF11 y FSTL3

Se generaron y compararon tres modelos de riesgo proporcional de Cox.

5

- GDF11: modelo de proteína univariado con proteína GDF11
- FSTL3: modelo de proteína univariado con proteína 3 relacionada con folistatina (FSTL3)
- GDF11.FSTL3: modelo de proteína combinatoria con GDF11 y FSTL3

10 Para la comparación entre modelos, ANOVA, Cociente de riesgos instantáneos Q4/Q1 de predictores lineales, se calculó el NRI de probabilidad de riesgo a 4 años y el AUC integrado en 4 años. El modelo GDF11.FSTL3 fue el mejor modelo con todos los métodos de evaluación.

15 Las muestras desviadas se excluyeron del análisis. Todos los modelos se calcularon con log transformado con base 10 y estandarizado.

Antes de combinar dos proteínas en el modelo, se aplicó la correlación de Spearman para verificar la relación entre GDF11 y FSTL3. La correlación entre dos proteínas es significativa ($p = 3,123 \cdot 10^{-12}$), pero la correlación de Spearman no es fuerte ($\rho = -0,2251$). La Tabla 15 muestra el resultado de la prueba de correlación de R.

20

Tabla 15: Prueba de correlación de Spearman entre GDF11 y FSTL3.

```

Correlación de rango de Spearman rho
datos: gdf11 y fstl3
S = 1,68e+08, Valor p = 3,123e-12
hipótesis alternativa: verdadero rho no es igual a 0
estimaciones de muestra:
rho
-0,2251
    
```

5 La correlación entre GDF11 y FSTL3 se muestra en la Figura 15. La RFU se convirtió en espacio log con base 10. La figura izquierda muestra la correlación de todas las muestras y la figura derecha muestra la correlación sin omitir una muestra, que tenía un valor de GDF11 alto. Los círculos negros y rojos significan muestras sin eventos y muestras de eventos, respectivamente.

10 Se generaron tres modelos de riesgo proporcional de Cox: GDF11, FSTL3 y GDF11.FSTL3. Los modelos de GDF11, FSTL3 son modelos Cox con una sola proteína y el GDF11.FSTL3 es un modelo combinado con dos proteínas. Antes de ajustar el modelo, se excluyeron los valores atípicos y los valores de RFU se transformaron y estandarizaron. Las tablas 16 y 17 muestran la comparación entre modelos individuales y modelo combinado con la tabla de desviación ANOVA. El modelo combinado se mejora significativamente de los modelos de proteínas individuales. Los valores p para GDF11 frente a GDF11.FSTL3 y FSTL3 frente a GDF11.FSTL3 son 2e-16 y 3.5e-06, respectivamente.

Tabla 16: La prueba ANOVA entre el modelo GDF11 y GDF11.FSTL3

```

Tabla de Análisis de la desviaciones
Modelo de Cox: la respuesta es s
Modelo 1: ~ GDF11.2765.4.3
Modelo 2: ~ GDF11.2765.4.3 + FSTL3.3438.10.2
loglik Chisq Df P(>|Chi|)
1 -2936
2 -2896 79 1 <2e-16 ***
---
Códigos signif.: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1
    
```

Tabla 17: La prueba ANOVA entre el modelo FSTL3 y GDF11.FSTL3

```

Tabla de Análisis de la desviaciones
Modelo de Cox: la respuesta es s
Modelo 1: ~ FSTL3.3438.10.2
Modelo 2: ~ GDF11.2765.4.3 + FSTL3.3438.10.2
loglik Chisq Df P(>|Chi|)
1 -2907
2 -2896 21,5 1 3,5e-06 ***
---
Códigos signif.: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1
    
```

20 La Tabla 18 muestra el cociente de riesgos instantáneos Q4/Q1 de predictores lineales para cada modelo. El modelo combinado muestra un cociente de riesgos instantáneos más alto que los modelos individuales. Los cuartiles están definidos por los predictores lineales de cada modelo de Cox.

25 Tabla 18: El cociente de riesgos instantáneos Q4/Q1

	Q4/Q1 HR	95 inferior	95 superior
GDF11	2,475	1,894	3,233
FSTL3	3,637	2,738	4,830
GDF11.FSTL3	4,080	3,068	5,426

30 La Figura 16 muestra las curvas de supervivencia de cada cuartil de todos los modelos. Los cuartiles primero a cuarto se describen con color negro, rojo, verde y azul (de arriba a abajo, las líneas son de color negro, rojo, verde, luego azul). El sombreado representa los intervalos de confianza del 95 %. La distancia entre el primer y el cuarto cuartiles del modelo GDF11.FSTL3 es más amplia que los modelos de proteínas individuales. Por otra parte, La distancia entre el segundo y tercer cuartiles del modelo GDF11.FSTL3 también es más amplia que los modelos de proteínas individuales.

En la Figura 3 se muestra una comparación entre las curvas de supervivencia para el modelo GDF11 y el modelo GDF11.FSTL3. La figura de la izquierda muestra la comparación de los grupos de bajo riesgo y la figura de la derecha muestra la comparación de los grupos de alto riesgo. El negro y el rojo representan el modelo GDF11 y el modelo GDF11.FSTL3, respectivamente. En este modelo, el grupo de bajo riesgo identificado por el modelo GDF11.FSTL3 tiene menos muestras de eventos que el grupo de bajo riesgo identificado por el modelo GDF11, y el grupo de alto riesgo identificado por el modelo GDF11.FSTL3 tenía más muestras de eventos que el grupo de alto riesgo identificado por el modelo GDF11. El modelo GDF11.FSTL3 fue, por lo tanto, más preciso para ambos grupos de muestras. La inclusión de FSTL3 mejoró el modelo para los grupos de alto y bajo riesgo.

El NRI se calculó entre los modelos de proteínas individuales GDF11 y FSTL3 y GDF11.FSTL3. La probabilidad se calculó en un plazo de 4 años y el riesgo basal se estimó con el estimador de Kaplan-Meier. Inferior y superior en la tabla son el intervalo de confianza del 95 % de NRI, que se estima con bootstrapping. GDF11 mejora el NRI de las muestras de eventos y FSTL3 mejora el NRI de las muestras sin eventos.

Tabla 19: NIR entre el modelo de proteína única y el modelo GDF11.FSTL3.

	GDF11 frente a GDF11.FSTL3				FSTL3 frente a GDF11.FSTL3		
	Estimación	Inferior	Superior		Estimación	Inferior	Superior
NRI	0,4326	0,28552	0,5772	NRI	0,31532	0,17287	0,4639
NRI+	0,1446	0,01143	0,2655	NRI+	0,22840	0,10174	0,3612
NRI-	0,2880	0,21878	0,3563	NRI-	0,08692	0,01022	0,1653
Pr(Up Case)	0,5721	0,50573	0,6329	Pr(Up Case)	0,61422	0,55081	0,6807
Pr(Down Case)	0,4276	0,36706	0,4942	Pr(Down Case)	0,38582	0,31923	0,4491
Pr (Abajo Ctrl)	0,6441	0,60935	0,6780	Pr (Abajo Ctrl)	0,54345	0,50504	0,5828
Pr (Arriba Ctrl)	0,3560	0,32190	0,3906	Pr (Arriba Ctrl)	0,45653	0,41738	0,4949

La Figura 18 muestra la comparación de la probabilidad de 4 años entre GDF11 y GDF11.FSTL3 (izquierda), y FSTL3 y GDF11.FSTL3 (derecha). Los puntos de color negro, rojo y verde describen las muestras de control, casos y censuradas el año 4.

La AUC integrada (Cíndice) en un plazo de 4 años se muestra en la Tabla 20. Se calcularon los intervalos de confianza del 95 % con bootstrapping, similar a NRI.

Tabla 20: La AUC integrada (Cíndice) en un plazo de 4 años

	Cíndice	95 inferior	95 superior
GDF11	0,5882	0,5555	0,6096
FSTL3	0,6038	0,5786	0,6321
GDF11.FSTL3	0,6286	0,6050	0,6566

Las curvas ROC al año 4 se muestran en la Figura 19. Los números en la leyenda se refieren a la AUC el año 4 (AUC no integrada).

Se compararon tres modelos de riesgo proporcional de Cox con varias estadísticas de evaluación, para los modelos de marcador único y el modelo de marcador combinado. El modelo combinatorio, que incluye GDF11 y FSTL3, funcionó mejor de acuerdo con todos los valores de evaluación.

Los siguientes cuadros muestran los tres modelos utilizados en este ejemplo.

<p>Call: coxph (fórmula = f, datos = x, x = T)</p> <p>n=937, número de eventos = 465</p> <table border="1"> <thead> <tr> <th></th> <th>coef</th> <th>exp(coef)</th> <th>se(coef)</th> <th>z</th> <th>Pr(> z)</th> </tr> </thead> <tbody> <tr> <td>GDF11,2765.4,3</td> <td>-0,3452</td> <td>0,7081</td> <td>0,0579</td> <td>-5,96</td> <td>2,5e-09 ***</td> </tr> </tbody> </table> <p>Códigos signif.: 0 '****' 0,001 '***' 0,01 0,05 0,1 ' ' 1</p> <table border="1"> <thead> <tr> <th></th> <th>exp(coef)</th> <th>exp(-coef)</th> <th>inferior.95</th> <th>superior.95</th> </tr> </thead> <tbody> <tr> <td>GDF11,2765.4,3</td> <td>0,708</td> <td>1,41</td> <td>0,632</td> <td>0,793</td> </tr> </tbody> </table> <p>Concordancia = 0,604 (se = 0,014) Rcuadrado = 0,04 (máximo posible = 0,998) Prueba de razón de verosimilitud = 38,1 en 1 df, p=6,66e-10 Prueba de Wald = 35,6 en 1 df, p=2,46e-09 Prueba de puntuación (rango log) = 28,4 en 1 df, p=9,95e-08</p>							coef	exp(coef)	se(coef)	z	Pr(> z)	GDF11,2765.4,3	-0,3452	0,7081	0,0579	-5,96	2,5e-09 ***		exp(coef)	exp(-coef)	inferior.95	superior.95	GDF11,2765.4,3	0,708	1,41	0,632	0,793											
	coef	exp(coef)	se(coef)	z	Pr(> z)																																	
GDF11,2765.4,3	-0,3452	0,7081	0,0579	-5,96	2,5e-09 ***																																	
	exp(coef)	exp(-coef)	inferior.95	superior.95																																		
GDF11,2765.4,3	0,708	1,41	0,632	0,793																																		
<p>Call: coxph (fórmula = f, datos = x, x = T)</p> <p>n=937, número de eventos = 465</p> <table border="1"> <thead> <tr> <th></th> <th>coef</th> <th>exp(coef)</th> <th>se(coef)</th> <th>z</th> <th>Pr(> z)</th> </tr> </thead> <tbody> <tr> <td>FSTL3.3438.10,2</td> <td>0,436</td> <td>1,547</td> <td>0,042</td> <td>10,4</td> <td><2e-16 ***</td> </tr> </tbody> </table> <p>---</p> <p>Códigos signif.: 0 '****' 0,001 '***' 0,01 0,05 0,1 ' ' 1</p> <table border="1"> <thead> <tr> <th></th> <th>exp(coef)</th> <th>exp(-coef)</th> <th>inferior.95</th> <th>superior.95</th> </tr> </thead> <tbody> <tr> <td>FSTL3.3438.10,2</td> <td>1,55</td> <td>0,646</td> <td>1,42</td> <td>1,68</td> </tr> </tbody> </table> <p>Concordancia = 0,634 (se = 0,014) Rcuadrado = 0,097 (máximo posible = 0,998) Prueba de razón de verosimilitud = 95,6 en 1 df, p = 0 Prueba de Wald = 108 en 1 df, p = 0 Prueba de puntuación (rango log) = 105 en 1 df, p = 0</p>							coef	exp(coef)	se(coef)	z	Pr(> z)	FSTL3.3438.10,2	0,436	1,547	0,042	10,4	<2e-16 ***		exp(coef)	exp(-coef)	inferior.95	superior.95	FSTL3.3438.10,2	1,55	0,646	1,42	1,68											
	coef	exp(coef)	se(coef)	z	Pr(> z)																																	
FSTL3.3438.10,2	0,436	1,547	0,042	10,4	<2e-16 ***																																	
	exp(coef)	exp(-coef)	inferior.95	superior.95																																		
FSTL3.3438.10,2	1,55	0,646	1,42	1,68																																		
<p>Call: coxph (fórmula = f, datos = x, x = T)</p> <p>n=937, número de eventos = 465</p> <table border="1"> <thead> <tr> <th></th> <th>coef</th> <th>exp(coef)</th> <th>se(coef)</th> <th>z</th> <th>Pr(> z)</th> </tr> </thead> <tbody> <tr> <td>GDF11.2765.4.3</td> <td>-0,2605</td> <td>0,7706</td> <td>0,0577</td> <td>-4,52</td> <td>6,2e-06 ***</td> </tr> <tr> <td>FSTL3.3438.10.2</td> <td>0,4064</td> <td>1,5014</td> <td>0,0434</td> <td>9,36</td> <td>< 2e-16 ***</td> </tr> </tbody> </table> <p>---</p> <p>Códigos signif.: 0 '****' 0,001 '***' 0,01 ' ' 0,05 0,1 ' ' 1</p> <table border="1"> <thead> <tr> <th></th> <th>exp(coef)</th> <th>exp(-coef)</th> <th>inferior.95</th> <th>superior.95</th> </tr> </thead> <tbody> <tr> <td>GDF11,2765.4,3</td> <td>0,771</td> <td>1,298</td> <td>0,688</td> <td>0,863</td> </tr> <tr> <td>FSTL3.3438.10,2</td> <td>1,501</td> <td>0,666</td> <td>1,379</td> <td>1,635</td> </tr> </tbody> </table> <p>Concordancia = 0,652 (se = 0,014) Rcuadrado = 0,117 (máximo posible = 0,998) Prueba de razón de verosimilitud = 117 en 2 df, p = 0 Prueba de Wald = 124 en 2 df, p = 0 Prueba de puntuación (rango log) = 119 en 2 df, p = 0</p>							coef	exp(coef)	se(coef)	z	Pr(> z)	GDF11.2765.4.3	-0,2605	0,7706	0,0577	-4,52	6,2e-06 ***	FSTL3.3438.10.2	0,4064	1,5014	0,0434	9,36	< 2e-16 ***		exp(coef)	exp(-coef)	inferior.95	superior.95	GDF11,2765.4,3	0,771	1,298	0,688	0,863	FSTL3.3438.10,2	1,501	0,666	1,379	1,635
	coef	exp(coef)	se(coef)	z	Pr(> z)																																	
GDF11.2765.4.3	-0,2605	0,7706	0,0577	-4,52	6,2e-06 ***																																	
FSTL3.3438.10.2	0,4064	1,5014	0,0434	9,36	< 2e-16 ***																																	
	exp(coef)	exp(-coef)	inferior.95	superior.95																																		
GDF11,2765.4,3	0,771	1,298	0,688	0,863																																		
FSTL3.3438.10,2	1,501	0,666	1,379	1,635																																		

Ejemplo 6: Modelo GDF11 y FSTL3 para grupos de eventos específicos

Los modelos Cox de GDF11, FSTL3 y GDF11.FSTL3 se ajustaron con muestras de CHF y muerte, y muestras de eventos trombóticos por separado, para determinar cómo funciona el modelo para cada tipo de evento CV. Cociente de riesgos instantáneos Q4/Q1 de predictores lineales del modelo, AUC integrada (Cíndice) en un plazo de 4 años, y se calcularon el NRI de la probabilidad de riesgo a 4 años. Para el cálculo de la probabilidad de riesgo, el estimador de Kaplan-Meier se utilizó como riesgo basal.

Se ajustó el modelo de riesgo proporcional de Cox con GDF11, FSTL3 y GDF11.FSTL3 con grupos de eventos específicos: grupo de CHF y muerte, y grupo de eventos trombóticos. El grupo de CHF y muerte incluye CHF (125), Muerte CVD (55), Muerte (135) y NINGUNA (472). El grupo de eventos trombóticos incluye IM (104), ICTUS(30), TIA (16) y NINGUNO (472). NINGUNO, que son muestras sin eventos, se usaron en ambos grupos. Para la evaluación de modelos, Cociente de riesgos instantáneos Q4/Q1 de predictores lineales, ka AUC integrada en un plazo de 4 años, y se calcularon NRI de probabilidad de riesgo de 4 años. La probabilidad de riesgo se calculó con el riesgo basal del estimador de Kaplan-Meier.

La Tabla 21 muestra el cociente de riesgos instantáneos Q4/Q1, la inversa del cociente de riesgos instantáneos y sus intervalos de confianza del 95 %. El cociente de riesgos instantáneos Q4/Q1 de GDF11 y FSTL3 no es significativamente diferente entre las muestras de DHF.DEATH y de eventos trombóticos, pero el cociente de riesgos instantáneos de GDF11.FSTL3 es mejor con los grupos de eventos trombóticos que con el grupo CHF.DEATH.

Tabla 21: Cociente de riesgos instantáneos Q4/Q1 de cada modelo y grupo.

\$GDF11				
	Q4/Q1 HR	Q1/Q4 HR	IC inferior	Superior del IC
Todos	2,475	0,4040	1,894	3,233
MUERTE CHF	2,726	0,3668	1,964	3,784
Evento trombótico	2,743	0,3645	1,698	4,432
\$ FSTL3				
	Q4/Q1 HR	Q1/Q4 HR	IC inferior	Superior del IC
Todos	3,637	0,2750	2,738	4,830
MUERTE CHF	4,478	0,2233	3,125	6,416
Evento trombótico	4,605	0,2171	2,731	7,765
\$ GDF11.FSTL3				
	Q4/Q1 HR	Q1/Q4 HR	IC inferior	Superior del IC
Todos	4,080	0,2451	3,068	5,426
MUERTE CHF	4,394	0,2276	3,069	6,291
Evento trombótico	5,493	0,1821	3,185	9,473

La Figura 1 muestra las curvas de supervivencia de los cuartiles del predictor lineal de cada grupo del modelo GDF11.FSTL3. Los cuartiles primero a cuarto se describen con color negro (línea superior), rojo (segunda línea hacia abajo), verde (tercera línea hacia abajo) y azul (línea inferior). El sombreado muestra los intervalos de confianza del 95 %. El primer cuartil del evento trombótico (grupo de bajo riesgo) muestra menos eventos. Esto sugiere que el modelo podría ser bastante sensible al evento trombótico.

La AUC integrada (Cíndice) en un plazo de 4 años y los intervalos de confianza del 95 % se muestran en la Tabla 22. Con Cíndice, no existen diferencias significativas entre el grupo CHF.DEATH y el grupo de eventos trombóticos, a pesar de que se encontró que el cociente de riesgos instantáneos Q4/Q1 era diferente entre los grupos.

Tabla 22: AUC integrada (Cíndice) en un plazo de 4 años

\$GDF11				
	Cíndice	Cíndice.IC 95 inferior	Cíndice. IC 95 superior	
Todos	0,5882			
MUERTE CHF	0,5892		0,5614	0,6165
Evento trombótico	0,6057		0,5582	0,6220
			0,5641	0,6503
\$ FSTL3				
	Cíndice	Cíndice.IC 95 inferior	Cíndice. IC 95 superior	
Todos	0,6038		0,5808	0,6344
MUERTE CHF	0,6018		0,5754	0,6482
Evento trombótico	0,5994		0,5667	0,6600

(continuación)

\$ GDF11.FSTL3	Cíndice	Cíndice.IC 95 inferior	Cíndice. IC 95 superior	
Todos	0,6286	0,6047	0,6558	
MUERTE CHF	0,6308	0,6020	0,6645	
Evento trombótico	0,6292	0,5939	0,6777	

En conclusión, el modelo Cox de GDF11, FSTL3 y GDF11.FSTL3 se generaron con grupos de muestra específicos. El modelo GDF11.FSTL3 muestra el mejor resultado con el cociente de riesgos instantáneos Q4/Q1 con el grupo de eventos trombóticos. Con Índice, todos los modelos mostraron resultados similares.

5

Ejemplo 7: Modelo GDF11 y GASP1/GASP2

La combinación de GDF11 con otras dos proteínas, también se probaron GASP1 (WFIKKN2, SwissProt Q8TEU8) y GASP2 (WFIKKN1, SwissProt Q96D09).

10

Se generaron los siguientes cuatro modelos de Cox: (1) GDF11, Modelo de Cox con proteína GDF11; (2) GDF11.WFIKKN1, Modelo Cox con GDF11 y GASP2; (3)

15

GDF11.WFIKKN2, Modelo Cox con GDF11 y GASP1; y (4) GDF11.WFIKKN1.WFIKKN2, Modelo Cox con GDF11, GASP1 y GASP2. Antes de crear los modelos, la medida de la proteína se estandarizó a Gaussiana (0,1).

Se calculó el cociente de riesgos instantáneos Q4/Q1 de predictores lineales para los modelos. El grupo Q1 se asume como grupo de bajo riesgo y el grupo Q4 se asume como grupo de alto riesgo. Se descubrió que la adición de GASP2 (WFIKKN1) no mejora el modelo GDF11, pero añadir WFIKKN2 mostró alguna mejora (de 2.432 a 2.719).

20

Los valores del cociente de riesgos instantáneos Q4/Q1 y las curvas de supervivencia de los cuartiles se muestran en la Tabla 23 y la Figura 21.

Tabla 23. Cociente de riesgos instantáneos Q4/Q1 de cada modelo.

##	Q4/Q1 HR	Q1/Q4 HR	IC inferior	Superior del IC
## GDF11	2,432	0,4111	1,864	3,175
## GDF11.WFIKKN1	2,392	0,4180	1,830	3,127
## GDF11.WFIKKN2	2,719	0,3678	2,071	3,569
## GDF11.WFIKKN1.WFIKKN2	2,758	0,3626	2,102	3,619

25 Además, los modelos se compararon con las tablas de desviación ANOVA. El resultado R de la comparación entre GDF11 y los modelos combinados se muestra a continuación. GDF11.WFIKKN2 y GDF11.WFIKKN1.WFIKKN2 fueron significativos en comparación con el modelo GDF11 (p = 3,1e-05, 0,00015, respectivamente). La adición de WFIKKN1 no mostró significancia (p = 0,38). Los valores de p se destacan a continuación.

30

- Comparación entre GDF11 y GDF11.WFIKKN1

```
## Tabla del Análisis de las desviaciones
## Modelo de Cox: la respuesta es s
## Modelo 1: ~ GDF11.2765.4.3
## Modelo 2: ~ GDF11.2765.4.3 + WFIKKN1.3191.50.2
##      loglik Chisq Df P(>|Chi|)
## 1      -2938
## 2      -2937  0,77  1      0,38
```

35

40

- Comparación entre GDF11 y GDF11.WFIKKN2

```
## Tabla del Análisis de las desviaciones
## Modelo de Cox: la respuesta es s
## Modelo 1: ~ GDF11.2765.4.3
## Modelo 2: ~ GDF11.2765.4.3 + WFIKKN2.3235.50.2
##      loglik Chisq Df P(> |Chi|)
## 1      -2938
## 2      -2929  17,4  1      3,1e-05 ***
## ---
## Códigos signif.: 0 '***' 0,001 '**' 0,01 '*' 0,05 0,1 ' ' 1
```

50

- Comparación entre GDF11 y GDF11.WFIKKN1.WFIKKN2

```

5  ## Tabla del Análisis de las desviaciones
  ## Modelo de Cox: la respuesta es s
  ## Modelo 1: ~ GDF11.2765.4.3
  ## Modelo 2: ~ GDF11.2765.4.3 + WFIKKN1.3191.50.2 + WFIKKN2.3235.50.2
  ## loglik Chisq Df P(>|Chi|)
  ## 1 -2938
  ## 2 -2929 17,6 2 0,00015 ***
10 ## ---
  ## Códigos signif.: 0 '****' 0,001 '***' 0,01 '**' 0,05 '1' '1'

```

Para evaluar los modelos, el cálculo de NRI también se realizó. La probabilidad se calculó en 4 años. La adición de GASP1 (WFIKKN2) mejoró el NRI (0,16), particularmente con muestras sin eventos (0,12). A partir de este resultado, GASP1 puede mejorar la tasa verdadera negativa. Por el contrario, GASP2 no mejoró NRI más de 0,1. El resultado R del NRI se muestra a continuación.

- NRI entre GDF11 y GDF11.WFIKKN1

20

##	##----- GDF11 frente a GDF11.WFIKKN1 ----	##	Estimación	Inferior	Superior
##	NRI		0,05855	-0,08109	0,20126
##	NRI+		0,09245	-0,02881	0,20405
##	NRI-		-0,03391	-0,11142	0,04671
##	Pn(Up Case)		0,54627	0,48586	0,60175
##	Pn(Down Case)		0,45382	0,39769	0,51467
##	Pn(Abajo Ctrl)		0,48303	0,44431	0,52336
##	Pn (Arriba Ctrl)		0,51694	0,47665	0,55573

- NRI entre GDF11 y GDF11.WFIKKN2

##	## ----- GDF11 frente a GDF11.WFIKKN2 -----	##	Estimación	Inferior	Superior
##	NRI		0,16315	0,02639	0,3063
##	NRI+		0,04422	-0,07236	0,1727
##	NRI-		0,11892	0,04351	0,1919
##	Pr(Up Case)		0,52206	0,46394	0,5861
##	Pr(Down Case)		0,47784	0,41338	0,5363
##	Pr (Abajo Ctrl)		0,55948	0,52176	0,5960
##	Pr (Arriba Ctrl)		0,44056	0,40405	0,4783

- NRI entre GDF11 y GDF11.WFIKKN1.WFIKKN2

25

##	## ----- GDF11 frente a GDF11.WFIKKN1.WFIKKN2 -----	##	Estimación	Inferior	Superior
##	NRI		0,13460	-0,01276	0,2759
##	NRI+		0,02732	-0,09310	0,1643
##	NRI-		0,10728	0,02863	0,1758
##	Pr(Up Case)		0,51364	0,45354	0,5820
##	Pr(Down Case)		0,48632	0,41779	0,5467
##	Pr (Abajo Ctrl)		0,55365	0,51428	0,5879
##	Pr (Arriba Ctrl)		0,44637	0,41211	0,4857

La probabilidad a 4 años entre modelos se muestra en la Figura 22.

- Por último, el cálculo de la AUC se realizó para la evaluación entre modelos. De acuerdo con los resultados

30

siguientes, ninguna proteína mejoró el modelo GDF11. Las curvas ROC para cada modelo se muestran en la Figura 23. Las curvas ROC para cada modelo fueron similares.

- 5 • GDF11


```
## -----
## Cíndice.95 inferior.95 superior
## 1 0,586 0,5579 0,6143
```
- 10 • GDF11.WFIKKN1


```
## -----
## Cíndice.95 inferior.95 superior
## 1 0,5849 0,5572 0,6133
```
- 15 • GDF11.WFIKKN2


```
## -----
## Cíndice.95 inferior.95 superior
## 1 0,5994 0,5717 0,6305
```
- 20 • GDF11.WFIKKN1.WFIKKN2


```
## -----
## Cíndice.95 inferior.95 superior
## 1 0,5988 0,5712 0,63
```

En resumen, GASP1 (WFKKN2) puede mejorar el modelo GDF11, pero la mejora es pequeña. GASP2 (WFKKN1) no mejoró el modelo GDF11.

30 El modelo de Cox utilizado en este ejemplo se muestra a continuación.

- GDF11


```
## Call:
## coxph (fórmula = f, datos = x, x = T)
##
## n=938, número de eventos = 465
##
##          coef      exp(coef)  se(coef)    z      Pr(>|z|)
## GDF11.2765.4.3 -0,3325  0,7171    0,0578  -5,75  8,7e-09 ***
## -----
## Códigos signif.: 0 '****' 0,001 '***' 0,01 0,05 0,1 ' ' 1
##
## exp(coef) exp(-coef) inferior.95 superior.95
## GDF11.2765.4.3 0,717 1,39 0,64 0,803
##
## Concordancia = 0,602 (se = 0,014)
## Rcuadrado = 0,037 (máximo posible = 0,998)
## Prueba de razón de verosimilitud = 35,6 en 1 df, p=2,42e-09
## Prueba de Wald = 33,1 en 1 df, p=8,75e-09
## Prueba de puntuación (rango log) = 26,6 en 1 df, p=2,53e-07
```

35

• GDF11.WFIKKN1

```
## Call:
## coxph (fórmula = f, datos = x, x = T)
##
## n=938, número de eventos = 465
##
##          coef      exp(coef)  se(coef)   z      Pr(>|z|)
## GDF11.2765.4.3 -0,3206  0,7257   0,0590  -5,43  5,6e-08 ***
## WFIKKN1.3191.50.2 -0,0409  0,9599   0,0466  -0,88  0,38
## ----
## Códigos signif.: 0      '****' 0,001 '***' 0,01 '**' 0,05 0,1 '*' 1
##
##                                     exp(coef) exp(-coef) inferior.95 superior.95
## GDF11.2765.4.3                    0,726    1,38      0,646      0,815
## WFIKKN1.3191.50.2                  0,960    1,04      0,876      1,052
##
## Concordancia = 0,601 (se = 0,014)
## Rcuadrado = 0,038 (máximo posible = 0,998)
## Prueba de razón de verosimilitud = 36,4 en 2 df, p=1,26e-08
## Prueba de Wald = 34,3 en 2 df, p=3,55e-08
## Prueba de puntuación (rango log) = 28,8 en 2 df, p=5,65e-07
```

5 • GDF11.WFIKKN2

```
## Call:
## coxph (fórmula = f, datos = x, x = T)
##
## n=938, número de eventos = 465
##
##          coef      exp(coef)  se(coef)   z      Pr(>|z|)
## GDF11.2765.4.3 -0,3369  0,7140   0,0575  -5,86  4,6e-09 ***
## WFIKKN2.3235.50.2 0,2014  1,2232   0,0484   4,16  3,2e-05 ***
## ---
## Códigos signif.: 0 '****' 0,001 '***' 0,01 0,05 0,1 '*' 1
##
##          exp(coef)  exp(-coef)  inferior.95  superior.95
## GDF11,2765.4,3    0,714    1,401      0,638      0,799
## WFIKKN2.3235.50.2 1,223    0,818      1,112      1,345
##
## Concordancia = 0,609 (se = 0,014)
## Rcuadrado = 0,055 (máximo posible = 0,998)
## Prueba de razón de verosimilitud = 53 en 2 df, p=3,18e-12
## Prueba de Wald = 50,7 en 2 df, p=9,63e-12
## Prueba de puntuación (rango log) = 42,1 en 2 df, p=7,26e-10
```

• GDF11.WFIKKN1.WFIKKN2

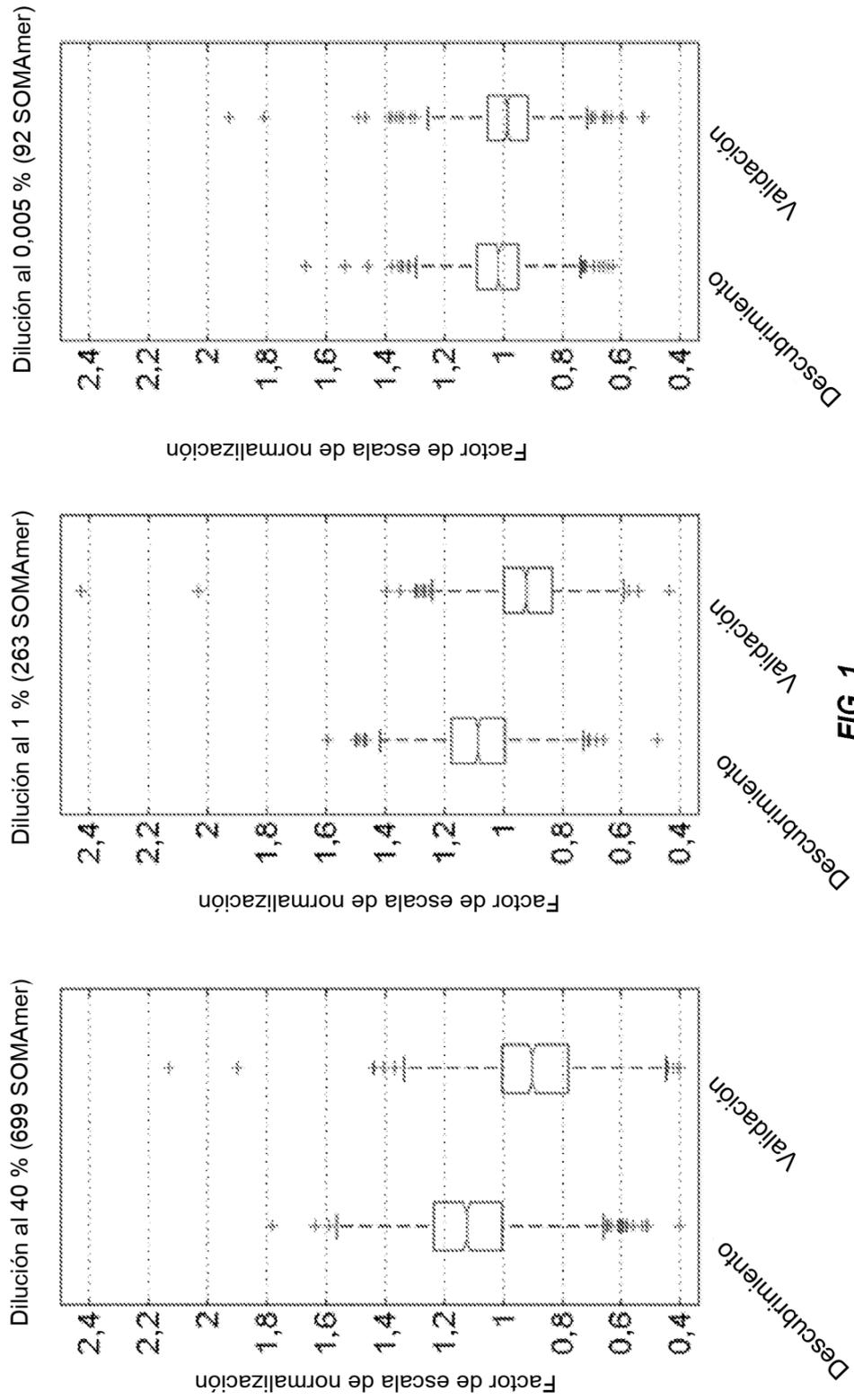
```

## Call:
## coxph (fórmula = f, datos = x, x = T)
5 ##
## n= 938, número de eventos = 465
##
##
##          coef      exp (coef)    se(coef)      z          Pr(>|z|)
## GDF11.2765.4.3    -0,3294      0,7193      0,0589     -5,60      2,2e-08 ***
## WFIKKN1.3191.50.2 -0,0256      0,9747      0,0466     -0,55          0,58
## WFIKKN2.3235.50.2  0,1989      1,2201      0,0486      4,10      4.2e-05 ***
## ---
## Códigos signif.: 0 '****' 0,001 '***' 0,01 0,05 0,1 ' '1
10 ##
##          exp (coef)    exp (-coef)    95 inferior    95 superior
## GDF11.2765.4.3      0,719      1,39      0,641      0,807
## WFIKKN1.3191.50.2  0,975      1,03      0,890      1,068
## WFIKKN2.3235.50.2  1,220      0,82      1,109      1,342
##
## Concordancia = 0,609 (se = 0,014)
## Rcuadrado = 0,055 (máximo posible = 0,998)
## Prueba de razón de verosimilitud = 53,2 en 3 df, p=1,62e-11
## Prueba de Wald = 51,4 en 3 df, p=4,1e-11
## Prueba de puntuación (rango log) = 43,7 en 3 df, p=1,73e-09

```

REIVINDICACIONES

1. Un método para la detección selectiva de un sujeto por el riesgo de un evento cardiovascular (CV) o la predicción de la probabilidad de que un sujeto sufra un evento CV, que comprende:
- 5 (a) formar un panel de biomarcadores que comprende biomarcadores de proteínas MMP12, angiopoyetina-2, complemento C7, troponina cardíaca I, proteína 4 relacionada con angiopoyetina, CCL18/PARC, complejo alfa-1-antiquimotripsina, GDF11 y alfa-2-antiplasmina; y
- 10 (b) detectar el nivel de cada uno de los biomarcadores de proteínas del panel en una muestra del sujeto.
2. Un método para la detección selectiva de un sujeto por el riesgo de un evento cardiovascular (CV) o la predicción de la probabilidad de que un sujeto sufra un evento CV, que comprende detectar el nivel del complejo alfa-1-antiquimotripsina, MMP12, angiopoyetina-2, complemento C7, troponina cardíaca I, proteína 4 relacionada con angiopoyetina, CCL18/PARC, GDF11 y alfa-2-antiplasmina en una muestra del sujeto.
- 15 3. El método de la reivindicación 1 o de la reivindicación 2, en el que la probabilidad de que el sujeto sufra un evento CV en un plazo de 4 años es alta si el nivel del complejo alfa-1-antiquimotripsina es mayor que un nivel de control, y el nivel de al menos cuatro, al menos cinco o los seis biomarcadores seleccionados del nivel de MMP12, angiopoyetina-2, complemento C7, troponina cardíaca I, proteína 4 relacionada con la angiopoyetina y CCL18/PARC
- 20 es más alto que un nivel de control de la proteína respectiva, y si el nivel de al menos un biomarcador o ambos biomarcadores seleccionados de GDF11 y alfa2-antiplasmina es más bajo que un nivel de control de la proteína respectiva.
4. El método de una cualquiera de las reivindicaciones anteriores, en el que el sujeto padece arteriopatía coronaria.
- 25 5. El método de una cualquiera de las reivindicaciones anteriores, en el que el sujeto no tiene un historial de eventos CV.
6. El método de una cualquiera de las reivindicaciones 1 a 4, en el que el sujeto ha sufrido al menos un evento CV.
- 30 7. El método de una cualquiera de las reivindicaciones anteriores, en el que el evento CV se selecciona de entre infarto de miocardio, ictus, insuficiencia cardíaca congestiva, ataque isquémico transitorio y muerte.
8. El método de una cualquiera de las reivindicaciones anteriores, en el que la muestra se selecciona entre una
- 35 muestra de sangre, una muestra de suero, una muestra de plasma y una muestra de orina.
9. El método de una cualquiera de las reivindicaciones anteriores, en donde el método comprende poner en contacto proteínas biomarcadoras de la muestra del sujeto con un conjunto de reactivos de captura de biomarcadores, en donde cada reactivo de captura de biomarcadores del conjunto de reactivos de captura de biomarcadores se une específicamente a una proteína biomarcadora diferente que se esté detectando.
- 40 10. El método de la reivindicación 9, en el que cada reactivo de captura de biomarcadores es un anticuerpo o un aptámero.
- 45 11. El método de la reivindicación 10, en el que dicho cada reactivo de captura de biomarcador es un aptámero.
12. El método de la reivindicación 11, en el que al menos un aptámero es un aptámero de tasa de disociación lenta, en donde cada aptámero de tasa de disociación lenta se une opcionalmente a su proteína diana con una tasa de disociación ($t_{1/2}$) de ≥ 30 minutos, ≥ 60 minutos, ≥ 90 minutos, ≥ 120 minutos, ≥ 150 minutos, ≥ 180 minutos, ≥ 210
- 50 minutos o ≥ 240 minutos.
13. El método de la reivindicación 12, en el que al menos un aptámero de tasa de disociación lenta comprende al menos un, al menos dos, al menos tres, al menos cuatro, al menos cinco, al menos seis, al menos siete, al menos ocho, al menos nueve o al menos 10 nucleótidos con modificaciones.
- 55



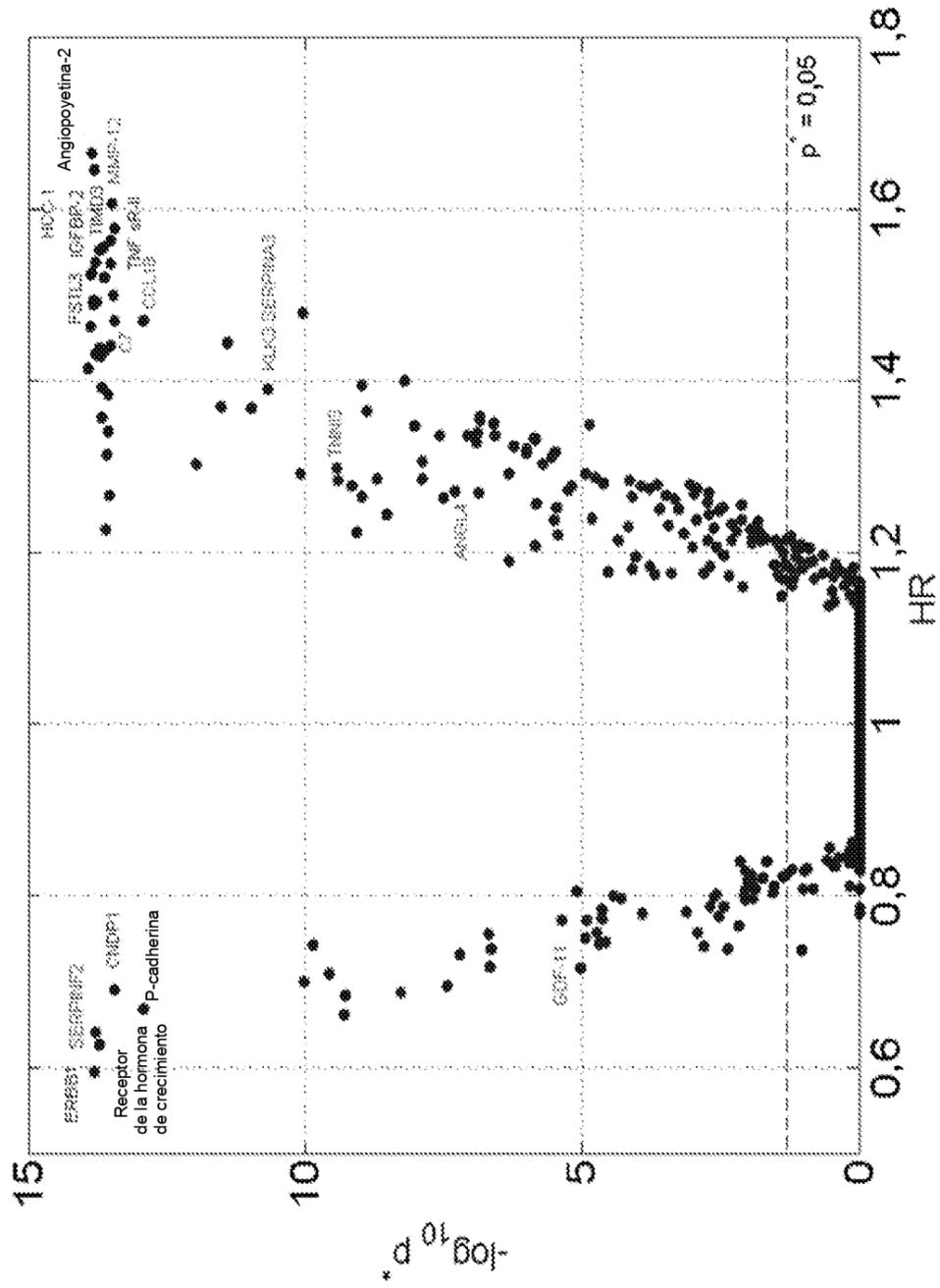


FIG. 2

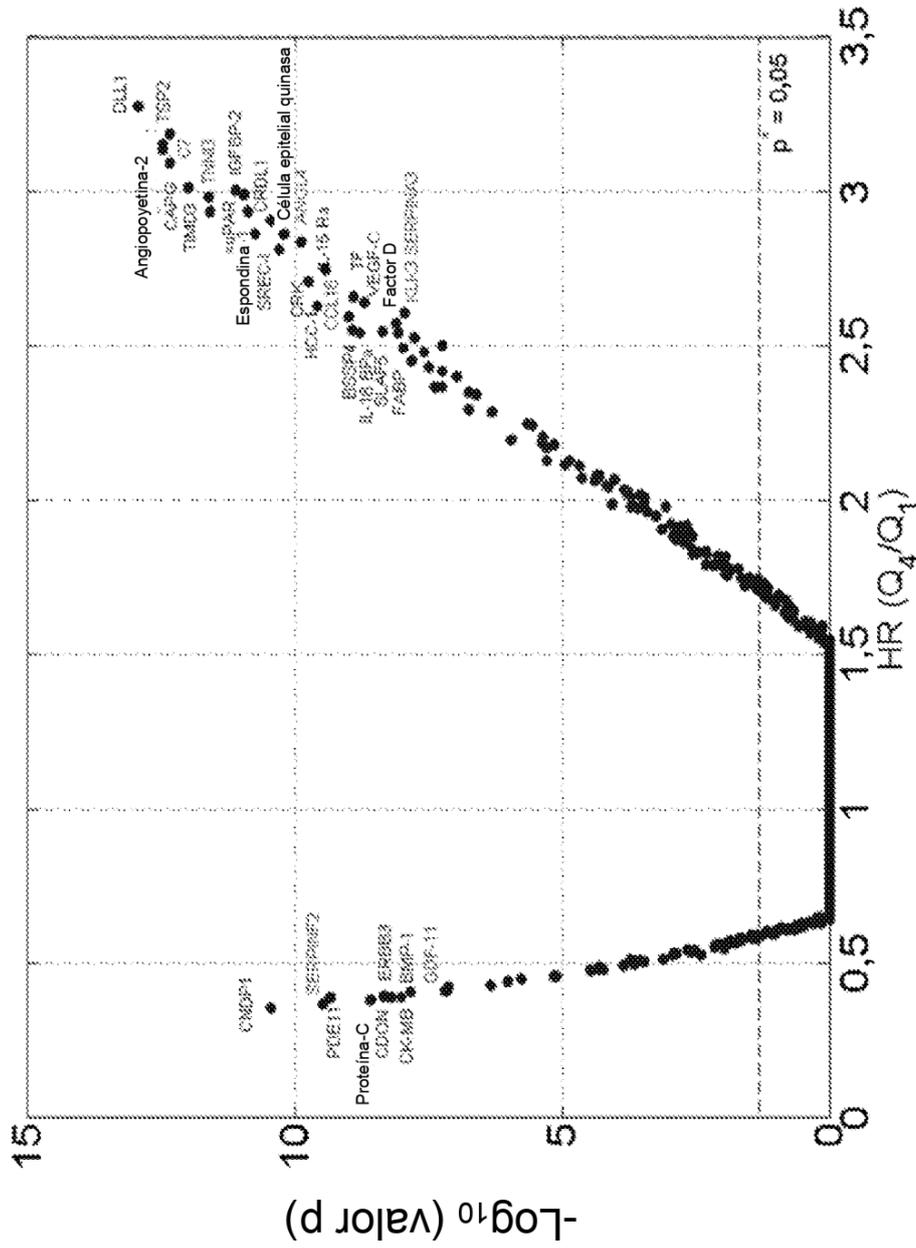


FIG. 2 (cont.)

Gráfico de concordancia de la intensidad media de la señal
(Sesgo de intensidad) 0,056

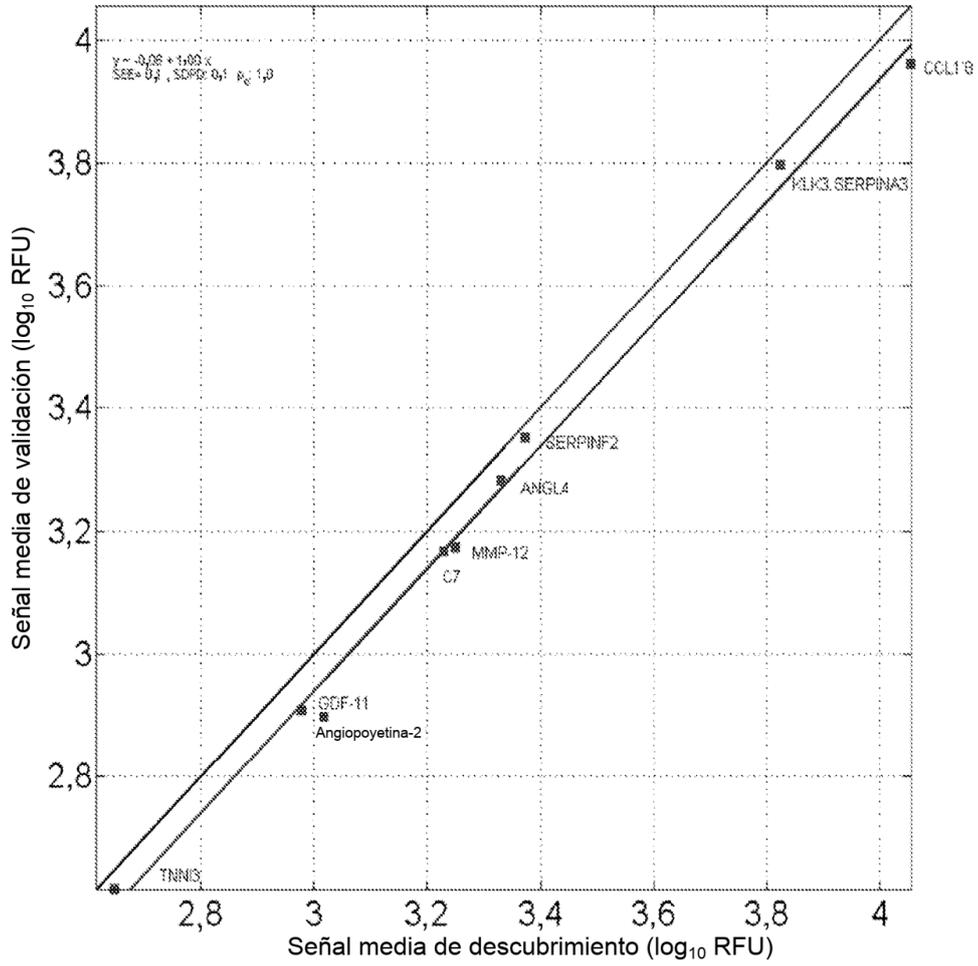


FIG. 3

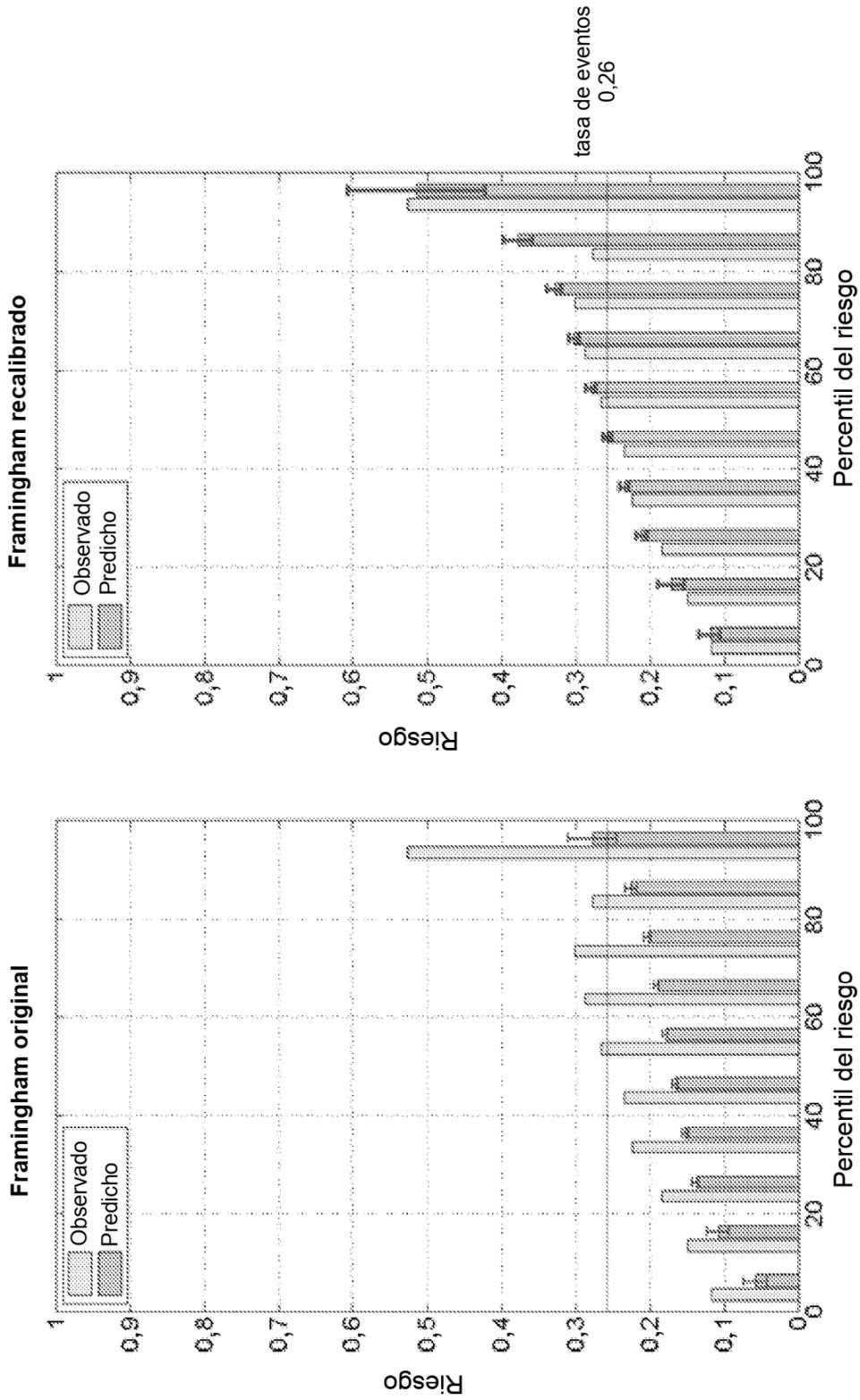


FIG. 4

Máx n.º Obs = 888

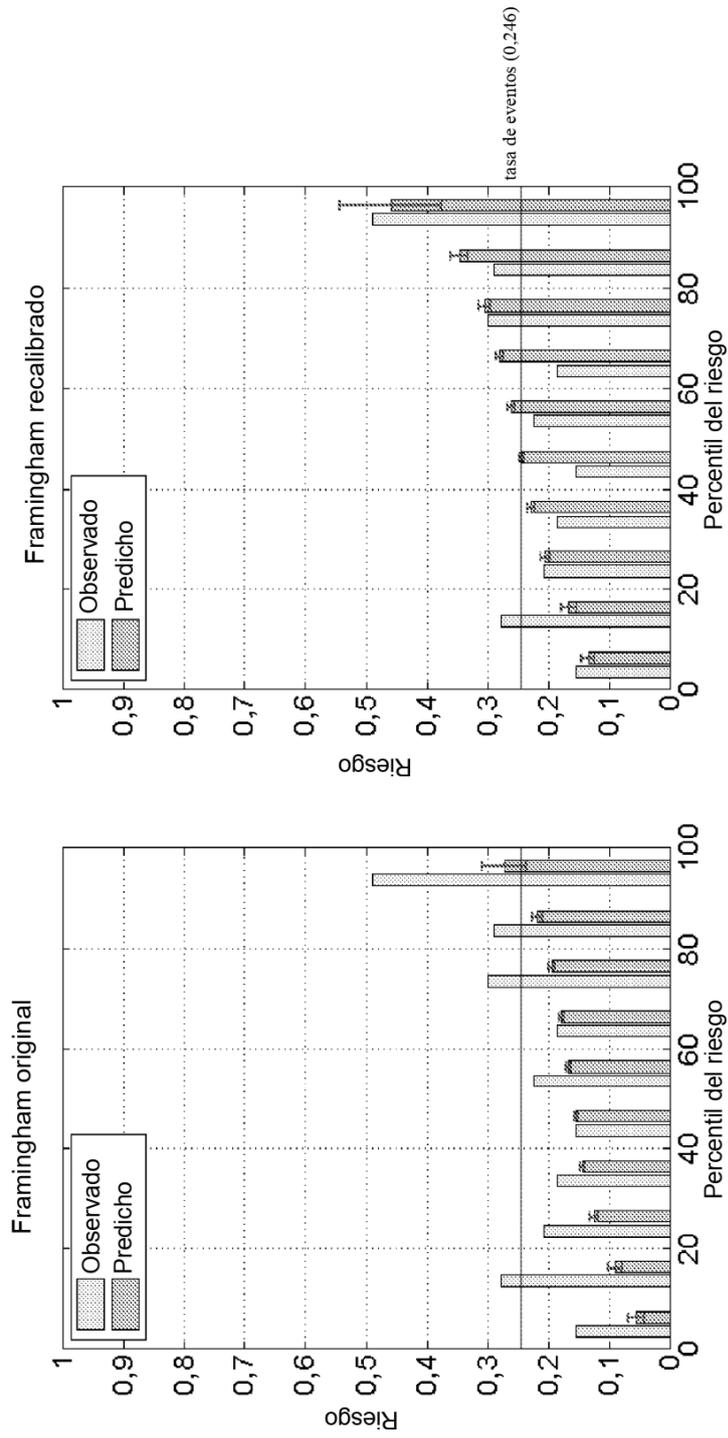


FIG. 5

Máx n.º Obs = 971

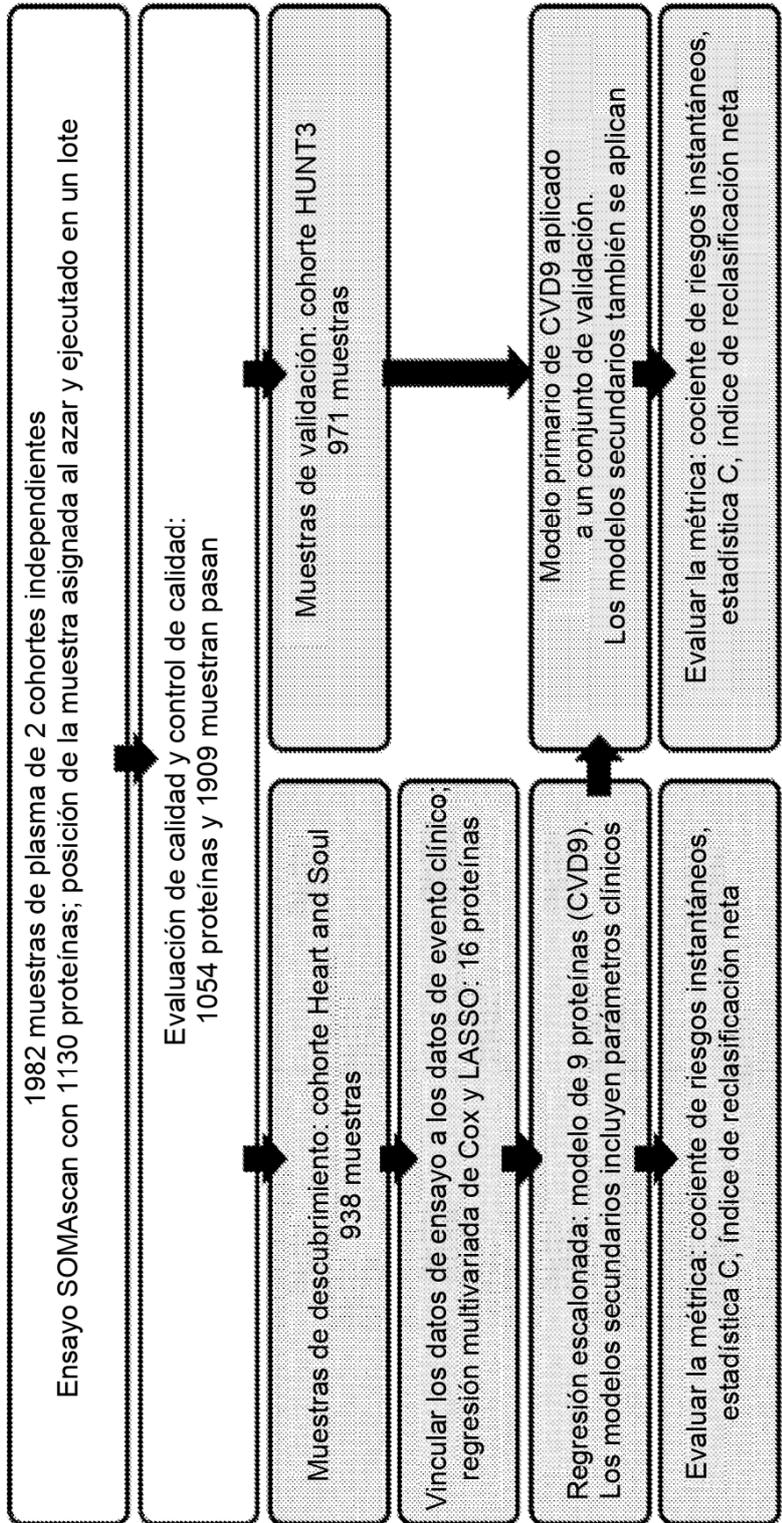


FIG. 6

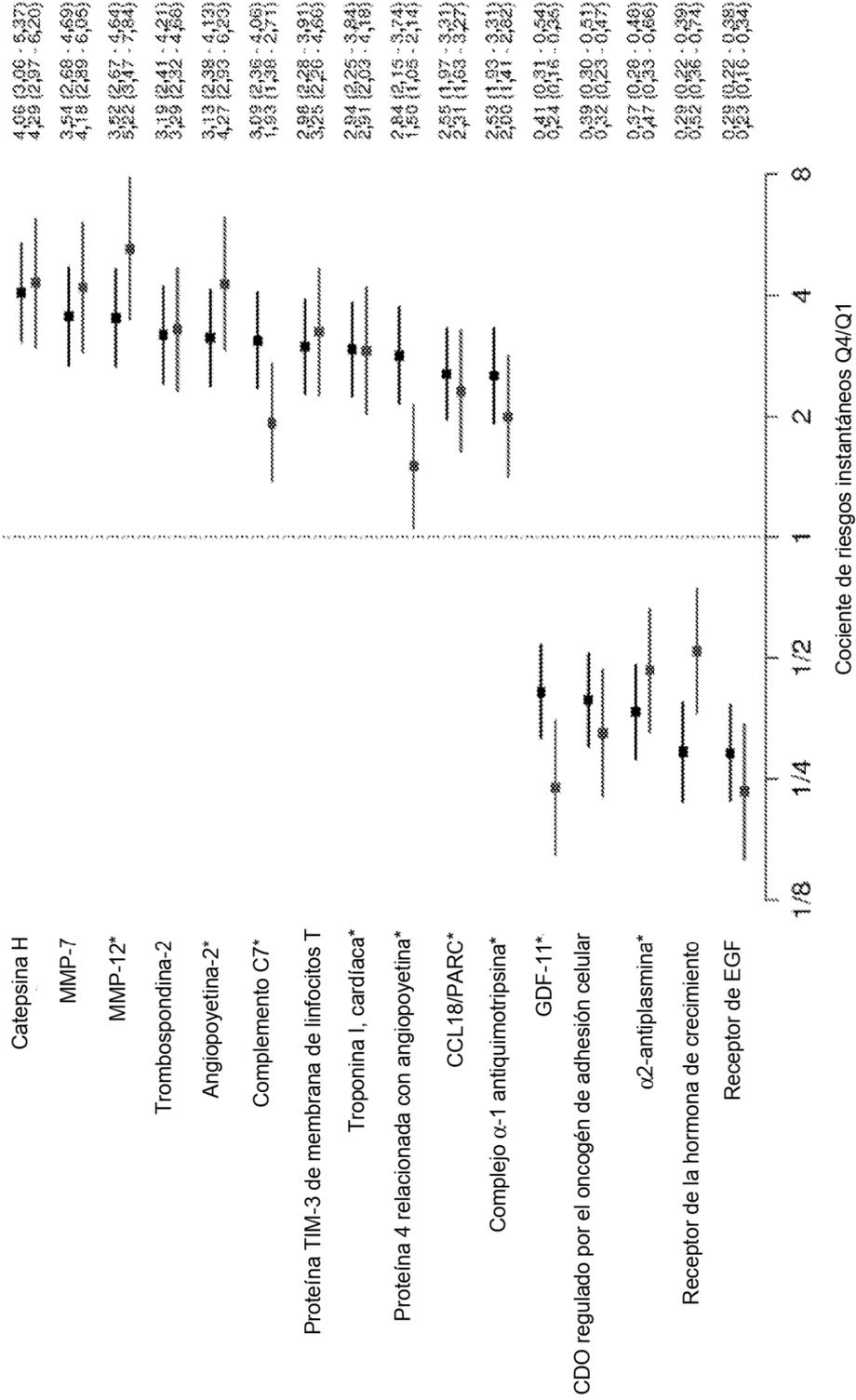


FIG. 7

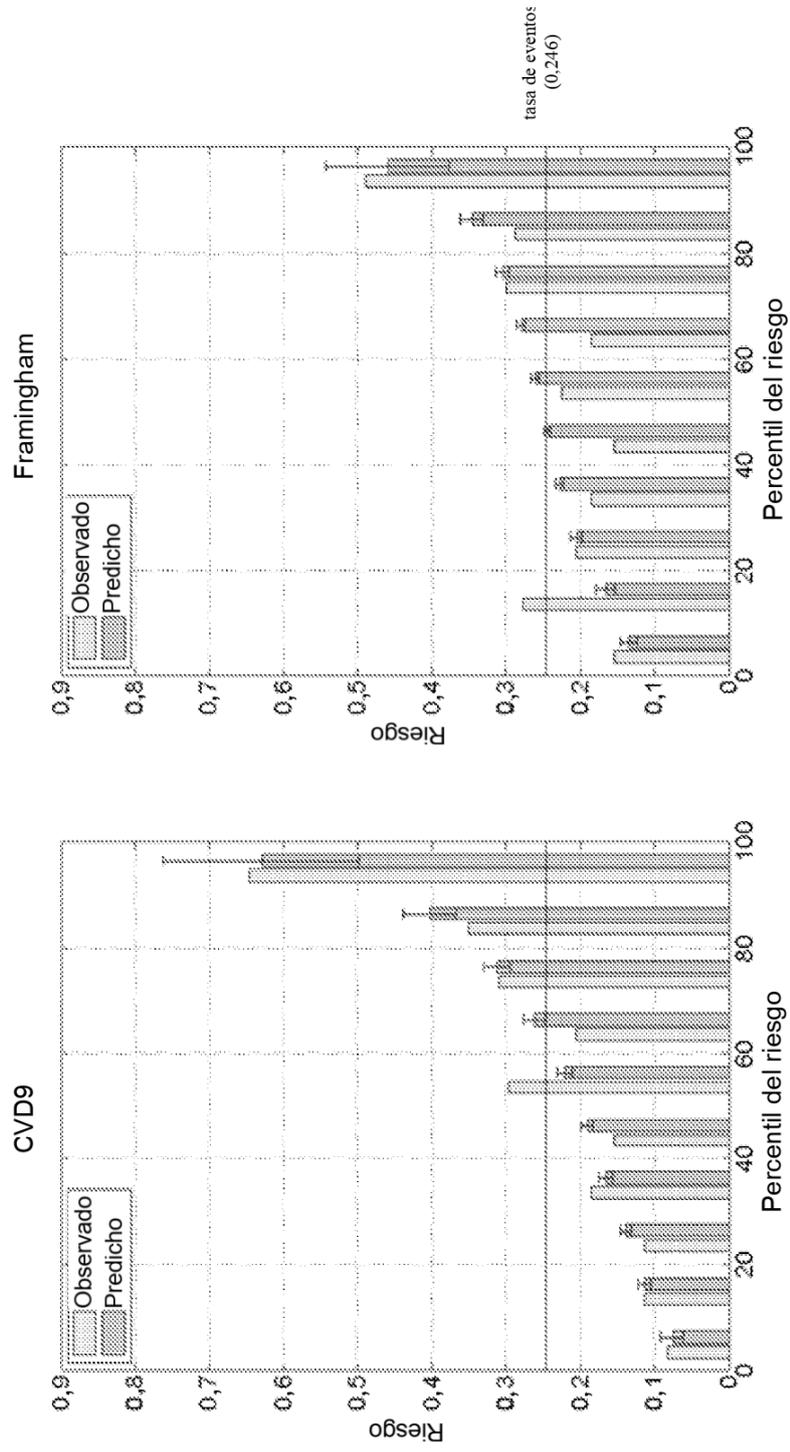
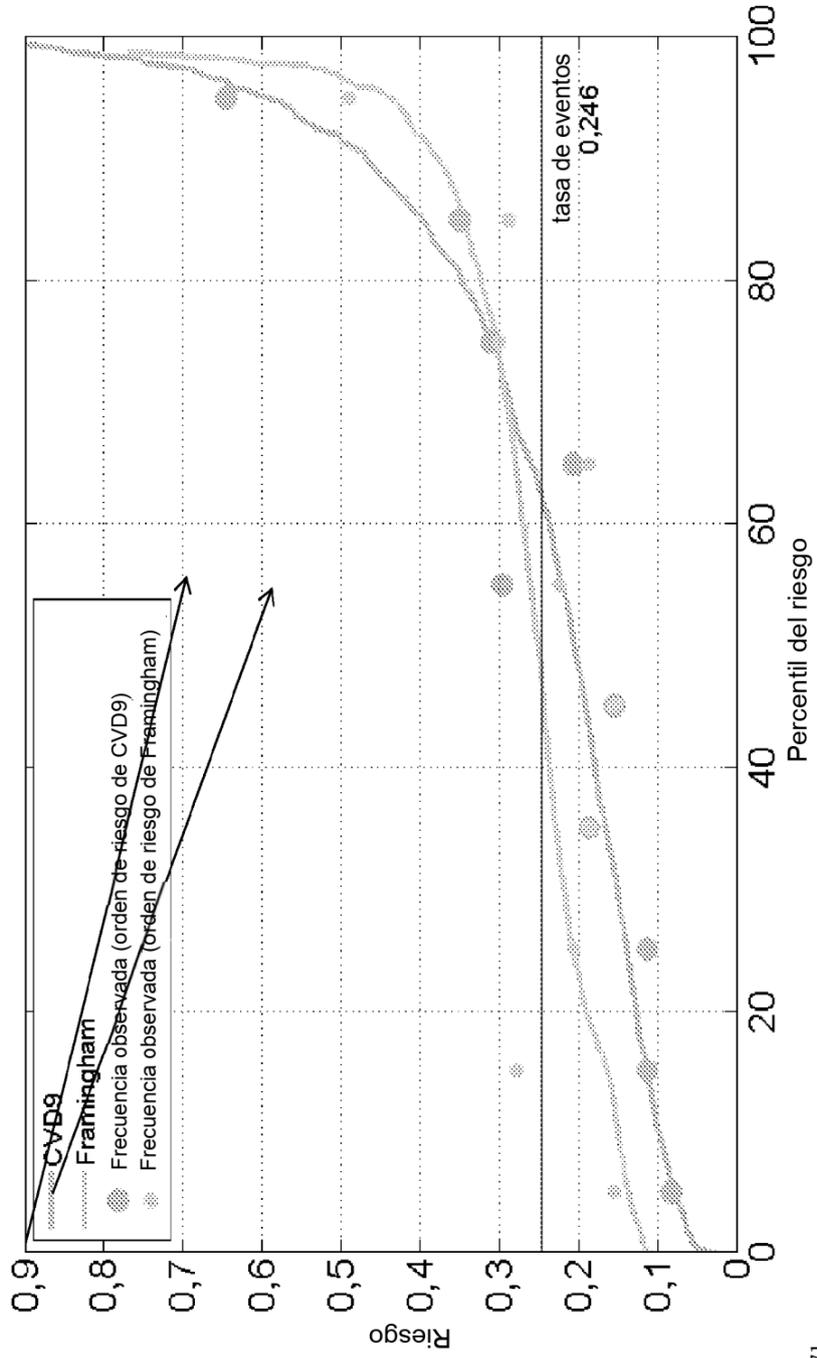


FIG. 8

Máx n.º Obs = 971

Eventos a 4 años (conjunto de validación)



Máx n.º Obs = 971

FIG. 9

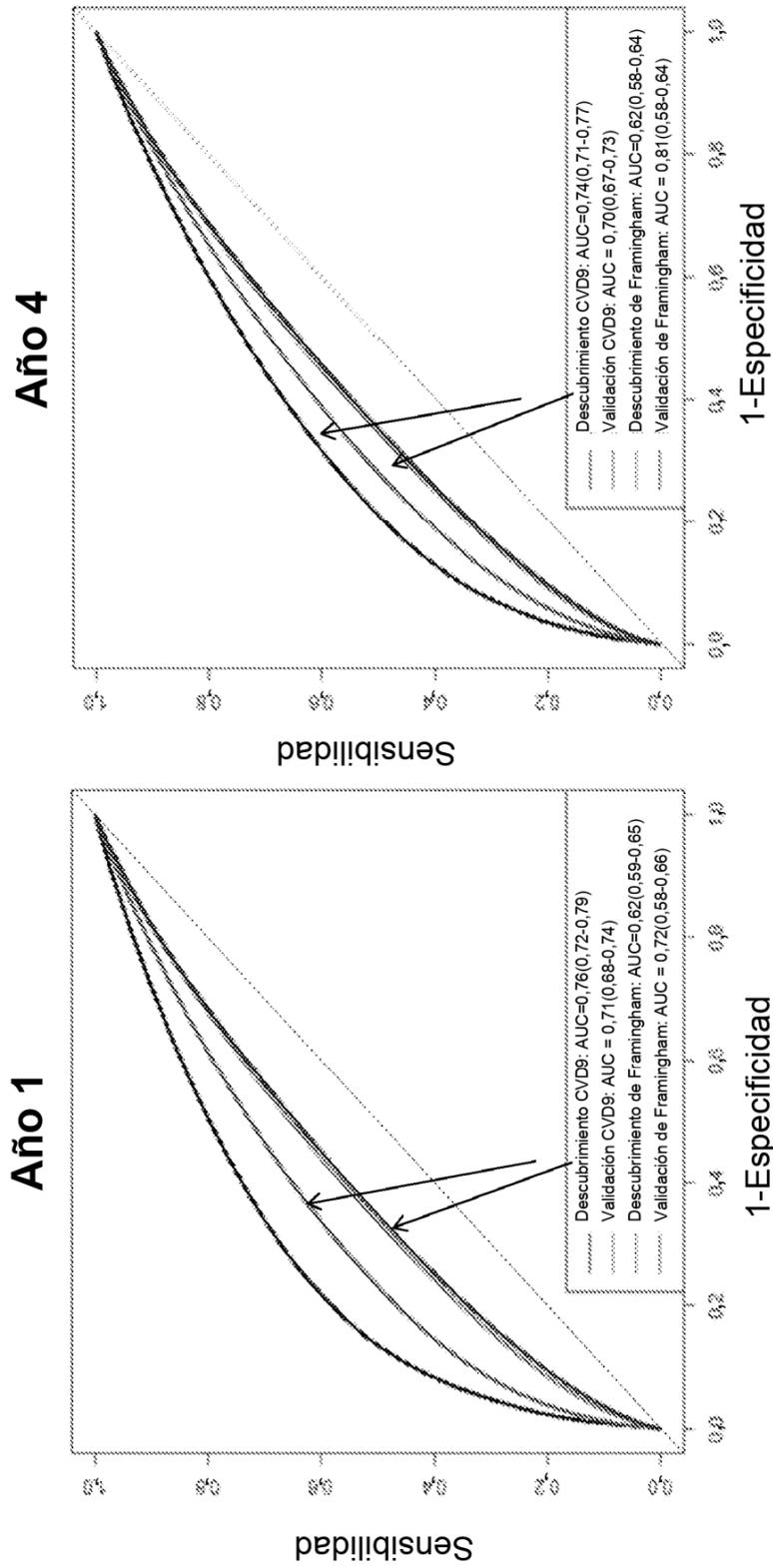


FIG. 10

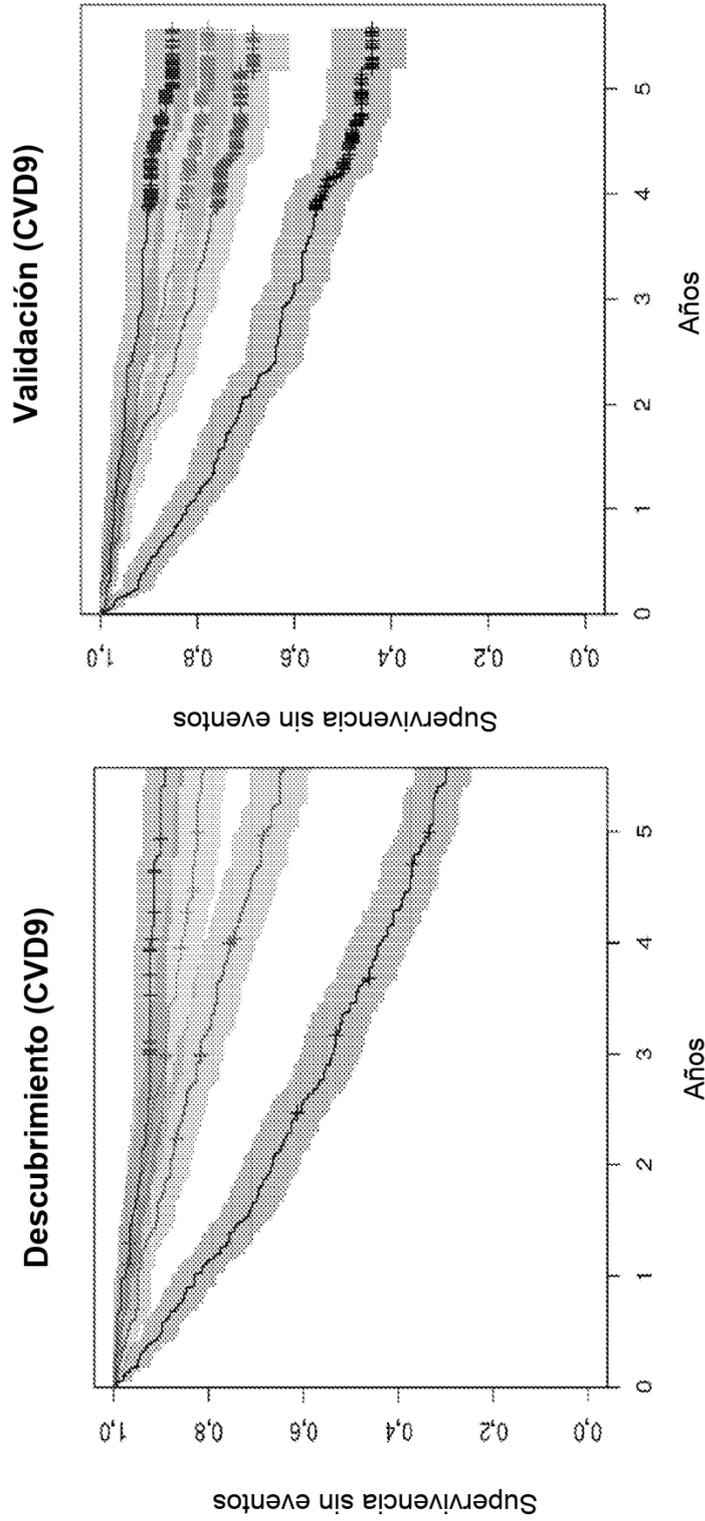


FIG. 11

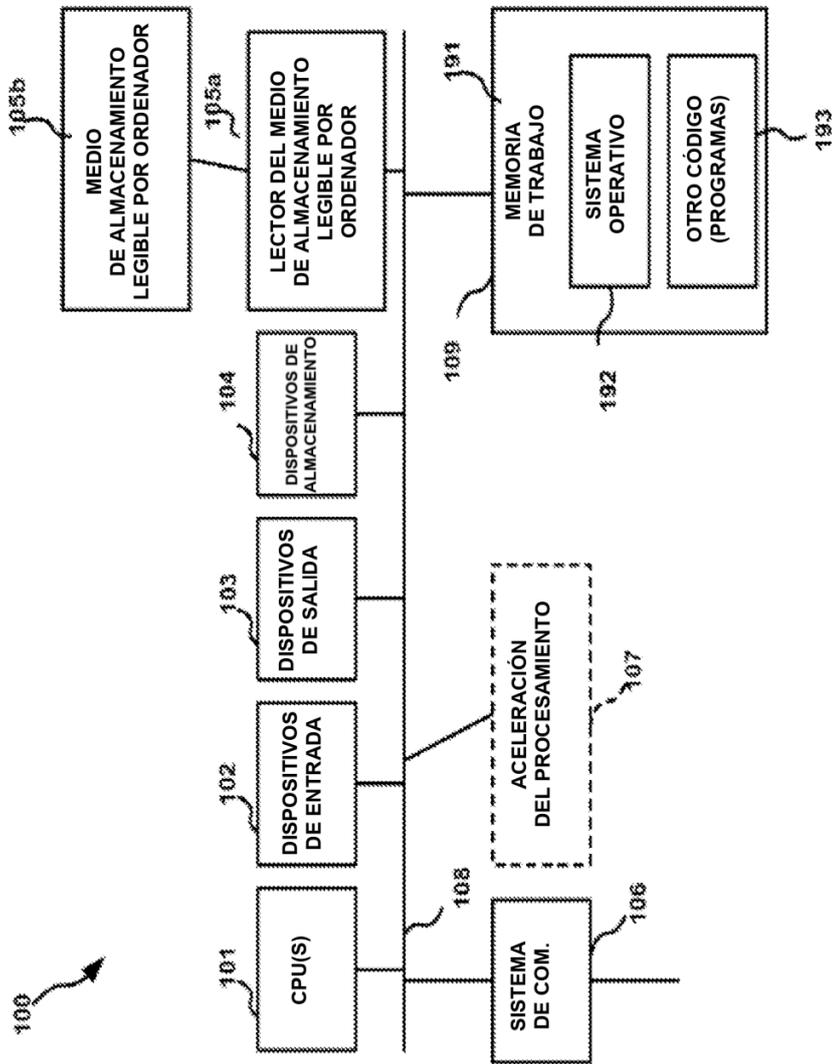


FIG. 12

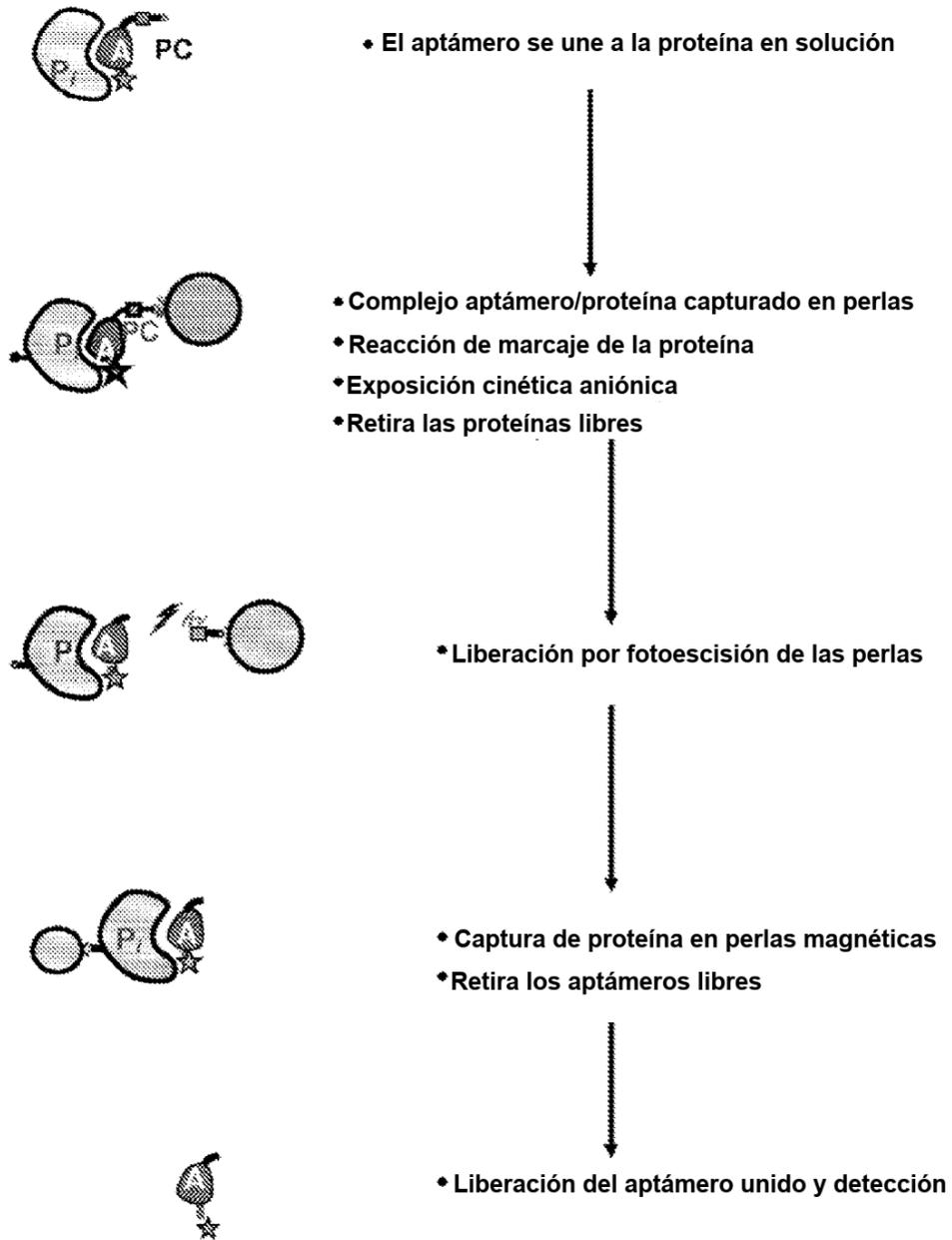
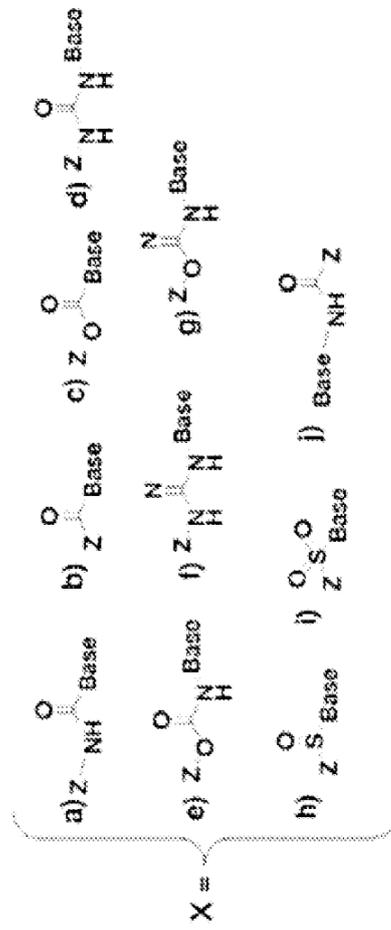


FIG. 13



Base = Uridina (U) o Citidina (C) (la unión está en la posición 5)

Z = R más grupo conector $(CH_2)_m$, donde $n = 0-3$

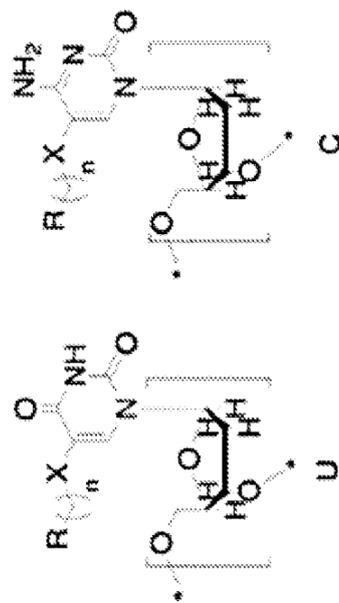


FIG. 14

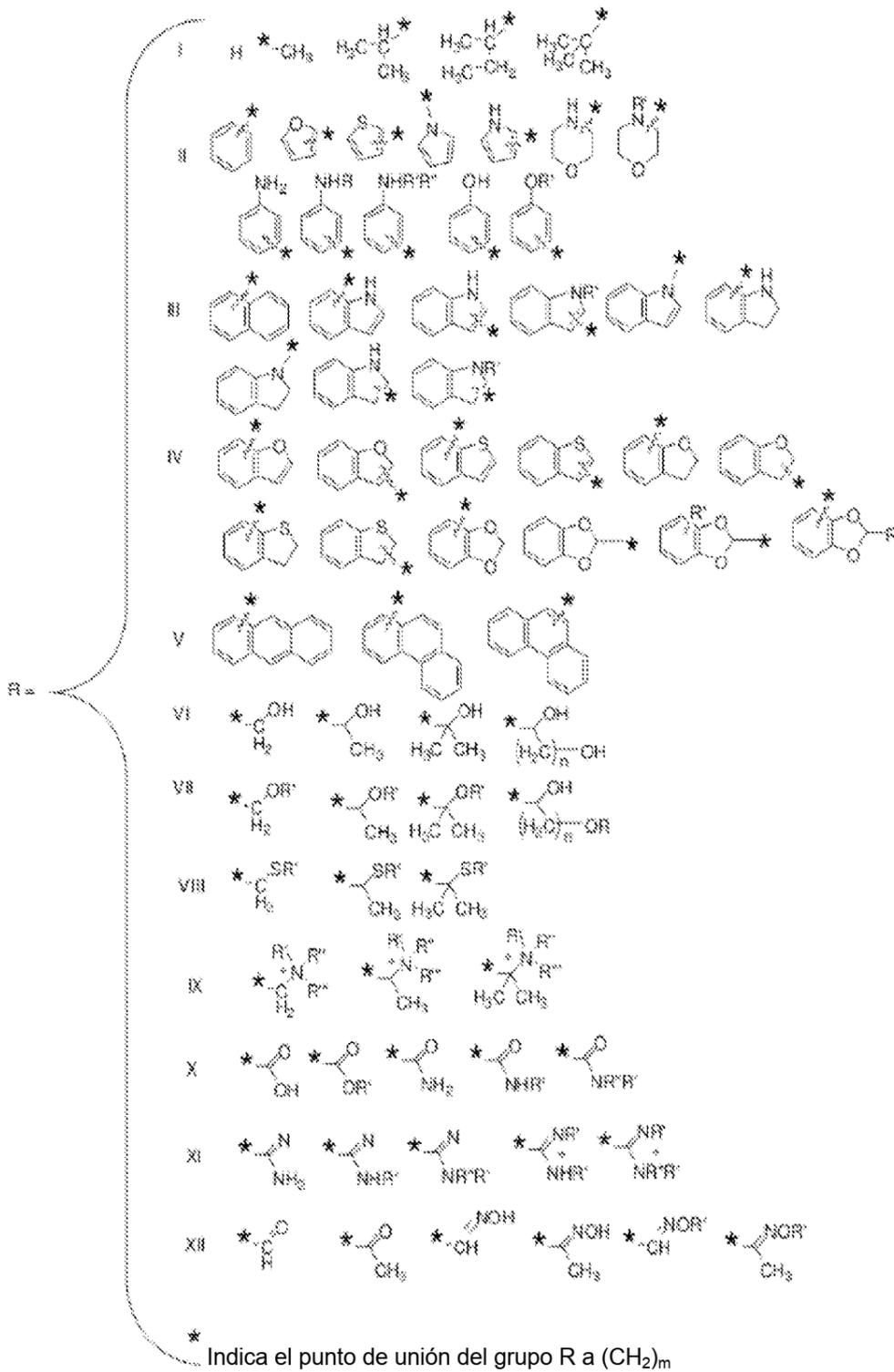


FIG. 14 (cont.)

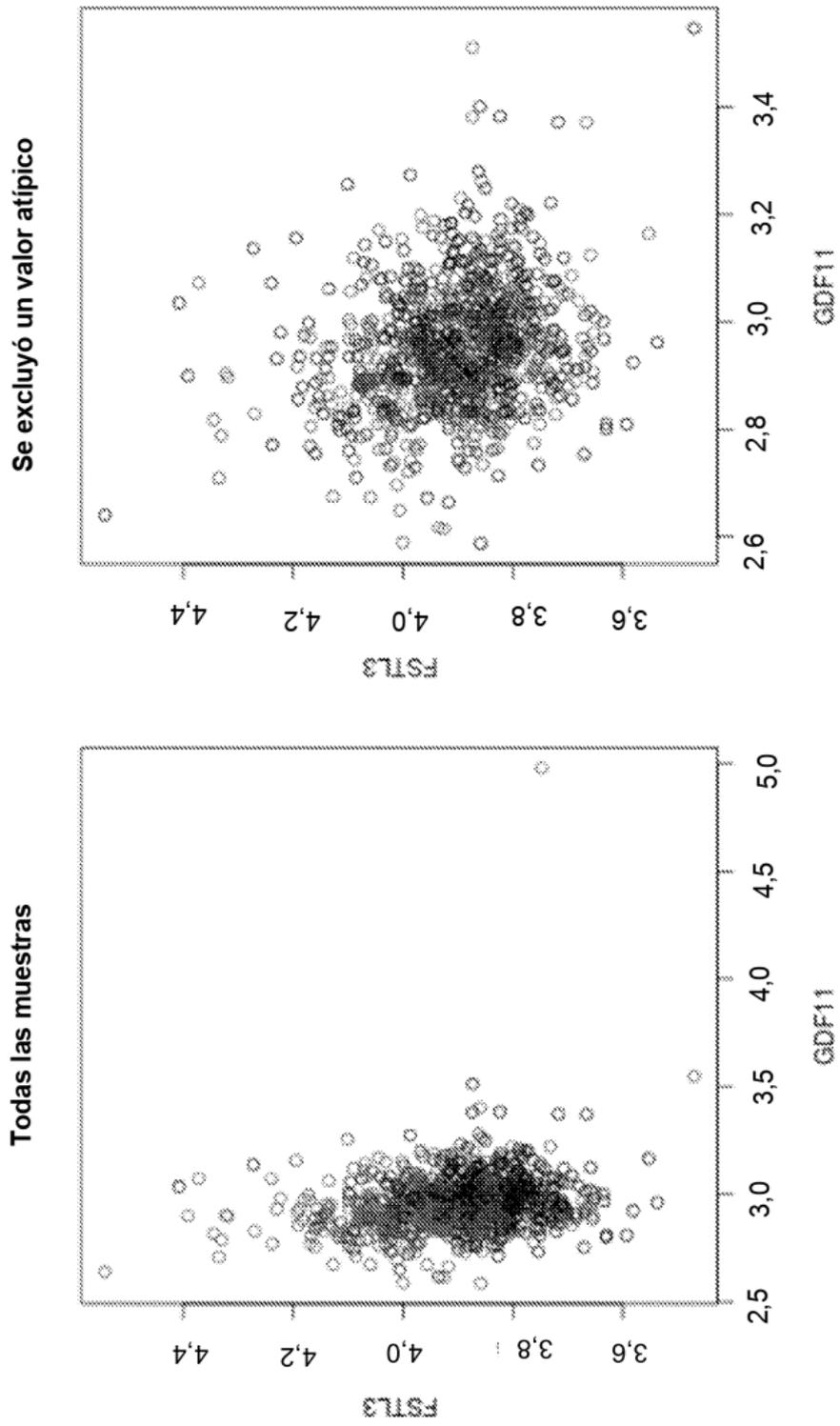


FIG. 15

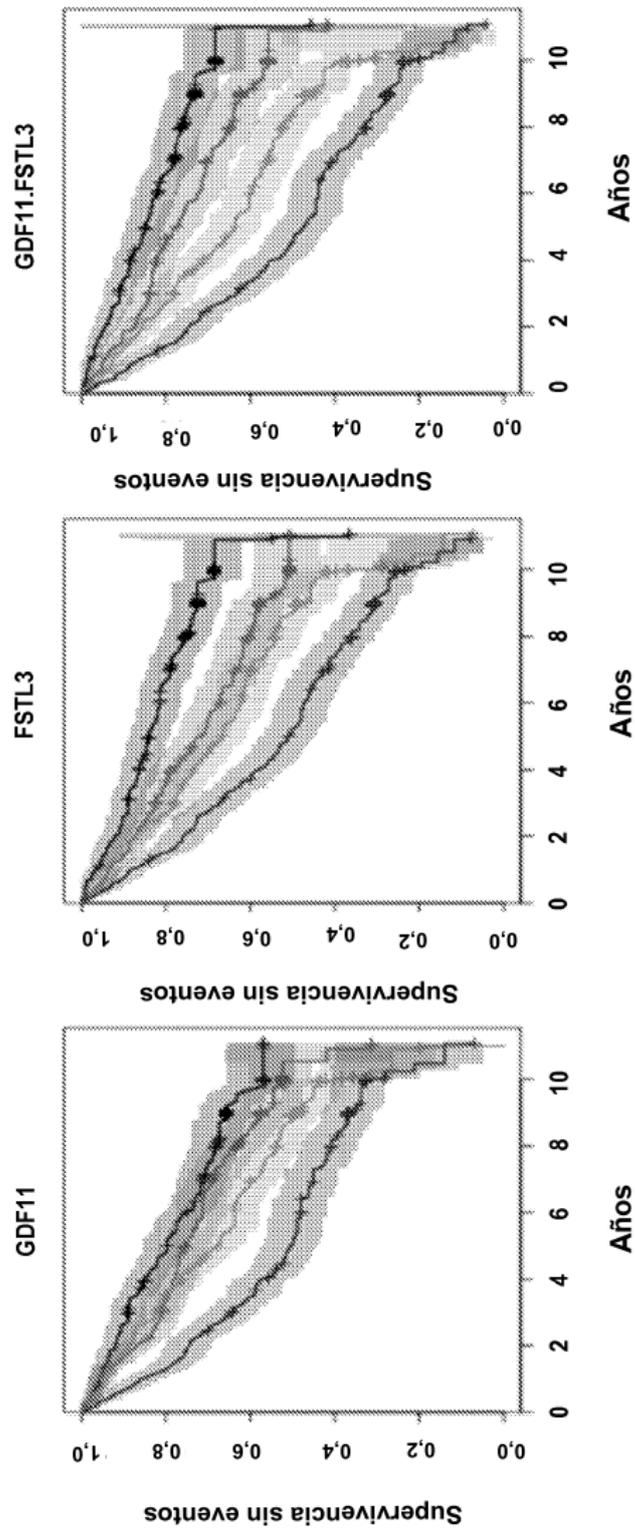


FIG. 16

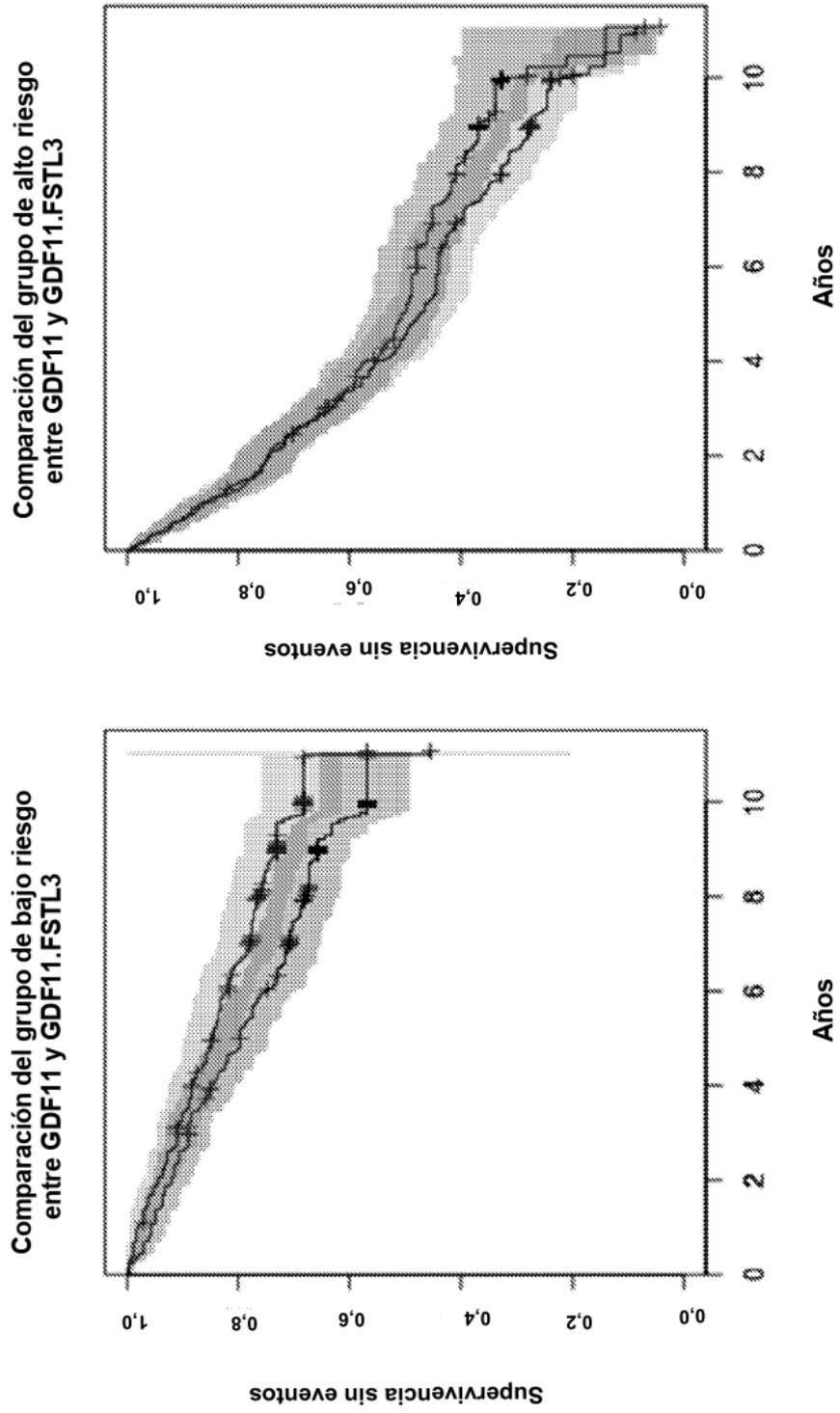
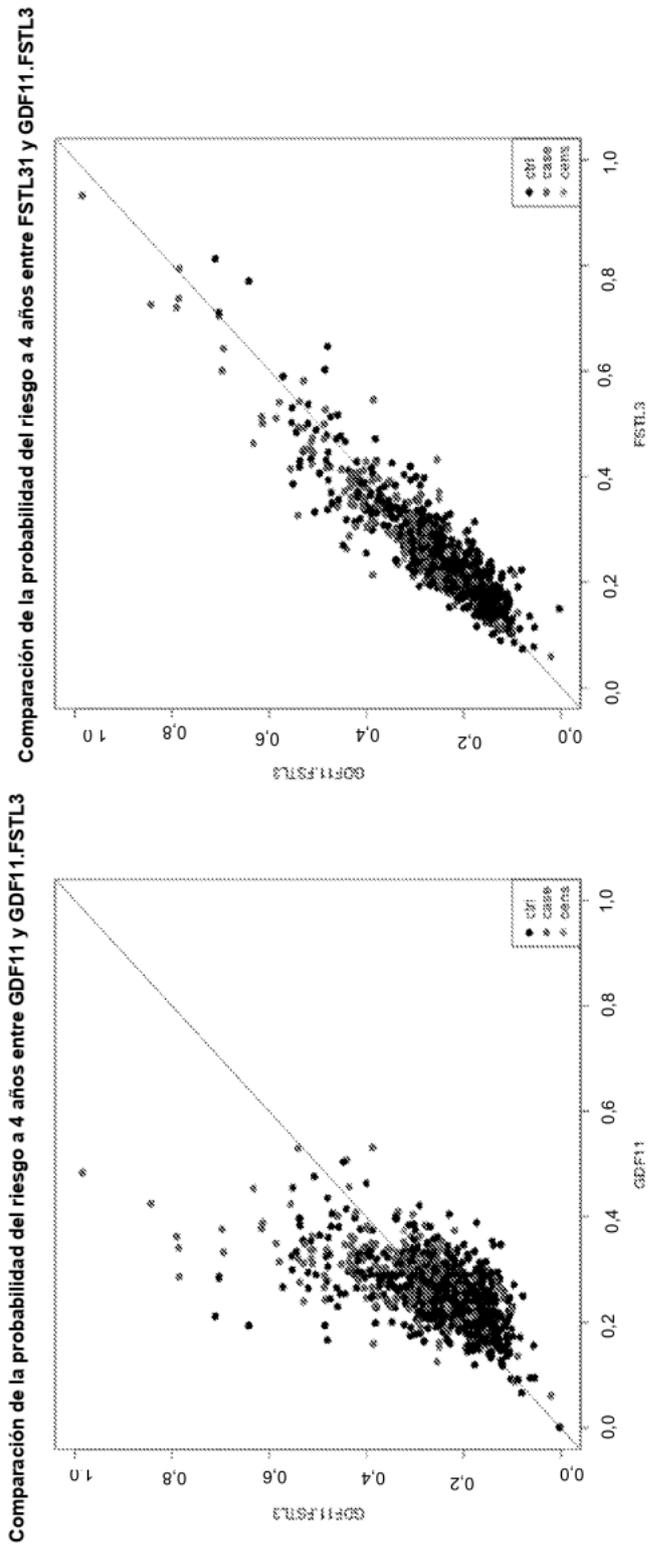


FIG. 17



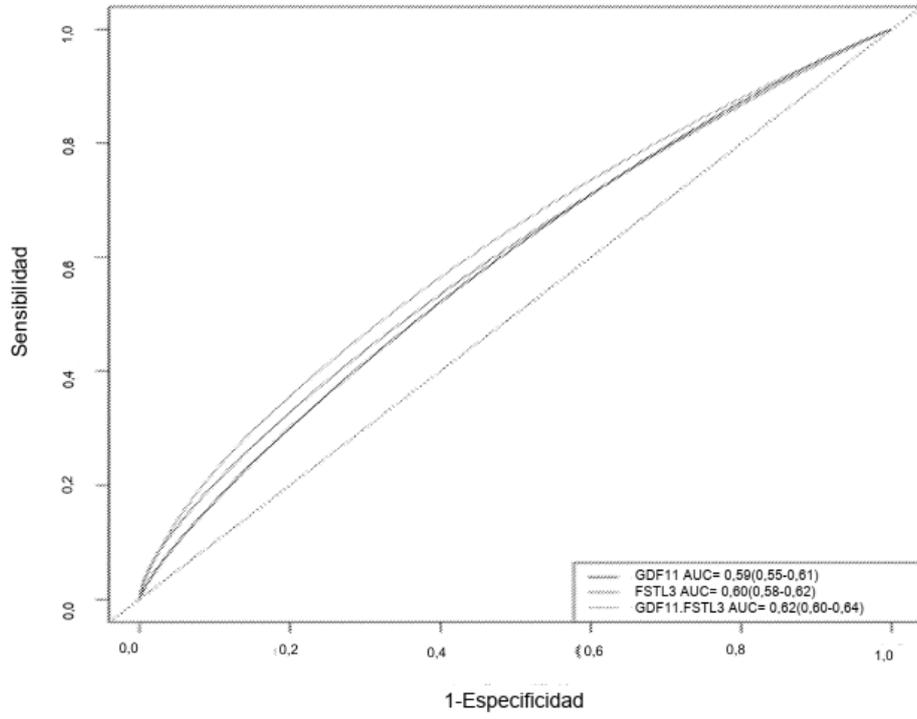


FIG. 19

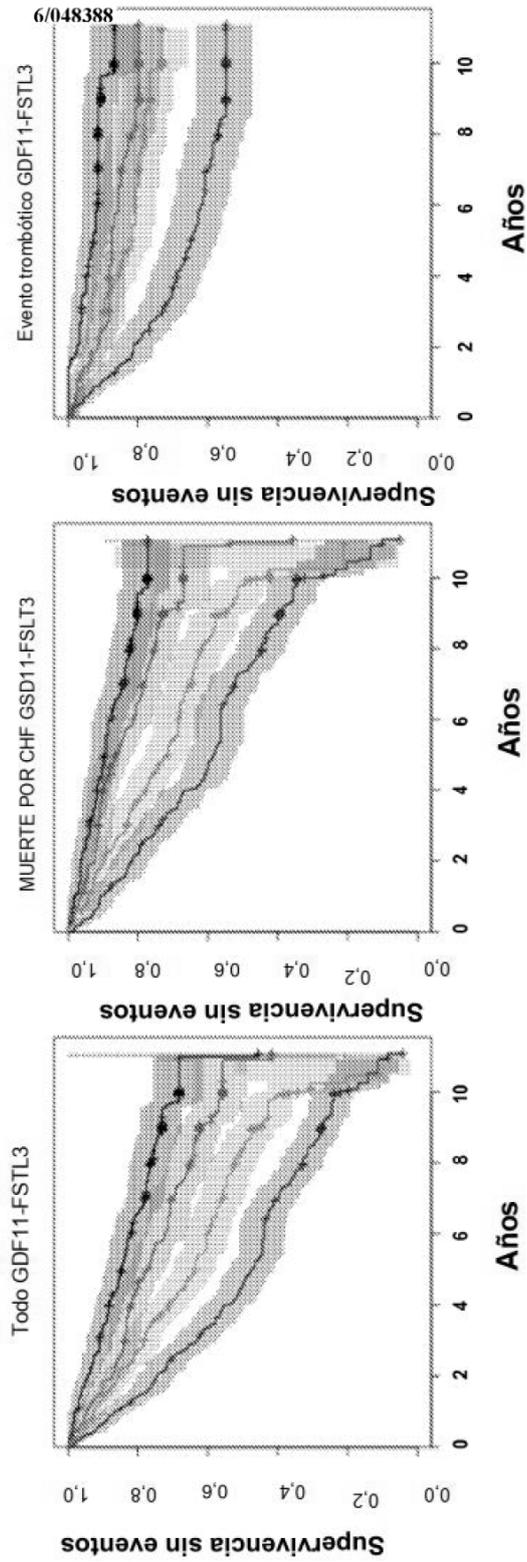


FIG. 20

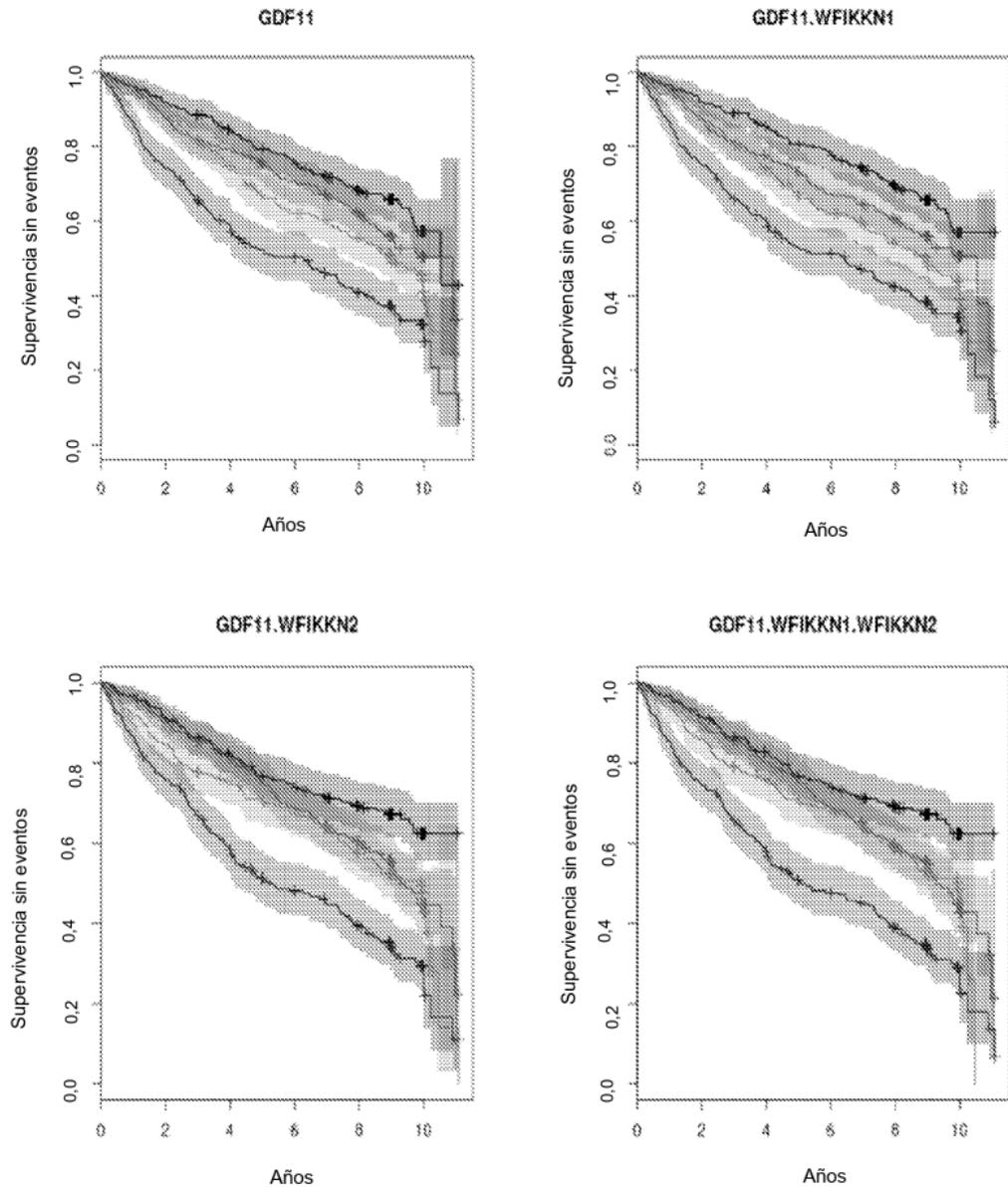


FIG. 21

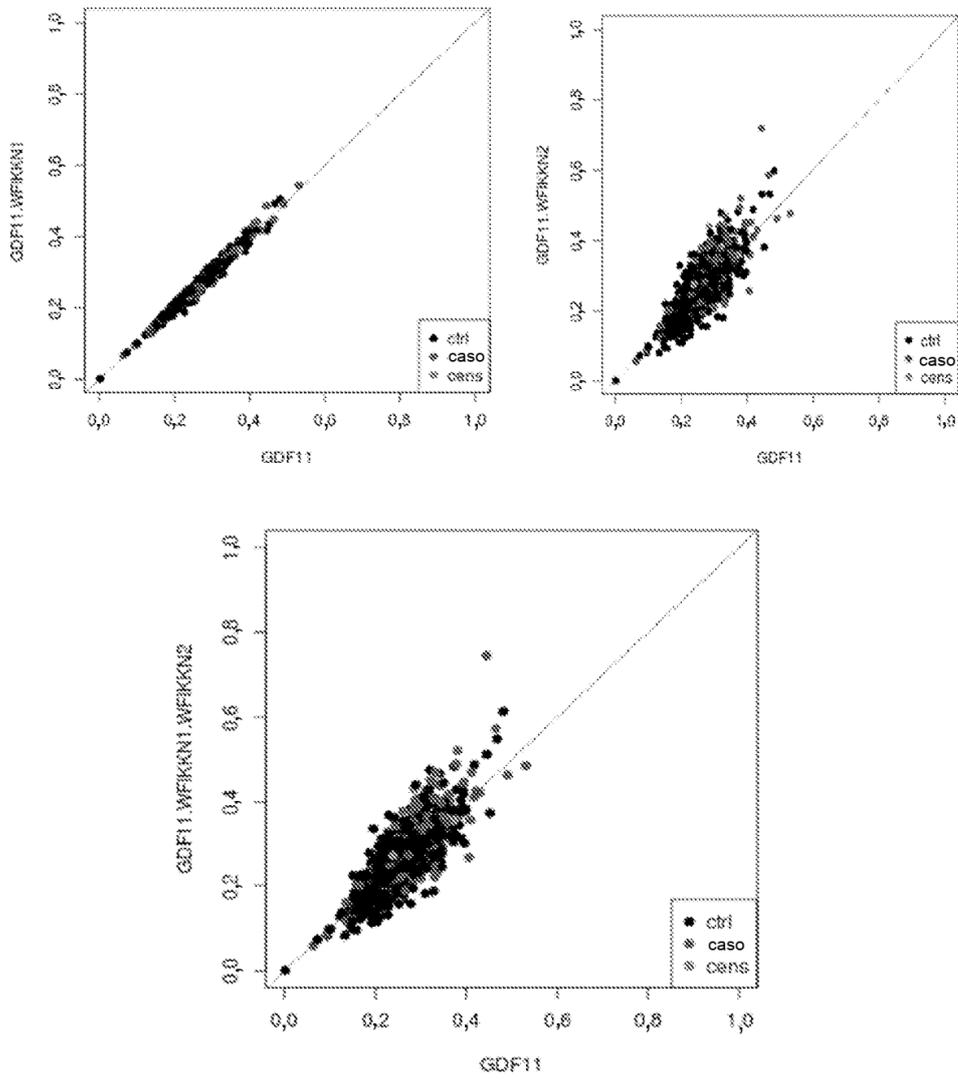


FIG. 22

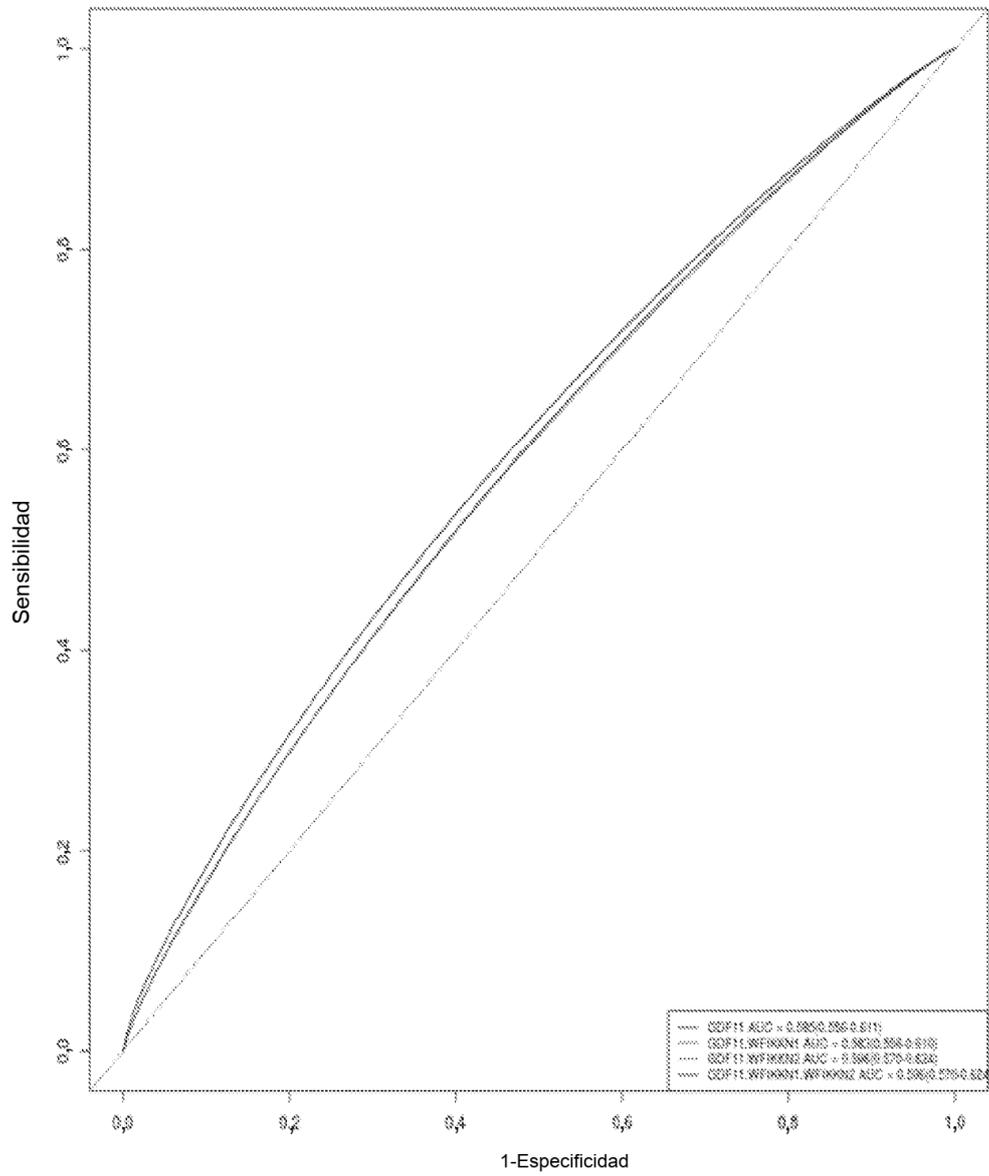


FIG. 23