(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 844 202**

(51) Int. CI.:

C12Q 1/6886 (2008.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **20.11.2014 PCT/US2014/066678**

(87) Fecha y número de publicación internacional: **28.05.2015 WO15077485**

(96) Fecha de presentación y número de la solicitud europea: **20.11.2014 E 14863459 (5)**

(97) Fecha y número de publicación de la concesión europea: **04.11.2020 EP 3071973**

(54) Título: **Triaje de pacientes que tienen hematuria asintomática usando biomarcadores genotípicos y fenotípicos**

(30) Prioridad:

21.11.2013 US 201361907013 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
21.07.2021

(73) Titular/es:

PACIFIC EDGE LIM (100.0%)
Anderson Lloyd, Level 10, Otago House, Cnr Moray Place and Princes Street
Dunedin, 9016, NZ

(72) Inventor/es:

DALPHIN, MARK;
DARLING, DAVID;
SUTTIE, JAMES, MILLER;
KAVALIERIS, LAIMONIS;
O'SULLIVAN, PAUL y
KUMAR, SATISH

(74) Agente/Representante:

PONS ARIÑO, Ángel

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

Triaje de pacientes que tienen hematuria asintomática usando biomarcadores genotípicos y fenotípicos

Reivindicación de prioridad

La presente solicitud reivindica el beneficio de prioridad respecto de la Solicitud Provisional de Patente de los Estados Unidos n.º 61/907.013, presentada el 21 de noviembre de 2013; Los inventores son David Darling, Satish Kumar, Mark Dalphin y Paul O'Sullivan.

Campo de la invención

La presente invención se refiere a la detección de pacientes que no están enfermos. Específicamente, la presente invención se refiere al uso de marcadores genéticos y marcadores fenotípicos para el triaje de pacientes que presentan hematuria sin cáncer. En particular, la presente invención se refiere al análisis de marcadores genéticos y marcadores fenotípicos en el triaje de pacientes con hematuria macroscópica o microscópica. Más particularmente, la presente invención se refiere al uso combinado de marcadores genéticos y fenotípicos para el triaje de pacientes con hematuria macroscópica o microscópica asintomática y para predecir si el estado de un paciente justifica procedimientos clínicos adicionales.

Antecedentes

La supervivencia de pacientes con cáncer mejora en gran medida cuando el cáncer se tratar de manera temprana. En el caso del cáncer de vejiga, los pacientes a los que se diagnostica enfermedad que está confinada en el sitio primario tienen una tasa de supervivencia a los 5 años del 73 %, en comparación con un 6 % para pacientes con enfermedad metastásica (Altekruse *et al*.). Por lo tanto, los desarrollos que propicien un diagnóstico temprano y preciso del cáncer de vejiga pueden dar lugar a un pronóstico mejorado para los pacientes. Para ayudar en la detección temprana del cáncer, se ha identificado una serie de marcadores específicos para el cáncer. Sin embargo, el uso de estos marcadores puede proporcionar resultados falsos positivos en pacientes que tienen enfermedades inflamatorias de vejiga y no cáncer de vejiga.

La hematuria asintomática ("HA") es una de las observaciones urológicas más frecuentes, con tasas de incidencia de entre un 2 % y un 30 %, dependiendo de la población (Schwartz G: *Proper evaluation of asymptomatic microscopic hematuria in the era of evidence-based medicine-progress is being made*. *Mayo Clin Proc*. 2013, 88(2); 123-125, McDonald M, Swagerty D, Wetzel L: *Assessment of Microscopic hematuria in adults*. *AFP* 2006 73:10, Grossfield G, Wolf J, Litwan M, Hricak H, Shuler C, Agerter D, *et al. Asymptomatic microscopic hematuria in adults*: *summary of AUA best practice policy recommendations*. *AFP* 2001:63:1145-54).

La HA es, sin embargo, indicativa de una gran variedad de patologías, con una incidencia de neoplasias malignas del tracto urinario en la población con HA en el intervalo del 1,9-7 %. Una serie completa de pruebas diagnósticas complementarias en todos los pacientes con HA representa una carga significativa para muchos sistemas sanitarios. Se ha explorado el uso de indicadores fenotípicos para discriminar entre pacientes de alto y bajo riesgo en un estudio reciente por Loo *et al*. (Loo R, Lieberman S, Slezak J, Landa H, Mariani A, Nicolaisen G, Aspera A y Jaconsen S: *Stratifying risk of urinary tract malignant tumors in patients with asymptomatic microscopic hematuria*. *Mayo Clin Proc*. 2013, 88(2); 129-138).

El estudio anterior efectuado en 4414 pacientes que presentaban HA confirmada demostró que un 73 % de los pacientes no tenían una causa identificada, mientras que en un 26 % de los pacientes estaba justificado efectuar algún tipo de prueba complementaria para identificar la causa. Aproximadamente un 2,5 % de los pacientes que presentaron HA fueron diagnosticados de neoplasia maligna urotelial, y los diagnósticos alternativos fueron otras afecciones, tales como infección del tracto urinario (ITU) (2,3 %), nefrolitiasis (16,2 %), sangrado de próstata (4 %) y contaminación (0,4 %) (Loo *et al*., anteriormente citado).

Sumario

Los presentes inventores han identificado un nuevo problema en el campo, específicamente cómo identificar pacientes que presentan hematuria que no tienen o que se encuentran en bajo riesgo de tener cáncer de vejiga. Esto resuelve el problema de que muchos pacientes con hematuria y sin cáncer de vejiga puedan someterse a costosas e invasivas pruebas complementarias adicionales cuando dichas pruebas no son necesarias. Por tanto, la invención es útil para descartar individuos de los riesgos y costes asociados con las pruebas complementarias completas para el cáncer de vejiga, cuando una combinación de información genética y fenotípica proporciona una identificación de pacientes que no tienen o que se encuentran en bajo riesgo de tener cáncer de vejiga y para efectuar un triaje eficaz de pacientes que no tienen cáncer, de aquellos que tienen afecciones cancerosas, incluyendo carcinomas uroteliales, carcinoma de células transicionales (CCT) y afecciones no cancerosas, entre las que se incluyen enfermedades inflamatorias. La presente invención representa un nuevo enfoque para un nuevo problema, que es inesperadamente útil, no para diagnosticar cáncer, sino para diagnosticar la ausencia de cáncer. El

uso de combinaciones de criterios genéticos y fenotípicos proporciona una inesperadamente mejor discriminación que las variables genéticas o fenotípicas de manera individual. Se obtuvieron datos de 541 observaciones procedentes de observaciones según los estándares CLIA y del ensayo de producto CURT +North Shore usando procedimientos de remuestreo con reposición para validación interna. Las variables fenotípicas incluyeron la presencia de: (1) historial de tabaquismo, (2) hematuria, (3) género y (4) edad. Las variables genéticas incluyeron el análisis de la expresión de IGF, HOXA13, MDK, CDC e IL8R. El modelo genético + fenotípico ("G+P") tuvo un rendimiento inesperadamente mejor que cualquiera de las variables genéticas o fenotípicas de manera individual.

La macrohematuria o el hallazgo de sangre identificada visualmente en la orina es un hallazgo común en pacientes con cáncer de vejiga. Para estos pacientes, normalmente es práctica común efectuar procedimientos diagnósticos adicionales para diagnosticar el cáncer de vejiga. Sin embargo, sigue siendo un problema identificar fácilmente pacientes con microhematuria y comprender las implicaciones de la microhematuria en los carcinomas uroteliales.

Los presentes inventores proporcionan en el presente documento métodos para determinar si un paciente que presenta macrohematuria o microhematuria puede ahorrarse procedimientos clínicos adicionales costosos e invasivos para detectar carcinomas uroteliales (CU), entre los que se incluyen cáncer de vejiga, en caso de que dicho paciente tenga una probabilidad lo suficientemente baja de tener cáncer de vejiga como para justificar no llevar a cabo procedimientos adicionales.

Se describen factores atribuibles a la alta probabilidad de carcinoma urotelial (CU). Los factores demográficos, tales como el género, la raza y la edad, además de factores ambientales, tales como historial de tabaquismo y la exposición ocupacional a aminas aromáticas, contribuyen significativamente al riesgo de desarrollar CU. La caracterización de pacientes en la evaluación sanitaria basándose en estos factores se usa rutinariamente según las circunstancias del caso. Por ejemplo, generalmente se acepta que un varón de 60 años con historial de tabaquismo que presente hematuria tiene más probabilidad de ser positivo para CU que una mujer de 35 años no fumadora que presente los mismos síntomas, aunque estas diferencias no se han cuantificado para contribuir a la probabilidad general de que el paciente tenga CU. La atribución de ponderaciones específicas a diversos factores genotípicos y fenotípicos y la combinación de estos con el resultado de una prueba diagnóstica pueden aumentar significativamente la precisión de la potencia diagnóstica de las pruebas no invasivas y aportan a los profesionales mayor certidumbre a la hora de separar pacientes basándose en su probabilidad de tener CU, definida por los resultados de la prueba clínica y de biomarcadores.

Aunque hay métodos disponibles para detectar la presencia de cáncer de vejiga, no hay métodos fiables y precisos para determinar si un paciente no tiene o se encuentra en bajo riesgo de tener cáncer de vejiga. Para abordar esta necesidad, los presentes inventores han desarrollado nuevos métodos analíticos para distinguir las afecciones cancerosas de las no cancerosas en pacientes que presentan hematuria, ya sea macrohematuria o microhematuria. En algunos aspectos de la presente invención, los presentes inventores combinan variables fenotípicas cuantificadas y la expresión cuantificada de marcadores genéticos para formar un índice de segregación combinado (el "ÍNDICE G + P") para discernir mediante triaje a los pacientes con HA con bajas probabilidades de tener CU de los pacientes con HA que tienen altas probabilidades de tener CU. Esta segregación define a aquellos pacientes que no necesitan pruebas complementarias urológicas completas de los que requieren pruebas complementarias completas y de este modo, evita pruebas complementarias innecesarias en pacientes con baja probabilidad de CU.

Ensayos fenotípicos

Las variables fenotípicas evaluadas en el ÍNDICE G+P incluyen la frecuencia de hematuria (HFREQ), la edad, el sexo, el historial de tabaquismo y el número de glóbulos rojos (RBC). Estos términos se definen más adelante en el presente documento. En el presente documento se definen variables fenotípicas para que incluyan hallazgos y observaciones clínicas.

Ensayos genotípicos

En general, los ensayos genotípicos preferidos desarrollados por Pacific Edge Ltd. incluyen la cuantificación de la expresión de los marcadores genéticos CDC2, HOXA13, MDK e IGFBP5 (un "ensayo de 4 marcadores"). En otro ensayo preferido, se cuantifican los 4 marcadores anteriores y un quinto marcador, IL8R (un ensayo de "5 marcadores" o Cxbladder®; una marca registrada de Pacific Edge Ltd., Dunedin, Nueva Zelanda) (Holyoake A, O'Sullivan P, Pollock R *et al*: *Development of a multiplex RNA urine test for the detection and stratification of transitional cell carcinoma of the bladder*. *Clin Cancer Res* 2008;14: 742, y O'Sullivan P, Sharples K, Dalphin M *et al*: *A Multigene Urine Test for the Detection and Stratification of Bladder Cancer in Patients Presenting with Hematuria*. *J Urol* 2012, Vol. 188 n.º 3; 746) y la patente WO2012/078053 (PACIFIC EDGE BIOTECHNOLOGY LIMITED), titulada "Novel Markers for Detection of Bladder Cancer".

En realizaciones preferidas, puede efectuarse un ensayo de 4 marcadores en orina no fraccionada usando amplificación mediante la PCR para cuantificar cuatro marcadores de ARNm (para CDC2, HOXA13, MDK e IGFBP5), que se sobreexpresan en el carcinoma urotelial. IL8R se sobreexpresa fuertemente en neutrófilos y por consiguiente, está elevada en afecciones inflamatorias no malignas. La inclusión de este 5.º marcador de ARNm

redujo significativamente el riesgo de detección de falsos positivos de carcinoma de células transicionales (CCT). Desde la perspectiva de los pacientes, la prueba es no invasiva y muy sencilla. Se toma una sola muestra de orina, normalmente pero no de manera exclusiva de orina del chorro medio y esto normalmente puede efectuarse en casa sin acudir a la clínica.

Se ha demostrado que el ensayo Cxbladder® es considerablemente más sensible que la citología en pacientes que presentan hematuria macroscópica. De manera más destacable, el ensayo Cxbladder® logró una sensibilidad del 100 % (con una especificidad predeterminada del 85 %) para todos los carcinomas uroteliales con un estadio mayor de Ta y del 97 % para todos los tumores de grado elevado. El ensayo Cxbladder® asigna un solo valor de puntuación que combina la expresión génica cuantitativa de cinco genes representados en la orina de los pacientes. La puntuación separa a los pacientes en tres clases basándose en la probabilidad de que el paciente tenga un carcinoma urotelial.

Para los pacientes que presentan hematuria, (ya sea macrohematuria o microhematuria), la presente invención ha demostrado que mejora estas herramientas genotípicas (ya sea el ensayo de 4 marcadores o el ensayo Cxbladder®) con la adición de variables fenotípicas tomadas del paciente a lo largo del mismo periodo de tiempo y la combinación de estas en una nueva herramienta, un índice que puede usarse para separar a los pacientes en tres clases de riesgo en relación con la probabilidad de que el paciente tenga carcinoma urotelial ("CU").

Aspectos

A continuación se ilustran aspectos de la presente invención. Puede entenderse que estos no son los únicos aspectos o realizaciones de la presente invención. Los expertos habituales en la materia pueden combinar uno o más aspectos para producir aspectos o realizaciones adicionales.

Un aspecto incluye un método para determinar, en un paciente que presente hematuria, el nivel de riesgo de que tenga cáncer urotelial, que comprende:

proporcionar una muestra de orina de dicho paciente;
cuantificar un valor, MI, que comprende cuantificar los niveles de expresión de MDK, CDC2, HOXA13 e IGFBP5 humano en dicha muestra;
evaluar las variables fenotípicas HFREQ, EdadMQ, el sexo, SMK y RBC de dicho paciente;
calcular el ÍNDICE G+P según:

la fórmula (i), ÍNDICE G+P = (1*HFREQ+3*Género+4*SMK) + (5*M1+2*IL-8),

o

la fórmula (ii), ÍNDICE G+P= (w1*HFREQ+w2*EdadMQ50+w3*Género+w4*SMK+w5*RBC) + (w6*M1+w7*IL-8),

o

ÍNDICE G+P = -8,46 + 0,79 IGF -1,60 HOXA + 2,10 MDK + 0,95 CDC - 0,38 IL8R + 0,98 SNS + 0,56 Hfreq +1,11 Género + 0,64 Edad;

y
determinar si el ÍNDICE G+P es mayor que un umbral que indica el nivel de riesgo de que el paciente tenga cáncer urotelial.

Los aspectos adicionales incluyen el método del otro aspecto, donde dicho umbral se selecciona entre el grupo de valores del ÍNDICE G+P de 0 a 5, de 6 a 10 o de 11-15, donde dicho valor de 0 a 5 indica bajo riesgo, de 6 a 10 indica riesgo moderado y de 11-15 indica alto riesgo.

Los aspectos adicionales incluyen el método de cualquier otro aspecto, donde en caso de que dicho umbral sea un valor del ÍNDICE G+P sea de 6-10, dicho paciente se somete a pruebas clínicas o de laboratorio adicionales.

Otros aspectos adicionales incluyen el método de cualquier aspecto anterior, donde en caso de que dicho umbral sea un valor del ÍNDICE G+P sea de 11-15, dicho paciente se somete a pruebas clínicas o de laboratorio adicionales.

Otros aspectos más incluyen el método de cualquier otro aspecto, donde en caso de que dicho umbral sea un valor del ÍNDICE G+P sea de 0-5, se asigna al paciente a una lista de observación para pruebas clínicas o de laboratorio adicionales.

Los aspectos adicionales incluyen el método de cualquier otro aspecto, donde el umbral se establece usando un

método estadístico.

Otros aspectos más incluyen el método de cualquier otro aspecto, en donde el método estadístico es uno cualquiera de análisis de discriminante lineal (LDA), regresión logística (LogReg), máquina de vectores de soporte (SVM), K 5 vecinos más próximos (KN5N) y clasificación de árbol de partición (TREE).

Los aspectos adicionales incluyen el método de cualquier otro aspecto, que además comprende cuantificar la expresión de un marcador genotípico adicional seleccionado entre la figura 6 o la figura 7.

Otros aspectos incluyen el método de cualquier aspecto anterior, donde dicha etapa de cuantificación de la expresión génica se lleva a cabo detectando los niveles de ARNm.

Los aspectos adicionales incluyen el método de cualquier otro aspecto, en donde dicha etapa de cuantificar la expresión génica se lleva a cabo detectando los niveles de ADNc.

Otros aspectos más incluyen el método de cualquier otro aspecto, donde dicha etapa de cuantificar la expresión genética se lleva a cabo usando un oligonucleótido complementario a al menos una porción de dicho ADNc.

Los aspectos adicionales incluyen el método de cualquier otro aspecto, donde dicha etapa de cuantificar la expresión genética se lleva a cabo usando un método de qRT-PCR usando un cebador directo y un cebador inverso.

Otros aspectos adicionales incluyen el método de cualquier otro aspecto, donde dicha etapa de cuantificar la expresión genética se lleva a cabo detectando los niveles de una proteína.

Otros aspectos más incluyen el método de cualquier otro aspecto, donde dicha etapa de cuantificar la expresión genética se lleva a cabo detectando los niveles de un péptido.

Los aspectos adicionales incluyen el método de cualquier otro aspecto, donde dicha etapa de cuantificar la expresión genética se lleva a cabo usando un anticuerpo dirigido contra dicho marcador.

Otros aspectos adicionales más incluyen el método de cualquiera de las reivindicaciones 1 a 8 o 13-15, donde dicha etapa de cuantificar la expresión genética se lleva a cabo usando un método de inmunoensayo de tipo sándwich o usando una microplaca de anticuerpos.

Otros aspectos más incluyen el método de cualquier otro aspecto, donde dicha cuantificación de la expresión genética se lleva a cabo usando un anticuerpo monoclonal.

Otros aspectos incluyen el método de cualquier otro aspecto, donde dicha cuantificación de la expresión genética se lleva a cabo usando un antisuero policlonal.

ÍNDICE G+P

Las variables fenotípicas y genotípicas descritas anteriormente se combinan en un ÍNDICE G+P de acuerdo con la siguiente relación:

ÍNDICE G+P = (w1*HFREQ+w2*EdadMQ50+w3*Género+w4*SMK+w5*RBC) + (w6*M1+w7*IL-8),

donde HFREQ significa la frecuencia de hallazgo de 3 o más glóbulos rojos por campo de alto aumento en un periodo de 6 meses; si la frecuencia es baja, entonces HFREQ = 0 y si es mayor de 3 glóbulos rojos por campo de alto aumento, entonces es 1. EdadMQ50 se refiere a la edad del sujeto, si es mayor que 50 años, entonces EdadMQ50 = 1 y si es menor de 50 años, entonces es 0. Al género se le asigna un valor de 1 para masculino y 0 para femenino. SMK significa si el sujeto es fumador o exfumador; en caso de que sea no fumador, entonces SMK = 0 y si es fumador, entonces es 1. RBC significa el número de glóbulos rojos; si es de 25 o más, entonces RBC se establece en 1 y si es menor de 25, entonces es 0. MI es una combinación de la expresión de los marcadores genéticos MDK, CDC, IGFBP5 y HOXA13; si M1 > 4,5 entonces se establece en 1, si es menor que 4,5, 0. IL-8 se refiere al nivel de expresión de ARN para IL-8; si IL-8 > 2,5 entonces IL-8 se establece a 1, si es menor que 2,5, 0. Los símbolos "*" significan el operador de multiplicación y los factores de ponderación, w1 - w7 son, respectivamente, las ponderaciones asignadas a cada una de las variables listadas anteriormente en el ÍNDICE G+P.

En otras realizaciones preferidas, (EdadMQ50 y RBC) pueden descartarse del modelo como se muestra a continuación:

ÍNDICE G+P = (1*HFREQ+3*Género+4*SMK) + (5*M1+2*IL-8)

El ÍNDICE G+P proporciona un valor de entre 0 y 15. Un paciente con un valor de ÍNDICE G+P de 11 a 15 se

considera en "alto riesgo" de cáncer de vejiga e indica la necesidad de pruebas complementarias adicionales para cáncer de vejiga. Un paciente con un valor de ÍNDICE G+P de 6 a 10 se considera en "riesgo moderado" de desarrollar cáncer de vejiga y está indicado realizar pruebas complementarias adicionales. Un paciente con un ÍNDICE G+P de 0 a 5 se considera en "bajo riesgo" de desarrollar cáncer de vejiga. Los pacientes en el grupo de "bajo riesgo" son asignados a una lista de espera de observación y en caso de que aparezcan síntomas adicionales o de que se produzcan episodios recurrentes de microhematuria, son evaluados nuevamente para valorar posibles pruebas complementarias adicionales.

Como se describe con más detalle en el ejemplo 3 (**FIG. 18** y **19**), los presentes inventores observaron que solamente la curva ROC para las variables fenotípicas cuantificadas producía un nivel modesto de potencia diagnóstica. La curva ROC para los marcadores genotípicos cuantificados, de manera individual, produjo un nivel significativo de potencia diagnóstica. Los presentes inventores observaron una potencia diagnóstica inesperadamente mejor cuando se combinaron variables tanto genotípicas como fenotípicas en un ÍNDICE G+P.

Cuantificación de la expresión genética

Las proteínas o ácidos nucleicos que se secretan o escinden de la célula o que se pierden por mecanismos apoptóticos, ya sea solos o en combinación entre sí, tienen utilidad como marcadores séricos o de fluidos corporales para el diagnóstico de enfermedades, incluyendo enfermedades inflamatorias en la vejiga y/o cáncer de vejiga o como marcadores para controlar la progresión de la enfermedad establecida. La detección de marcadores proteínicos y celulares puede llevarse a cabo usando métodos conocidos en la técnica e incluyen el uso de RT-PCR, qRT-PCR, anticuerpos monoclonales, antisueros policlonales y similares.

Específicamente, la presente invención proporciona métodos de triaje de pacientes que presentan hematuria, (ya sea macrohematuria o microhematuria), que comprenden: (i) proporcionar una muestra biológica; (ii) detectar uno o más marcadores de tumor de vejiga (BTM) en dicha muestra. Los marcadores de tumor de vejiga de interés particular incluyen MDK, CDC2, HOXA13 e IGFBP5 (un "ensayo de 4 marcadores"). Opcionalmente, también pueden detectarse los niveles de marcador de neutrófilos humanos, receptor B de interleucina 8 (IL8Rb) en la muestra (ensayo Cxbladder®). Puede determinarse la presencia del cáncer comparando los niveles de uno o más BTM con los niveles en pacientes normales, pacientes que tienen cáncer de vejiga en estadio temprano y/o pacientes que tienen una enfermedad inflamatoria. Por ejemplo, puede determinarse la presencia de cáncer comparando la expresión de BTM frente a un umbral de expresión. El umbral puede ser en el orden de expresión que es al menos 1,1, 1,2, 1,3, 1,4, 1,5, 1,6, 1,7, 1,8, 1,9, 2, 3, 4, 5, 6, 7, 8, 9 o 10, 100, 1000 o hasta 10.000 veces el nivel de expresión en un grupo de pacientes que no tienen cáncer. En otros aspectos, una alta expresión de IL8Rb sin alteración en la expresión de un marcador de tumor de vejiga puede ser indicativa de una enfermedad inflamatoria en lugar de cáncer.

Los métodos de la presente invención pueden usarse junto con cualquier marcador adecuado para detectar el cáncer de vejiga. Los ejemplos de marcadores adecuados para su uso en la invención se exponen en las FIG. 6 o 7. La presente invención incluye el uso de uno cualquiera o más de los marcadores expuestos en las FIG. 6 o 7.

Opcionalmente, en otras realizaciones preferidas, la presente invención puede incluir cualquier combinación de IL8Rb con uno o más de los marcadores MDK, CDC2, HOXA13 e IGFBP5, que también pueden estar en combinación con uno o más marcadores diferentes adecuados para detectar el cáncer de vejiga, por ejemplo, uno cualquiera o más de los marcadores expuestos en las FIG. 6 o 7. Más específicamente, la presente invención incluye la cuantificación de la expresión de uno cualquiera o más de una combinación de marcadores: IL8Rb / MDK, IL8Rb / CDC2, IL8Rb / HOXA13, IL8Rb / IGFBP5, IL8Rb / MDK / CDC2, IL8Rb / MDK / HOXA13, IL8Rb / MDK / IGFBP5, IL8Rb / CDC2 / HOXA13, IL8Rb / CDC2 / IGFBP5, IL8Rb / HOXA13 / IGFBP5, IL8Rb / MDK / CDC2 / HOXA13, IL8Rb / MDK / CDC2 / IGFBP5, IL8Rb / CDC2 / HOXA13 / IGFBP5 e IL8Rb / MDK / CDC2 / HOXA13 / IGFBP5. Estas combinaciones pueden incluir, opcionalmente, uno o más marcadores adicionales adecuados para detectar el cáncer de vejiga, por ejemplo, uno cualquiera o más de los marcadores expuestos en las FIG. 6 o 7.

La presente invención también proporciona un método para detectar afecciones inflamatorias de la vejiga, que comprende: (i) proporcionar una muestra biológica de un paciente; y (ii) detectar los niveles del marcador de neutrófilos humanos, receptor B de interleucina 8 (IL8Rb) en dicha muestra. La presencia de afecciones inflamatorias de la vejiga se establece comparando los niveles de IL8Rb con los niveles en pacientes normales, pacientes que tienen hematuria y pacientes que tienen una afección inflamatoria de la vejiga. Por ejemplo, la presencia de una afección inflamatoria de la vejiga puede determinarse comparando la expresión del marcador IL8Rb frente a un umbral. El umbral puede ser en el orden de expresión que es al menos 1,1, 1,2, 1,3, 1,4, 1,5, 1,6, 1,7, 1,8, 1,9, 2, 3, 4, 5, 6, 7, 8, 9, 10, 100, 1000 o hasta 10.000 veces el nivel de expresión en otro grupo de pacientes.

Los métodos genotípicos preferidos de la presente invención pueden llevarse a cabo detectando cualquier marcador adecuado de la expresión génica, por ejemplo, determinando los niveles de ARNm, ADNc, una proteína o péptido utilizando cualquier método adecuado.

Puede establecerse un diagnóstico mediante el uso de un sistema clasificador, por ejemplo, análisis de discriminantes lineales (LDA), regresión logística (LogReg), máquina de vectores de soporte (SVM), K 5 vecinos más próximos (KN5N) y clasificación de árbol de partición (TREE).

Breve descripción de las figuras

La presente invención se describe haciendo referencia a realizaciones específicas de la misma y haciendo referencia a las figuras, en las que:

La **FIG. 1** representa las secuencias de proteína y ARNm de IL8Rb (también conocido como CXCR2).

La **FIG. 2** representa una gráfica de gráficas de dispersión que muestran el efecto de IL8Rb a la hora de diferenciar el CCT de enfermedades no malignas (enfermedad prostática, cistitis, infección del tracto urinario y urolitiasis). En las **FIG. 2c** y **2f** se han sustituido diferentes marcadores de ARN de cáncer de vejiga por IL8Rb. (a). MDK/IGFBP5; (b). MDK/HOXA13; (c) MDK/IL8Rb; (d) CDC2/IGFBP5; (e). CDC2/HOXA13; (f). CDC2/IL8Rb.

La **FIG. 3** representa el análisis de curvas ROC (sensibilidad frente a especificidad) que muestra el efecto de incluir IL8Rb en los algoritmos diagnósticos obtenidos usando análisis de discriminante lineal (LD) y regresión lineal (LR). Las curvas ROC se obtuvieron de pacientes con CCT y cánceres del tracto urinario superior (n=61) y las enfermedades no malignas cistitis, infección del tracto urinario y urolitiasis (n=61). (a). LD1 (continua) y LD2 (discontinua). (b) LR1 (continua) y LR2 (discontinua). IL8Rb se incluye en LD2 y LR2.

La **FIG. 4** representa el análisis de curvas ROC extendido que muestra el efecto de incluir IL8Rb en los algoritmos diagnósticos obtenidos usando análisis de discriminante lineal (LD) y regresión lineal (LR). Las curvas ROC se obtienen de pacientes con CCT (n=56) y, a diferencia de la **FIG. 3**, cualquier enfermedad no maligna en la cohorte (n=386). (a). LD1 (continua) y LD2 (discontinua), (b) LR1 (discontinua) y LR2 (continua). IL8Rb se incluye en LD2 y LR2.

La **FIG. 5** representa diagramas de cajas que muestran la acumulación de ARNm de IL8Rb en la orina de pacientes con enfermedad urológica no maligna. El ARN se ha cuantificado mediante qRT-PCR usando el método de delta-Ct (Holyoake *et al*., 2008). Con este método, un menor Ct refleja mayores niveles de ARN. BPH: hiperplasia benigna de próstata; ITU: infección del tracto urinario; próstata NE: enfermedades de próstata no específicas; Próstata Vasc.: próstata vascular; warfarina: hematuria secundaria al uso de warfarina. Las observaciones en pacientes con cistitis/ITU son significativamente diferentes (p=0,001) con respecto a las otras presentaciones no malignas mostradas.

La **FIG. 6** representa marcadores que se sabe que se sobreexpresan en cáncer de vejiga y son adecuados para su uso en la presente invención.

La **FIG. 7** representa marcadores que se sabe que se subexpresan en cáncer de vejiga y son adecuados para su uso en la presente invención.

La **FIG. 8** representa un diagrama de flujo para los procedimientos de reclutamiento de pacientes y los números para el ejemplo 2.

La **FIG. 9** representa las características clínicas y demográficas iniciales de los pacientes según el estado de enfermedad a los 3 meses.

La **FIG. 10** representa la sensibilidad y la especificidad general de las pruebas de orina.

La **FIG. 11** representa diversas curvas ROC; La **FIG. 11a** representa curvas ROC para el ELISA de NMP22 y uRNA-D (prueba que comprende los cuatro marcadores MDK+CDC2+IGFBP5+HOXA13); y la **FIG. 11b** representa la curva ROC para los cinco marcadores MDK, CDC2, HOXA13, IGFBP5 e IL8Rb.

La **FIG. 12** representa la sensibilidad de las pruebas de orina según el estadio, el grado, la ubicación del tumor, la multiplicidad del tumor, el estado de hematuria, la creatinina de la muestra de orina y el sexo. Las tablas muestran los números y el porcentaje con una prueba de orina positiva entre aquellas con CCT.

La **FIG. 13** representa la especificidad de las pruebas de orina según el diagnóstico, la macrohematuria o, la creatinina y el sexo. Las tablas muestran el número y el % con un resultado de prueba de orina negativo entre aquellos sin CCT.

La **FIG. 14** representa curvas ROC para las combinaciones de marcadores: (a) MDK, (b) CDC, (c) IGFBP5, (d) HOXA13, (e) MDK+CDC2, (f) MDK+IGFBP5, (g) MDK+HOXA13, (h) CDC2+IGFBP5, (i) CDC+HOXA13, (j) IGF+HOXA13, (k) MDK+CDC2+IGFBP5, (l) MDK+CDC2+HOXA13, (m) MDK+IGFBP5+HOXA13, (n) CDC2+IGFBP5+HOXA13, (o) MDK+CDC2+IGFBP5+HOXA13, más o menos IL8Rb, usando cinco modelos de clasificación diferentes (i) análisis de discriminantes lineales (LDA), (ii) regresión logística (LogReg), (iii) máquina de vectores de soporte (SVM), (iv) K 5 vecinos más próximos (KN5N) y (v) clasificación de árbol de partición (TREE).

La **FIG. 15** representa los resultados de estudios de sensibilidad y selectividad. La **FIG. 15a** representa el "área bajo la curva" (ABC) para una tasa de falsos positivos de hasta el 20 % (a una especificidad del 80 %) de las curvas ROC de la **FIG. 14** y la **FIG. 15b** muestra la diferencia de la ABC resultante de la inclusión de IL8Rb.

La **FIG. 16** representa gráficas de la sensibilidad de las combinaciones de los cuatro marcadores, MDK, CDC2, IGFBP5 y HOXA13, más o menos IL8Rb, usando cinco modelos de clasificación diferentes (i) análisis de discriminantes lineales (LDA), (ii) regresión logística (LogReg), (iii) máquina de vectores de soporte (SVM), (iv) K 5 vecinos más próximos (KN5N) y (v) clasificación de árbol de partición (TREE), a diferentes especificidades predeterminadas; (a) 80 %, (b) 85 %, (c) 90 %, (d) 95 %, (e) 98 %.

La **FIG. 17** representa los aumentos de sensibilidad resultantes de añadir IL8Rb a diferentes especificidades predeterminadas (a) 80 %, (b) 85 %, (c) 90 %, (d) 95 %, (e) 98 % y los aumentos resultantes en la especificidad

resultantes de añadir IL8Rb a diferentes especificidades predeterminadas (f) 80 %, (g) 85 %, (h) 90 %, (i) 95 %, (j) 98 %.

La **FIG. 18** representa curvas ROC para la evaluación de pacientes que tienen hematuria estudiados usando o bien solo pruebas genéticas, solo evaluación del fenotipo y/o tanto prueba genética como evaluación fenotípica.

La **FIG. 19** representa una gráfica de razones de probabilidad (eje horizontal) para las variables de género, el historial de tabaquismo y la HFREQNEW de la presente invención.

Las FIG. 20A y **20B** representan diagramas de flujo para estándares de comunicación de la precisión estadística.

La FIG. 20A representa un diagrama de flujo para pacientes con macrohematuria a lo largo de las tres cohortes en el presente estudio.

La **FIG. 20B** representa un diagrama de flujo para comunicar la precisión diagnóstica en pacientes con microhematuria incluido en este estudio.

La **FIG. 21** representa curvas ROC que representan los tres modelos de clasificación. ÍNDICE P (línea de puntos), ÍNDICE G (línea discontinua) e ÍNDICE G+P (línea continua).

La **FIG. 22** representa NPV frente a la proporción de pacientes con hematuria con prueba negativa según cada uno de los modelos. ÍNDICE P (línea de puntos), ÍNDICE G (línea discontinua) e ÍNDICE G+P (línea continua).

La **FIG. 23** representa una gráfica de la relación entre los resultados detectados (eje horizontal) frente a los resultados del triaje (eje vertical).

La **FIG. 24** representa una gráfica del ÍNDICE G2 (eje horizontal) frente al ÍNDICE G1+P (eje vertical).

Descripción detallada

Definiciones

Antes de describir en detalle las realizaciones de la invención, resultará útil proporcionar algunas definiciones de términos como se usan en el presente documento.

El término "marcador" se refiere a una molécula que se asocia cuantitativa o cualitativa con la presencia de un fenómeno biológico. Los ejemplos de "marcadores" incluyen un polinucleótido, tal como un gen o fragmento génico, ya sea codificante o no codificante, ADN o fragmento de ADN, ARN o fragmento de ARN, ya sea codificante o no codificante; o un producto génico, incluyendo un polipéptido, tal como un péptido, oligopéptido, proteína o fragmento de proteína; o cualquiera de los metabolitos relacionados, subproductos o cualquier otra molécula identificativa, tal como anticuerpos o fragmentos de anticuerpos, estén relacionados directa o indirectamente con un mecanismo subyacente al fenómeno. Los marcadores de la invención incluyen las secuencias de nucleótidos (por ejemplo, secuencias de GenBank) como se divulgan en el presente documento, en particular, las secuencias de longitud completa, cualquier secuencia codificante, cualquier fragmento, cualquier sonda posible (por ejemplo, creada a lo largo de un límite de exón-exón), incluyendo aquellas con motivos de captura, horquillas o fluoróforos o cualquier complemento de las mismas y cualquier marcador medible de las mismas como se ha definido anteriormente.

Como se usan en el presente documento, "anticuerpos" y términos relacionados se refiere a moléculas de inmunoglobulina y a porciones inmunológicamente activas de moléculas de inmunoglobulina (Ig), es decir, moléculas que contienen un sitio de unión a antígeno que se unen específicamente a (inmunorreaccionan con) un antígeno. Estas incluyen, pero sin limitación, policlonales, monoclonales, quiméricos, monocatenarios, fragmentos Fc, Fab, Fab' y Fab_2 y una biblioteca de expresión de Fab. Las moléculas de anticuerpo se refieren a cualquiera de las clases IgG, IgM, IgA, IgE e IgD, que difieren entre sí por la naturaleza de la cadena pesada presente en la molécula. Estas también incluyen subclases, tales como IgG1, IgG2 y otras. La cadena ligera puede ser una cadena kappa o una cadena lambda. La referencia en el presente documento a anticuerpos incluye una referencia a todas las clases, subclases y tipos. También se incluyen anticuerpos quiméricos, por ejemplo, anticuerpos monoclonales o fragmentos de los mismos que son específicos para más de una fuente, por ejemplo, una secuencia de ratón o humana. Además se incluyen anticuerpos de camélido, anticuerpos de tiburón o nanocuerpos.

Los términos "cáncer" y "canceroso" se refieren a o describen la afección fisiológica en mamíferos que se caracteriza típicamente por un crecimiento celular anormal o no regulado. El cáncer y la patología del cáncer pueden estar asociados, por ejemplo, con metástasis, interferencia con el funcionamiento normal de las células vecinas, liberación de citocinas u otros productos de secreción a niveles anormales, supresión o agravamiento de la respuesta inflamatoria o inmunológica, neoplasia, neoplasia premaligna, neoplasia maligna, invasión de tejidos u órganos circundantes o distantes, tales como ganglios linfáticos, etc.

El término "tumor" se refiere a todos los crecimientos celulares neoplásicos y a la proliferación, ya sea maligna o benigna y a todas las células y tejidos precancerosas y cancerosas.

La expresión "cáncer de vejiga" se refiere a un tumor que se origina en la vejiga. Estos tumores tienen la capacidad de metastatizar a cualquier órgano.

El término "BTM" o la expresión "marcador de tumor de vejiga" o "miembro de la familia de BTM" significa un marcador tumoral (MT) que se asocia con cánceres uroteliales, cáncer de vejiga, carcinoma de células transicionales de vejiga (CCT), carcinomas epidermoides y adenocarcinomas de vejiga. El término BTM también incluye combinaciones de marcadores individuales, cuya combinación mejora la sensibilidad y especificidad de detección del cáncer de vejiga. Ha de entenderse que el término BTM no requiere que el marcador sea específico solo para tumores de vejiga. Más bien, la expresión de BTM puede estar alterada en otros tipos de

células, células enfermas, tumores, incluyendo tumores malignos.

La expresión "subexpresión de BTM" significa un marcador que muestra menor expresión en tumores de vejiga que en tejido de vejiga no maligno.

La expresión "sobreexpresión de BTM" significa un marcador que muestra mayor expresión en tumores de vejiga que en tejido no maligno.

Las expresiones "expresado de manera diferencial", "expresión diferencial", y expresiones similares, se refieren a un gen marcador cuya expresión se activa a un mayor o menor nivel en un sujeto (por ejemplo, muestra de ensayo) que tiene una afección, específicamente cáncer, tal como melanoma, en relación con su expresión en un sujeto de control (por ejemplo, muestra de referencia). Los términos también incluyen marcadores cuya expresión se activa a un nivel mayor o menor en diferentes estadios de la misma afección; en enfermedades con un buen o mal pronóstico; o en células con mayores o menores niveles de proliferación. Un marcador expresado de manera diferencial puede estar activado o inhibido a nivel de polinucleótido o a nivel de polipéptido o puede someterse a corte y empalme alternativo para dar como resultado un producto de polipéptido diferente. Dichas diferencias pueden evidenciarse por un cambio en los niveles de ARNm, expresión superficial, secreción u otra partición de un polipéptido, por ejemplo.

La expresión diferencial puede incluir una comparación de la expresión entre dos o más marcadores (por ejemplo, genes o sus productos génicos); o una comparación de las relaciones de la expresión entre dos o más marcadores (por ejemplo, genes o sus productos génicos); o una comparación de dos productos procesados de manera diferencial (por ejemplo, transcritos o polipéptidos) del mismo marcador, que difieren entre sujetos normales y sujetos enfermos; o entre diversos estadios de la misma enfermedad; o entre enfermedades que tienen un buen o mal pronóstico; o entre células con mayores y menores niveles de proliferación; o entre tejido normal y tejido enfermo, específicamente cáncer o melanoma. La expresión diferencial incluye diferencias tanto cuantitativas como cualitativas en el patrón de expresión temporal o celular en un gen o sus productos de expresión entre, por ejemplo, células normales y enfermas o entre células que han sufrido diferentes procesos patológicos o estadios de la enfermedad o células con diferentes niveles de proliferación.

El término "expresión" incluye la producción de polinucleótidos y polipéptidos, en particular, la producción de ARN (por ejemplo, ARNm) a partir de un gen una porción de un gen e incluye la producción de un polipéptido codificado por un ARN o gen o porción de un gen y la aparición de un material detectable asociado con la expresión. Por ejemplo, la formación de un complejo, por ejemplo, a partir de una interacción entre polipéptidos, una interacción entre un polipéptido y un nucleótido o similares, se incluye en alcance del término "expresión". Otro ejemplo es la unión de un ligando de unión, tal como una sonda de hibridación o anticuerpo, a un gen u otro polinucleótido u oligonucleótido, un polipéptido o un fragmento de proteína y la visualización del ligando de unión. Por tanto, la intensidad de un punto en una micromatriz, en una transferencia de hibridación, tal como una transferencia de Northern o en una inmunotransferencia, tal como una transferencia de Western o en una matriz de perlas o mediante análisis por la PCR, se incluye dentro del término "expresión" de la molécula biológica subyacente.

Las expresiones "umbral de expresión génica" y "umbral de expresión definido" se usan indistintamente y se refieren al nivel de un marcador en cuestión, fuera del cual el nivel de expresión del polinucleótido o el polipéptido sirve como marcador predictivo para una afección en el paciente. Por ejemplo, la expresión de IL8Rb por encima de un determinado umbral sirve para diagnosticar que el paciente tiene una afección inflamatoria. También puede usarse un umbral cuando se evalúa a un paciente por sospecha de cáncer de vejiga, usando marcadores de cáncer de vejiga. Los niveles de expresión por encima de un umbral indican que el paciente tiene una afección inflamatoria de vejiga, que tiene probabilidad de causar un resultado falso positivo en una prueba para el cáncer, mientras que un nivel de expresión de IL8Rb por debajo de un umbral predice que el paciente no tiene una afección inflamatoria de vejiga. Al incluir la medición de IL8Rb, cualquier resultado de la expresión de los marcadores tumorales de vejiga puede basarse en si los niveles de IL8Rb se encuentran por debajo del umbral (es decir, es probable que un resultado positivo sea positivo para que el paciente tenga cáncer en vez de que los niveles aumentados de los marcadores tumorales de vejiga sean realmente el resultado de la exfoliación de células no malignas de la mucosa a causa de inflamación).

La expresión "umbral diagnóstico" se refiere a un umbral en el que puede decirse que un paciente ha sido diagnosticado con o sin una afección dada, por ejemplo, cáncer de vejiga. En general, se establece un umbral diagnóstico para lograr una sensibilidad y especificidad deseadas, dependiendo de factores tales como la población, la prevalencia y el posible resultado clínico. En general, el umbral diagnóstico puede calcularse y/o determinarse usando algoritmos y/o análisis de datos computarizado.

El umbral exacto dependerá de la población y también de cualquier modelo que se use para predecir la enfermedad (modelo predictivo). Un umbral se establece experimentalmente a partir de estudios clínicos, tales como los descritos en los ejemplos más adelante. Dependiendo del modelo de predicción empleado, puede establecerse el umbral de expresión para lograr máxima sensibilidad o máxima especificidad o mínimo error (tasa de clasificación máxima). Por ejemplo, puede establecerse un mayor umbral para lograr errores mínimos, pero esto puede dar como resultado una menor sensibilidad. Por lo tanto, para cualquier modelo predictivo determinado, se usarán estudios clínicos para establecer un umbral de expresión que normalmente logra la máxima sensibilidad a la vez que tiene una tasa de error mínima. En general, es probable que el error se encuentre en el orden de expresión que es

El término "sensibilidad" significa la proporción de individuos con la enfermedad que (según el modelo) tienen un resultado positivo en la prueba. Por tanto, la sensibilidad aumentada significa menos resultados de la prueba falsos negativos.

El término "especificidad" significa la proporción de individuos sin la enfermedad que (según el modelo) tienen un resultado negativo en la prueba. Por tanto, especificidad aumentada significa menos resultados de la prueba falsos positivos.

La expresión "característica operativa del receptor" ("curva ROC") significa una gráfica de la tasa de auténticos positivos (sensibilidad) frente a la tasa de falsos positivos (especificidad) para diferentes puntos de corte para un marcador o prueba particular. Cada punto en la curva ROC representa un punto de sensibilidad/especificidad específico que corresponderá a un umbral determinado. Las curvas ROC pueden ser importantes para establecer un umbral para proporcionar un resultado deseado. El área bajo una curva ROC (expresado como un análisis de área bajo la curva (ABC)) puede ser una medida de en qué medida un marcador determinado o una prueba que consiste en diversos marcadores puede distinguir entre dos o más resultados diagnósticos. También pueden usarse curvas ROC para comparar la precisión de dos pruebas diferentes.

El término "oligonucleótido" se refiere a un polinucleótido, normalmente una sonda o cebador, lo que incluye, sin limitación, desoxirribonucleótidos monocatenarios, ribonucleótidos mono o bicatenarios, híbridos de ARN:ADN y ADN bicatenarios. Los oligonucleótidos, tales como oligonucleótidos de sonda de ADN monocatenarios, normalmente se sintetizan mediante métodos químicos, por ejemplo, usando sintetizadores de oligonucleótidos automatizados que se encuentran disponibles comercialmente o mediante diversos métodos diferentes, entre los que se incluyen sistemas de expresión *in vitro*, técnicas recombinantes y expresión en células y organismos.

El término "sobreexpresión" o "sobreexpresado" se refiere a un nivel de expresión de un gen o marcador en un paciente que se encuentra por encima del observado en tejido normal. Puede considerarse que la expresión está sobreexpresada si es 1,1, 1,2, 1,3, 1,4, 1,5, 1,6, 1,7, 1,8, 1,9, 2, 10, 100, 1000 o hasta 10.000 veces la expresión en tejido normal o en tejidos de otro grupo de pacientes.

El término "polinucleótido", cuando se usa en singular o plural, se refiere generalmente a cualquier polirribonucleótido o polidesoxirribonucleótido, que puede ser ARN o ADN no modificado o ARN o ADN modificado. Esto incluye, sin limitación, ADN mono y bicatenario, ADN que incluye regiones mono y bicatenarias, ARN mono y bicatenario y ARN que incluye regiones mono y bicatenarias, moléculas híbridas que comprenden ADN y ARN que pueden ser monocatenarias o, más habitualmente, bicatenarias o incluir regiones mono y bicatenarias. También se incluyen regiones tricatenarias que comprenden ARN o ADN o tanto ARN como ADN. Se incluyen específicamente ARNm, ADNc y ADN genómicos y cualquier fragmento de los mismos. El término incluye ADN y ARN que contienen una o más bases modificadas, tales como bases tritiadas o bases infrecuentes, tales como inosina. Los polinucleótidos de la invención pueden abarcar secuencias codificantes y no codificantes o secuencias sentido o antisentido. Se entenderá que cada referencia a un "polinucleótido" o un término similar, en el presente documento, incluirá las secuencias de longitud completa, así como cualquier fragmento, derivado o fragmento de las mismas.

El término "fenotípico" significa un rasgo que es observable en una situación clínica o en una visita clínica o en el historial de un paciente. Cuando se usa en una fórmula para calcular el índice G+P, "fenotípico" o "P" significa la edad del paciente, el sexo, la incidencia de hematuria y el historial de tabaquismo.

"Polipéptido", como se usa en el presente documento, se refiere a un oligopéptido, péptido o secuencia de proteína o un fragmento del mismo y a moléculas de origen natural, recombinantes, sintéticas o semisintéticas. Cuando en el presente documento se cita un "polipéptido" para hacer referencia a una secuencia de aminoácidos de una molécula de proteína de origen natural, "polipéptido" y términos similares, no pretenden limitar la secuencia de aminoácidos a la secuencia de aminoácidos completa natural para la molécula de longitud completa. Se entenderá que cada referencia a un "polipéptido" o un término similar, en el presente documento, incluirá la secuencia de longitud completa, así como cualquier fragmento, derivado o fragmento de las mismas.

El término "qPCR" o "QPCR" se refiere a la reacción en cadena de la polimerasa cuantitativa descrita, por ejemplo, en *PCR Technique: Quantitative PCR*, J.W. Larrick, ed., Eaton Publishing, 1997 y *A-Z of Quantitative PCR*, S. Bustin, ed., IUL Press, 2004.

La expresión "retrotranscripción" significa un proceso en el que un oligorribonucleótido, incluyendo un ARN mensajero ("ARNm") se usa como molde para la síntesis bioquímica de un oligodesoxirribonucleótido complementario ("ADNc") usando una enzima ("retrotranscriptasa"), que se une al ARN molde y cataliza una serie de reacciones de adición que acoplan secuencialmente bases de desoxirribonucleótido para formar una hebra de oligodesoxirribonucleótido que es complementaria al molde de ARN.

El término "hematuria" se define como la presencia de sangre en la orina. Puede estar presente como hematuria macroscópica (trazas visibles de células sanguíneas) o hematuria microscópica (trazas microscópicas de sangre) en la orina. Una indicación confirmada de microhematuria se define como la presencia de 3 o más glóbulos rojos por campo de microscopía de alto aumento (HPF) en un mínimo de 3 muestras de orina recogidas de manera adecuada. La microhematuria también puede detectarse mediante tira reactiva de orina (estimación de comparación colorimétrica) en la consulta médica. La hematuria (ya sea microscópica o macroscópica) puede ser asintomática (sin síntomas adicionales asociados con hematuria) o sintomática. Los síntomas adicionales incluyen disuria (dolor al orinar), sensación de vaciado incompleto de la vejiga o aumento de la frecuencia de micción.

La "rigurosidad" de las reacciones de hibridación puede determinarse fácilmente por un experto habitual en la materia y en general, es un cálculo empírico que depende de la longitud de la sonda, la temperatura de lavado y la concentración de sal. En general, las sondas de mayor longitud requieren mayores temperaturas para una hibridación adecuada, mientras que las sondas más cortan necesitan temperaturas más bajas. La hibridación depende generalmente de la capacidad para que el ADN desnaturalizado para volver a hibridarse cuando hay presencia de hebras complementarias en un ambiente por debajo de su temperatura de fusión. Cuanto mayor

sea el grado de homología deseado entre la sonda y la secuencia hibridada, mayor será la temperatura relativa que pueda usarse. Como resultado, se deduce que las temperaturas relativas más elevadas tenderán a hacer que las condiciones de reacción sean más rigurosas, mientras que las temperaturas más bajas harán lo contrario. Pueden encontrarse detalles adicionales y una explicación de la rigurosidad de las reacciones de hibridación, por ejemplo, en Ausubel *et al.*, *Current Protocols in Molecular Biology*, Wiley Interscience Publishers, (1995).

Las "condiciones rigurosas" o "condiciones de alta rigurosidad", como se definen en el presente documento, normalmente: (1) emplean una baja fuera iónica y alta temperatura para el lavado. Por ejemplo, cloruro de sodio 0,015 M/citrato de sodio 0,0015 M/dodecilsulfato de sodio al 0,1 % a 50 °C; (2) emplean un agente de desnaturalización durante la hibridación, tal como formamida, por ejemplo, formamida al 50 % (v/v) con seroalbúmina bovina al 0,1 %/Ficol 1 al 0,1 %/polivinilpirrolidona al 0,1 %/fosfato de sodio 50 mM, tampón a pH 6,5 con cloruro de sodio 750 mM, citrato de sodio 75 mM a 42 °C; o (3) emplean formamida al 50 %, SSC 5X (NaCl 0,75 M, citrato de sodio 0,075 M), fosfato de sodio 50 mM (pH 6,8), pirofosfato de sodio al 0,1 %, 5X, solución de Denhardt, ADN de esperma de salmón tratado con ultrasonidos (50 ug/ml), SDS al 0,1 % y sulfato de dextrano al 10 % a 42 °C, con lavados a 42 °C en SSC 0,2X (cloruro de sodio/citrato de sodio) y formamida al 50 % a 55 °C, seguido de un lavado de alta rigurosidad que comprende SSC 0,1X que contiene EDTA a 55 °C.

Las "condiciones de rigurosidad moderada" pueden identificarse como se describe por Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Nueva York: Cold Spring Harbor Press, 1989 e incluyen el uso de solución de lavado y condiciones de hibridación (por ejemplo, temperatura, fuerza iónica y % de SDS) menos rigurosas que las descritas anteriormente. Un ejemplo de condiciones de rigurosidad moderada es incubación durante una noche a 37 °C en una solución que comprende: formamida al 20 %, SSC 5X (NaCl 150 mM, citrato trisódico 15 mM), fosfato de sodio 50 mM (pH 7,6), solución de Denhardt 5X, sulfato de dextrano al 10 % y 20 mg/ml de ADN de esperma de salmón cortado y desnaturalizado, seguido de lavado de los filtros en SSC 1X a aproximadamente 37-50 °C. El experto en la materia conocerá cómo ajustar la temperatura, fuerza iónica, etc. según sea necesario para adaptarse a factores tales como la longitud de la sonda y similares.

La expresión "IL8Rb" significa el receptor B de interleucina 8 marcador de neutrófilos (también conocido como receptor 2 de quimiocinas (motivo C-X-C) [CXCR2]) (FIG. 1; SEQ ID NO: 1 y 2) e incluye el marcador IL8Rb. El término incluye un polinucleótido, tal como un gen o fragmento génico, ARN o fragmento de ARN; o un producto génico, incluyendo un polipéptido, tal como un péptido, oligopéptido, proteína o fragmento de proteína; o cualquiera de los metabolitos relacionados, subproductos o cualquier otra molécula identificativa, tales como anticuerpos o fragmentos de anticuerpo.

El término "fiabilidad" incluye la baja incidencia de falsos positivos y/o falsos negativos. Por tanto, con una mayor fiabilidad de un marcador, se asocian menos falsos positivos y/o falsos negativos con los diagnósticos efectuados usando dicho marcador. "Precisión" es la proporción de resultados verdaderos (auténticos positivos más auténticos negativos) dividida entre el número total de casos en la población, de acuerdo con la fórmula:

$$\frac{\text{Auténticos positivos + Auténticos negativos}}{\text{Número total de mediciones}}$$

El término "triaje" significa distinguir pacientes con hematuria que tienen bajas probabilidades de tener cáncer de vejiga de aquellos pacientes con hematuria que tienen una probabilidad razonable de tener cáncer de vejiga y que requieran pruebas clínicas complementarias adicionales, incluyendo cistoscopía u otro procedimiento clínico.

Realizaciones

Por lo tanto, en determinados aspectos, se proporciona una combinación de marcadores genéticos y marcadores fenotípicos que permite diferenciar pacientes que tienen bajas probabilidades de tener cáncer de vejiga de aquellos pacientes con un riesgo suficiente de tener cáncer de vejiga para justificar pruebas clínicas complementarias adicionales, entre las que posiblemente se incluyen cistoscopia y otros procedimientos. En otros aspectos, se proporcionan marcadores que tienen una fiabilidad mayor de aproximadamente un 70 %; en otras realizaciones, mayor de aproximadamente un 73 %, en otras realizaciones más, mayor de aproximadamente un 80 %, en más realizaciones adicionales, mayor de aproximadamente un 90 %, en otras más, mayor de aproximadamente un 95 %, en más realizaciones adicionales, mayor de aproximadamente un 98 % y en ciertas realizaciones, aproximadamente un 100 % de fiabilidad.

Para el análisis genético, la práctica de la presente invención empleará, a menos que se indique otra cosa, técnicas convencionales de biología molecular (entre las que se incluyen técnicas recombinantes), microbiología, biología celular y bioquímica, que se encuentran entre las capacidades de la técnica. Dichas técnicas se explican completamente en las referencias, tales como, *Molecular Cloning: A Laboratory Manual*, 2.ª edición, Sambrook *et al.*, 1989; *Oligonucleotide Synthesis*, MJ Gait, ed., 1984; *Animal Cell Culture*, R.I. Freshney, ed., 1987; *Methods in Enzymology*, Academic Press, Inc.; *Handbook of Experimental Immunology*, 4.ª edición, D.M. Weir y CC. Blackwell, eds., Blackwell Science Inc., 1987; *Gene Transfer Vectors for Mammalian Cells*, J.M. Miller y M.P. Calos, eds., 1987; *Current Protocols in Molecular Biology*, F. M. Ausubel *et al.*, eds., 1987; y *PCR: The Polymerase Chain Reaction*, Mullis *et al.*, eds., 1994.

Ha de entenderse que los términos anteriores pueden hacer referencia a proteínas, secuencias de ADN y/o secuencias de ARN. También ha de entenderse que los términos anteriores también se refieren a proteínas, ADN y/o ARN no humano que tiene secuencias homólogas como se representan en el presente documento.

Realizaciones de la presente invención

Normalmente, los pacientes con hematuria derivados al urólogo son citados para observación en un consultorio específico para hematuria. Normalmente se da prioridad a los pacientes con hematuria macroscópica frente a aquellos con microhematuria. La información proporcionada por el médico de atención primaria en el momento de la derivación puede ser muy variable, lo que hace que sea difícil estratificar de manera precisa a los pacientes según la probabilidad de que tengan un cáncer urotelial. Normalmente se solicita de manera rutinaria una citología de orina antes de auscultar al paciente y, en caso de que sea positiva, se usa para aumentar la prioridad del paciente para efectuar pruebas clínicas complementarias completas, aunque la citología de orina, aun siendo altamente específica, tiene muy baja sensibilidad y por tanto, tiene poco valor práctico con su alta frecuencia de falsos negativos[(6)].

Análisis fenotípico y genotípico de pacientes con hematuria que no tienen cáncer de vejiga

ÍNDICE G+P

Las variables fenotípicas y genotípicas descritas anteriormente se combinan en un ÍNDICE G+P de acuerdo con la siguiente relación:

ÍNDICE G+P = (w1*HFREQ+w2*EdadMQ50+w3*Género+w4*SMK+w5*RBC) + (w6*M1+w7*IL-8),

donde HFREQ significa la frecuencia de hallazgo de 3 o más glóbulos rojos por campo de alto aumento en un periodo de 6 meses; si la frecuencia es baja, entonces HFREQ = 0 y si es mayor de 3 glóbulos rojos por campo de alto aumento, entonces es 1. EdadMQ50 se refiere a la edad del sujeto, si es mayor que 50 años, entonces EdadMQ50 = 1 y si es menor de 50 años, entonces es 0. Al género se le asigna un valor de 1 para masculino y 0 para femenino. SMK significa si el sujeto es fumador o exfumador; en caso de que sea no fumador, entonces SMK = 0 y si es fumador, entonces es 1. RBC significa el número de glóbulos rojos; si es de 25 o más, entonces RBC se establece en 1 y si es menor de 25, entonces es 0. MI es una combinación de la expresión de los marcadores genéticos MDK, CDC, IGFBP5 y HOXA13; si M1 > 4,5 entonces se establece en 1, si es menor que 4,5, 0. IL-8 se refiere al nivel de expresión de ARN para IL-8; si IL-8 > 2,5 entonces IL-8 se establece a 1, si es menor que 2,5, 0. Los símbolos "*" significan el operador de multiplicación y los factores de ponderación, w1 - w7 son, respectivamente, las ponderaciones asignadas a cada una de las variables listadas anteriormente en el ÍNDICE G+P.

En otras realizaciones preferidas, (EdadMQ50 y RBC) pueden descartarse del modelo como se muestra a continuación:

ÍNDICE G+P = (1*HFREQ+3*Género+4*SMK) + (5*M1+2*IL-8)

El ÍNDICE G+P proporciona un valor de entre 0 y 15. Un paciente con un valor de ÍNDICE G+P de 11 a 15 se considera en "alto riesgo" de cáncer de vejiga e indica la necesidad de pruebas complementarias adicionales para cáncer de vejiga. Un paciente con un valor de ÍNDICE G+P de 6 a 10 se considera en "riesgo moderado" de desarrollar cáncer de vejiga y está indicado realizar pruebas complementarias adicionales. Un paciente con un ÍNDICE G+P de 0 a 5 se considera en "bajo riesgo" de desarrollar cáncer de vejiga. Los pacientes en el grupo de "bajo riesgo" son asignados a una lista de espera de observación y en caso de que aparezcan síntomas adicionales o de que se produzcan episodios recurrentes de microhematuria, son evaluados nuevamente para valorar posibles pruebas complementarias adicionales.

Las variables genotípicas útiles para diferenciar pacientes con y sin cáncer de vejiga incluyen la expresión de los marcadores de ARN "M1" que es una combinación de MDK, CDC, IGFBP5 (IGBP5) y HOXA13. Otra variable genotípica es la expresión de ARN para IL8R. Se muestran coeficientes para estas variables genotípicas en la **tabla 7** a continuación. Se usó un umbral de 4,5 y 2,5 para M1 e IL8R, respectivamente y se asignó un coeficiente de 5 y 2, respectivamente.

El ÍNDICE G+P proporciona un valor de entre 0 y 15. Un valor del ÍNDICE G+P de 11 a 15 se considera "alto riesgo" de cáncer de vejiga e indica la necesidad de pruebas complementarias adicionales para cáncer de vejiga. Un valor del ÍNDICE G+P de 6 a 10 se considera "riesgo moderado" de desarrollar cáncer de vejiga y está indicado realizar pruebas complementarias adicionales. Un ÍNDICE G+P de 0 a 5 se considera "bajo riesgo" de desarrollar cáncer de vejiga y estos pacientes son asignados a una lista de espera y en caso de que aparezcan síntomas adicionales o de que se produzcan episodios recurrentes de microhematuria, son evaluados nuevamente para valorar pruebas complementarias más completas.

Como se describe con más detalle en el ejemplo 3 (**FIG. 18** y **19**), los presentes inventores observaron que

solamente la curva ROC para los datos fenotípicos cuantificados producía un nivel modesto de potencia diagnóstica. La curva ROC para los datos genotípicos, de manera individual, produjo un nivel significativo de potencia diagnóstica. Los presentes inventores observaron una potencia diagnóstica inesperadamente mejor cuando se combinaron datos tanto genotípicos como genotípicos en un ÍNDICE G+P.

Análisis genético de pacientes que no tienen cáncer de vejiga

En algunas realizaciones preferidas, la presente invención combina el uso de un ensayo de 4 marcadores o un ensayo Cxbladder® (variables genotípicas) y uno o más factores de riesgo clave (variables fenotípicas) para producir un índice de selección que puede usarse para el triaje de pacientes con hematuria microscópica o macroscópica en cuanto a su riesgo potencial de tener cáncer urotelial. Aunque no excluye la necesidad de una cistoscopia flexible, los pacientes considerados en riesgo de cáncer urotelial según las variables fenotípicas y genotípicas pueden ser atendidos antes, mejorando potencialmente el resultado general para el paciente.

Los marcadores genotípicos pueden usarse como herramientas para detectar a pacientes sin cáncer o para seleccionar grupos de pacientes que se encuentran en un riesgo bajo, medio o elevado de tener una enfermedad. Por ejemplo, los marcadores pueden expresarse diferencialmente entre tejido enfermo y tejido no enfermo correspondiente. En esta situación, la detección de expresión diferencial se asocia con la presencia de la enfermedad. Como alternativa, el marcador puede asociarse directamente con cambios que se producen en los tejidos enfermos o cambios causados por la enfermedad. Las enfermedades inflamatorias se asocian con un aumento de los neutrófilos. Se ha observado que el marcador de neutrófilos, receptor B de interleucina 8 (IL8Rb; **fig. 1**; SEQ ID NO: 1 y 2) puede proporcionar un buen marcador para la presencia de neutrófilos en una muestra y por tanto, puede usarse como marcador diagnóstico para la detección de enfermedad inflamatoria en una muestra y en particular, en la detección de enfermedad inflamatoria de la vejiga.

Como se muestra en la **FIG. 5**, la acumulación de IL8Rb en la orina indica la presencia de enfermedad inflamatoria en la vejiga. Específicamente, la **FIG. 5** muestra la acumulación de IL8Rb en la orina de pacientes que tienen las afecciones; hiperplasia benigna de próstata, infección del tracto urinario, enfermedades de próstata no específicas, próstata vascular y secundarias al uso de warfarina. Sin embargo, se apreciará, que el uso de IL8Rb no se limita únicamente a la detección de estas enfermedades, sino que estos ejemplos muestran que IL8Rb no aumenta en muestras de pacientes que tienen una enfermedad inflamatoria de la vejiga. Es decir, IL8Rb puede usarse como marcador de inflamación asociada con enfermedad de vejiga y por tanto, es adecuado para su uso en la detección de cualquier afección asociada con la inflamación. Por lo tanto, la detección de la cantidad de IL8Rb puede usarse como marcador para una enfermedad inflamatoria de la vejiga. Más particularmente, IL8Rb puede usarse para detectar enfermedades inflamatorias de la vejiga asociadas con la acumulación de neutrófilos.

Las pruebas de orina para CCT se basan en gran medida en la presencia de marcadores en la orina procedentes de células tumorales exfoliadas. La capacidad para detectar estas células puede enmascararse por la presencia de grandes números de células contaminantes, tales como, sangre y células inflamatorias. Además, la inflamación del revestimiento de la vejiga puede dar como resultado el aumento de la exfoliación de células no malignas de la mucosa. Como resultado, las pruebas de orina que usan marcadores procedentes de células transicionales de vejiga tienen más probabilidades de proporcionar un resultado falso positivo de las muestras de orina tomadas de pacientes con cistitis, infección del tracto urinario u otras afecciones que dan como resultado inflamación del tracto urinario o exfoliación de células transicionales, tales como, urolitiasis (Sanchez-Carbayo *et al*.).

Una forma para tratar de evitar dichos resultados falsos positivos ha sido seleccionar marcadores con baja expresión relativa en la sangre o células inflamatorias. El uso de dichos marcadores da como resultado menos falsos positivos en pacientes con CCT que presentan afecciones inflamatorias no malignas. Sin embargo, la baja expresión de los marcadores en las células de origen hematológico no logra compensar la frecuencia aumentada de exfoliación de células transicionales no malignas.

Se ha descubierto que el impacto negativo que tienen las células transicionales exfoliadas de tejido inflamado en la precisión de las pruebas de orina para cáncer de vejiga puede minimizarse mejorando la identificación de pacientes con afecciones inflamatorias del tracto urinario. En el presente estudio, se ha descubierto sorprendentemente que el uso del marcador IL8Rb en combinación con uno o más marcadores tumorales de vejiga (BTM) posibilita una detección más precisa del cáncer de vejiga. En particular, una prueba de cáncer de vejiga basada en marcadores que incluya el marcador IL8Rb es menos susceptible a los resultados falsos positivos, lo que pueda dar como resultado pacientes que padecen una afección inflamatoria no cancerosa.

En general, la presencia o ausencia de una afección inflamatoria se determina por tener un umbral de expresión génica, por encima del cual la expresión de IL8Rb es indicativa de una afección inflamatoria. Por ejemplo, la expresión de IL8Rb por encima de un determinado umbral sirve para diagnosticar que el paciente tiene una afección inflamatoria (véanse los umbrales descritos anteriormente).

Cuando se usa IL8Rb junto con uno o más marcadores predictivos para la presencia de cáncer de vejiga, la presencia de expresión elevada de los uno o más marcadores de tumor de vejiga y la expresión de IL8Rb, por

encima de un cierto umbral, es predictivo de que el paciente tenga una afección inflamatoria y no cáncer. Además, si la prueba se efectúa en orina del paciente, este resultado predice que el paciente tiene una afección inflamatoria de vejiga. Los altos niveles de los marcadores tumorales de vejiga son con gran probabilidad el resultado de células no malignas que proceden de la mucosa como resultado de la inflamación. Es decir, el paciente, aunque tenga altos niveles de los uno o más marcadores tumorales de vejiga, en realidad no tiene cáncer de vejiga, un falso positivo.

Como alternativa, en caso de que el paciente tenga niveles anormalmente altos o niveles diagnósticos de uno o marcadores de tumor de vejiga, pero el nivel de IL8Rb se encuentra por debajo de un umbral, es probable que el paciente tenga cáncer y en particular, cáncer de vejiga o urotelial. Este es especialmente el caso, si la prueba se efectúa en orina del paciente. Este resultado es especialmente beneficioso para el prestador de servicios sanitarios ya que pueden estar seguros de que el paciente tiene cáncer y pueden comenzar inmediatamente el tratamiento y despreocuparse de que el resultado esté de hecho causado por una afección inflamatoria que arroje un resultado falso positivo.

Sorprendentemente, se ha demostrado que la cuantificación del ARN a partir del gen que codifica el marcador de neutrófilos, receptor B de interleucina 8 (IL8Rb) mejora el rendimiento general de detección de pacientes con CCT, usando marcadores de CCT y BTM conocidos. Las secuencias de referencia para IL8Rb se muestran en la FIG. 1 y en las SEQ ID NO: 1 y 2). Además de su papel en la detección del CCT, se ha evaluado si IL8Rb se puede usar como marcador de orina para ayudar en el diagnóstico de enfermedades inflamatorias (FIG. 5).

El marcador IL8Rb puede usarse de manera aislada para la detección de afecciones inflamatorias de la vejiga utilizando métodos conocidos para detectar los niveles de expresión génica. A continuación se exponen ejemplos de métodos para detectar la expresión génica.

Como alternativa, IL8Rb puede combinarse con uno o más BTM para detectar el cáncer de vejiga. Se ha demostrado que mediante el uso del marcador de enfermedad inflamatoria, IL8Rb, como parte de la prueba para cáncer de vejiga, se minimiza la influencia del tejido inflamado a la hora de crear un resultado falso positivo. El marcador IL8Rb puede usarse asociado con cualesquiera marcadores de cáncer de vejiga o como alternativa, puede usarse con dos o más marcadores, como parte de una firma, para detectar el cáncer de vejiga.

La reducción del número de resultados falsos positivos significa que se someterá a menos pacientes que no tienen cáncer de vejiga a procedimientos potencialmente innecesarios, entre los que se incluye la cistoscopia, que comporta sus propios riesgos. La reducción del número de resultados falsos negativos significa que es más probable que se detecte un paciente con cáncer de vejiga y por tanto, puedan efectuarse análisis complementarios para el cáncer.

La acción de IL8Rb a la hora de mejorar la detección del cáncer de vejiga es el resultado de la capacidad para diferenciar afecciones no malignas de pacientes que tienen cáncer de vejiga. Esto se logra debido a que un aumento de IL8Rb indica un aumento de la presencia de neutrófilos en una muestra. Por lo tanto, la capacidad de IL8Rb no depende del marcador de tumor de vejiga empleado. Como se muestra en las FIG. 2 y 12 a 15, cuando se combina con diversos marcadores de tumor de vejiga y combinaciones de marcadores de tumor de vejiga, IL8Rb tuvo el efecto general de aumentar la especificidad de la capacidad de los uno o más marcadores para detectar el cáncer en los sujetos.

Un ejemplo de una firma de acuerdo con la presente invención es el uso de IL8Rb en combinación con MDK, CDC2, IGFBP5 y HOXA13, que también pueden estar en combinación con uno o más marcadores diferentes adecuados para detectar el cáncer de vejiga, por ejemplo, uno cualquiera o más de los marcadores expuestos en las FIG. 6 o 7. Como se muestra en las FIG. 14 y 15, IL8Rb puede usarse en cualquier combinación de los marcadores, específicamente, las combinaciones IL8Rb / MDK, IL8Rb / CDC2, IL8Rb / HOXA13, IL8Rb / IGFBP5, IL8Rb / MDK / CDC2, IL8Rb / MDK / HOXA13, IL8Rb / MDK / IGFBP5, IL8Rb / CDC2 / HOXA13, IL8Rb / CDC2 / IGFBP5, IL8Rb / HOXA13 / IGFBP5, IL8Rb / MDK / CDC2 / HOXA13, IL8Rb / MDK / CDC2 / IGFBP5, IL8Rb / CDC2 / HOXA13 / IGFBP5 e IL8Rb / MDK / CDC2 / HOXA13 / IGFBP5. Como se muestra en las FIG. 14 y 15, la inclusión de IL8Rb aumenta la capacidad del marcador o de la combinación de marcadores para diagnosticar con precisión el cáncer de vejiga en un sujeto. La presente invención no ha de limitarse a estas combinaciones específicas y opcionalmente puede incluir uno o más marcadores adicionales adecuados para detectar el cáncer de vejiga, por ejemplo, uno cualquiera o más de los marcadores expuestos en las **FIG. 6** o **7.** La **tabla 1** a continuación muestra los identificadores para los marcadores específicos MDK, CDC2, IGFBP5 y HOXA13 e IL8Rb.

Tabla 1: Identificadores para marcadores de tumor de vejiga

Nombre del gen PE	Nombre del gen HGNC (oficial)	RefSeq del NCBI	ID del gen de Entrez del NCBI	URL de HGNC
MDK	MDK	NM_002391	4192	http://www.genenames.org/data/hgnc_data.php?hgnc_id=6972
CDC	CDK1	NM_001170406	983	http://www.genenames.org/data/hgnc_data.php?hgnc_id=1722

(continuación)

Nombre del gen PE	Nombre del gen HGNC (oficial)	RefSeq del NCBI	ID del gen de Entrez del NCBI	URL de HGNC
IGF	IGFBP5	NM_000599	3488	http://www.genenames.org/data/hgnc_data.php?hgnc_id=5474
HOXA	HOXA13	NM_000522	3209	http://www.genenames.org/data/hgnc_data.php?hgnc_id=5102
IL8Rb	CXCR2	NM_001168298	3579	http://www.genenames.org/data/hgnc_data.php?hgnc_id=6027

Las FIG. 2 a **4** y **12** a **17** muestran el efecto del uso de IL8Rb en combinación con cuatro marcadores conocidos representativos de cáncer de vejiga; MDK, CDC2, IGFBP5 y HOXA13. Los resultados muestran que mediante la incorporación del uso de IL8Rb individualmente con cada marcador (**FIG. 2, 14** y **15** a **17**), pero también cuando se usa con todas las posibles combinaciones de los cuatro marcadores de BTM como firma, hay una mejora en la capacidad para diferenciar las muestras de pacientes con CCT de aquellas con afecciones no malignas.

Como se muestra en las **FIG. 10** a **13,** la inclusión de IL8Rb con los cuatro marcadores, MDK, CDC2, IGFBP5 y HOXA13 (uRNA-D) no solo aumentó el rendimiento general de la prueba en comparación con solo los cuatro marcadores, sino que comparativamente, la prueba fue muy superior a otras pruebas conocidas, NMP22®, "una marca registrada de Matritech, Inc., de Massachusetts, Estados Unidos" Elisa, NMP22 BladderChek® (una marca registrada de Matritech, Inc., de Massachusetts, Estados Unidos) y citología. **Las FIG. 14** a **17** también muestran el efecto de IL8Rb en las diversas combinaciones de los cuatro marcadores, MDK, CDC2, IGFBP5 y HOXA13.

La **FIG. 14** muestra las curvas ROC para todas las combinaciones de los cuatro marcadores, MDK, CDC2, IGFBP5 y HOXA13, con y sin IL8Rb, calculadas usando cinco modelos de clasificación diferentes (i) análisis de discriminantes lineales (LDA), (ii) regresión logística (LogReg), (iii) máquinas de vectores de soporte (SVM), (iv) K 5 vecinos más próximos (KN5N) y (v) clasificación de árbol de partición (TREE). La **FIG. 15** muestra una tabla del área bajo la curva (ABC) para los 5 clasificadores y las 15 combinaciones de los 4 biomarcadores, con y sin IL8Rb. Este cálculo de la ABC está restringido al área desde una tasa de falso positivo de 0 hasta una tasa de falso positivo del 20 %, lo que abarca los intervalos útiles de especificidad (80 - 100 %). La ABC cuantifica las diferencias visibles en las curvas ROC de la **FIG. 14.** La **FIG. 16** muestra la sensibilidad de todas las combinaciones de los cuatro marcadores, medidas con y sin IL8Rb a niveles de especificidad de (a) 80 %, (b) 85 %, (c) 90 %, (d) 95 % y (e) 98 %. La **FIG. 17** muestra en una tabla los cambios en la sensibilidad (dirección vertical en las curvas ROC; un resultado mejor es "ascendente") o la especificidad (dirección horizontal en la curva ROC; un resultado mejor es hacia la izquierda) a los niveles de especificidad fijos de (a, f) 80 %, (b, g) 85 %, (c, h) 90 %, (d, I) 95 % y (e, j) 98 %, respectivamente.

Estos resultados muestran que IL8Rb, en general, mejora la capacidad de los biomarcadores (MDK, CDC, IGFBP5 y HOXA13), de manera individual o en combinación, para diferenciar las muestras tumorales de las normales.

En general, estos resultados muestran que IL8Rb fue capaz de aumentar la precisión con la que la prueba puede detectar el cáncer de vejiga o urotelial. Las mejoras de mayor magnitud se observaron o bien con marcadores que no tuvieron tan buen rendimiento con la inclusión de IL8Rb o con clasificadores que no tuvieron tan buen rendimiento. Se observaron mejoras de menor magnitud para marcadores y/o clasificadores que tuvieron buen rendimiento antes de añadir IL8Rb y por tanto, había menos espacio de mejora. Resulta importante destacar que los resultados muestran un análisis basado en población y el beneficio de la incorporación de IL8Rb podría ser mayor cuando se diagnostica a pacientes individuales, especialmente aquellos para los que no esté claro el diagnóstico basado en los marcadores de BTM.

Estos resultados demuestran que IL8Rb no solo puede usarse para detectar una enfermedad inflamatoria de la vejiga, sino también que cuando se usa en combinación con marcadores para cáncer de vejiga, da como resultado una detección mejorada del cáncer de vejiga, que surge de una reducción en los resultados "falsos positivos".

Estos resultados también muestran la utilidad de IL8Rb en tanto que afecta al rendimiento general de las diversas combinaciones de marcadores y confirman la capacidad de IL8Rb para mejorar el rendimiento de uno o más marcadores de cáncer de vejiga para detectar de manera precisa el cáncer en un paciente. Además, **las FIG. 14** y **15** muestran que pueden lograrse los mismos resultados usando un intervalo de modelos de clasificación y demuestran que el resultado no depende de un modelo o algoritmo de clasificación, sino de la combinación de marcadores usada. Estos resultados confirman que puede usarse cualquier modelo o algoritmo de clasificación en la presente invención. En particular, las **FIG. 14** y **15** muestran que IL8Rb tiene un mayor efecto a los mayores valores de especificidad y en particular, en los intervalos de mayor utilidad clínica.

Por lo tanto, mediante el uso del Índice G+P de la presente invención, se tiene la capacidad de efectuar un triaje de pacientes con hematuria basándose en variables fenotípicas y genéticas en grupos que se encuentran en "alto riesgo" de tener cáncer urotelial en los que está justificado efectuar inmediatamente pruebas complementarias adicionales, "en riesgo" y en los que está justificado efectuar inmediatamente pruebas complementarias y aquellos

con "bajo riesgo", que pueden ponerse en una lista de observación para su evaluación posterior.

Detección de marcadores genéticos en muestras corporales

En varias realizaciones preferidas, las pruebas para el cáncer pueden efectuarse de manera deseable en muestras obtenidas de sangre, plasma, suero, líquido peritoneal obtenido, por ejemplo, usando lavados peritoneales u otros fluidos corporales, tales como orina, linfa, líquido cefalorraquídeo, fluido gástrico o muestras de heces. Para la detección de afecciones inflamatorias de la vejiga o de cáncer de vejiga, la prueba se efectúa idealmente en una muestra de orina.

Específicamente, los presentes métodos para detectar una enfermedad inflamatoria de la vejiga o cáncer de vejiga pueden efectuarse en cualquier muestra adecuada del cuerpo que pueda ser indicativa de la orina, pero idealmente, el nivel de IL8Rb y cualquier marcador de cáncer adicional se determina directamente a partir de una muestra de orina.

Una prueba puede llevarse a cabo o bien directamente en una muestra de orina o puede estabilizarse la muestra mediante la adición de cualquier compuesto o tampón adecuado conocido en la técnica para estabilizar y prevenir la degradación del ARN y/o las proteínas en la muestra, de tal forma que pueda analizarse en una fecha posterior o incluso para asegurarse de que el ARN y/o las proteínas estén estabilizados durante el análisis.

La determinación del nivel de proteína y/o ARN en la orina del sujeto puede efectuarse directamente en la orina o puede tratarse la orina para purificar y/o concentrar el ARN y/o las proteínas. Se conocen bien en la técnica diversos métodos para extraer y/o concentrar proteínas y/o ARN y pueden usarse en la presente invención.

Se apreciará que en la técnica se conocen diversos métodos para determinar el nivel de expresión de un gen particular, ya sea a nivel de ARN y/o proteína y puede usarse cualquier método adecuado en la presente invención. Más adelante se describen en líneas generales algunos métodos frecuentes, aunque la invención no está restringida a estos métodos y es adecuado cualquier método para cuantificar los niveles de proteína y/o ARN para su uso en la presente invención.

Estrategias generales para la detección de enfermedades y cáncer usando marcadores genotípicos

A continuación se describen brevemente metodologías generales para determinar los niveles de expresión, aunque se apreciará que resultaría adecuado cualquier método para determinar los niveles de expresión.

PCR cuantitativa (qPCR)

La PCR cuantitativa (qPCR) puede llevarse a cabo en muestras de tumor, en suero y plasma usando cebadores y sondas específicos. En reacciones controladas, la cantidad de producto formada en una reacción PCR (Sambrook, J., E Fritsch, E. y T Maniatis, *Molecular Cloning: A Laboratory Manual* 3.ª ed. Cold Spring Harbor Laboratory Press: Cold Spring Harbor (2001)) se correlaciona con la cantidad de molde inicial. La cuantificación del producto de la PCR puede llevarse a cabo deteniendo la reacción PCR cuando se encuentra en fase logarítmica, antes de que los reactivos sean el factor limitante. Posteriormente se someten a electroforesis los productos de la PCR en geles de agarosa o poliacrilamida, se tiñen con bromuro de etidio o una tinción para ADN comparable y se mide mediante densitometría la intensidad de la tinción. Como alternativa, puede medirse el progreso de una reacción PCR usando máquinas para PCR tales como Prism 7000™ de Applied Biosystems (una marca registrada de Applera Corporation, Connecticut, Estados Unidos) o LightCycler™ de Roche (una marca registrada de Roche Molecular Systems, Inc., California, Estados Unidos), que miden la acumulación de producto en tiempo real. La PCR en tiempo real mide o bien la fluorescencia de los colorantes intercalados en el ADN, tales como Sybr Green en el producto de la PCR sintetizado o la fluorescencia liberada por una molécula indicadora cuando se escinde de una molécula inactivadora; las moléculas indicadoras e inactivadoras se incorporan en una sonda oligonucleotídica que se hibrida con la molécula de ADN diana después de la extensión de la hebra de ADN a partir de los cebadores oligonucleotídicos. La sonda oligonucleotídica se desplaza y degrada mediante la acción enzimática de la polimerasa Taq en el siguiente ciclo de la PCR, lo que libera el indicador de la molécula inactivadora. En una variación, conocida como Scorpion, la sonda está unida covalentemente al cebador.

PCR de retrotranscripción (RT-PCR)

La RT-PCR puede usarse para comparar niveles de ARN en diferentes poblaciones de muestra, en tejidos normales y tumorales, con o sin tratamiento farmacológico, para caracterizar los patrones de expresión, pada diferenciar entre ARN estrechamente relacionados y para analizar la estructura del ARN.

Para la RT-PCR, la primera etapa es el aislamiento del ARN a partir de una muestra diana. El material de partida es normalmente ARN total aislado de tumores humanos o líneas celulares tumorales humanas y tejidos o líneas celulares normales correspondientes, respectivamente. El ARN puede aislarse de diversas muestras, tales como muestras tumorales de tejidos de mama, pulmón, colon (por ejemplo, intestino grueso o delgado), colorrectal,

gástrico, esofágico, anal, rectal, prostático, cerebral, hepático, renal, pancreático, de bazo, timo, testículo, ovario, útero, vejiga, etc., de tumores primarios o de líneas celulares tumorales y de muestras agrupadas de donantes sanos. En caso de que la fuente de ARN sea un tumor, el ARN puede extraerse, por ejemplo, de muestras de tejido congeladas o de archivo fijadas (por ejemplo, fijadas con formol) e incluidas en parafina.

La primera etapa en la elaboración de perfiles de expresión génica mediante RT-PCR es la retrotranscripción del molde de ARN en ADNc, seguido de su amplificación exponencial en una reacción PCR. Las retrotranscriptasas más frecuentemente empleadas son la retrotranscriptasa del virus de la mieloblastosis aviar (AMV-RT) y la retrotranscriptasa del virus de la leucemia murina de Moloney (MMLV-RT). La etapa de retrotranscripción normalmente se ceba usando cebadores específicos, hexámeros aleatorios o cebadores oligo-dT, dependiendo de las circunstancias y del objetivo de la elaboración del perfil de expresión. Por ejemplo, el ARN extraído puede retrotranscribirse usando un kit para PCR de ARN GeneAmp® (Perkin Elmer, CA, EE. UU.), siguiendo las instrucciones del fabricante. El ADNc obtenido puede usarse posteriormente como molde en la reacción de PCR posterior.

Aunque la etapa de la PCR puede emplear diversas ADN polimerasas dependientes de ADN termoestables, normalmente emplea la ADN polimerasa Taq, que tiene actividad nucleasa 5'-3' pero carece de actividad de endonucleasa de comprobación 3'-5'. Por tanto, la qPCR TaqMan® (una marca registrada de Roche Molecular Systems, Inc., California, Estados Unidos) normalmente utiliza la actividad nucleasa 5' de la polimerasa Taq o Tth para hidrolizar una sonda de hibridación unida a su amplicón diana, pero puede usarse una enzima con actividad nucleasa 5' equivalente.

Se usan dos cebadores oligonucleotídicos para generar un amplicón típico de una reacción PCR. Se diseña un tercer oligonucleótido o sonda para detectar la secuencia de nucleótidos ubicada entre los dos cebadores de la PCR. La sonda es no extensible mediante la enzima ADN polimerasa Taq y se marca con un colorante fluorescente indicador y un colorante fluorescente inactivador. Cualquier emisión inducida mediante láser por el colorante indicador se inactiva mediante el colorante inactivador cuando los dos colorantes se encuentran juntos al estar sobre la sonda. Durante la reacción de amplificación, la enzima ADN polimerasa Taq escinde la sonda de una manera dependiente del molde. Los fragmentos de sonda resultantes se disocian en solución y la señal del colorante indicador liberado se libera del efecto inactivador del segundo fluoróforo. Se libera una molécula de colorante indicador por cada nueva molécula sintetizada y la detección del colorante indicador inactivado proporciona la base para la interpretación cuantitativa de los datos.

La RT-PCR TaqMan® (una marca registrada de Roche Molecular Systems, Inc., California, Estados Unidos) puede llevarse a cabo usando equipos disponibles en el mercado, tales como, por ejemplo, el sistema de detección de secuencias ABI PRISM 7700™ (una marca comercial de Applera Corporation, Connecticut, Estados Unidos) (Perkin-Elmer-Applied Biosystems, Foster City, CA, EE. UU) o Lightcycler™ (una marca registrada de Roche Molecular Systems, Inc., California, Estados Unidos (Roche Molecular Biochemicals, Mannheim, Alemania). En una realización preferida, el procedimiento de nucleasa 5' se efectúa en un dispositivo de PCR cuantitativa en tiempo real, tal como el sistema de detección de secuencias ABI PRISM 7700™. El sistema consiste en un termociclador, un láser, un dispositivo de carga acoplada (CCD), una cámara y un ordenador. El sistema amplifica muestras en un formato de 96 pocillos en un termociclador. Durante la amplificación, se recoge en tiempo real la señal fluorescente inducida por láser mediante cables de fibra óptica para los 96 pocillos y se detecta en el CCD. El sistema incluye un programa informático para controlar el instrumento y para analizar los datos.

Los datos del ensayo de nucleasa 5' se expresan inicialmente como Cp o el ciclo umbral. Como se ha analizado anteriormente, los valores de fluorescencia se registran durante cada ciclo y representan la cantidad de producto amplificado hasta ese punto en la reacción de amplificación. El punto cuando se registra por primera vez la señal fluorescente como estadísticamente significativa es el ciclo umbra, Cp.

PCR cuantitativa en tiempo real (qRT-PCR)

Una variación más reciente de la técnica de RT-PCR es la PCR cuantitativa en tiempo real, que mide la acumulación de producto de la PCR mediante una sonda fluorigénica doblemente marcada (es decir, la sonda TaqMan®). La PCR en tiempo real es compatible tanto con la PCR competitiva cuantitativa como con la PCR comparativa cuantitativa. La primera usa un competidor interno para cada secuencia diana para la normalización, mientras que la última usa un gen de normalización contenido en la muestra o un gen constitutivo para la RT-PCR. Se proporcionan detalles adicionales, por ejemplo, por Held *et al.*, *Genome Research* 6: 986-994 (1996).

Los niveles de expresión pueden determinarse usando tejidos fijados incluidos en parafina como fuente de ARN. De acuerdo con un aspecto de la presente invención, los cebadores de la PCR se diseñan para flanquear secuencias intrónicas presentes en el gen que se va a amplificar. En esta realización, la primera etapa en el diseño del cebador/sonda es la de delinear las secuencias intrónicas en los genes. Esto puede efectuarse mediante programas informáticos disponibles de manera pública, tal como el programa informático DNA BLAT desarrollado por Kent, W. J., *Genome Res.* 12 (4): 656-64 (2002) o mediante el programa informático BLAST, lo que incluye sus variaciones. Las etapas posteriores siguen métodos bien establecidos de diseño del cebador y la sonda para la PCR.

Para evitar señales no específicas, es útil enmascarar las secuencias repetitivas dentro de los intrones cuando se diseñan los cebadores y las sondas. Esto puede lograrse fácilmente usando el programa Repeat Masker disponible en línea a través de la Escuela de Medicina de Baylor, que explora sistemáticamente las secuencias de ADN frente a una biblioteca de elementos repetitivos y devuelve una secuencia de consulta en la que enmascaran los elementos repetitivos. Las secuencias enmascaradas pueden usarse posteriormente para diseñar secuencias de cebador y sonda usando cualquier paquete de diseño de cebador/sonda comercial o disponible al público de otro modo, tal como Primer Express (Applied Biosystems); MGB assay-by-design (Applied Biosystems); Primer3 (Steve Rozen y Helen J. Skaletsky (2000) *Primer3 on the VIMNV for general users and for biologist programmers* en: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols*: *Methods in Molecular Biology*. Humana Press, Totowa, NJ, págs. 365-386).

Los factores más importantes tomados en consideración en el diseño del cebador para la PCR incluyen la longitud del cebador, la temperatura de fusión (Tm) y el contenido de G/C, la especificidad, las secuencias de cebador complementarias y la secuencia terminal 3'. En general, los cebadores para la PCR óptimos tienen normalmente 1730 bases de longitud y contienen aproximadamente un 20-80 %, tal como, por ejemplo, aproximadamente un 50-60 % de bases de G+C. Normalmente se prefieren temperaturas de fusión de entre 50 y 80 °C, por ejemplo, de aproximadamente 50 a 70 °C. Para guías adicionales para el diseño del cebador y la sonda para la PCR, véase, por ejemplo, Dieffenbach, C. W. *et al.*, *General Concepts for PCR Primer Design* en: *PCR Primer, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Nueva York, 1995, págs. 133-155; Innis y Gelfand, *Optimization of PCRs* en: *PCR Protocols, A Guide to Methods and Applications*, CRC Press, Londres, 1994, págs. 5-11; y Plasterer, T. N. *Primerselect: Primer and probe design. Methods Mol.* Biol. 70: 520-527 (1997), cuya divulgación completa se incorpora expresamente al presente documento por referencia.

Análisis de micromatrices

La expresión diferencial también puede identificarse o confirmarse usando la técnica de micromatrices. Por tanto, puede medirse el perfil de expresión de marcadores específicos de enfermedad o bien en tejido tumoral fresco o incluido en parafina, usando la tecnología de micromatrices. En este método, se emplacan o disponen en una matriz las secuencias polinucleotídicas de interés (incluyendo ADNc y oligonucleótidos), sobre un sustrato de microplaca. Las secuencias dispuestas en forma de matriz (es decir, sondas de captura) se hibridan posteriormente con polinucleótidos específicos de las células o tejidos de interés (es decir, dianas). Como en el método de la RT-PCR, la fuente de ARN normalmente es ARN total aislado de tumores o líneas celulares tumorales humanas y los tejidos o líneas celulares normales correspondientes. Por tanto, el ARN puede aislarse de diversos tumores primarios o líneas celulares tumorales. En caso de que la fuente del ARN sea un tumor primario, el ARN puede extraerse, por ejemplo, de muestras de tejido congeladas o de archivo fijadas incluidas en parafina (FFPE) y muestras de tejido fijadas (por ejemplo, fijadas con formol), que se preparan y conservan rutinariamente en la práctica clínica diaria.

En una realización específica de la técnica de micromatrices, se aplican insertos de clones de ADN amplificados mediante la PCR a un sustrato. El sustrato puede incluir hasta 1, 2, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o 75 secuencias de nucleótidos. En otros aspectos, el sustrato puede incluir al menos 10.000 secuencias de nucleótidos. Las secuencias dispuestas en micromatriz, inmovilizadas sobre la micromatriz, son adecuadas para la hibridación en condiciones rigurosas. En cuanto otras demás realizaciones, las dianas para las micromatrices pueden tener al menos 50, 100, 200, 400, 500, 1000 o 2000 bases de longitud; o 50-100, 100-200, 100-500, 100-1000, 100-2000 o 500-5000 bases de longitud. En cuanto a realizaciones adicionales, las sondas de captura para las micromatrices pueden tener al menos 10, 15, 20, 25, 50, 75, 80 o 100 bases de longitud; o 10-15, 10-20, 10-25, 10-50, 10-75, 10-80 o 20-80 bases de longitud.

Las sondas de ADNc marcadas fluorescentemente pueden generarse mediante la incorporación de nucleótidos fluorescentes mediante retrotranscripción de ARN extraído de tejidos de interés. Las sondas de ADNc marcadas aplicadas a la microplaca se hibridan específicamente a cada punto de ADN en la matriz. Después de un lavado riguroso para retirar sondas no unidas específicamente, se escanea el chip mediante microscopía láser confocal o mediante otro método de detección, tal como una cámara CCD. La cuantificación de la hibridación de cada elemento dispuesto en la matriz permite la evaluación de la abundancia del ARNm correspondiente. Con la fluorescencia de dos colores, se hibridan por pares a la matriz las sondas de ADNc marcadas por separado a partir de dos fuentes de ARN. De este modo, se determina simultáneamente la abundancia relativa de los transcritos de las dos fuentes correspondientes a cada gen especificado.

La escala miniaturizada de la hibridación permite una evaluación rápida y conveniente del patrón de expresión para grandes números de genes. Se ha demostrado que estos métodos tienen la sensibilidad necesaria para detectar transcritos escasos, de los que se expresan pocas copias por cada célula y para detectar de manera reproducible diferencias de aproximadamente dos veces en los niveles de expresión (Schena *et al.*, *Proc. Natl. Acad. Sci. USA* 93 (2): 106-149 (1996)). El análisis de micromatrices puede efectuarse mediante equipos disponibles comercialmente, siguiendo los protocolos del fabricante, tal como usando la tecnología GenChip® de Affymetrix, la tecnología de micromatrices de Illumina o la tecnología de micromatrices de INcyte. El desarrollo de métodos de micromatrices para el análisis a gran escala de la expresión génica hace posible buscar sistemáticamente marcadores moleculares

de clasificación del cáncer y la predicción de resultados en diversos tipos de tumor.

Aislamiento, purificación y amplificación de ARN

Los métodos generales para la extracción del ARNm se conocen bien en la técnica y se divulgan en manuales convencionales de biología molecular, entre los que se incluyen Ausubel *et al.*, *Current Protocols of Molecular Biology*, John Wiley and Sons (1997). Se divulgan métodos de extracción de ARN a partir de tejidos incluidos en parafina, por ejemplo, en Rupp y Locker, *Lab Invest*. 56: A67 (1987) y De Sandres *et al.*, *BioTechniques* 18: 42044 (1995). En particular, el aislamiento de ARN puede efectuarse usando un kit de purificación, un conjunto de tampones y proteasas de fabricantes comerciales, tales como Qiagen, de acuerdo con las instrucciones del fabricante. Por ejemplo, puede aislarse el ARN total de células en cultivo usando minicolumnas RNeasy® de Qiagen "una marca registrada de Qiagen GmbH, Hilden, Alemania". Otros kits de aislamiento de ARN disponibles comercialmente incluyen el kit de purificación de ADN y ARN completo MasterPure™ (EPICENTRE (D, Madison, WI) y el kit de aislamiento de ARN de bloque de parafina (Ambion, Inc.). El ARN total de las muestras de tejido puede aislarse usando ARN Stat-60 (Tel-Test). El ARN preparado de tumores puede aislarse, por ejemplo, mediante centrifugación en gradiente de densidad de cloruro de cesio.

Se proporcionan las etapas de un protocolo representativo para determinar el perfil de expresión génica usando tejidos fijados incluidos en parafina como fuente de ARN, que incluye aislamiento de ARNm, purificación, extensión del cebador y amplificación en varios artículos publicados en revistas científicas (por ejemplo: T. E. Godfrey *et al. J. Molec. Diagnostics* 2: 84-91 (2000); K. Specht *et al.*, *Am. J. Pathol*. 158: 419-29 (2001)). Brevemente, un proceso representativo comienza con el corte de secciones con un espesor de aproximadamente 10 micrómetros de muestras de tejido tumoral incluido en parafina. A continuación se extrae el ARN y se eliminan las proteínas y el ADN. Después del análisis de la concentración de ARN, se pueden incluir las etapas de reparación y/o amplificación del ARN, en caso necesario y el ARN se retrotranscribe usando promotores específicos de genes, seguido de RT-PCR. Finalmente, los datos se analizan para identificar las mejores opciones de tratamiento disponibles para el paciente basándose en el patrón de expresión génica característico identificado en la muestra tumoral examinada.

Inmunohistoquímica y proteómica

Los métodos de inmunohistoquímica también son adecuados para detectar los niveles de expresión de los marcadores de proliferación de la presente invención. Por tanto, para detectar la expresión, se usan anticuerpos o antisueros, preferentemente antisueros policlonales y lo más preferentemente, anticuerpos monoclonales específicos para cada marcador. Los anticuerpos pueden detectarse mediante marcaje directo de los anticuerpos en sí, por ejemplo, con marcadores radiactivos, marcadores fluorescentes, marcadores de haptenos, tales como, biotina o una enzima, tal como peroxidasa de rábano picante o fosfatasa alcalina. Como alternativa, el anticuerpo primario no marcado se usa junto con un anticuerpo secundario marcado, lo que comprende antisueros, antisueros policlonales o un anticuerpo monoclonal específico para el anticuerpo primario. Los protocolos y kits de inmunohistoquímica se conocen bien en la técnica y se encuentran disponibles comercialmente.

Puede usarse proteómica para analizar los polipéptidos presentes en una muestra (por ejemplo, tejido, organismo o cultivo celular) en un punto de tiempo determinado. En particular, pueden usarse técnicas de proteómica para evaluar los cambios globales de expresión de polipéptido en una muestra (también denominadas proteómica de expresión). El análisis de proteómica normalmente incluye: (1) separación de polipéptidos individuales en una muestra mediante electroforesis en gel de poliacrilamida en 2-D (PAGE en 2-D); (2) identificación de los polipéptidos individuales recuperados del gel, por ejemplo, mediante espectrometría de masas o secuenciación N-terminal y (3) análisis de los datos usando bioinformática. Los métodos de proteómica son complementos valiosos para otros métodos de elaboración de perfiles de expresión génica y pueden usarse, individualmente o en combinación con otros métodos, para detectar los productos de los marcadores de proliferación de la presente invención.

Métodos de hibridación usando sondas de ácido nucleico selectivas para un marcador

Estos métodos implican la unión de la sonda de ácido nucleico a un soporte e hibridación en condiciones adecuadas con ARN o ADNc procedente de la muestra de ensayo (Sambrook, J., E Fritsch, E. y T Maniatis, *Molecular Cloning: A Laboratory Manual* 3.ª ed. Cold Spring Harbor Laboratory Press: Cold Spring Harbor (2001)). Estos métodos pueden aplicarse a marcadores procedentes de una muestra de tejido tumoral o de fluido. Las preparaciones de ARN o ADNc se marcan típicamente con una molécula fluorescente o radiactiva para posibilitar la detección y cuantificación. En algunas aplicaciones, el ADN de hibridación puede marcarse con una estructura ramificada marcada fluorescentemente para potenciar la intensidad de la señal (Nolte, F.S., *Branched DNA signal amplification for direct quantitation of nucleic acid sequences in clinical specimens*. *Adv. Clin. Chem.* 33, 201-35 (1998)). El marcador no hibridado se elimina mediante lavado minucioso en soluciones con baja salinidad, tales como SSC 0,1x, SDS al 0,5 %, antes de cuantificar la cantidad de hibridación mediante detección de la fluorescencia o densitometría de imágenes del gel. Los soportes pueden ser sólidos, tales como membranas de nylon o nitrocelulosa o consisten en microesferas o perlas que se hibridan cuando se encuentran en suspensión líquida. Para posibilitar el lavado y la purificación, las perlas pueden ser magnéticas (Haukanes, B-1 y Kvam, C., *Application of magnetic beads in bioassays*. *Bio/Technology* 11, 60-63 (1993)) o se marcan fluorescentemente para posibilitar la citometría de flujo

(véase, por ejemplo: Spiro, A., Lowe, M. y Brown, D., *A Bead-Based Method for Multiplexed Identification and Quantitation of DNA Sequences Using Flow Cytometry*. *Appl. Env. Micro.* 66, 4258-4265 (2000)).

Una variación de la técnica de hibridación es el ensayo QuantiGene Plex® (una marca registrada de Panomics, de California, Estados Unidos) (Genospectra, Fremont) que combina un soporte de perlas fluorescentes con amplificación de la señal de ADN ramificado. Otra variación más de la tecnología de hibridación es el ensayo de ARNm Quantikine® (R&D Systems, Minneapolis). La metodología es como se describe en las instrucciones del fabricante. Brevemente, el ensayo usa sondas de hibridación oligonucleotídicas conjugadas a digoxigenina. La hibridación se detecta usando anticuerpos dirigidos contra digoxigenina acoplados a fosfatasa alcalina en ensayos colorimétricos.

Se conocen bien en la técnica métodos adicionales y no es necesario describirlos adicionalmente en el presente documento.

Ensayos inmunológicos ligados a enzimas (ELISA)

Brevemente, en los ensayos de ELISA sándwich, se une un anticuerpo policlonal o monoclonal contra el marcador a un soporte sólido (Crowther, J.R. *The ELISA guidebook*. Humana Press: Nueva Jersey (2000); Harlow, E. y Lane, D., *Using antibodies: a laboratory manual*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor (1999)) o suspensión de perlas. Se conocen otros métodos en la técnica y no es necesario describirlos adicionalmente en el presente documento. Los anticuerpos monoclonales pueden obtenerse a partir de hibridomas o seleccionarse de fagotecas para anticuerpos (Hust M. y Dubel S., *Phage display vectors for the in vitro generation of human antibody fragments*. *Methods Mol Biol*. 295:71-96 (2005)). Los sitios de unión no específica se bloquean con preparaciones de proteína no diana y detergentes. Después, el anticuerpo de captura se incuba con una preparación de muestra o tejido del paciente que contiene el antígeno. La mezcla se lava antes de incubar el complejo de anticuerpo/antígeno con un segundo anticuerpo que detecta el marcador diana. El segundo anticuerpo se conjuga normalmente a una molécula fluorescente u otra molécula indicadora que pueda o bien detectarse en una reacción enzimática o con un tercer anticuerpo conjugado a un indicador (Crowther, anteriormente citado). Como alternativa, en los ELISA directos, la preparación que contiene el marcador puede unirse al soporte o perla y el antígeno diana se detecta directamente con un conjugado de anticuerpo-indicador (Crowther, anteriormente citado).

Los métodos para producir anticuerpos monoclonales y antisueros policlonales se conocen bien en la técnica y no es necesario describirlos con más detalle en el presente documento.

Inmunodetección

Los métodos también pueden usarse para la inmunodetección de miembros de la familia de marcadores en sueros o plasma de pacientes de cáncer de vejiga tomados antes y después de la cirugía para eliminar el tumor, la inmunodetección de miembros de la familiar de marcadores en pacientes con otros cánceres, entre los que se incluyen, pero sin limitación, colorrectal, pancreático, ovárico, melanoma, de hígado, esofágico, de estómago, endometrial y de cerebro y la inmunodetección de miembros de la familia de marcadores en orina y heces de pacientes con cáncer de vejiga.

Los marcadores de enfermedad también pueden detectarse en tejidos o muestras usando otras técnicas de inmunodetección estándar, tales como inmunotransferencia o inmunoprecipitación (Harlow, E. y Lane, D., *Using antibodies: a laboratory manual*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor (1999)). En la inmunotransferencia, las preparaciones de proteína procedentes de tejidos o fluidos que contienen el marcador se someten a electroforesis a través de geles de poliacrilamida en condiciones desnaturalizantes o no desnaturalizantes. Después, las proteínas se transfieren a un soporte de membrana, tal como nylon. Después, se hace reaccionar el marcador directa o indirectamente con anticuerpos monoclonales o policlonales, como se ha descrito para la inmunohistoquímica. Como alternativa, en algunas preparaciones, las proteínas pueden transferirse sobre membranas sin separación previa por electroforesis. La señal puede cuantificarse mediante densitometría.

En la inmunoprecipitación, se incuba una preparación soluble que contiene el marcador con un anticuerpo monoclonal o policlonal contra el marcador. Después, se incuba la reacción con perlas inertes hechas de agarosa o poliacrilamida con proteína A o proteína G unida covalentemente. Las perlas de proteína A o G. Las perlas de proteína A o G interactúan específicamente con los anticuerpos, formando un complejo inmovilizado de anticuerpo-marcador-antígeno unido a la perla. Después del lavado, el marcador unido puede detectarse y cuantificarse mediante inmunotransferencia o ELISA.

Determinación de un diagnóstico basándose en análisis genotípico

Una vez que se ha obtenido el nivel de expresión de IL8Rb y opcionalmente de uno o más marcadores de cáncer adicionales, se puede determinar un diagnóstico para dicho sujeto. En caso de que la expresión de IL8Rb se encuentre por encima de la expresión observada en sujetos que no tienen una enfermedad inflamatoria de vejiga y/o sea coherente con el nivel de expresión en sujetos que se sabe que tienen una enfermedad inflamatoria de vejiga,

se diagnosticará que el sujeto tiene una enfermedad inflamatoria de vejiga. Como alternativa, en caso de que la expresión no se encuentre por encima de la expresión observada en sujetos que no tienen una enfermedad inflamatoria de vejiga y/o de que se encuentre por debajo de los niveles de expresión en sujetos que se sabe que tienen una enfermedad inflamatoria de vejiga, se diagnosticará que el sujeto no tiene una enfermedad inflamatoria de vejiga.

En la situación donde la IL8Rb se usa junto con uno o más marcadores para cáncer de vejiga, el nivel de expresión de IL8Rb se comparará con el nivel de expresión de sujetos sin una enfermedad inflamatoria de vejiga y/o sujetos que se sabe que tienen una enfermedad inflamatoria de vejiga. Los uno o más marcadores de cáncer se comparan con el nivel de expresión en sujetos sin cáncer de vejiga y/o sujetos que se sabe que tienen cáncer de vejiga. En caso de que el nivel de expresión de IL8Rb sea coherente con un sujeto que no tiene una enfermedad inflamatoria de vejiga (menor que en un sujeto que tiene una enfermedad inflamatoria de vejiga) y el nivel de expresión de los uno o más marcadores de cáncer de vejiga sea coherente con un sujeto que tiene cáncer de vejiga (a diferencia de un sujeto que no tiene cáncer de vejiga), se diagnostica que el sujeto tiene cáncer de vejiga. En caso de que el nivel de expresión de IL8Rb sea mayor que en un sujeto que no tiene una enfermedad inflamatoria de vejiga (coherente con un sujeto que tiene una enfermedad inflamatoria de vejiga) y el nivel de expresión de los uno o más marcadores de cáncer de vejiga sea coherente con un sujeto que tiene cáncer de vejiga (a diferencia de un sujeto que no tiene cáncer de vejiga), se diagnostica que el sujeto tiene una enfermedad inflamatoria de vejiga. En caso de que el nivel de expresión de IL8Rb sea coherente con un sujeto que no tiene una enfermedad inflamatoria de vejiga (menor que en un sujeto que tiene una enfermedad inflamatoria de vejiga) y el nivel de expresión de los uno o más marcadores de cáncer de vejiga sea coherente con un sujeto que no tiene cáncer de vejiga (a diferencia de un sujeto que tiene cáncer de vejiga), se diagnostica que el sujeto no tiene ni cáncer de vejiga ni una enfermedad inflamatoria de vejiga.

Debido a que normalmente hay un solapamiento en los niveles de expresión entre la expresión normal y de enfermedad de un marcador diagnóstico, para determinar un diagnóstico para un sujeto es típico determinar un umbral de clasificación. Un umbral de clasificación es un valor o umbral que separa a los sujetos en las categorías de enfermo o no enfermo. Normalmente se evalúa un umbral mediante el uso de una curva de características operativas de receptor (ROC), que representa la sensibilidad frente a la especificidad para todos los umbrales conocidos.

Determinación de umbrales diagnósticos

Para pruebas que usan marcadores de enfermedad, pueden obtenerse umbrales diagnósticos que permiten decir que una muestra es positiva o negativa para la enfermedad, por ejemplo, cáncer de vejiga. Estos umbrales diagnósticos se determinan mediante el análisis de cohortes de pacientes en los que se investiga la presencia de cáncer de vejiga o enfermedad inflamatoria de vejiga. Los umbrales diagnósticos pueden variar para diferentes aplicaciones de pruebas; por ejemplo, los umbrales diagnósticos para su uso en la exploración sistemática de poblaciones se determinan usando cohortes de pacientes que en gran medida están libres de síntomas urológicos y estos umbrales diagnósticos pueden ser diferentes a los usados en pruebas para pacientes sometidos a vigilancia para recurrencia del cáncer de vejiga. Puede seleccionarse un umbral diagnóstico para proporcionar un nivel práctico de especificidad de la prueba en la situación clínica necesaria; es decir, una especificidad que permite una sensibilidad razonable sin números excesivos de pacientes que reciben resultados falsos positivos. Esta especificidad puede encontrarse en el intervalo del 80-100 %.

Un umbral diagnóstico se determina aplicando un algoritmo que combina los niveles de expresión genotípicos de cada marcador a cada muestra a partir de un ensayo clínico prospectivo.

Las muestras usadas son de pacientes con cáncer de vejiga y una serie de trastornos urológicos no malignos. Un umbral diagnóstico se selecciona determinando la puntuación del algoritmo que dio como resultado la especificidad deseada. Por ejemplo, en algunas aplicaciones, se desea una especificidad del 85 %. Posteriormente se determina un umbral diagnóstico seleccionando una puntuación del algoritmo que da como resultado la clasificación correcta de un 85 % de los pacientes sin cáncer de vejiga como negativos para el cáncer. En otras aplicaciones (tales como la exploración sistemática de poblaciones), se prefiere una mayor especificidad, tal como del 90 %. Para establecer un umbral para esta aplicación, se selecciona una puntuación del algoritmo que da como resultado la clasificación correcta de un 90 % de los pacientes sin cáncer de vejiga como negativos para el cáncer. Se explican ejemplos de uso de un algoritmo en la sección de ejemplos.

Como alternativa a los umbrales individuales, la prueba puede usar intervalos de ensayo que proporcionan diferentes grados de probabilidad de presencia de enfermedad y que tienen diferentes consecuencias clínicas asociadas con los mismos. Por ejemplo, una prueba puede tener tres intervalos; uno asociado con un alto riesgo (por ejemplo, 90 %) de presencia de cáncer de vejiga, un segundo asociado con un bajo riesgo de cáncer de vejiga y un tercero considerado como sospechoso de tener enfermedad. El intervalo "sospechoso" puede asociarse con una recomendación de repetir la prueba en un periodo de tiempo definido.

Análisis de datos

Una vez que se ha completado el método para evaluar la cantidad de ARN y/o proteína, se tienen que analizar los datos para determinar la distribución de los valores de biomarcador asociados con muestras tumorales y no tumorales. Esto implica normalmente normalizar los datos en bruto, es decir, retirar el "ruido" de fondo, etc. y promediar cualquier duplicado (o más), comparación con patrones y establecer valores de corte o umbrales para separar de manera óptima las dos clases de muestras. Se conocen muchos métodos para hacer esto y el método exacto dependerá del método específico para determinar la cantidad de ARN y/o proteína usada.

A continuación se muestra un ejemplo de cómo puede efectuarse el análisis de datos cuando se usa qRT-PCR. Sin embargo, se apreciará que puede adaptarse el proceso general para su uso en otros métodos para determinar el contenido de ARN y/o proteína o pueden establecerse otros métodos por un experto en la materia para lograr los mismos resultados.

Datos

Las mediciones de fluorescencia se toman a longitudes de onda ω_i i=1,2 en cada ciclo de la PCR. Por tanto, para cada pocillo se observa un par de curvas de fluorescencia, indicadas por $f_t(\omega_i)$, donde t=1,⋯,k indica el número de ciclo e i = 1,2 indica las longitudes de onda.

Las curvas de fluorescencia tienen una forma sigmoidal que comienza con un valor inicial prácticamente horizontal y aumenta uniformemente hasta una asíntota superior. Se usará la ubicación de un punto C_p donde la curva de fluorescencia se desvía del valor inicial lineal para caracterizar la concentración del gen diana. Se proporciona más adelante una definición precisa de C_p.
A continuación se muestra un ejemplo de un esquema para procesar estos datos.

- Compensar el solapamiento de fluorescencia entre las bandas de frecuencia,
- Estimar un modelo uniforme para cada curva de fluorescencia para estimar C_p
- Combinar los datos de pocillos duplicados.
- Estimar las curvas patrón
- Calcular una concentración en relación con el patrón.

Cada muestra biológica proporciona concentraciones relativas de 5 genes, que son las entradas para la función discriminante.

Compensación del color

Indica el nivel de fluorescencia del colorante j en el ciclo t y la frecuencia ω mediante $W_{tj}(\omega)$. En un ensayo multiplexado, la respuesta medida a cualquier frecuencia ω es la suma de los aportes de todos los colorantes a dicha frecuencia, por lo que para cada ciclo.

$$f_t(\omega) = W_{t1}(\omega) + W_{t2}(\omega) + \ldots$$

El objetivo de la compensación del color es extraer las aportaciones individuales $W_{tj}(\omega)$, a partir de las mezclas observadas $f_t(\omega)$.

En una situación ideal, la fluorescencia $W_{tj}(\omega_0)$, causada por el colorante j a una frecuencia ω es proporcional a su fluorescencia $W_{tj}(\omega_0)$ a la frecuencia de referencia ω_0, independientemente del nivel de $W_{tj}(\omega_0)$. Esto sugiere la relación lineal

$$\begin{bmatrix} f_t(\omega_1) \\ f_t(\omega_2) \end{bmatrix} = \begin{bmatrix} W_{t1}(\omega_1) + W_{t2}(\omega_1) \\ W_{t1}(\omega_2) + W_{t2}(\omega_2) \end{bmatrix} = \begin{bmatrix} 1 & A_{12} \\ A_{21} & 1 \end{bmatrix} \begin{bmatrix} W_{t1}(\omega_1) \\ W_{t2}(\omega_2) \end{bmatrix}$$

para algunas constantes de proporcionalidad A_{12} y A_{21} que se van a determinar,

En realidad, hay efectos adicionales, que se modelan eficazmente introduciendo términos lineales en este sistema, por tanto

$$\begin{bmatrix} f_t(\omega_1) \\ f_t(\omega_2) \end{bmatrix} = \begin{bmatrix} 1 & A_{12} \\ A_{21} & 1 \end{bmatrix} \begin{bmatrix} W_{t1}(\omega_1) \\ W_{t2}(\omega_2) \end{bmatrix} + \begin{bmatrix} a_1 + b_1 t \\ a_2 + b_2 t \end{bmatrix}$$

Después de estimar los parámetros de "compensación del color" A_{12} y A_{21} se pueden recuperar $W_{t1}(\omega_1)$ y $W_{t2}(\omega_2)$, aunque distorsionados por un valor inicial lineal, mediante la multiplicación de matriz:

$$\begin{bmatrix} W_{t1}(\omega_1) \\ W_{t2}(\omega_2) \end{bmatrix} = \begin{bmatrix} 1 & A_{12} \\ A_{21} & 1 \end{bmatrix}^{-1} \begin{bmatrix} f_t(\omega_1) \\ f_t(\omega_2) \end{bmatrix} + \begin{bmatrix} a_1^* + b_1^* t \\ a_2^* + b_2^* t \end{bmatrix}$$

$W_{t1}(\omega_1)$ y $W_{t2}(\omega_2)$ se denominan datos con "color compensado". Las distorsiones lineales $a_i^* + b_i^* t$ en el último término de esta expresión se acomodarán en la estimación del valor inicial cuando se estima un modelo para los datos con color compensado 2 a continuación. No tiene influencia en la estimación de C_p.

La estimación de los coeficientes de compensación del color requiere un ensayo por separado usando sondas individuales (en contraposición con las dúplex). Por tanto $W_{t2}(\omega_2) = 0$ lo que da:

$$\begin{bmatrix} f_t(\omega_1) \\ f_t(\omega_2) \end{bmatrix} = \begin{bmatrix} 1 & A_{12} \\ A_{21} & 1 \end{bmatrix} \begin{bmatrix} W_{t1}(\omega_1) \\ 0 \end{bmatrix} + \begin{bmatrix} a_1 + b_1 t \\ a_2 + b_2 t \end{bmatrix}$$

Por tanto,

$$f_t(\omega_2) = A_{21} f_t(\omega_1) + a^* + b^* t$$

El coeficiente A_{2l} puede estimarse mediante regresión lineal convencional de $f_t(\omega_2)$ en $f_t(\omega_1)$ y el ciclo de la PCR t para $t = 1,\cdots,k$.

Estimación del modelo

En esta sección, y_t $t=1,\cdots,k$ indica una curva de fluorescencia con el color compensado.

Amplificación

Los modelos se estiman únicamente para curvas de fluorescencia que muestran amplificación no trivial. Los presentes inventores definen el término "amplificación" como una desviación no trivial del valor inicial lineal de la curva de fluorescencia con el color compensado. Se usa la relación de señal a ruido (SNR) para cuantificar la amplificación. En este caso, la SNR se define como la relación de la varianza de la señal a la varianza del ruido. La varianza del ruido se establece como una parte de la calibración del procedimiento del ensayo y permanece constante: para este fin, se usa la variante residual de un modelo lineal para el valor inicial a partir de los pocillos que no tienen amplificación, es decir, pocillos sin ARN. Para cada curva de fluorescencia, se estima la varianza de la señal como la varianza residual a partir de la recta con mejor ajuste ("mejor" ha de interpretarse en el sentido de mínimos cuadrados).

- En caso de que la SNR sea menor que un umbral específico, la curva de fluorescencia es prácticamente lineal y no hay presencia de amplificación. Por tanto, no hay puntos de desviación respecto del valor inicial y puede declararse que la concentración en la muestra es cero.
- En caso de que la SNR se encuentre por encima del umbral, hay presencia de amplificación y puede estimarse una concentración.

Los umbrales para la SNR (adimensional) se seleccionan para proporcionar una discriminación clara entre las curvas "amplificadas" y "no amplificadas". Por ejemplo, los siguientes intervalos para los umbrales son eficaces para los marcadores.

Flúor	Gen	Intervalo
JOE	MDK	40-120
JOE	CDC	35-70
JOE	IL8R	30-60
FAM	IGF	50-80
FAM	HOXA	50-150
FAM	XENO	50-80

Modelo

Estimación de un modelo sigmoidal para cada curva de fluorescencia. Puede usarse cualquier forma paramétrica adecuada del modelo, pero esta ha de ser capaz de modelar las siguientes características:

- valor inicial lineal que pueda tener una pendiente distinta de cero,
- asimetrías aproximadamente en el punto intermedio.
- asíntotas en los niveles inferior y superior
- aumento uniforme desde el valor inicial hasta la asíntota superior

Un ejemplo de un modelo que cumple con estos requisitos es

$$g_t(\theta) = A + A_s t + \frac{D}{\left(1 + \left(\frac{t}{B}\right)^E\right)^F}$$

Los presentes inventores lo denominan el "modelo 6PL". El vector paramétrico θ= *[A,As,D,B,E,F]* está sujeto a las siguientes restricciones para garantizar que $g_t(\theta)$ es una función creciente de *t* y tiene propiedades empíricas de una curva de fluorescencia.

$$D > 0, B > 0, E < 0, F < 0$$

Los otros dos parámetros determinan el valor inicial $A + A_s t$, y estos parámetros no necesitan restricciones explícitas aunque *A* siempre es un valor positivo y el parámetro de la pendiente A_s siempre es pequeño. El parámetro D determina el nivel de amplificación por encima del valor de referencia. Los demás parámetros, *B.E.F* no tienen por sí mismos una interpretación intrínseca, pero determinan la forma de la curva. Estos parámetros también son los únicos parámetros que tienen influencia en la estimación de C_p. Cuando $A_s = 0$, esto se conoce como la función logística de cinco parámetros (5PL) y si, además, $F = 1$ este modelo se reduce a un modelo logístico de cuatro parámetros (4PL), Gottschalk y Dunn (2005), Spiess *et al.* (2008).

Inicialización

Los valores iniciales para la estimación no lineal se establecen como

- $A_s = 0$, $F = 1$
- A = media de $(y, \cdots, y_5)$
- D = intervalo de $(y_1, \cdots, yk)$
- B = ciclo correspondiente a la semialtura
- E se inicializa convirtiendo $g_t(\theta)$ en una forma lineal que tiene fijos los valores de los parámetros restantes a sus valores iniciales definidos anteriormente. La linealización obtiene

$$E \log\left(1 + \frac{t}{B}\right) = \log\left(\frac{D}{y_t - A}\right)$$

A continuación se estima E mediante regresión de log $\log\left(\frac{D}{y_t - A}\right)$ en $\log\left(1 + \frac{t}{B}\right)$ para *t* seleccionado de tal forma que

$$A + \frac{D}{10} < y_t < A + \frac{9D}{10}$$

Una forma alternativa de este modelo que proporciona un análisis prácticamente idéntico (con su propia inicialización) es:

$$A + A_s t + \frac{D}{\left(1 + \exp\left(-\frac{t - B}{E}\right)\right)^F}$$

Cuando $A_s = 0$ esto se conoce en ocasiones como la ecuación de Richards, Richards (1959).

Criterios de estimación

Estimación de parámetros para minimizar una suma penalizada del criterio de cuadrados:

$$\sum_t (y_t - g_t(\theta))^2 + \lambda(\theta)$$

En este caso, $\lambda(\theta)$ es una función no negativa que penaliza los valores elevados de algunos (o todos) los parámetros en θ. Este método se conoce como regularización o regresión contraída (Hoerl, 1962) y puede obtenerse desde un punto de vista Bayesiano determinando una distribución previa adecuada para el vector paramétrico θ. Una elección satisfactoria para la penalización es:

$$\lambda(\theta) = \lambda(B^2 + D^2 + E^2 + F^2)$$

Los valores elevados de λ sesgan las estimaciones paramétricas hacia cero y reducen la varianza de las estimaciones paramétricas. Por el contrario, λ reducida (o cero) proporciona estimaciones de parámetros inestables y complicaciones de convergencia en los algoritmos de minimización. La elección de λ es un equilibrio entre el sesgo y la varianza o la estabilidad. Las pruebas empíricas demuestran que puede lograrse un equilibrio satisfactorio entre el sesgo y la varianza si λ se selecciona dentro del intervalo:

$$0{,}01 > \lambda > 0{,}0001.$$

Esta elección también garantiza la convergencia del algoritmo de optimización.

Elección del algoritmo

Para cualquier elección de λ en el intervalo anterior, la descripción en el párrafo anterior define completamente las estimaciones del parámetro. Se ha usado con éxito un procedimiento no lineal de mínimos cuadrados basado en el procedimiento clásico de Gauss-Newton (tal como el algoritmo de Levenberg-Marquardt implementado en More, 1978) y es una estrategia adecuada. También se han evaluado en este contexto algoritmos de optimización generales, tales como Nelder y Mead, 1965 o el algoritmo de Boryden-Fletcher, implementado por Byrd *et al*., 1995).

Estimación de C_p

C_p es el punto en tiempo *t* que maximiza la segunda derivada de $g_t(\theta)$. Cada curva de fluorescencia proporciona un C_p que caracteriza la concentración del gen diana. La media de los C_p estimados para cada conjunto de duplicados técnicos se calcula y usa en el análisis posterior.

Curvas patrón

Las concentraciones absolutas o relativas se obtienen a partir de una comparación con curvas patrón en la misma placa de PCR. Modelado de una dilución en serie usando el modelo lineal:

$$C_p = R + S \log_{10} Conc$$

donde *Conc* es una concentración absoluta o relativa del patrón. Los parámetros de punto de corte y pendiente son específicos para la placa. Modelado de la variabilidad entre placas en los parámetros de punto de corte y pendiente estableciendo modelos de población

$$R \sim N(\mu_R, \sigma_R^2)$$

$$S \sim N(\mu_S, \sigma_S^2)$$

donde los parámetros μ_R, σ_R^2, μ_S, σ_S^2 se establecen basándose en los datos anteriores como se describen más adelante. Por tanto, para una placa determinada, *R* y S pueden interpretarse como observaciones a partir de estas poblaciones.

Para un duplicado *i* del patrón a una concentración $Conc_j$, puede usarse el siguiente modelo:

$$C_p(i,j) = R + S \log_{10} Conc_j + \epsilon_{ij}$$

donde $\epsilon_{ij} \sim N(0, \sigma_j^2)$. Obsérvese que la varianza de los valores residuales depende de C_p. Se proporcionan estimaciones empíricas de $Var(\epsilon_{ij})$ en la **tabla 2**. Estimación de los parámetros *R* y S usados maximizando la función de probabilidad. Se interpreta el parámetro de pendiente en términos de eficiencia del proceso de la PCR mediante la expresión:

$$S = -\frac{1}{\log_{10} Eficacia}$$

Este modelo tiene una interpretación Bayesiana: Se proporcionan distribuciones anteriores vagas (no informativas) para los parámetros μ_R, σ_R^2, μ_S, σ_S^2. Por tanto, los modelos de población para *R* y S y para $C_p(i,j)$ determinan por completo un modelo de probabilidad para los datos anteriores. Un algoritmo de Monte Carlo basado en cadenas de Markov (MCMC) (Lunn *et al*., 2009) permite la estimación de μ_R, σ_R^2, μ_S, σ_S^2.
En caso de que se omita la distribución anterior, se obtiene como resultado una interpretación frecuentista tradicional. Siguiendo este procedimiento de estimación, es posible obtener las estimaciones del parámetro de población dependiente del gen en la **tabla 3**.

Tabla 2: Varianza de los valores residuales

C_p	σ^2	C_p	σ^2
12	0,0100	21	0,0466
13	0,0108	22	0,0625
14	0,0119	23	0,0860
15	0,0134	24	0,1212
16	0,0155	25	0,1750
17	0,0184	26	0,2591
18	0,0224	27	0,3931
19	0,0279	28	0,6112
20	0,0356	29	0,9741

Tabla 3: Parámetros de población para las pendientes y puntos de corte de las curvas patrón

	μ_R	σ_R^2	μ_S	σ_S^2
MDK	19,49	0,5112	-3,426	0,0481
CDC	18,91	0,2343	-3,414	0,0198
IL8R	31,43	0,0919	-3,192	0,0017
IGF	20,63	0,3835	-3,275	0,0247
HOXA	22,51	0,1544	-3,270	0,0037

Las estimaciones del punto de corte y la pendiente de la curva patrón se indican mediante $\hat{R}$ y $\hat{S}$.

Concentraciones relativas ΔC_p

Uso de la curva patrón para calcular $C_{p(REF)}$ a la concentración $Conc_{REF}$ a partir de la expresión: $C_{p(REF)} = \hat{R} + \hat{S} \log_{10} Conc_{REF}$. La concentración relativa de una muestra se proporciona mediante la expresión:

$$\Delta C_p = \frac{C_p - C_{p(REF)}}{\hat{S}} = \log_{10} \frac{Conc_{MUESTRA}}{Conc_{REF}}$$

Como alternativa, $\hat{S}$ puede estimarse aproximadamente a un nivel fijo correspondiente a una eficiencia de la PCR de 2. Por tanto, $\hat{S} = -1/\log_{10}(2) = -3{,}32$. Se emplea la misma notación ΔC_P para cualquiera de las dos elecciones. Las estimaciones de ΔC_P resultantes, una para cada gen, son datos de entrada para la función discriminante en la etapa siguiente.

Función discriminante

Los valores de ΔC_P corresponden a un valor de biomarcador relativo, eliminándose la variación entre placas. El examen de los 5 valores de ΔC_P comparados entre sí (por ejemplo, véase la FIG. 2), muestra cómo las muestras de tumor tienen normalmente diferentes valores de biomarcador que las muestras no de tumor. Además, aunque sigue habiendo solapamiento en las áreas para tumor y normal, se separa bien de manera efectiva un gran número de muestras. En estas circunstancias, pueden usarse muchos clasificadores estadísticos diferentes para separar las muestras normales de las tumorales. En este caso se demuestra que una muestra de varios clasificadores funciona para separar estas muestras. Se usaron 5 métodos de clasificación diferentes: 1) análisis de discriminantes lineales (LDA), 2) regresión lineal (LogReg); 3) máquinas de vectores de soporte (SVM); 4) K vecino más cercano (KNN) basado en 5 vecinos (KN5N); y 5) árboles de partición recursivos (TREE) (Cite: Venables y Ripley y Dalgaard).

La creación de un clasificador requiere un conjunto de datos que contiene los valores de biomarcador para un gran número de muestras que deben representar en última instancia la población que se va a evaluar mediante el clasificador. Por ejemplo, en caso de que un clasificador se vaya a usar para la detección sistemática de una población en riesgo (por ejemplo, de 50 años o más, fumadores), el conjunto de datos necesario para crear el clasificador (denominado el "conjunto de entrenamiento") debe imitar a dicha población y contener únicamente muestras de personas de más de 50 años que fumen. Normalmente, para obtener precisión de la medición con un error menor del 10 % para los parámetros tales como sensibilidad y especificidad, el conjunto de entrenamiento ha de ser mayor de 300 muestras.

La estimación de la eficacia de un clasificador puede efectuarse usando validación cruzada, validación cruzada (Wikipedia: Validación cruzada), el conjunto de datos se divide en un pequeño número de particiones de igual tamaño (normalmente de 3 a 10). Se descarta una sección y las demás secciones se usan para construir un clasificador; posteriormente se analiza la sección descartada mediante el nuevo clasificador y se anotan sus predicciones. Esto se efectúa a su vez para cada sección y todas las predicciones se combinan y analizan para calcular las características del clasificador: Sensibilidad, Especificidad, etc. En caso de que la validación cruzada se efectúa particionando los datos en 10 partes, se denomina validación cruzada de factor 10; de manera similar, en 3 partes sería una validación cruzada de factor 3. En caso de que los datos se particionen en tantas clases como

muestras, esto se denomina "validación cruzada descartando uno". Al evaluar datos no usados para construir el clasificador, este método proporciona una estimación del rendimiento del clasificador en ausencia de muestras adicionales.

Se han construido clasificadores usando las 15 combinaciones de 4 biomarcadores, MDK, IGFBP5, CDC2 y HOXA13, todas con y sin el biomarcador de IL8Rb, usando el conjunto de datos de del ensayo clínico descrito en otra parte del presente documento (ejemplo 1) y se han evaluado estos 30 clasificadores usando validación cruzada de factor 10. Esto se efectuó para cada uno de los 5 tipos de clasificador listados anteriormente y las curvas ROC calculadas. Todos los trabajos se efectuaron usando el ambiente de programación estadística R (CITE). Estos resultados **(FIG. 14)** muestran que en la mayoría de casos, el clasificador con IL8Rb es más sensible para valores de especificidad que son útiles desde el punto de vista diagnóstico (tasa de falsos positivos del 0 al 20 %; especificidad del 100 al 80 %). El área bajo la curva (ABC) para la región con utilidad diagnóstica de las especificidades se usa para cuantificar el buen rendimiento de los clasificadores, en donde los valores mayores indican un mejor rendimiento del clasificador. La **FIG. 15a** muestra en una tabla la ABC para clasificador y combinación de biomarcadores, mientras que la figura 15b muestra la cantidad de aumento en la ABC para cada condición cuando se añade IL8Rb. En la mayoría de los casos, la adición de IL8Rb mejora la capacidad para efectuar diagnósticos precisos. Los valores de sensibilidad específicos para los valores de especificidad útiles desde el punto de vista diagnóstico se presentan en forma de tabla para todos los clasificadores en la **FIG. 16.** Además, la **FIG. 17** presenta en forma de tabla la cantidad de aumento de la sensibilidad o la especificidad que proporciona la adición de IL8Rb.

La utilidad del clasificador se crea cuando, habiéndolo creado y evaluado, se usa para evaluar una nueva muestra. Para simplificar la interpretación de los resultados, se determina una puntuación de corte o umbral; las muestras a un lado del punto de corte se consideran positivas y las del otro lado, negativas para los tumores. Pueden determinarse valores de corte adicionales, por ejemplo, para indicar niveles crecientes de certidumbre de los resultados. En este caso, se ha establecido un corte que proporciona una tasa de falsos positivos del 15 % en el conjunto de entrenamiento. Usando estas curvas ROC validadas de manera cruzada, se puede estimar de este modo la sensibilidad. Normalmente, también se estableció un corte con un valor predictivo positivo del 75 %. Para usar estos cortes, se estableció un resultado "negativo" para puntuaciones menores que el valor de corte establecido por la especificidad del 85 %. Las puntuaciones mayores de un PPV del 75 % se denominan "positivas" y una puntuación entre las dos se denomina "intermedia" o "sospechosa".

Anticuerpos dirigidos contra IL8Rb

En aspectos adicionales, la presente divulgación incluye la producción de anticuerpos dirigidos contra IL8Rb. El marcador IL8Rb puede producirse en una cantidad suficiente para que sea adecuado para provocar una respuesta inmunológica. En algunos casos, puede usarse IL8Rb de longitud completa y en otros, puede ser suficiente un fragmento peptídico de IL8Rb como inmunógeno. El inmunógeno puede inyectarse en un hospedador adecuado (por ejemplo, ratón, conejo, etc.) y si se desea, puede inyectarse un adyuvante, tal como adyuvante completo de Freund o adyuvante incompleto de Freund, para aumentar la respuesta inmunitaria. Puede apreciarse que la producción de anticuerpos es rutinaria en la técnica de inmunología y no es necesario describirla adicionalmente en el presente documento. Como resultado, se pueden producir anticuerpos, incluyendo anticuerpos monoclonales o de presentación en fagos, dirigidos contra IL8Rb.

En otros aspectos adicionales, pueden producirse anticuerpos contra la proteína o el núcleo de proteína de los marcadores tumorales identificados en el presente documento o contra una secuencia oligonucleotídica única para un IL8Rb. Aunque algunas proteínas pueden estar glucosiladas, en algunos casos, las variaciones en el patrón de glucosilación pueden provocar una detección errónea de formas de IL8Rb que carecen de los patrones de glucosilación habituales. Por tanto, en ciertos aspectos de la presente divulgación, los inmunógenos de IL8Rb pueden incluir IL8Rb desglucosilado o fragmentos de IL8Rb desglucosilados. La desglucosilación puede lograrse usando una o más glucosidasas conocidas en la técnica. Como alternativa, puede expresarse el ADNc de IL8Rb en líneas celulares deficientes para glucosilación, tales como líneas celulares procariotas, entre las que se incluyen *E. coli* y similares.

Pueden producirse vectores de expresión que contienen oligonucleótidos codificantes de IL8Rb. Muchos de estos vectores pueden estar basados en vectores estándar conocidos en la técnica. Pueden usarse vectores para transfectar diversas líneas celulares para producir líneas celulares productoras de IL8Rb, que pueden usarse para producir las cantidades deseadas de IL8Rb para desarrollar anticuerpos específicos u otros reactivos para la detección de IL8Rb o para estandarizar ensayos desarrollados para IL8Rb.

Kits

Basándose en los descubrimientos de la presente invención, pueden preverse y producirse varios tipos de kits de prueba. En primer lugar, pueden producirse kits que tienen un dispositivo de detección precargado con una molécula de detección (o "reactivo de captura"). En los aspectos para la detección del ARNm de IL8Rb, dichos dispositivos pueden comprender un sustrato (por ejemplo, vidrio, silicio, cuarzo, metal, etc.) sobre el que hibridan los oligonucleótidos como reactivos de captura al que se une el ARNm que se va a detectar. En algunos aspectos, la

detección directa del ARNm puede lograrse hibridando ARNm (marcador con cy3, cy5, radiomarcador u otro marcador) a los oligonucleótidos sobre el sustrato. En otros aspectos, la detección del ARNm puede lograrse produciendo en primer lugar ADN complementario (ADNc) para el ARNm deseado. Posteriormente, el ADNc marcado puede hibridarse a los oligonucleótidos sobre el sustrato y se detecta.

Los anticuerpos también pueden usarse en kits como reactivos de captura. En algunos aspectos, un sustrato (por ejemplo, una placa multipocillo) puede tener un IL8Rb y BM específicos y reactivos de captura unidos a los mismos. En algunos aspectos, un kit puede tener incluido un reactivo de bloqueo. Los reactivos de bloqueo pueden usarse para reducir la unión no específica. Por ejemplo, puede reducirse la unión a oligonucleótidos no específicos usando ADN en exceso de cualquier fuente conveniente que no contiene oligonucleótidos de IL8Rb y BTM, tales como ADN de esperma de salmón. Puede reducirse la unión de anticuerpos no específica usando un exceso de una proteína de bloqueo, tal como seroalbúmina. Puede apreciarse que se conocen en la técnica numerosos métodos para detectar oligonucleótidos y proteínas y puede usarse cualquier estrategia que pueda detectar específicamente moléculas asociadas con marcadores y se consideran dentro del alcance de la presente invención.

También pueden usarse anticuerpos cuando están unidos a un soporte sólido, por ejemplo, usando una placa de anticuerpos, que puede permitir la detección de múltiples marcadores con una sola placa.

Además de un sustrato, un kit de ensayo puede comprender reactivos de captura (tales como sondas), soluciones de lavado (por ejemplo, SSC, otras sales, tampones, detergentes y similares), así como restos de detección (por ejemplo, cy3, cy5, radiomarcadores y similares). Los kits también pueden incluir instrucciones de uso y un embalaje.

La detección de IL8Rb y los BTM en una muestra puede efectuarse usando cualquier técnica adecuada y puede incluir, pero sin limitación, sondas oligonucleotídicas, qPCR o anticuerpos generados contra los marcadores de cáncer.

Se apreciará que la muestra que se va a analizar no está restringida a una muestra del tejido sospechoso de ser una enfermedad inflamatoria o un tumor. El marcador puede secretarse en el suero u otro fluido corporal. Por lo tanto, una muestra puede incluir una muestra corporal e incluye biopsias, sangre, suero, lavados peritoneales, líquido cefalorraquídeo, orina y muestras de heces.

También se apreciará que la presente invención no está restringida a la detección del cáncer en seres humanos, sino que es adecuada para la detección del cáncer en cualquier animal, lo que incluye, pero sin limitación, perros, gatos, caballos, ganado bovino, ovejas, ciervos, cerdos y cualquier otro animal que se sepa que puede contraer cáncer.

Pruebas generales para marcadores de enfermedad inflamatoria o cáncer en fluidos corporales

En general, los métodos de ensayo para nucleótidos, proteínas y péptidos en estos fluidos se conocen en la técnica. La detección de oligonucleótidos puede llevarse a cabo usando métodos de hibridación, tales como transferencias de Northern, transferencias de Southern o métodos de micromatriz o qPCR. Los métodos para detectar proteínas incluyen ensayos inmunosorbentes ligados a enzimas (ELISA), placas de proteínas que tienen anticuerpos, radioinmunoensayo (RIA) de perlas en suspensión, transferencia de Western y unión a lectina. Sin embargo, a modo ilustrativo, dichos niveles de marcadores de enfermedad pueden cuantificarse usando un ensayo inmunosorbente ligado a enzimas (ELISA) de tipo sándwich. Para los ensayos en plasma, se añaden a los pocillos de una placa de microtitulación una alícuota de a 5 ul de una muestra diluida de manera adecuada o de marcador patrón diluido en serie y 75 ul de anticuerpo dirigido contra marcador humano conjugado con peroxidasa. Después de un periodo de incubación de 30 minutos a 30 °C, se lavan los pocillos con Tween 20 al 0,05 % en solución salina tamponada con fosfato (PBS) para retirar el anticuerpo no unido. Los complejos unidos de marcador y anticuerpo dirigido contra el marcador se incuban posteriormente con H_2O_2 que contiene o-fenilendiamina durante 15 minutos a 30 °C. La reacción se detiene añadiendo H_2SO_4 1 M y se mide la absorbancia a 492 nm con un lector de placas de microtitulación.

Puede apreciarse que los anticuerpos dirigidos contra IL8Rb pueden ser anticuerpos monoclonales o antisueros policlonales. También puede apreciarse que puede estudiarse adecuadamente cualquier otro fluido corporal.

No es necesario que un marcador sea secretado, en un sentido fisiológico, para que sea útil. Más bien, puede ser eficaz cualquier mecanismo mediante una proteína o gen marcador entre en el suero para producir un nivel detectable y cuantificable del marcador. Por tanto, la secreción de proteínas solubles por las células, la descamación de proteínas de membrana de las membranas plasmáticas, la secreción de formas de corte y empalme alternativo del ARNm o las proteínas expresadas a partir del mismo, y la muerte celular (por ejemplo, apoptótica) pueden producir niveles suficientes para que sean útiles del marcador.

Cada vez hay más pruebas que respaldan el uso de marcadores séricos como herramientas para diagnosticar y/o evaluar la eficacia de la terapia para diversos tipos de cáncer.

Ejemplos

Los ejemplos descritos en el presente documento tienen la finalidad de ilustrar las realizaciones de la invención. Otras realizaciones, métodos y tipos de análisis se encuentran dentro del alcance de los expertos habituales en la técnica de diagnóstico molecular y no es necesario describirlos detalladamente en el presente documento. Otras realizaciones dentro del alcance de la técnica se consideran parte de la presente invención.

Ejemplo 1: Análisis genotípico del cáncer de vejiga

Métodos

Pacientes: Entre abril de 2008 y septiembre de 2009, se inscribieron 485 pacientes que presentaban hematuria macroscópica pero no antecedentes de neoplasia maligna del tracto urinario, en once clínicas de urología en Nueva Zelanda y Australia. Cada paciente proporcionó una muestra de orina inmediatamente antes de someterse a la cistoscopia y cualquier procedimiento diagnóstico adicional. Se efectuó un diagnóstico a los tres meses después de la inscripción en el estudio. De estos 485 pacientes, se obtuvieron datos de expresión génica de los cinco genes del estudio en 442 pacientes usando los métodos descritos a continuación. Las características de estos pacientes se muestran en la **tabla 4.**

Tabla 4: Características de la población de estudio I

Diagnóstico	**Número**
Hiperplasia benigna de próstata	18
Cistitis	18
Hematuria inducida por ejercicio	3
Nefropatía inducida por ejercicio	3
Neoplasia no específica	3
Enfermedad prostática no específica	63
Próstata vascular	49
Otro cáncer urológico (distinto de CCT)	5
Vasos superficiales	3
Estenosis uretral	6
Infección del tracto urinario	18
Urolitiasis	25
Uso de warfarina	10
Etiología desconocida	155
Miscelánea	7
CCT	56
Total	**442**

La **tabla 4** muestra el número de pacientes en cada una de las categorías de diagnóstico principales a los tres meses después de que el paciente presentase inicialmente hematuria macroscópica.

Análisis de orina: Se analizaron muestras de orina por un laboratorio de citología centralizado de revisión (Southern Community Laboratories, Dunedin, Nueva Zelanda). Se llevaron a cabo las pruebas diagnósticas NMP22 BladderChek® (Matritech) y NMP22 ELISA (Matritech) de acuerdo con las instrucciones del fabricante en el centro sanitario (BladderChek®) o por Southern Community Laboratories (NMP22 ELISA).

Cuantificación de ARN: se mezclaron 2 ml de orina de cada paciente con tampón de extracción de ARN que contenía tiocianato de guanidina 5,64 M, sarcosil al 0,5 % y NaoAc 50 mM a pH 6,5. Posteriormente se extrajo el ARN total mediante extracción con Trizol (Invitrogen) y el procedimiento de RNeasy (Qiagen), como se ha descrito anteriormente. El ARN se eluyó de las columnas en 35 ul de agua y se usaron 3 ul en cada ensayo de reacción en cadena de la polimerasa de retrotranscripción (qRT-PCR) monoplexado o multiplexado. Cada 16 ul de reacción de qRT-PCR contenía 0,3 U de RNAse-OUT (Invitrogen), 0,225 uM de cada sonda Taqman, 1,25 U de Superscript III (Invitrogen), 0,275 uM de cada cebador, 1,5 U de polimerasa Taq Fast Start (Roche), DTT 10 mM, dNTPs 0,375 mM, MgSO4 4,5 mM, 1,6 ul de tampón para PCR Fast Start 10x (Roche) y 2,6 ul de solución rica en GC (Roche). Los cebadores y las sondas doblemente marcadas se obtuvieron de Integrated DNA Technologies (Coralville, EE. UU.) para cada uno de los cinco genes del estudio: MDK, CDC2, HOXA13, IGFBP5 e IL8Rb. Las secuencias de cebador/sonda se muestran en la tabla 2. Las reacciones se establecieron en placas de 96 pocillos y se ciclaron como se indica a continuación en un dispositivo Light Cycler® 480 de Roche: 50 °C, 15 min; 95 °C 8 min; 10 ciclos de 95 °C 15 s, 60 °C 2 min y 30 ciclos de 95 °C 15 s, 60 °C 1 min. Se incluyeron curvas patrón de diluciones en serie de 1/16 de un ARN de referencia (procedente de ARN de líneas celulares agrupados) en cada placa para generar un intervalo de 0,3 pg/µl a 20 ng/µl. Los datos se recogieron en la fase de extensión de los 30 ciclos térmicos finales y se exportaron en forma de un archivo de texto en bruto. La tabla 5 a continuación representa las secuencias de cebador y sonda usadas para la cuantificación por qRT-PCR de los cinco marcadores de ARN.

Tabla 5

Marcador	Sec. directa	Sec. inversa	Sonda
MDK	TGC ACC CCC AAG ACC AAA (SEQ ID NO. 3)	TGA TTA AAG CTA ACG AGC AGA CAG AA (SEQ ID NO. 4)	CCT TCC CTT TCT TGG CTT TGG CCT TT (SEQ ID NO. 5)
IGFBP5	CGT TGT ACC TGC CCA ATT GTG A (SEQ ID NO. 6)	GGG ACG CAT CAC TCA ACG TT (SEQ ID NO. 7)	AAG AGA AAG CAG TGC AAA CCT TCC CGT (SEQ ID NO. 8)
CDC2	GCC GCC GCG GAA TAA T (SEQ ID NO. 9)	TGT CTA CCC TTA TAC ACA ACT CCA TAG G (SEQ ID NO. 10)	AGC CGG GAT CTA CCA TAC CCA TTG ACT AAC T (SEQ ID NO. 11)
HOXA13	TGG AAC GGC CAA ATG TAC TG (SEQ ID NO. 12)	TGG CGT ATT CCC GTT CAA GT (SEQ ID NO.13)	ACT CTG CCC GAC GTG GTC TCC CA (SEQ ID NO.14)
IL8Rb	CCT TGA GGC ACA GTG AAG ACA TC (SEQ ID NO. 15)	CCT GTA GGA CAC CTC CAG AAG AG (SEQ ID NO.16)	TGG CCA CTC CAA TAA CAG CAG GTC ACA (SEQ ID NO. 17)

Análisis de datos de la qRT-PCR

Los datos de fluorescencia en bruto se exportaron del dispositivo LightCycler® 480 de Roche como un archivo delimitado por tabuladores que contenía el número de ciclo frente a dos canales de datos de fluorescencia para todos los pocillos en la placa. Los datos se procesaron usando un programa en R que aplicaba compensación de color ([Bernard 1999]) a los datos para corregir respecto del filtrado de un canal fluorescente en otro. Posteriormente se ajusta a un modelo logístico de 5 puntos para estimar el Cp usando el máximo de la segunda derivada ([Spiess 2008]).

Todas las muestras y controles se aplicaron por duplicado a las placas de la PCR. Los valores de Cp de los pocillos duplicados se promediaron antes de su uso. En caso de que la diferencia entre los dos valores de Cp superase 3 unidades, se repitió dicha muestra. Para proporcionar estandarización entre placas de PCR, los C_P se expresaron como ΔC_P en relación con un ARN de referencia (procedente de los ARN de línea celular agrupados) a 20 ng/µl:

$$\Delta C_P = C_P \text{ (muestra)} - C_P \text{ (ARN de referencia)}$$

Análisis estadístico

Los valores de ΔC_P de la qRT-PCR de MDK, CDC2, HOXA13, IGFBP5 e IL8Rb se usaron para generar clasificadores para separar muestras que contenían CCT de muestras que no contenían CCT, basándose en un análisis de discriminantes lineales o regresión logística ([Venables 2002]). En ambos casos, se permitieron interacciones entre los genes en los modelos de clasificador. La generación del LDA se efectuó mediante procedimientos estándar, como se describen, por ejemplo, en "*Modern Applied Statistics with S*, 4.ª edición" por W.N. Venables y B.D. Ripley (2002), Springer. Se limpiaron los datos incompletos del conjunto de datos del estudio y después se usó el ambiente estadístico en R (R Development Core Team (2009) y la función "lda" del paquete MASS (Venables y Ripley (2002)) para generar y evaluar el discriminante lineal en los datos del ensayo clínico.

La generación del clasificador de regresión logística se efectuó de un modo similar a la generación del LDA. De nuevo, se limpiaron los datos incompletos de los datos del estudio. Se creó un clasificador de regresión logística usando R; no se necesitaron paquetes adicionales. La regresión logística se efectuó como se describe por Dalgaard (2008). La comparación entre clasificadores se efectuó usando curvas ROC, usando el paquete de R, ROCR (Sing *et al.* 2009). Se generaron intervalos de confianza para las curvas ROC usando los métodos de Macskassy *et al* ([Macskassy 2005]). Se generaron los siguientes algoritmos:

Clasificador de discriminantes lineales

El primer clasificador, un discriminante lineal, (denominado LDA-3), se basa en valores de cinco genes

(normalizados a un valor de referencia restando el valor de referencia) permitiendo múltiples interacciones entre los genes. El clasificador se construyó en R usando la función "lda()" del paquete denominado "MASS". (R versión 2.9.1; MASS versión 7.2-49). El clasificador se construyó usando la siguiente ecuación:

lda3 <- lda(CCT.SN ~ MDK * IGF * CDC * HOXA * IL8R, datos=uRNA.Trial)

donde lda3 es el modelo creado; CCT.SN es el valor auténtico para la "presencia de CCT en orina" (sí o no) determinado mediante cistoscopia; MDK, IGF, CDC, HOXA e IL8R son el valor de Cp del gen normalizado; y uRNA.Trial es un archivo de datos que contiene los valores de Cp para cada uno de los 5 genes y CCT.SN (sí o no) del ensayo clínico. El uso del ASTERISCO, '*', en la fórmula significa multiplicación. La evaluación de la puntuación del clasificador toma como entrada un nuevo marco de datos que contiene los cinco valores génicos, así como el clasificador, lda3, para proporcionar una puntuación de clasificador:

puntuación <- c(predecir(lda3, nuevos.datos)$x),

donde "puntuación" es el resultado usado del clasificador para predecir la presencia de TCC; "lda3" es el clasificador creado anteriormente y "nuevos.datos" es un ARCHIVO de datos que contiene los valores medidos de los cinco genes denominados con los mismos nombres usados en la creación del clasificador. La sintaxis, '$x' y "c(...)" está presente para extraer la puntuación específicamente de la gran cantidad de información devuelta por la función predict. AL establecer el corte de puntuación en 0,112 y más, determina una especificidad del 85 % para la presencia de CCT en la muestra de orina. Los coeficientes para LDA-3 se muestran en la **tabla 6**:

Tabla 6

MDK.d.R100	5,333639e+00
IGF.d.R100	3,905978e+00
CDC.d.R100	6,877143e-01
HOXA.d.R100	6,073742e+00
IL8R.d.R100	-1,229466e+00
MDK.d.R100:IGF.d.R100	-7,420480e-01
MDK.d.R100:CDC.d.R100	-2,611158e-01
IGF.d.R100:CDC.d.R100	-1,965410e-01
MDK.d.R100:HOXA.d.R100	-8,491556e-01
IGF.d.R100:HOXA.d.R100	-4,037102e-01
CDC.d.R100:HOXA.d.R100	-3,429627e-01
MDK.d.R100:IL8R.d.R100	1,903118e-01
IGF.d.R100:IL8R.d.R100	2,684005e-01
CDC.d.R100:IL8R.d.R100	-1,229809e-01
HOXA.d.R100:IL8R.d.R100	2,909062e-01
MDK.d.R100:IGF.d.R100:CDC.d.R100	4,108895e-02
MDK.d.R100:IGF.d.R100:HOXA.d.R100	7,664999e-02
MDK.d.R100:CDC.d.R100:HOXA.d.R100	4,832034e-02
IGF.d.R100:CDC.d.R100:HOXA.d.R100	2,116340e-02
MDK.d.R100:IGF.d.R100:IL8R.d.R100	-3,750854e-02
MDK.d.R100:CDC.d.R100:IL8R.d.R100	1,664612e-02
IGF.d.R100:CDC.d.R100:IL8R.d.R100	2,089442e-03
MDK.d.R100:HOXA.d.R100:IL8R.d.R100	-1,539486e-02
IGF.d.R100:HOXA.d.R100:IL8R.d.R100	-3,894153e-02
CDC.d.R100:HOXA.d.R100:IL8R.d.R100	6,295032e-03
MDK.d.R100:IGF.d.R100:CDC.d.R100:HOXA.d.R100	-4,359738e-03
MDK.d.R100:IGF.d.R100:CDC.d.R100:IL8R.d.R100	-2,019317e-04
MDK.d.R100:IGF.d.R100:HOXA.d.R100:IL8R.d.R100	3,746882e-03
MDK.d.R100:CDC.d.R100:HOXA.d.R100:IL8R.d.R100	-2,902150e-03
IGF.d.R100:CDC.d.R100:HOXA.d.R100:IL8R.d.R100	4,799489e-04
MDK.d.R100:IGF.d.R100:CDC.d.R100:HOXA.d.R100:IL8R.d.R100	7,512308e-05

Clasificador de regresión logística

Se obtuvo un segundo clasificador basado en regresión logística a partir del mismo conjunto de datos limpiado que LDA-3. En lugar de usar la función lda(), sin embargo, se usó la función glm() del paquete stats (incluido en una instalación básica de R) como se muestra a continuación:

lr1 <- glm(CCT.SN ~ CDC * IGF * HOXA * IL8R * MDK,

familia=binomial("logit"), datos=uRNA.Trial),

donde "lr1" es el clasificador creado y los otros parámetros son como se describen para el discriminante lineal. Una vez más, la interacción completa se especifica usando el operador "*". La clasificación se efectúa de un modo muy

similar al de LDA-3:

puntuación <- predecir(lr1, nuevos.datos, tipo='respuesta'),

donde "puntuación" es el valor usado para clasificar las muestras de orina basándose en la medida de los cinco genes en "nuevos.datos", como anteriormente. El corte para lr1 se establece en 0,102 para lograr una especificidad del 85 %; los valores próximos al corte se consideran positivos para CCT. Los coeficientes para el clasificador son: -103,0818143 + 3,9043769 * CDC.d.R100 + 13,1120675 * IGF.d.R100 + 17,4771819 * HOXA.d.R100 + -10,7711519 * IL8R.d.R100 + 21,1027595 * MDK.d.R100 + -0,5938881 * CDC.d.R100 * IGF.d.R100 + -1,0736184 * CDC.d.R100 * HOXA.d.R100 + -1,3340189 * IGF.d.R100 * HOXA.d.R100 + 0,3126461 * CDC.d.R100 * IL8R.d.R100 + 1,4597355 * IGF.d.R100 * IL8R.d.R100 + 1,8739459 * HOXA.d.R100 * IL8R.d.R100 + -1,035054 * CDC.d.R100 * MDK.d.R100 + -2,5885156 * IGF.d.R100 * MDK.d.R100 + -2,7013483 * HOXA.d.R100 * MDK.d.R100 + 1,4546134 * IL8R.d.R100 * MDK.d.R100 + 0,0767503 * CDC.d.R100 * IGF.d.R100 * HOXA.d.R100 + -0,0663361 * CDC.d.R100 * IGF.d.R100 * IL8R.d.R100 + -0,1015552 * CDC.d.R100 * HOXA.d.R100 * IL8R.d.R100 + -0,2110656 * IGF.d.R100 * HOXA.d.R100 * IL8R.d.R100 + 0,1361215 * CDC.d.R100 * IGF.d.R100 * MDK.d.R100 + 0,1601118 * CDC.d.R100 * HOXA.d.R100 * MDK.d.R100 + 0,259745 * IGF.d.R100 * HOXA.d.R100 * MDK.d.R100 + -0,0106468 * CDC.d.R100 * IL8R.d.R100 * MDK.d.R100 + -0,1947899 * IGF.d.R100 * IL8R.d.R100 * MDK.d.R100 + -0,185286 * HOXA.d.R100 * IL8R.d.R100 * MDK.d.R100 + 0,0136603 * CDC.d.R100 * IGF.d.R100 * HOXA.d.R100 * IL8R.d.R100 + -0,0151368 * CDC.d.R100 * IGF.d.R100 * HOXA.d.R100 * MDK.d.R100 + 0,0056651 * CDC.d.R100 * IGF.d.R100 * IL8R.d.R100 * MDK.d.R100 + 0,0030538 * CDC.d.R100 * HOXA.d.R100 * IL8R.d.R100 * MDK.d.R100 + 0,0232556 * IGF.d.R100 * HOXA.d.R100 * IL8R.d.R100 * MDK.d.R100 + -0,000867 * CDC.d.R100 * IGF.d.R100 * HOXA.d.R100 * IL8R.d.R100 * MDK.d.R100

Resultados

Análisis por qRT-PCR de muestras de orina

Para obtener una vista de conjunto del efecto de IL8Rb en la detección del CCT, se construyeron gráficas de dispersión bidimensionales usando datos de qRT-PCR obtenidos de la orina de pacientes o bien con CCT (n=56) o las afecciones no malignas urolitiasis (n=25), infección del tracto urinario (n=18) o cistitis (n=18). Las gráficas de dispersión se construyeron usando pares de genes de una firma genética (MDK, CDC2, HOXA13, IGFBP5). Posteriormente se sustituyó un gen de cada par por IL8Rb y se volvieron a representar los datos. Estas gráficas se muestran en las **FIG. 2a-f**. La sustitución de IGFBP5 y HOXA13 por IL8Rb en las gráficas con MDK (**FIG. 2a-c**) mostró una mejora de la separación entre muestras de pacientes con CCT y aquellos con afecciones no malignas. Se observó la misma tendencia en gráficas con CDC2, en las que se sustituyeron IGFBP5 y HOXA13 por IL8Rb (**FIG. 2d-f**).

La aportación de IL8Rb al diagnóstico correcto de CCT en pacientes que presentan hematuria macroscópica se cuantificó posteriormente mediante análisis de curvas ROC. Los datos de la qPCR para cada gen en la firma (MDK, CDC2, IGFBP5 y HOXA13) e IL8Rb se usaron para desarrollar algoritmos de discriminante lineal que maximizan la discriminación entre los pacientes con CTT y aquellos sin CCT. Se desarrollaron dos algoritmos de discriminante lineal usando la cohorte completa de 442 muestras: LD1, que usó los datos de qRT-PCR de MDK, CDC2, HOXA13 e IGFBP5 y LD2, que usó MDK, CDC2, HOXA13, IGFBP5 e IL8Rb. Posteriormente se usaron LD1 y LD2 para generar curvas ROC que muestran la sensibilidad y especificidad de la detección del CCT en el grupo de pacientes con CCT confirmado (n=56) o las afecciones no malignas urolitiasis (n=25), infección del tracto urinario (n=18) o cistitis (n=18). La figura 3a muestra las curvas ROC para LD1 y LD2. El área bajo la curva ROC para LD1 fue del 78 % en comparación con un 84 % para LD2.

Como alternativa al análisis de discriminantes lineales, se usó regresión lineal como método independiente para desarrollar un algoritmo para la discriminación entre pacientes con CCT y aquellos con enfermedad no maligna. En cuanto al análisis de discriminantes lineales, se desarrollaron los algoritmos de regresión logística usando la cohorte completa de 442 muestras. Las curvas ROC obtenidas usando regresión logística y las 56 muestras de TCC y las 61 no malignas descritas anteriormente se muestran en la **FIG. 3b**. El área bajo la curva ROC para LR1 (obtenida usando los datos de qRT-PCR de MDK, CDC2, HOXA13 e IGFBP5) fue del 80 % en comparación con un 86 % para LR2 (obtenida usando los datos de qRT-PCR de MDK, CDC2, HOXA13, IGFBP5 e IL8Rb). Estos datos ilustran claramente que la inclusión de IL8Rb en los métodos para la detección del CCT usando muestras de orina pueden dar lugar a una discriminación mejorada entre pacientes con CCT y enfermedades no malignas, tales como cistitis, infección del tracto urinario y urolitiasis.

Para confirmar la precisión mejorada proporcionada por IL8Rb para la discriminación entre pacientes con CCT y urolitiasis, infección del tracto urinario o cistitis se mantuvo en una cohorte no seleccionada de pacientes que comprendía un gran número y diversidad de pacientes sin neoplasias malignas, se repitieron los análisis de curvas ROC con la cohorte completa de 442 muestras descritas en la **tabla 1**. En este análisis, el área bajo la curva para LD1 y LD2 fue del 86 y el 89 %, respectivamente (**FIG. 4a**). De manera similar, el área bajo la curva para LR1 fue del 87 % y para LR2 del 91 % (**FIG. 4b**). Este resultado confirma que IL8Rb proporciona una precisión mejorada en la detección del CCT usando muestras de orina.

Esta mejora en la detección del cáncer a causa de la inclusión de IL8Rb se ilustró adicionalmente aplicando LD1/LD2 y LR1/LR2 a la cohorte de 442 pacientes y después determinando la sensibilidad de detección solo de CCT en estadio Ta. Los tumores en estadio Ta son tumores más pequeños y más diferenciados que normalmente son más difíciles de detectar que los tumores en un estadio superior. LD1 detectó 18/31 (58 %) de los tumores Ta en comparación con 19/31 (61 %) para LD2 a una especificidad del 85 %. LR1 detectó 21/31 (68 %) en comparación con 24/31 (77 %) para LR2 (especificidad del 85 %). Estos datos demuestran que la inclusión de IL8Rb en los algoritmos LD y LR aumentó la sensibilidad de la detección de tumores en estadio Ta hasta un 9 %. EN comparación con estas muestras de ARN, las otras tres pruebas para cáncer de vejiga en este estudio mostraron una precisión notablemente menor para la detección de tumores Ta: citología de orina (sensibilidad del 39 %, especificidad del 94 %), NMP22 ELISA (sensibilidad del 35 %, especificidad del 88 %) y NMP22 (BladderChek® "una marca registrada de Matritech, Inc. de Massachusetts, Estados Unidos") (sensibilidad del 39 %, especificidad del 96 %).

IL8Rb como ayuda en el diagnóstico de la inflamación del tracto urinario

Para determinar la capacidad de IL8Rb para su uso en el diagnóstico de pacientes con inflamación del tracto urinario por causas tales como cistitis o infecciones del tracto urinario, se determinaron mediante qRT-PCR los niveles en orina de ARNm de IL8Rb en pacientes con hematuria diagnosticados con hiperplasia benigna de próstata, enfermedad prostática no específica, próstata vascular, hematuria secundaria al uso de warfarina y cistitis/infección del tracto urinario. Los niveles medios de ΔCt de IL8Rb para cada una de estas afecciones fueron -3,12, -3,10, -2,84, -1,98 y -5,27, respectivamente. Se determinó que la diferencia entre la media del nivel de IL8Rb en los pacientes con cistitis/infección del tracto urinario y las otras afecciones no malignas combinadas era significativa (p=0,001) usando la prueba de suma de rangos de Wilcoxon. En la **FIG. 5** se muestran diagramas de cajas que presentan estos datos. Estos datos muestran una elevación de los niveles de IL8Rb en la mayoría de pacientes a los que se diagnostica cistitis o infección del tracto urinario, en comparación con las otras afecciones no malignas examinadas. Es posible explicar el solapamiento entre las gráficas por una combinación de tres factores: (i) la incapacidad de la práctica clínica convencional para diagnosticar correctamente cada afección, (ii) comorbilidades (por ejemplo, infección e hiperplasia benigna de próstata) y (iii) la asociación normal de elevados recuentos de neutrófilos en un subconjunto de pacientes con hiperplasia benigna de próstata, enfermedad prostática no específica, próstata vascular o hematuria secundaria al uso de warfarina. Independientemente, dada la estricta asociación entre inflamación y números de neutrófilos, la cuantificación de IL8Rb en orina proporciona un método preciso para detectar inflamación en el tracto urinario, ya sea como consecuencia de la infección o asociada con otras afecciones no malignas.

Ejemplo 2

Métodos

Población de estudio

Se inscribió de manera prospectiva una serie consecutiva de pacientes sin antecedentes de CCT de nueve clínicas de urología en Nueva Zelanda y dos en Australia entre el 28 de abril de 2008 y el 11 de agosto de 2009. El conjunto de pacientes incluyó los pacientes usados en el ejemplo 1, pero incluyó 46 pacientes adicionales, cuyos datos no estaban disponibles para el primer análisis. El estudio adicional también incluye el análisis adicional de los resultados obtenidos. Se recogieron muestras y se extrajo y analizó el ARN como se describe en el **ejemplo 1.**

Desarrollo de la prueba de ARN

uRNA® consta de cuatro marcadores de ARNm, *CDC2, HOXA13, MDK* e *IGFBP5.* Estos marcadores se seleccionaron basándose en su baja expresión en sangre y células inflamatorias y su sobreexpresión en el CCT.[2] En este estudio de cohortes, se especificó prospectivamente un algoritmo de discriminantes lineales (uRNA-D) que combinó los cuatro marcadores en una sola puntuación. uRNA-D era independiente, habiéndose desarrollado en un conjunto de datos anterior. Sin embargo, no se obtuvo usando un grupo de pacientes estrictamente caracterizado que representase la población diana prevista para la prueba. Como consecuencia, el protocolo del estudio también definió el desarrollo de un nuevo algoritmo (*Clasificador-D*) para el uso de los cinco marcadores *CDC2*, *HOXA13, MDK, IGFBP5* e *IL8Rb* usando datos obtenidos de los pacientes inscritos en el presente estudio de cohortes.

Además de *Clasificador-D,* se obtuvo un segundo algoritmo (*Clasificador-S*) usando los datos del estudio de cohortes para posibilitar la identificación de tumores que estaban o bien en estadio avanzado (estadio ≥1) o de mayor grado (clasificación de la OMS/ISUP de 1998). El algoritmo S comprendía los cinco marcadores, entre los que se incluyen *CDC2* y *HOXA13* que anteriormente se ha demostrado que se expresan diferencialmente entre tumores en estadio Ta y aquellos en estadio ≥1.

Desarrollo del clasificador

El desarrollo de dos clasificadores para el uso de los cinco marcadores *CDC2, HOXA13, MDK, IGFBP5* e *IL8Rb (Clasificador-D* Y *Clasificador-S)* se basó en datos obtenidos en este estudio, de acuerdo con los métodos descritos en la presente memoria descriptiva. Brevemente, se elaboraron modelos de regresión logística usando el ambiente

de programación estadística, R (R Development Core Team (2011), *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Viena, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/). Se evaluaron modelos elaborados usando los valores de ΔC_P para los cinco marcadores y sus interacciones de dos vías (por ejemplo, *MDK* x *CDC2, MDK* x *IGFBP5,* etc.) respecto de su capacidad para clasificar; aquellos con los valores de AIC más bajos se evaluaron en un procedimiento de validación cruzada descartando uno respecto de su sensibilidad cuando la especificidad se estableció al 85 %. Varios modelos demostraron un rendimiento comparable para cada uno de *Clasificador-D* y *Clasificador-S,* seleccionándose el modelo con el menor número de parámetros.

Métodos estadísticos

Cuando se especificó una prueba diagnóstica en el protocolo, se calcularon las proporciones y los intervalos de confianza del 95 % para la sensibilidad y la especificidad. Se representaron curvas de características operativas de receptor (ROC) y se compararon usando los comandos Stata roctab y coccomp (Statacorp y Delong). Para *Clasificador-D*, no son adecuados los intervalos de confianza, pero se usó la prueba exacta de Fisher o de Chi cuadrado (cuando lo permitiese el tamaño de la muestra) para evaluar una asociación entre el CTT o las características del paciente y las probabilidades de resultados auténticos positivos o falsos positivos. Se usaron modelos de regresión logística para explorar factores asociados con resultados falsos positivos y falsos negativos. Todos los análisis se efectuaron en Stata versión 11.2.

Resultados

Se inscribió inicialmente un total de 517 pacientes en el estudio. Se excluyó a un 4 % de los pacientes ya que se observó que no eran elegibles (n=10), no se sometieron a cistoscopia (n=9), no se indicó su estado de CCT (n=2) o no proporcionaron una muestra de orina aceptable (n=2) (figura 8). Se excluyó a 10 pacientes adicionales del análisis debido a que no tenían resultados para una o más de las pruebas de orina. En la **FIG. 9** se muestran la demografía inicial y las características clínicas de los 485 pacientes restantes.

La prevalencia del CCT en la cohorte fue del 13,6 %. En dos faltaba un estadio de revisión (ambos eran Ta según la revisión local) y a dos no se les asignó un grado de revisión (uno era grado 1 según el patólogo local, el otro bajo). De los 66 tumores, 55 eran superficiales (estadio Ta, T1 o Tis) y 11 eran invasivos en el músculo (T2). Ninguno de los pacientes tenía metástasis detectables o afectación de los ganglios linfáticos próximos. Usando el sistema de graduación de 1973, 24 se clasificaron como grado 3, 38 como grado 2, tres como grado 1 y uno como desconocido. Con el sistema WHO98, 29 se clasificaron como grado elevado, cuatro como mixto, 32 como bajo grado y uno desconocido. Además de los CTT, se diagnosticó a dos pacientes de papiloma y a siete de otras neoplasias (cinco de estas urológicas).

El corte para la prueba uRNA-D se determinó en la cohorte de estudio, con la especificidad establecida al 85 %. Con este corte, uRNA-D detectó 41 de los 66 casos de CCT (sensibilidad del 62 %), en comparación con NMP22™ ELISA (50 %), Bladderchek® (38 %) y citología (56 %). La prueba de ARN desarrollada en los datos de cohorte Clasificador-D detectó 54 de los casos de CCT (82 %) a un nivel de especificidad del 85 % y 48 (73 %) a una especificidad del 90 %. Los valores de uRNA-D y NMP22™ ELISA pueden compararse directamente, ya que ambas pruebas estaban completamente especificadas antes del estudio. La **FIG. 21** muestra las curvas ROC; las áreas bajo las curvas (ABC) son 0,81 y 0,73, respectivamente (p=0,03). La curva ROC para *Clasificador-D* fue de 0,87 (**FIG. 21**) y la mejora en el rendimiento en relación con uRNA-D parece encontrarse principalmente en el intervalo de especificidades clínicamente relevantes (por encima del 80 %).

En general, Clasificador-D detectó un 97 % de los tumores de grado alto/grado 3, en comparación con uRNA-D (83 %), citología (83 %), NMP22 ELISA (69 %) y Bladderchek® (38 %). *Clasificador-D* también fue más sensible para la detección de tumores de grado bajo (69 %), encontrándose para las otras pruebas en el intervalo del 28-41 % (**FIG**. **12**). *Clasificador-D* fue positivo para todos los casos de CCT en estadio ≥1 más ambos Ti, pero la sensibilidad fue del 68 % para el estadio Ta (p=0,016, **FIG. 12**). Esta fue sustancialmente mayor que en las otras pruebas, siendo uRNA-D la más elevada al 41 %. Los pacientes con CCT con macrohematuria o microhematuria evidente en su muestra de orina tenían más probabilidades de que se detectase su CCT al incluir IL8Rb que aquellos sin macrohematuria o microhematuria (p<0,0005), aunque esto es al menos parcialmente el resultado de la mayor proporción de CCT en estadio y grado elevado entre aquellos con macrohematuria o microhematuria. Los números fueron insuficientes para explorar esto adicionalmente en los análisis de regresión.

De los 12 casos no detectados mediante Clasificador-D, todos fueron de estadio Ta y todos excepto uno eran de bajo grado (OMS ISUP 1998). Solo dos de los doce (ambos de bajo grado, CCT de estadio Ta) fueron detectados por otra prueba (uno tanto mediante NMP22™ ELISA como BladderChek® y uno mediante uRNA-D). De los 12 casos no detectados mediante *Clasificador-D,* todos fueron de estadio Ta y todos excepto uno eran de bajo grado (OMS ISUP 1998). Solo dos de los doce (ambos de bajo grado, CCT de estadio Ta) fueron detectados por otra prueba (uno tanto mediante NMP22™ ELISA como BladderChek® y uno mediante uRNA-D). La citología no detectó ninguno de los CCT no detectados por *Clasificador-D*.

Paciente A: tumor T2 de la pelvis renal del grado elevado, sin Ti concurrentes, no se proporciona tamaño.
Paciente B: T3a de vejiga de grado elevado sin Ti concurrentes, 2x 3 cm
Paciente C: un tumor de grado elevado que mide 4,8 x 5,6 cm con invasión extensa de la capa estromal y la lámina propria muscular, que se extiende hasta la grasa perivesical sin evidencias de metástasis.

La especificidad de las pruebas de orina entre aquellos con diagnósticos alternativos y de acuerdo con las características de la muestra de orina se muestran en la **FIG. 13.** Los pacientes de control con macrohematuria o microhematuria tenían mayores probabilidades de tener resultados de la prueba falsos positivos que aquellos sin macrohematuria o microhematuria (p=0,002) y también había una tendencia hacia pacientes con cálculos, aunque las diferencias en la especificidad según el diagnóstico no eran estadísticamente significativas en general (p=0,12). Hubo cinco pacientes con otros cánceres urológicos; solo uno de estos proporcionó un resultado de la prueba Clasificador-D positivo. Los resultados del ajuste de modelos de regresión logística fueron similares. En un modelo de regresión logística con diagnóstico y macrohematuria o microhematuria, la asociación con el estado de macrohematuria o microhematuria se mantuvo significativo (p=0,006) y, cuando se comparó directamente con la ausencia de diagnóstico, aquellos con cálculos tenían una probabilidad aumentada 2,7 veces de un resultado falso positivo en la prueba (IC del 95 % (de 1,1 a 6,4), p=0,03). La edad no afectó a la especificidad de la prueba.

La macrohematuria o microhematuria detectada en la muestra de orina era el único factor claramente asociado con la sensibilidad del ensayo. El valor predictivo de una prueba positiva en esta cohorte fue del 63 % para aquellos con macrohematuria o microhematuria y del 24 % para aquellos sin estas, lo que refleja en gran medida la mayor prevalencia del CCT en los pacientes con macrohematuria o microhematuria (39 % frente al 6 %).

Hubo 54 con CTT en los que la prueba Clasificador-D fue positiva. Se clasificó a estos pacientes en CCT grave y menos grave usando Clasificador-S. El CCT grave se definió como de estadio ≥1 o de grado 3 en cualquier estadio. A una especificidad del 90 %, el Clasificador-S clasificó correctamente a 32/35 (91 %) de los casos de CCT grave.

Ejemplo 3: Análisis genotípico y fenotípico combinado de pacientes con hematuria I

Este estudio se centra en pacientes que presentan hematuria microscópica asintomática (HA) confirmada que se someten a pruebas clínicas complementarias completas para la investigación de un posible cáncer urotelial (CU). Se inscriben aproximadamente 500 pacientes para participar en el estudio.

Como se usan en el presente documento, a continuación se definen términos de los ejemplos adicionales.

Objetivos

Los objetivos son determinar: (1) la eficacia de un algoritmo genotípico y fenotípico en pacientes que presentan microhematuria en los que está programado efectuar pruebas clínicas complementarias urológicas completas, (2) las características de rendimiento (la sensibilidad, la especificidad, el área bajo la curva ROC, los valores predictivos positivos y negativos) del algoritmo genotípico y fenotípico ÍNDICE G+P para la detección del CU primario en pacientes que presentan hematuria microscópica confirmada y (3) el número de pacientes diagnosticados correctamente como negativos para CU mediante una herramienta genotípica y fenotípica, el ÍNDICE G+P y por tanto, no necesitan cistoscopia exploratoria.

Población de estudio

La población de estudio consiste en pacientes que presentan microhematuria confirmada que cumplen con los requisitos del estudio. Se inscribe a los pacientes a través de sus médicos de cabecera que los derivan a las clínicas de urología.

Consentimiento informado

Se contacta con los pacientes con una cistoscopia exploratoria programada para evaluar su posible participación. Se informa a los pacientes de la naturaleza del estudio y se obtiene su consentimiento. Los pacientes proporcionan información demográfica, ocupacional y de tabaquismo y garantizan que comprenden completamente la información para el paciente y los formularios de consentimiento antes de proporcionar muestras de orina. Los coordinadores del estudio rellenan una página de CRF que detalla las entradas relevantes para el índice genotípico y fenotípico y la transferencia de datos para el análisis.

Criterios de inclusión

Pacientes que se someten a investigación cistoscópica para carcinoma urotelial después de un hallazgo clínico confirmado de hematuria microscópica (mínimo de 3 RBC por campo de alto aumento (HPF)) en 2 o 3 especímenes de orina adecuadamente recogidos [(3)].

Los pacientes se muestran dispuestos a cumplir con los requisitos del estudio.

Los pacientes tienen más de 18 años.

Criterios de exclusión

Historial previo de carcinoma urotelial (CU).

Presenta hematuria macroscópica.
Historial previo (en los últimos 12 meses de un episodio de hematuria macroscópica con diagnóstico confirmado (maligno o de otro tipo),

Índice G+P

Los presentes inventores desarrollaron un nuevo índice, el "índice G+P", que comprende una combinación de datos genotípicos y fenotípicos. El componente genotípico ("G") utiliza información de expresión de biomarcadores de ARN junto con cinco factores clínicos tomados del paciente en el mismo espacio de tiempo (datos fenotípicos, "P") para determinar el riesgo de CU en pacientes con HA.

En todos los pacientes se efectúan pruebas clínicas complementarias convencionales para determinar el resultado clínico verdadero y los resultados de este estudio son resultados simulados basados en los datos clínicos recogidos y los datos genotípicos recogidos de las muestras de orina de los pacientes. No se altera en sí la atención al paciente como resultado de las observaciones del estudio. Los pacientes proporcionan muestras de orina, que se remiten para análisis genético.

Triaje de pacientes

No se efectúan cambios en el tratamiento habitual general para los pacientes que participan en el estudio. Todos los pacientes con una batería completan de pruebas complementarias urológicas programada se someten a las pruebas adecuadas según el tratamiento habitual actual.

Datos del estudio

La demografía y la información de factores de riesgo son datos de entrada para el índice genotípico y fenotípico. Los datos de la patología final (determinada mediante cistoscopia flexible y seguimiento) se cotejan (ϕ) con los resultados y la información demográfica y se someten a análisis estadístico.

Determinación del índice G+P

Usando los conjuntos de datos obtenidos de muestras recogidas de un gran número de sitios en Nueva Zelanda y Australia de aproximadamente 500 pacientes, se desarrolló un modelo de entrenamiento para predecir la probabilidad de "CCT = Sí".

Datos recogidos para las variables usadas en las poblaciones de entrenamiento y validación

Variables fenotípicas

Los hallazgos clínicos son: género, edad, historial de tabaquismo y HFREQNEW (<=1 indicada como baja; >1 indicada como alta).

Variables genotípicas

Las variables genotípicas incluyen la expresión de marcadores de ARN: M1 (= MDK + CDC + IGBP5 - HOXA13) e IL8R. La tabla 7 a continuación muestra estimaciones de los coeficientes de cada uno de los factores en la validación del ÍNDICE G+P:

Tabla 7

Análisis de las estimaciones de máxima probabilidad							
Parámetro		**DF**	**Estimación**	**Error estándar**	**Chi cuadrado de Wald**	**Pr > ChiSq**	**Exp(Est)**
Punto de corte		1	-4,8445	0,6532	55,0092	<0,0001	0,008
género	**2**	1	-1,8544	0,6750	7,5484	0,0060	0,157
tabaquismo	**2**	1	-0,9049.	0,4537	3,9775	0,0461	0,405
tabaquismo	**3**	1	-1,3407	0,4936	7,3756	0,0066	0,262
HFREQNEW	**Baja**	1	-07015	0,3478	4,0681	0,0437	0,496

Análisis de las estimaciones de máxima probabilidad							
Parámetro		**DF**	**Estimación**	**Error estándar**	**Chi cuadrado de Wald**	**Pr > ChiSq**	**Exp(Est)**
M1		1	1,2221	0,1553	61,9328	<0,0001	3,394
IL8R		1	-0,2408	0,1789	1,8112	0,1784	0,786

La FIG. 18 representa curvas ROC para el índice G+P. La **FIG. 18** es una gráfica de la sensibilidad (eje vertical) frente a 1-especificidad (eje horizontal) para los resultados de acuerdo con una realización de la presente invención. Por comparación, una línea diagonal representa el modelo. Los resultados basados únicamente en información fenotípica se muestran como la línea discontinua-de puntos, el resultado basado únicamente en información genotípica se muestra como la línea discontinua y los resultados basados en el índice G+P se muestran como la línea continua. Estos resultados indican que la combinación de información genotípica y fenotípica proporciona una mejora sustancial e inesperada en la predicción del resultado.

Los modelos exploratorios tomaron en consideración siete variables fenotípicas, pero EdadMQ50 mostró un efecto insignificante cuando había datos insuficientes para RBC, por lo que se descartaron estas variables del modelo final. Basándose en el nivel de significación de las cinco variables fenotípicas restantes y los dos marcadores de ARN, se construyó un índice. Usando la relación entre M1, IL8R y CCT (=sí) en el conjunto de datos de entrenamiento, se usó un umbral de 4,5 y 2,5 para M1 e IL8R, respectivamente. Se asignó una puntuación de 5, 4, 3, 2 y 1 para M1, fumadores, varones, IL8R y HFREQ, lo que da como resultado una puntuación del índice en el intervalo de 0 a 15. El algoritmo integrado basado en coeficientes se proporciona a continuación como el índice G+P combinado:

ÍNDICE G+P = (1*HFREQ+3*Género+4*SMK) + (5*M1+2*IL-8)

Las razones de probabilidad para los diferentes factores clínicos que se conservaron en el modelo final se muestran en la **FIG. 19**. Puede interpretarse que una razón de probabilidad tiene un efecto nocivo o protector en el sujeto dependiendo del grado de desviación respecto de 1 (es decir, sin efecto). Las razones de probabilidad cuyos límites de confianza excluyan 1 son estadísticamente significativas. En general, a los factores con una razón de probabilidad mayor (por ejemplo, SMK, género) se les asignan ponderaciones mayores en comparación con los factores con menor razón de probabilidad (por ejemplo, HFREQ).

La tabla de clasificación para el modelo completo se presenta a continuación en la **tabla 8**.

Tabla 8

Tabla de clasificación									
Nivel de probabilidad	**Correcto**		**Incorrecto**		**Porcentajes**				
	Evento	**No evento**	**Evento**	**No evento**	**Correcto**	**Sensibilidad**	**Especificidad**	**Falso POS**	**Falso NEG**
0,000	66	0	415	0	13,7	100,0	0,0	86,3	.
0,050	61	245	170	5	63,6	92,4	59,0	73,6	2,0
0,100	54	319	96	12	77,5	81,8	76,9	64,0	3,6
0,150	47	358	57	19	84,2	71,2	86,3	54,8	5,0
0,200	45	376	39	21	87,5	68,2	90,6	46,4	5,3
0,250	40	381	34	26	87,5	60,6	91,8	45,9	6,4
0,300	37	391	24	29	89,0	56,1	94,2	39,3	6,9
0,350	36	398	17	30	90,2	54,5	95,9	32,1	7,0
0,400	35	402	13	31	90,9	53,0	96,9	27,1	7,2
0,450	35	403	12	31	91,1	53,0	97,1	25,5	7,1
0,500	31	407	8	35	91,1	47,0	98,1	20,5	7,9
0,550	29	408	7	37	90,9	43,9	98,3	19,4	8,3
0,600	29	408	7	37	90,9	43,9	98,3	19,4	8,3
0,650	28	409	6	38	90,9	42,4	98,6	17,6	8,5
0,700	24	409	6	42	90,0	36,4	98,6	20,0	9,3
0,750	20	409	6	46	89,2	30,3	98,6	23,1	10,1
0,800	12	410	5	54	87,7	18,2	98,8	29,4	11,6
0,850	11	411	4	55	87,7	16,7	99,0	26,7	11,8
0,900	6	412	3	60	86,9	9,1	99,3	33,3	12,7
0,950	1	413	2	65	86,1	1,5	99,5	66,7	13,6
1,000	0	415	0	66	86,3	0,0	100,0	.	13,7

Estudio de validación preliminar del índice G+P

Para evaluar adicionalmente el uso del índice G+P, se llevó a cabo otro estudio. Basándose en la significación estadística de diversos marcadores clínicos y de ARN, se construyó un índice. Había 98 sujetos de los que se disponía de su estado de CCT (sí o no), así como de las variables del ÍNDICE G+P.

Se asignó una puntuación de 5, 4, 3, 2 y 1 para M1 (pruebas genéticas), fumadores, varones, IL8R y HFREQ, lo que da como resultado una puntuación del índice en el intervalo de 0 a 15. El número de auténticos positivos y auténticos negativos fue de 6 y 84, respectivamente. De manera similar, el número de falsos positivos y falsos negativos fue de 5 y 3, respectivamente. Por tanto, la precisión general del índice propuesto fue de 0,92.

Implicaciones y seguimiento basado en el índice G+P

En caso de que el resultado del ÍNDICE G+P indique un "alto riesgo" de CU definido como **una puntuación de** 11-15 o más, se da prioridad al paciente para someterse a una cistoscopia flexible y a una ecografía abdominal, según esté clínicamente indicado.

En caso de que el resultado del ÍNDICE G+P indique un "riesgo moderado" de CU, definido como una puntuación de 6-10, se revisa al paciente y se efectúa un seguimiento conforme a la práctica clínica. Ha de valorarse el uso de citología, uretroscopia y/o un escáner TC.

En caso de que el resultado del ÍNDICE G+P indique un "bajo riesgo" de CU, definido como una puntuación de 0-5, se empleará el tratamiento habitual normal para el paciente y se le asignará a la lista de espera correspondiente.

Referencias

Las siguientes referencias se refieren a la divulgación anterior.

Agresti, A. y Coull, B. A. (1998) *Approximate is better than "exact" for interval estimation of binomial proportions*. *The American Statistician*, 52, 119-126.

Altekruse SF, Kosary CL, Krapcho M, Neyman N, Aminou R, Waldron W, Ruhl J, Howlader N, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Cronin K, Chen HS, Feuer EJ, Stinchcomb DG, Edwards BK (eds). *SEER Cancer Statistics Review*, 1975-2007, National Cancer Institute. Bethesda, Maryland, http://seer.cancer.gov/csr/1975_2007/, basado en el envío de datos SEER de noviembre de 2009, publicado en el sitio web de SEER, 2010.

Brown, L. D. Cai, T. T. y DasGupta, A. (2001) *Interval estimation for a binomial proportion*. *Statistical Science*, 16, 101-133.

Buckhaults *et al.*, Cancer Research 61:6996-7001 (2002) describe ciertos genes secretados y de la superficie celular expresados en tumores colorrectales.

Byrd, R. H., Lu, P., Nocedal, J. y Zhu, C. (1995) *A limited memory algorithm for bound constrained optimization*. *SIAM J. Scientific Computing*, 16, 1190-1208.

Dalgaard (2008). "*Introductory Statistics with R*, 2.ª edición"; Peter Dalgaard, Capítulo 13 (2008), Springer, ISBN 978-0-387-79053-4.

DeLong, E. R., D. M. DeLong y D. L. Clarke-Pearson. 1988. *Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach*. *Biometrics* 44: 837-845.

Gottschalk, P.G. y Dunn, J.R. (2005) *The five-parameter logistic: A characterization and comparison with the four-parameter logistic*. *Analytic Biochemistry*, 343, 54-65. Doi:10.1016/j.ab.2005.04.035

Grossfield G, Wolf J, Litwan M, Hricak H, Shuler C, Agerter D, *et al. Asymptomatic microscopic hematuria in adults*: *summary of AUA best practice policy recommendations*. *AFP* 2001:63:1145-54.

Hall *et al.*, (*Laryngoscope* 113(1):77-81 (2003) (PMID: 12679418) (Resumen) describe el valor predictivo de la tiroglobulina sérica en el cáncer de tiroides.

Hoerl, A.E. (1962) *Application of ridge analysis to regression problems*. *Chemical Engineering Progress*, 58, 54-59.

Holyoake A, O'Sullivan P, Pollock R, Best T, Watanabe J, Kajita Y, *et al. Development of a multiplex RNA urine test for the detection and stratification of transitional cell carcinoma of the bladder*. *Clin Cancer Res*. 1 de febrero de 2008;14(3):742-9.

Hotte *et al.*, (*AJ. American Cancer Society* 95(3):507-512 (2002) describe la osteopontina plasmática como una proteína detectable en fluidos corporales humanos y se asocia con ciertas neoplasias malignas.

Kim *et al.*, (*JAMA* 287(13):1671-1679 (2002) describe la osteopontina como un biomarcador diagnóstico potencial para el cáncer de ovario.

Koopman *et al.*, (*Cancer Epidemiol*. *Biomarkers Prey* 13(3):487-491 (2004) (resumen) describe la osteopontina como un biomarcador para el adenocarcinoma pancreático.

Kuo *et al*. (*Clin. Chim. Acta.* 294(1-2):157-168 (2000) (resumen) describe la metaloproteinasa de matriz-2 y -9 sérica en pacientes infectados por HCF y VHB.

Leman *et al.*, (*Urology*, 69(4) 714-20 (2007) (resumen) describe EPCA-2 como marcador sérico para el cáncer de páncreas.

Loo R, Lieberman S, Slezak J, Landa H, Mariani A, Nicolaisen G, Aspera A y Jaconsen S: *Stratifying risk of urinary tract malignant tumors in patients with asymptomatic microscopic hematuria*. *Mayo Clin Proc*. 2013, 88(2); 129-138.

Lunn, D., Spiegelhalter, D., Thomas, A. y Best, N. (2009) *The BUGS project: Evolution, critique and future directions (with discussion)*, *Statistics in Medicine* 28: 3049-3082.

Marchi *et al.*, (*Cancer* 112, 1313-1324 (2008) (resumen) describe la proapolipoproteína A1 como marcador sérico

de metástasis cerebral en pacientes con cáncer de pulmón.
Martin *et al.*, (*Prostate Cancer Prostatic Dis.* 9 de marzo de 2004 (PMID: 15007379) (resumen) describe el uso de calicreína 2 humana, antígeno específico de próstata (PSA) y PSA libre como marcadores para la detección del cáncer de próstata.
Mazzaferri *et al.*, (*J. Clin. Endocrinol. Metab.* 88(4): 1433-1441 (2003) (resumen) describe la tiroglobulina como método de control potencial para pacientes con carcinoma de tiroides.
McDonald M, Swagerty D, Wetzel L: *Assessment of Microscopic hematuria in adults*. *AFP* 2006 73:10.
Melle *et al.*, (*Clin. Chem.* 53(4), 629-635 (2007) (resumen) describe HSP27 como un marcador sérico para el adenocarcinoma pancreático.
Moré, J.J. (1978) *The Levenberg-Marquardt algorithm: implementation and theory*, en *Lecture Notes in Mathematics 630: Numerical Analysis*, G.A. Watson (Ed.), Springer-Verlag: Berlín, págs. 105-116.
Nelder, J. A. y Mead, R. (1965) *A simplex algorithm for function minimization. Computer Journal* 7, 308-31.
O'Sullivan P, Sharples K, Dalphin M *et al*: *A Multigene Urine Test for the Detection and Stratification of Bladder Cancer in Patients Presenting with Hematuria*. *J Urol* 2012, Vol 188 n.º 3; 746
Pellegrini *et al.*, (*Cancer Immunol. Immunother.* 49(7):388-394 (2000) (resumen) describe la medición del antígeno carcinoembrionario soluble y de TIMP 1 como marcadores del cáncer colorrectal preinvasivo.
(*R Development Core Team* (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Viena, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/)
Richards, F. J. (1959) *A flexible growth function for empirical use*. *Journal of Experimental Botany*. 10, 290-300.
Rudland *et al.*, (*Cancer Research* 62: 3417-3427 (2002) describe la osteopontina como una proteína asociada con la metástasis en el cáncer de mama en humanos.
Sing *et al* (2009). (Tobias Sing, Oliver Sander, Niko Beerenwinkel y Thomas Lengauer (2009). *ROCR: Visualizing the performance of scoring classifiers*. Paquete R versión 1.0-4. http://CRAN.R-project.org/package=ROCR).
Speiss, A.-N., Feig, C. y Ritz, C. (2008) *Highly accurate sigmoidal fitting of realtime RCR data by introducing a parameter for asymmetry*. *BMC Bioinformatics*, 9, 221. Doi:10.1186/1471-2105-9-211.
Sprenger H, Lloyd AR, Lautens LL, Bonner TI, Kelvin DJ. *Structure, genomic organization, and expression of the human interleukin-8 receptor B gene. J Biol Chem.* 15 de abril de 1994;269(15): 11065-72.
StataCorp. 2009. Stata Statistical Software: Release 11. College Station, TX: StataCorp LP.
Schwartz G: *Proper evaluation of asymptomatic microscopic hematuria in the era of evidence-based medicine-progress is being made*. *Mayo Clin Proc*. 2013, 88(2); 123-125.
Trivedi C y Messing EM: *Commentary: the role of cytologic analysis of voided urine in the work-up of asymptomatic microhematuria*. *BMC Urology* 2009, 9:13
Tsigkou *et al.*, (*I Clin Endocrinol Metab*, 92(7) 2526-31 (2007) (resumen) describe la inhibina total como un potencial marcador sérico para el cáncer de ovarios.
Tsigkou *et al.*, (*I Clin Endocrinol Metab*, 92(7) 2526-31 (2007) (resumen) describe la inhibina total como un potencial marcador sérico para el cáncer de ovarios.
Venables, W. N. y Ripley, B. D. (2002). *Modern Applied Statistics with S*. Cuarta edición. Springer, Nueva York. ISBN 0-387-95457-0).
Whitley *et al*., (*Dim Lab. Med.* 24(1):29-47 (2004) (resumen) describe la tiroglobulina como marcador sérico para el carcinoma de tiroides.
Yoshikawa *et al.*, (*Cancer Letters*, 151: 81-86 (2000) describe el inhibidor tisular de la metaloproteinasa-1 de matriz en el plasma de pacientes con cáncer gástrico.

Ejemplo 4: Triaje de pacientes que presentan hematuria usando el índice G+P II

El ÍNDICE G+P indica un positivo cuando adopta valores en el intervalo de 11 a 15.

Definiciones

Como se usan en el presente documento, se usan las siguientes definiciones en este y los siguientes ejemplos.

"AMH" significa microhematuria asintomática;
"AUA" significa la American Urological Association;
"ABC" significa área bajo la curva;
"IC" significa intervalo de confianza;
"TC" significa tomografía computarizada;
"ELISA" significa ensayo inmunosorbente ligando a enzimas;
"FISH" significa fluorescencia con hibridación *in situ*;
"HPF" significa campo de alto aumento;
"logOR" significa el logaritmo de la razón de probabilidad;
"Hfreq" significa la frecuencia diaria media de hematuria durante el episodio de hematuria más reciente;
"ISUP" significa la International Society of Urological Pathology;
"IRM" significa formación de imágenes por resonancia magnética;
"NPV" valor predictivo negativo;
"OR" significa razón de probabilidad;
"QC" significa control de calidad;

"QoL" significa calidad de vida;
"Fenotípico" se usa para definir las características del pronóstico clínico y para distinguirlas de los biomarcadores basados en la expresión génica que se han definido ampliamente como variables "genotípicas".
"RBC" significa glóbulos rojos;
"ROC" significa curva operativa de receptor;
"RT-qPCR" significa reacción en cadena de la polimerasa cuantitativa de retrotranscripción;
"STARD" significa estándares para comunicar la precisión diagnóstica;
"CU" significa carcinoma urotelial;
"OMS" significa Organización Mundial de la Salud.

Introducción

La hematuria, que con frecuencia se asocia con causas tales como el agrandamiento benigno de la próstata, infección o cálculos renales, pero también es un síntoma del carcinoma urotelial (CU), tiene una incidencia estimada de entre un 1 y un 22 % de los pacientes en una población general [1,2]. La (macro)hematuria macroscópica se caracteriza por un cambio de color visible en la orina de los pacientes, mientras que la (micro)hematuria microscópica se define de manera más precisa como la presencia de ≥3 glóbulos rojos por campo de alto aumento (RBC/HPF) en tres muestras de orina recogidas de manera sucesiva [2]. Se ha comunicado que la prevalencia general de CU en pacientes con microhematuria es de aproximadamente un 4 %, mientras que varios estudios han demostrado de manera fiable que la prevalencia del CU es mucho mayor en pacientes con macrohematuria, en el intervalo de aproximadamente el 12-23 % [2-6], aunque se somete a evaluación urológica a cuatro veces más pacientes con microhematuria que con macrohematuria [7]. De forma notable, dado que los cambios recientes en las guías de la American Urological Association (AUA) [2] han visto reducido el umbral para microhematuria asintomática (AMH) a ≥3 RBC/HPF en una sola muestra y se han propuesto umbrales incluso más bajos (≥1 RBC/HPF) [8], se espera un aumento consiguiente en el número de pacientes con hematuria que se someterán a pruebas complementarias urológicas para investigar un potencial CU y un aumento correspondiente en la carga clínica y financiera general de estos pacientes en los sistemas sanitarios.

Dichas derivaciones relacionadas con la hematuria representan una carga clínica significativa para los urólogos, ya que todos los pacientes han de someterse a una batería completa de pruebas complementarias que con frecuencia proporcionan un diagnóstico no concluyente. Además, las pruebas diagnósticas existentes, muchas de las cuales son invasivas o tienen altas cargas de radiación, pueden tener un efecto perjudicial para la calidad de vida (QoL) del paciente, especialmente si el paciente se somete a cistoscopias sucesivas según figura en las guías actuales [2]. Se ha informado que las cistoscopias efectuadas sin profilaxis con antibióticos, un 22 % de los pacientes tuvieron bacteriuria asintomática y un 1,9 % de los pacientes desarrolló infección del tracto urinario (ITU) con fiebre antes de 30 días [9]. Otros estudios también han informado de una alta prevalencia de macrohematuria, dolor al miccionar y disfunción eréctil transitoria en hombres después de la cistoscopia [10,11].

Los sistemas sanitarios también soportan una carga financiera significativa como resultado de someter a los pacientes con hematuria a una batería completa de pruebas complementarias urológicas [12,13] y se ha concluido que la citología de orina representa un coste que no ofrece un beneficio diagnóstico significativo [14-16]. Por consiguiente, la integración de una prueba no invasiva y precisa a las pruebas clínicas complementarias principales de los pacientes que presentan hematuria permite a los médicos efectuar un triaje eficaz de los pacientes con hematuria, reduciendo de este modo el número de pacientes que se somete a una batería completa de pruebas complementarias urológicas y a cistoscopia exploratoria para CU y ofrece beneficios significativos, tanto para los pacientes como para los sistemas sanitarios [15-19].

Como factores de riesgo para el CU en pacientes con hematuria se encuentran varias características de pronóstico clínico que incluyen la edad, el género, el historial de tabaquismo y el grado de la hematuria [3,20-22]. Recientemente, varios grupos han tratado de desarrollar modelos basados en características de pronóstico clínico para predecir el riesgo de CU en pacientes con hematuria [20-22], pero desde un punto de vista crítico, estos modelos tienen una precisión limitada y se han centrado principalmente en la detección de pacientes con CU en lugar de descartar pacientes que no tienen la enfermedad. Estos modelos basados en detección han sido por tanto insuficientes para identificar de manera fiable a pacientes con enfermedad durante una evaluación primaria, incluso usándolos en combinación con una citología de orina [20-22].

A pesar de la mayor incidencia de CU en pacientes que presentan macrohematuria, varios estudios demuestran que no hay diferencias significativas en la distribución del CU según el grado y el estadio en pacientes que presentan microhematuria en comparación con aquellos que presentan macrohematuria [5,23-25]. Por lo tanto, la AUA recomienda que todos los pacientes con macrohematuria o AMH sean derivados a un urólogo para la elaboración de pruebas complementarias urológicas completas, ya que la gravedad de la hematuria no es lo suficientemente predictiva de la presencia de CU [2]. Sin embargo, dado que los pacientes con hematuria pueden someterse únicamente a un análisis de orina limitado en una evaluación primaria, que consiste en citología y en algunos casos estudios por imagen, tal como ecografía, normalmente es necesaria la elaboración de pruebas complementarias urológicas completas para detectar o descartar de manera concluyente el CU. Aunque en las guías actuales se especifica la citología de orina y se usa rutinariamente en pacientes con sospecha de CU, los resultados de la

citología son con frecuencia no concluyentes, con hallazgos atípicos o sospechosos y también adolece de un bajo rendimiento diagnóstico causado por un riesgo relativamente elevado de resultados falsos negativos para pacientes con hematuria relacionada con CU [2,26,27]. Por consiguiente, puede ser difícil descartar las causas benignas de la hematuria, ya sea macrohematuria o AMH, durante una evaluación primaria, especialmente si la hematuria relacionada con CU es intermitente y parece resolverse después del tratamiento para una causa benigna [12].

Diversos estudios basados en genes se han propuesto elaborar perfiles de biomarcadores de orina en pacientes con CU y estos biomarcadores pueden ser útiles por méritos propios para detectar la enfermedad [28,29]. También existe la oportunidad de efectuar un triaje de pacientes basándose en sus características clínicas y el perfil de expresión génica. La combinación de las pruebas de ensayo inmunosorbente ligado a enzimas (ELISA) NMP22 o de un panel de genes marcadores con características clínicas ha demostrado mejorar la precisión diagnóstica en comparación con solo las características clínicas, pero estos modelos combinados aún no han proporcionado avances significativos en la precisión diagnóstica general, especialmente cuando se pretende identificar a pacientes de bajo riesgo [30,31]. Sin embargo, se considera que la incorporación de factores clínicos y la expresión de genes específicos en un algoritmo combinado es clave para proporcionar la mejor orientación para diagnosticar y controlar a pacientes con hematuria o CU [32].

Cxbladder™ Detect (Pacific Edge Ltd., Dunedin, Nueva Zelanda), una prueba multigénica efectuada en orina no fraccionada, ha demostrado previamente ser más sensible que la citología de orina y NMP22 para detectar el CU en pacientes con macrohematuria [33] y más preciso que la citología de orina, NMP22 y la fluorescencia con hibridación *in situ* (FISH) en un análisis comparativo (Kasabov, Darling, Breen, ***et al.***, observaciones no publicadas). Cxbladder Detect emplea la tecnología de reacción en cadena de la polimerasa cuantitativa de retrotranscripción (RT-qPCR) para cuantificar cinco marcadores de ARNm, cuatro marcadores que se sobreexpresan en CU junto con un quinto marcador que está elevado en las afecciones inflamatorias no malignas y ofrece un alto nivel de especificidad y sensibilidad cuando se usa para detectar el CU en pacientes que presentan hematuria [33]. Se barajó la hipótesis de que un modelo integrado que combina biomarcadores genéticos de alto rendimiento con variables fenotípicas recogidas de muestras del paciente proporcionará una resolución clínica superior usando alta sensibilidad (es decir, una baja probabilidad de que un paciente con CU reciba un resultado falso negativo), un alto valor predictivo negativo (es decir, es verdadera una alta proporción de todos los resultados negativos) y una elevada tasa de pruebas negativas para permitir el triaje preciso de pacientes que tienen una baja probabilidad de CU. Cuando se combinan estas variables genotípicas y fenotípicas en un nuevo modelo de segregación permite que se identifiquen y se sometan a triaje pacientes con hematuria que tienen bajas probabilidades de CU, en contraposición con someterlos a pruebas complementarias urológicas completas.

Métodos

Selección de pacientes

Se ha analizado una muestra prospectiva de 695 pacientes, donde el verdadero resultado clínico se determinó usando una evaluación clínica convencional. La muestra del estudio consiste en una cohorte inicial de pacientes con hematuria que otorgaron consentimiento y proporcionaron una muestra como se ha descrito anteriormente [33], donde se inscribió de manera prospectiva a una serie consecutiva de 517 pacientes con historial reciente de macrohematuria, edad ≥45 años y sin antecedentes de CU, en nueve clínicas de urología de Australia y Nueva Zelanda. Se efectuó un seguimiento de estos pacientes durante tres meses para determinar el estado de CU o de un diagnóstico alternativo después de un análisis multigénico de muestras de orina, basándose el diagnóstico positivo para CU en el aspecto en la cistoscopia y el examen histopatológico. El estadio de la enfermedad se clasificó según los criterios de estadificación de TNM determinados mediante estudios de patología y de diagnóstico por imagen y el grado del tumor se clasificó según la práctica de anatomopatología local, usando la clasificación consenso de 1998 de la Organización Mundial de la Salud (OMS)/International Society of Urological Pathology (ISUP) [34].

Posteriormente se inscribieron cohortes adicionales de 94 y 84 que se sometieron a pruebas urológicas después de un evento de macrohematuria de dos centros en Nueva Zelanda entre marzo de 2012 y abril de 2013 y se incluyeron en el desarrollo de los modelos. Los centros participantes se seleccionaron basándose en su experiencia previa participando en el estudio inicial y la voluntad de evaluar el producto Cxbladder Detect con configuraciones clínicas individuales.

Se recogió prospectivamente a un conjunto de prueba adicional de 45 pacientes que presentan hematuria y se usaron para la validación adicional del ÍNDICE G+P, como se expone a continuación.

Los criterios de elegibilidad fueron similares a los de [33], salvo por que eran elegibles para la inscripción pacientes con una edad ≥18 años y aquellos que se habían sometido previamente a cistoscopia para investigar el CU que habían tenido un resultado negativo. Además, como en [33], se excluyeron aquellos pacientes que mostraron síntomas de ITU o de cálculos de vejiga o renales.

La aprobación ética para este estudio fue concedida por todos los centros participantes y se obtuvo consentimiento informado de todos los pacientes que proporcionaron muestras.

Recogida y evaluación de muestras de orina

Para proporcionar datos de expresión génica, se recogió una sola muestra de orina del chorro medio de los participantes usando el sistema de muestreo de Pacific Edge. El análisis multigénico de las muestras de todos los estudios se efectuó de acuerdo con el procedimiento operativo estándar, empleado para la prueba multigénica comercial Cxbladder Detect. Todas las muestras de orina (4,5 ml) de la cohorte inicial se recogieron en una clínica antes de la cistoscopia y se transfirieron a un líquido estabilizador mediante aspiración por vacío y se remitieron a Pacific Edge antes de 48 horas. Posteriormente se almacenaron las muestras a -80 °C hasta que se necesitaron para el análisis por lotes. Las muestras de las cohortes posteriores se recogieron de un modo similar, pero se enviaron a Pacific Edge a temperatura ambiente y se procesaron antes de 7 días desde la recogida de la muestra de acuerdo con los límites de control de calidad (QC) revisados y las pruebas de tolerancia efectuadas en el laboratorio de diagnósticos de Pacific Edge.

Análisis estadístico

Se usó regresión logística de una variable para estimar los coeficientes del logaritmo de la razón de probabilidad (logOR) sin ajustar (en bruto) para cuatro variables fenotípicas binarias asociadas con CU: la edad, el sexo, el historial de tabaquismo y la frecuencia diaria media de hematuria durante el episodio de hematuria más reciente del paciente (Hfreq; véase la **tabla 9**).

Tabla 9: Definiciones de las variables fenotípicas binarias asociadas con el CU y sus puntuaciones correspondientes

Parámetro fenotípico	**Puntuación**	
	0	**1**
Género	Femenino	Masculino
Edad	<60 años	≥ 60 años
Historial de tabaquismo	Nunca ha fumado	Fumador o exfumador
Hfreq	≤ 1 episodio/día	>1 episodio/día

Se usó regresión logística de múltiples variables en las cuatro variables fenotípicas para generar coeficientes de logOR ajustados en el modelo fenotípico (ÍNDICE P).

El ÍNDICE G se desarrolló usando regresión logística para determinar la asociación entre el CU y las concentraciones de ARNm para los cinco genes de Cxbladder® Detect (*IGFBP5, HOXA13, MDK, CDK1* y *CXCR2*) en las muestras de orina. Se generó un modelo genotípico-fenotípico (ÍNDICE G+P) de múltiples variables usando una combinación de las nueve variables del ÍNDICE G y el ÍNDICE P. Estos modelos lineales determinaron el logOR a partir del que se obtuvo la probabilidad de que un paciente tuviese CU.

El rendimiento relativo de cada modelo se ilustró en curvas operativas de receptor (ROC) que representan la tasa de falso positivo frente a la tasa de auténtico positivo cuando se evalúa el CU, determinado por cada modelo. El área bajo la curva (ABC) se usó para comparar la eficacia relativa de cada modelo, considerándose óptima una ABC próxima a 1.

Para reducir el potencial sesgo cuando se efectúan la estimación y la predicción del modelo en el mismo conjunto de datos, se calculó una ABC con sesgo corregido para cada uno de los tres modelos de regresión logístico usando remuestreo con reposición [35]. La diferencia entre la ABC nominal para la muestra original y la ABC media de las muestras con reposición es una estimación del sesgo de la muestra y las ABC nominales se ajustaron de manera correspondiente. También se obtuvieron las estimaciones con reposición de los intervalos de confianza (CI) con el sesgo corregido [36].

Además, un criterio de diseño para este ensayo clínico fue que las características de rendimiento de cada modelo deben superar un NPV umbral de 0,97, con una sensibilidad tan alta como sea posible con la restricción adicional de tener una elevada tasa de prueba negativa. La tasa de prueba negativa se selecciona para proporcionar una elevada resolución clínica cuando se somete a triaje a pacientes que presentan hematuria que tienen bajas probabilidades de tener CU. Se efectuaron comparaciones entre el ÍNDICE G, el ÍNDICE P y el ÍNDICE G+P y se determinó el rendimiento de cada modelo en términos de sensibilidad y NPV con una tasa de pruebas negativas lo suficientemente elevada como para proporcionar una herramienta eficaz para el triaje de pacientes con hematuria que tienen bajas probabilidades de CU.

Resultados

Demografía de la muestra

De los 695 pacientes con macrohematuria registrados entre las tres cohortes, se consideró que 23 no eran elegibles

y se excluyeron muestras de 85 pacientes más después de la inscripción debido a la ausencia de datos suficientes o de que las muestras no cumpliesen los estándares de QC (véase la **FIG. 20A**). En total, se dispuso de 587 pacientes para elaborar el modelo que comprendían 72 muestras positivas para CU y 515 negativas para CU.

De las 45 muestras de pacientes con microhematuria proporcionadas, 40 fueron adecuadas para análisis, considerándose que 5 pacientes no eran elegibles y se excluyeron del análisis (véase la **FIG. 20B**). Los 45 pacientes se habían sometido a una evaluación urológica completa y la verdad clínica se confirmó como negativo para CU. Los datos demográficos completos para ambas poblaciones de muestra se presentan en la **tabla 10**.

Tabla 10: Demografía de la población de muestras para pacientes con macrohematuria y microhematuria con datos completos

Parámetro		**Pacientes con macrohematuria (N=587), n (%)**	**Pacientes con microhematuria (N=40), n (%)**
Edad, años	0-49	65 (11,1)	21 (52,5)
	50-59	111 (18,9)	
	60-69	145 (24,7)	19 (47,5)
	70-79	175 (29,8)	
	80-100	91 (15,5)	
Género	Femenino	113 (19,3)	25 (62,5)
	Masculino	474 (80,7)	15 (37,5)
Historial de tabaquismo	Nunca ha fumado	246 (41,9)	25 (62,5)
	Fumador o exfumador	341 (58,1)	15 (37,5)
Hfreq (episodios/día)	≤1	332 (56,6)	40 (100)
	>1	255 (43,4)	-
Estadio del tumor	Normal	515 (87,7)	40 (100)
	T1	16 (2,7)	-
	T2	11 (1,9)	-
	T3	2 (0,3)	-
	Ta	40 (6,8)	-
	Tis	3 (0,5)	-

Relación entre las variables fenotípicas y el riesgo de CU en pacientes con macrohematuria

Los análisis de regresión logística de una variable sin ajustar de cada una de las cuatro variables fenotípicas binarias indicaron que la edad ≥60 años, género masculino, historial de tabaquismo y elevada frecuencia de macrohematuria se asociaban con un riesgo aumentado de CU (**tabla 11**).

Tabla 11: OR ajustadas y no ajustadas para CU según los datos fenotípicos y genotípicos para pacientes con hematuria

				OR no ajustada (IC del 95 %)	**OR de la variable P ajustada (IC del 95 %)**	**OR de la variable G+P ajustada (IC del 95 %)**
Variables fenotípicas		Control	CU			
Edad, años	<60	151	11	2,30 (1,22-4,73)	2,24 (1,18-4,65)	1,89 (0,85-4,64)
	≥60	364	61			
Género	Femenino	105	8	2,05 (1,01-4,75)	1,58 (0,76-3,72)	3,03 (1,12-9,36)
	Masculino	410	64			
Historial de tabaquismo	Nunca ha fumado	227	19	2,20 (1,29-3,91)	2,19 (1,27-3,92)	2,67 (1,34-5,64)
	Fumador o exfumador	288	53			
Hfreq (promedio de episodios/día)	≤1	300	32	1,74 (1,06-2,88)	1,80 (1,08-3,00)	1,76 (0,93-3,35)
	>1	215	40			
Variables genotípicas				**OR no ajustada (IC del 95 %)**	**OR de la variable G ajustada (IC del 95 %)**	**OR de la variable G+P ajustada (IC del 95 %)**

(continuación)

	OR no ajustada (IC del 95 %)	**OR de la variable P ajustada (IC del 95 %)**	**OR de la variable G+P ajustada (IC del 95 %)**
IGFBP5	7,34 (4,59-12,33)	2,15 (1,03-4,58)	2,21 (1,03-4,83)
HOXA13	6,27 (3,92-10,34)	0,33 (0,13-0,83)	0,20 (0,07-0,56)
MDK	7,10 (4,73-11,10)	4,76 (1,74-13,62)	8,14 (2,64-26,60)
CDK1	7,80 (5,11-12,39)	3,47 (1,39-9,13)	2,59 (0,98-7,18)
CXCR2	1,69 (1,36-2,10)	0,65 (0,45-0,92)	0,69 (0,47-0,98)

Las OR ajustadas de las variables ÍNDICE P, ÍNDICE G e ÍNDICE G+P son los coeficientes expresados como potencia del ÍNDICE P, el ÍNDICE G y el ÍNDICE G+P, respectivamente.

Los coeficientes del logOR ajustados se calcularon en el modelo de regresión logística multivariable.

ÍNDICE P = -3,78 + 0,81 × Edad + 0,46 × Género + 0,78 × Historial de tabaquismo + 0,59 × Hfreq

donde a cada variable fenotípica se le asigna una puntuación binaria de 0 o 1, como se indica en la tabla 9 y los intervalos de confianza para los coeficientes se presentan en la tabla 11. La estimación con el sesgo corregido para la ABC para el ÍNDICE P es 0,66 (IC del 95 %: 0,55-0,67; FIG. 21).

Relación entre las variables genotípicas y el riesgo de CU en pacientes con macrohematuria

Se estimó el ÍNDICE G mediante regresión logística usando el log de las concentraciones de ARNm de los cinco genes *IGFBP5, HOXA13, MDK, CDK1* y *CXCR2* en muestras de orina para predecir la incidencia de CU.

ÍNDICE G = -6,22 + 0,77 × *IGFBP5* - 1,11 × *HOXA13* + 1,56 × *MDK* + 1,24 × *CDK1* - 0,43 × *CXCR2*

El ÍNDICE G proporciona una ABC con el sesgo corregido de 0,83 (IC del 95 %: 0,74-0,89; FIG. 21).

Relación entre las variables genotípicas y fenotípicas y el riesgo de CU en pacientes con macrohematuria

Posteriormente se combinaron las cinco variables genotípicas continuas con las cuatro variables fenotípicas binarias para estimar el ÍNDICE G+P usando regresión logística de múltiples variables.

ÍNDICE G+P = -8,46 + (0,79 × *IGF* - 1,60 × *HOXA* + 2,10 × *MDK* + 0,95 × *CDC* - 0,38 × *IL8R*) + (0,64 × Edad + 1,11 × Género + 0,98 × Historial de tabaquismo + 0,56 × Hfreq)

El ÍNDICE G+P proporciona una ABC con el sesgo corregido de 0,86 (IC del 95 %: 0,80-0,91).

Comparación entre el ÍNDICE G y el ÍNDICE G+P

Hay solapamiento entre los intervalos de confianza para el ÍNDICE G y el ÍNDICE G+P, por lo que se construyó una versión con reemplazo de una prueba emparejada determinando la diferencia en la ABC para el ÍNDICE G y el ÍNDICE G+P para cada muestra con reemplazo. Se generaron diez mil muestras con reemplazo con un tamaño de muestra de n=587 mediante muestreo aleatorio con reemplazo a partir de las 587 muestras originales disponibles para análisis. El IC al 95 % resultante para la diferencia entre modelos fue de 0,01-0,08. Por tanto, la probabilidad de que la diferencia real entre las dos ABC sea menor de 0,01 es <0,025, lo que indica que hay elevadas probabilidades de que la ABC para el ÍNDICE G+P sea significativamente mayor que la ABC para el ÍNDICE G.

NPV y sensibilidad de los modelos

El ÍNDICE G+P generó un NPV >0,97 a lo largo del intervalo de frecuencias de prueba negativa de 0,2 a 0,7 y prácticamente siempre fue mayor que el NPV para el modelo del ÍNDICE G (**FIG. 22**). El ÍNDICE G+P ofreció características de rendimiento de sensibilidad de 0,95 y NPV de 0,98 cuando la frecuencia de prueba negativa era de 0,4 (**tabla 12**; **FIG. 22**). Por el contrario, el ÍNDICE G logró una sensibilidad de tan solo 0,86 y un NPV de 0,96 cuando la frecuencia de prueba negativa era de 0,4 (**tabla 12**).

Tabla 12: Características de rendimiento de cada modelo cuando se establecen umbrales para diversas

tasas de pruebas negativas

Umbral (logOR)	Frecuencia de prueba negativa (IC del 95 %)	NPV (IC del 95 %)	Sensibilidad (IC del 95 %)	Especificidad (IC del 95 %)
ÍNDICE G+P				
-2,54	0,25 (0,21-0,28)	0,97 (0,92-0,99)	0,93 (0,85-0,98)	0,27 (0,23-0,31)
-2,52	0,38 (0,34-0,42)	0,95 (0,91-0,97)	0,83 (0,74-0,91)	0,41 (0,37-0,45)
-2,39	0,42 (0,37-0,45)	0,95 (0,91-0,97)	0,82 (0,72-0,90)	0,45 (040-0,49)
-1,95	0,51 (0,47-0,54)	0,92 (0,89-0,95)	0,68 (0,56-0,78)	0,53 (0,49-0,57)
-1,93	0,51 (0,46-0,55)	0,92 (0,89-0,95)	0,68 (0,56-0,78)	0,53 (0,49-0,58)
-1,73	0,82 (0,79-0,85)	0,90 (0,87-0,92)	0,32 (0,22-0,43)	0,84 (0,81-0,87)
ÍNDICE G				
-3,46	0,20 (0,17-0,23)	0,94 (0,88-0,97)	0,90 (0,80-0,95)	0,22 (0,18-0,25)
-3,23	0,30 (0,26-0,34)	0,95 (0,91-0,98)	0,89 (0,80-0,95)	0,33 (0,28-0,37)
-3,04	0,40 (0,36-0,44)	0,96 (0,92-0,98)	0,86 (0,77-0,93)	0,44 (0,40-0,48)
-2,86	0,50 (0,46-0,54)	0,97 (0,94-0,98)	0,86 (0,77-0,93)	0,55 (0,51-0,59)
-2,63	0,60 (0,56-0,63)	0,96 (0,94-0,98)	0,82 (0,71-0,90)	0,66 (0,62-0,69)
-2,41	0,70 (0,66-0,73)	0,96 (0,94-0,98)	0,78 (0,65-0,86)	0,77 (0,73-0,80)
ÍNDICE G+P				
-4,02	0,20 (0,17-0,23)	0,97 (0,93-0,99)	0,96 (0,88-0,99)	0,22 (0,19-0,26)
-3,67	0,30 (0,26-0,33)	0,98 (0,94-0,99)	0,94 (0,87-0,99)	0,33 (0,29-0,37)
-3,33	0,40 (0,36-0,44)	0,98 (0,95-1,00)	0,95 (0,86-0,98)	0,45 (0,40-0,49)
-2,99	0,50 (0,46-0,54)	0,98 (0,96-0,99)	0,92 (0,83-0,97)	0,56 (0,52-0,60)
-2,71	0,60 (0,56-0,64)	0,97 (0,95-0,99)	0,86 (0,76-0,93)	0,67 (0,63-0,71)
-2,37	0,70 (0,66-0,73)	0,97 (0,94-0,98)	0,80 (0,70-0,88)	0,77 (0,73-0,80)

Aplicación del ÍNDICE G+P en pacientes con microhematuria

Aunque se ha desarrollado el ÍNDICE G+P usando datos de pacientes con macrohematuria, se evaluó su robustez en 40 muestras adicionales de pacientes con microhematuria (Hfreq=0). Se esperaba una mayor frecuencia de pruebas negativas en una población con microhematuria, ya que la incidencia de CU es menor en esta población y usando una frecuencia de prueba negativa de 0,4, 32 pacientes (80 %) tuvieron una prueba negativa y se efectuará el triaje de manera correcta, por lo que no será necesario efectuar pruebas urológicas completas para determinar el CU.

Análisis

Este estudio define una herramienta clínica que brinda a los profesionales médicos la capacidad de descartar mediante triaje la necesidad de efectuar una batería completa de pruebas complementarias urológicas para detectar CU en pacientes que presentan hematuria. El estudio presenta un modelo genotípico-fenotípico validado internamente, ÍNDICE G+P, con estimaciones del IC basadas en remuestreo con reposición, que ofrece una combinación de alta sensibilidad y alto NPV (es decir, una baja probabilidad de que un paciente con CU tenga un resultado falso negativo y una alta proporción de resultados negativos auténticos) que no se obtiene mediante modelos obtenidos exclusivamente de datos genotípicos o fenotípicos individuales. Esto dota a los profesionales médicos de una oportunidad única para descartar mediante triaje a pacientes con microhematuria y macrohematuria, en particular, identificando pacientes con bajo riesgo de que tengan CU que no requieren pruebas complementarias urológicas completas.

En el contexto de una alta sensibilidad, una elevada tasa de pruebas negativas es una consideración importante para una prueba de descarte mediante triaje eficaz que tenga como objetivo evitar que los pacientes con baja probabilidad de CU se sometan a pruebas clínicas complementarias completas [37]. Por consiguiente, con una frecuencia de pruebas negativas de 0,4, la sensibilidad del ÍNDICE G+P presentado en este documento maximiza tanto la sensibilidad como el NPV (0,95 y 0,98, respectivamente). Esto puede compararse con el mejor ajuste a partir del modelo genotípico publicado en [33] (sensibilidad=0,82; NPV=0,97) y también es comparable con la sensibilidad y NPV tanto de la cistoscopia (sensibilidad=0,89-0,98; NPV=0,99) como en la cistoscopia virtual usando escáneres de tomografía computarizada (TC) o formación de imágenes por resonancia magnética (IRM) (sensibilidad=0,94 y 0,91, respectivamente) [38-40].

Se reconoce que la población de muestra usada para obtener el ÍNDICE G, el ÍNDICE P y el ÍNDICE G+P en este

caso consistía en pacientes con macrohematuria. Sin embargo, la alta sensibilidad de esta prueba con una tasa de pruebas negativas de 0,4 en pacientes con macrohematuria permite la aplicación del ÍNDICE G+P entre poblaciones tanto con macrohematuria como con microhematuria. Suponiendo que los pacientes con y sin UC estén distribuidos de manera similar entre las poblaciones de pacientes con microhematuria y macrohematuria, pero con una prevalencia de CU esperada del 4 % en la población con microhematuria, también puede esperarse un NPV elevado en la población de pacientes con microhematuria.

Al aplicar el ÍNDICE G+P a la población de muestra de pacientes con microhematuria que no tienen CU, se demostró que se debería haber descartado mediante triaje a un 80 % de los pacientes basándose en el resultado. Solo se derivaría a un 20 % para efectuar una batería completa de pruebas urológicas complementarias. Esto contrasta con las guías convencionales que en la actualidad derivarían a todos los pacientes (100 %) con microhematuria que no puedan asignarse a una causa benigna para efectuar pruebas urológicas complementarias completas, lo que causa costes significativos innecesarios y afecta negativamente a la QoL del paciente.

La gravedad de la hematuria se correlaciona con la probabilidad de que un paciente tenga CU, pero no el estadio o el grado de cualquier tumor y aproximadamente un 96 % y un 77-88 % de los pacientes con microhematuria y macrohematuria, respectivamente, derivados a un urólogo no tendrá CU [2-6]. Por lo tanto, evitar pruebas urológicas complementarias innecesarias para los pacientes con hematuria tiene varios beneficios. La cistoscopia puede asociarse con efectos adversos, tales como dolor al miccionar, sangrado, ITU, disfunción sexual masculina y la ansiedad que acompaña a un diagnóstico de CU no definitivo o no confirmado [9-11]. De manera más destacable, esta nueva estrategia tiene el potencial de reducir la carga de recursos y el coste financiero asociado con una batería completa de pruebas urológicas complementarias en pacientes negativos para CU. Por ejemplo, en el Reino Unido, si se evita la cistoscopia en pacientes con hematuria con una citología negativa inicial y/o pruebas de biomarcadores tumorales, se ha estimado que se ahorraría aproximadamente 770$ por paciente (483£ por paciente) evaluado [13]. El ÍNDICE G+P descrito en el presente documento proporciona una alternativa eficaz al uso de la citología de orina cuando se usa en una situación de evaluación primaria. Esto es particularmente relevante en situaciones donde la evaluación primaria se efectúa por médicos de atención primaria.

Basándose en esto, si se asigna un "coste nominal" arbitrario de 4.500$ para cada batería completa de pruebas urológicas complementarias, el coste total de efectuar dichas pruebas en 1.000 pacientes con microhematuria sería de aproximadamente 4,5 millones de dólares. Por el contrario, si se descarta mediante triaje a un 80 % de pacientes con microhematuria usando el ÍNDICE G+P con un coste nominal arbitrario de 2.500$, el coste total directo de las baterías de pruebas urológicas completas para el 20 % restante de los pacientes tendría un coste total de 3,4 millones de dólares. Esto proporciona un ahorro neto inicial de costes directos de aproximadamente 1,1 millones de dólares por cada 1.000 pacientes con microhematuria.

Aunque el algoritmo genotípico desarrollado por O'Sullivan *et al.* [33] comprendía los mismos constituyentes genotípicos que el ÍNDICE G+P presentado en este documento, se calibró el equilibrio entre sensibilidad y especificidad para la detección primaria óptima del CU en pacientes sintomáticos (es decir, que presentan hematuria) que se sometieron a una batería completa de pruebas urológicas complementarias. Por el contrario, el ÍNDICE G+P en este estudio también incorporó variables fenotípicas y se ha optimizado para alta sensibilidad y alto NPV, para descartar a aquellos pacientes con hematuria que no necesitaban una batería completa de pruebas urológicas complementarias para sospecha de CU. No se efectúa ningún intento de definir o seleccionar a pacientes con CU. En cambio, el objetivo es descartar con confianza a aquellos que no tienen CU y por tanto, todos los pacientes no descartados se someterán a una batería completa de pruebas urológicas complementarias.

Aunque varios estudios han tratado anteriormente de desarrollar modelos predictivos que toman en consideración el fenotipo cuando se evalúa el riesgo de CU en pacientes que presentan hematuria, la precisión de los modelos dependientes de fenotipo de manera individual parece estar limitada. Por ejemplo, Loo *et al.* [21] investigaron de manera prospectiva si se podían usar parámetros fenotípicos para identificar pacientes con microhematuria que pueden no haber necesitado una derivación a urología y una batería completa de pruebas complementarias y determinaron que la edad, el género masculino y un diagnóstico reciente de macrohematuria eran factores de predicción significativos de CU. Un historial de tabaquismo y >25 RBC/HPF en un análisis de orina reciente no fueron factores de predicción estadísticamente significativos de CU, de manera aislada, pero incluso cuando se incluyeron en su "índice de riesgo de hematuria" para mejorar la precisión predictiva, este índice dio como resultado una ABC de 0,809 [21]. Curiosamente, las OR fenotípicas en este estudio y las identificadas por Loo *et al.* son comparables, con IC del 95 % solapantes para el historial de tabaquismo y el género y aunque la edad, el género y el historial de tabaquismo tienen ponderaciones similares en cada modelo, la influencia del componente genotípico del ÍNDICE G+P presentado en este documento representa probablemente la mayor ABC [2].

Igualmente, Cha *et al.* [20] informaron que la edad, el historial de tabaquismo y el grado de hematuria, pero no el género, estaban correlacionados de manera significativa con la presencia de CU en pacientes con hematuria asintomática y usaron un modelo de múltiples variables para desarrollar un nomograma formado por datos fenotípicos y de citología de orina para predecir el CU. Como en Loo *et al.* [21], las OR fenotípicas comunicadas son comparables con las comunicadas en el presente documento, pero incluso tras incorporar la citología de orina en el nomograma, la ABC de 0,831 comunicada en [20] fue menor que la del ÍNDICE G+P.

En otro estudio, Tan *et al.* [22] estratificaron de manera retrospectiva a pacientes con hematuria que habían sido derivados a una clínica especializada en urología en grupos de alto y bajo riesgo usando un nomograma obtenido a partir de la edad del paciente, el sexo, el historial de tabaquismo y el grado de hematuria. Aunque las comparaciones con este estudio deben efectuarse con cuidado debido a la elevada proporción de pacientes excluidos debido a la falta de datos (80 de los 405 pacientes), la ABC de 0,804, la sensibilidad de 0,900 y el NPV de 0,953 fueron todos menores que en el ÍNDICE G+P descrito en el presente documento.

También se han efectuado varios intentos de mejorar la precisión de los modelos fenotípicos complementándolos con los resultados de las pruebas de biomarcadores de orina. Cuando se usa el ensayo de proteómica de punto de atención de la matriz de proteínas nucleares NMP22 de manera aislada para detectar el CU, este tiene una sensibilidad de 0,557 y un NPV de 0,968 [17]. Lotan *et al.* [41] publicaron un algoritmo de múltiples variables que comprendía factores fenotípicos, NMP22 y citología de orina con una ABC para predecir el CU de 0,826 que posteriormente se validó retrospectivamente con una ABC de 0,802 [31]. Sin embargo, es importante destacar que este modelo pretendía diferenciar entre pacientes de alto riesgo que tenían y no tenían CU, en contraposición con maximizar la sensibilidad y el NPV para descartar mediante triaje a pacientes con baja probabilidad de CU.

La precisión mejorada obtenida con los algoritmos que comprenden datos tanto genotípicos como fenotípicos se ha demostrado con anterioridad en cáncer de mama, en particular [42-45]. Igualmente, Mitra *et al.* [30] usaron una combinación de marcadores moleculares y de intensidad de tabaquismo para calcular un modelo de múltiples variables que era mejor que los parámetros clínicos y patológicos rutinarios para predecir la supervivencia en pacientes con CU. Sin embargo, el presente estudio es el primero en demostrar que pueden combinarse factores de riesgo fenotípicos con datos genotípicos para aumentar la precisión de un modelo para separar a pacientes con hematuria en categorías que requieren niveles diferenciales de seguimiento urológico y atención clínica en lugar de predicción de supervivencia.

Cuando se combinan datos fenotípicos con datos genotípicos en un modelo, es probable que la resolución de los datos tenga impacto en la precisión del modelo. Por ejemplo, el tabaquismo es un factor de riesgo bien comprendido para CU y se incluye en la mayoría de modelos fenotípicos para detectar el CU. En Cha *et al.* [20], Tan *et al.* [22], Lotan *et al.* [31,41] y el presente estudio, se usaron los discriminantes binarios de no fumador y fumador/exfumador, mientras que Mitra *et al.* [30] calcularon la intensidad del tabaquismo basándose en los años de tabaquismo y el número de cigarros que se fuma al día y Loo *et al.* [21] clasificaron a los fumadores en no fumadores, fumadores pasivos, exfumadores y fumadores. Aunque se sabe que el riesgo de CU aumenta sustancialmente con la exposición al humo del tabaco [46], la definición arbitraria de variables fenotípicas puede limitar la precisión general y la utilidad de los modelos fenotípicos. Por el contrario, no sería inesperada una interacción entre las variables genotípicas y fenotípicas de un paciente. Sin embargo, la combinación del impacto de los factores fenotípicos y las variables genéticas en una sola herramienta mejoró la precisión del modelo descrito en este estudio. También se aplica un principio similar a la descripción del fenotipo de hematuria. Los pacientes que presentan microhematuria o macrohematuria se encuentran esencialmente en un espectro biológico continuo y tienen diferentes probabilidades de tener CU [2-6,21]. Por consiguiente, a pesar de que todos los pacientes con microhematuria en este estudio que tenían una puntuación de Hfreq de 0, la gravedad de su hematuria, en combinación con otros factores fenotípicos, es probable que se haya tenido en cuenta indirectamente en el componente fenotípico del ÍNDICE G+P.

Referencias

Los artículos citados inmediatamente a continuación se refieren al ejemplo 4 y todos se incorporan al presente documento por referencia en su totalidad.

1. Kelly JD, Fawcett DP, Goldberg LC: *Assessment and management of non-visible haematuria in primary care*. *BMJ* 2009, 338:a3021.
2. Davis RE, Jones JS, Barocas DA, Castle EP, Lang EK, Leveillee RJ, Messing EM, Miller SD, Peterson AC, Turk TMT, Weitzel W: *Diagnosis, evaluation and follow-up of asymptomatic microhematuria (AMH) in adults: AUA guideline*. *J Urol* 2012, 188(6 Supl):2473-2481.
3. Sutton JM. *Evaluation of hematuria in adults*. *JAMA* 1990, 263:2475-2480.
4. Khadra MH, Pickard RS, Charlton M, Powell PH, Neal DE: *A prospective analysis of 1,930 patients with hematuria to evaluate clinical practice*. *J Urol* 2000, 163:524-527.
5. Davidson P: *Re-design of a haematuria clinic: Assessment of 2346 haematuria patients*. *J Urol* 2011,185(4S):e495.
6. Price SJ, Shephard EA, Stapley SA, Barraclough K, Hamilton WT: *Non-visible versus visible haematuria and bladder cancer risk: A study of electronic records in primary care. Br J Gen Pract* 2014, 64:e584-e589.
7. Buteau A, Seideman CA, Svatek RS, Youssef RF, Chakrabati G, Reed G, Bhat D, Lotan Y: *What is evaluation of hematuria by primary care physicians? Use of electronic medical records to assess practice patterns with intermediate follow-up*. *Urol Oncol* 2014, 32:128-134.
8. Jimbo M: *Evaluation and management of hematuria*. *Prim Care* 2010, 373:461-472.
9. Herr HW: *The risk of urinary tract infection after flexible cystoscopy in bladder tumor patients who did not receive prophylactic antibiotics*. *J Urol*, 18 de julio de 2014. pii: S0022-5347(14)03963-9.

10. Burke DM, Shackley DC, O'Reilly PH: *The community-based morbidity of flexible cystoscopy*. *BJU Int* 2002, 89:347-349.
11. Stav K, Leibovici D, Goren E, Livshitz A, Siegel YI, Lindner A, Zisman A: *Adverse effects of cystoscopy and its impact on patients' quality of life and sexual performance*. *Isr Med Assoc J* 2004, 6:474-478.
12. Rao PK, Jones JS: *How to evaluate 'dipstick hematuria': What to do before you refer*. *Cleve Clin J Med* 2008, 75:227-233.
13. Rodgers M, Nixon J, Hempel S, Aho T, Kelly J, Neal D, Duffy S, Ritchie G, Kleijnen J, Westwood M: *Diagnostic tests and algorithms used in the investigation of haematuria: systematic reviews and economic evaluation*. *Health Technol Assess* 2006, 10:iii-iv, xi-259.
14. Falebita OA, Lee G, Sweeney P: *Urine cytology in the evaluation of urological malignancy revisited: is it still necessary? Urol Int* 2010, 84:45-49.
15. Feifer AH, Steinberg J, Tanguay S, Aprikian AG, Brimo F, Kassouf W: *Utility of urine cytology in the workup of asymptomatic microscopic hematuria in low-risk patients*. *Urology* 2010, 75:1278-1282.
16. Svatek RS, Hollenbeck BK, Holmang S, Lee R, Kim S, Stenzl A, Lotan Y. *The economics of bladder cancer: Costs and considerations of caring for this disease*. *Eur Urol* 2014, 66:253-262.
17. Grossman HB, Messing E, Soloway M, Tomera K, Katz G, Berger Y, Shen Y: *Detection of bladder cancer using a point-of-care proteomic assay*. *JAMA* 2005, 293:810-816.
18. Friedlander DF, Resnick MJ, You C, Bassett J, Yarlagadda V, Penson DF, Barocas DA: *Variation in the intensity of hematuria evaluation: a target for primary care quality improvement*. *Am J Med* 2014, 127:633-640.
19. Shinagare AB, Silverman SG, Gershanik EF, Chang SL, Khorasani R: *Evaluating hematuria: impact of guideline adherence on urologic cancer diagnosis*. *Am J Med* 2014, 127:625-632.
20. Cha EK, Tirsar LA, Schwentner C, Hennenlotter J, Christos PJ, Stenzl A, Mian C, Martini T, Pycha A, Shariat SF, Schmitz-Drager BJ: *Accurate risk assessment of patients with asymptomatic hematuria for the presence of bladder cancer*. *World J Urol* 2012, 30:847-852.
21. Loo RK, Lieberman SF, Slezak JM, Landa LM, Mariani AJ, Nicolaisen G, Aspera AM, Jacobsen SJ: *Stratifying risk of urinary tract malignant tumors in patients with asymptomatic microscopic hematuria*. *Mayo Clin Proc* 2013, 88:129-138.
22. Tan GH, Shah SA, Ann HS, Hemdan SN, Shen LC, Galib NAFA, Singam P, Kong CHC, Hong GE, Bahadzor B, Zainuddin ZM: *Stratifying patients with haematuria into high or low risk groups for bladder cancer: a novel clinical scoring system*. *Asian Pac J Cancer Prev* 2013, 14:6327-6330.
23. Sultana S, Goodman C, Bryne D, Baxby K: *Microscopic haematuria: urological investigation using a standard protocol*. *Br J Urol* 1996, 78:691-698.
24. Sugimura K, Ikemoto S-I, Kawashima H, Nishisaka N, Kishimoto T: *Microscopic hematuria as a screening marker for urinary tract malignancies*. *Int J Urol* 2001, 8:1-5.
25. Viswanath S, Zelhof B, Ho E, Sethia K, Mills R: *Is routine urine cytology useful in the haematuria clinic? Ann R Coll Surg Engl* 2008, 90:153-155.
26. Steiner H, Bergmeister M, Verdorfer I, Granig T, Mikuz G, Bartsch G, Stoehr B, Brunner A: *Early results of bladder-cancer screening in a high-risk population of heavy smokers*. *BJU Int* 2008, 102:291-296.
27. Yeung C, Dinh T, Lee J: *The health economics of bladder cancer: An updated review of the published literature*. *Pharmacoeconomics* 2014, 32:1093-1104.
28. Holyoake A, O'Sullivan P, Pollock R, Best T, Watanabe J, Kajita Y, Matsui Y, Ito M, Nishiyama H, Kerr N, da Silva Tatley F, Cambridge L, Toro T, Ogawa O, Guilford P: *Development of a multiplex RNA urine test for the detection and stratification of transitional cell carcinoma of the bladder*. Clin Cancer Res 2008, 14:742-749.
29. Sapre N, Anderson PD, Costello AJ, Hovens CM, Corcoran NM: *Gene-based urinary biomarkers for bladder cancer: an unfulfilled promise? Urol Oncol* 2014, 32:48.e9-e17.
30. Mitra AP, Castelao JE, Hawes D, Tsao-Wei DD, Jiang X, Shi SR, Datar RH, Skinner EC, Stein JP, Groshen S, Yu MC, Ross RK, Skinner DG, Cortessis VK, Cote RJ: *Combination of molecular alterations and smoking intensity predicts bladder cancer outcome*. *Cancer* 2013, 119:756-765.
31. Lotan Y, Svatek RS, Krabbe LM, Xylinas E, Klatte T, Shariat SF: *Prospective external validation of model for bladder cancer detection*. *J Urol* 2014, 192:1343-1348.
32. Abogunrin F, O'Kane HF, Ruddock MW, Stevenson M, Reid CN, O'Sullivan JM, Anderson NH, O'Rourke D, Duggan B, Lamont JV, Boyd RE, Hamilton P, Nambirajan T, Williamson KE: *The impact of biomarkers in multivariate algorithms for bladder cancer diagnosis in patients with hematuria*. *Cancer* 2012, 118:2641-2650.
33. O'Sullivan P, Sharples K, Dalphin M, Davidson P, Gilling P, Cambridge L, Harvey J, Toro T, Giles N, Luxmanan C, Felipe Alves C, Yoon HS, Hinder V, Masters J, Kennedy-Smith A, Beaven T, Guilford P: *A multigene urine test for the detection and stratification of bladder cancer in patients presenting with hematuria*. *J Urol* 2012, 188:741-747.
34. Epstein JI, Amin MB, Reuter VR, Mostofi FK: *The World Health Organization/International Society of Urological Pathology consensus classification of urothelial (transitional cell) neoplasms of the urinary bladder*. *Bladder Consensus Conference Committee*. *Am J Surg Pathol* 1998, 22:1435-1448.
35. Efron B, Tibshirani RJ: *An introduction to the bootstrap*. Nueva York: Chapman Hill; 1993.
36. DiCiccio TJ, Tibshirani RJ: *Bootstrap confidence intervals*. *Statist Sci* 1996, 11:189-228.
37. Van't Hoog AH, Cobelens F, Vassall A, Van Kampen S, Dorman SE, Alland D, Ellner J: *Optimal triage test characteristics to improve the cost-effectiveness of the Xpert MTB/RIF assay for TB diagnosis: A decision analysis*. *PLoS ONE* 2013, 8:e82786.
38. Qu X, Huang X, Wu L, Huang G, Ping X, Yan W: *Comparison of virtual cystoscopy and ultrasonography for*

bladder cancer detection: A meta-analysis. *Eur J Radiol* 2010, 80:188-197.
39. Mowatt G, N'Dow J, Vale L, Nabi G, Boachie C, Cook JA, Fraser C, Griffiths TR; *Aberdeen Technology Assessment Review (TAR) Group: Photodynamic diagnosis of bladder cancer compared with white light cystoscopy: Systematic review and meta-analysis. Int J Technol Assess Health Care* 2011, 27:3-10.
40. Blick CG, Nazir SA, Mallett S, Turney BW, Onwu NN, Roberts IS, Crew JP, Cowan NC: *Evaluation of diagnostic strategies for bladder cancer using computed tomography (CT) urography, flexible cystoscopy and voided urine cytology: results for 778 patients from a hospital haematuria clinic*. *BJU Int* 2012, 110:84-94.
41. Lotan Y, Capitanio U, Shariat SF, Hutterer GC, Karakiewicz PI: *Impact of clinical factors, including a point-of-care nuclear matrix protein-22 assay and cytology, on bladder cancer detection*. *BJU Int* 2009, 103:1368-1374.
42. Tyrer J, Duffy SW, Cuzick J: *A breast cancer prediction model incorporating familial and personal risk factors*. *Stat Med* 2004, 23:1111-1130.
43. Dubsky P, Brase JC, Jakesz R, Rudas M, Singer CF, Greil R, Dietze O, Luisser I, Klug E, Sedivy R, Bachner M, Mayr D, Schmidt M, Gehrmann MC, Petry C, Weber KE, Fisch K, Kronenwett R, Gnant M, Filipits M, en representación del Austrian Breast and Colorectal Cancer Study Group (ABCSG): *The EndoPredict score provides prognostic information on late distant metastases in ER+/HER2- breast cancer patients*. *Br J Cancer* 2013, 109:2959-2964.
44. Brentnall AR, Evans DG, Cuzick J: *Distribution of breast cancer risk from SNPs and classical risk factors in women of routine screening age in the UK*. *Br J Cancer* 2014, 110:827-828.
45. Filipits M, Nielsen TO, Rudas M, Greil R, Stöger H, Jakesz R, Bago-Horvath Z, Dietze O, Regitnig P, Gruber-Rossipal C, Müller-Holzner E, Singer CF, Mlineritsch B, Dubsky P, Bauernhofer T, Hubalek M, Knauer M, Trapl H, Fesl C, Schaper C, Ferree S, Liu S, Cowens JW, Gnant M, Austrian Breast and Colorectal Cancer Study Group: *The PAM50 risk-of-recurrence score predicts risk for late distant recurrence after endocrine therapy in postmenopausal women with endocrine-responsive early breast cancer*. Clin Cancer Res 2014, 20:1298-1305.
46. Krabbe LM, Svatek RS, Shariat SF, Messing E, Lotan Y: *Bladder Cancer Risk: Use of the PLCO and NLST to Identify as Suitable Screening Cohort*. Urol Oncol. 16 de julio de 20149): pii: S1078-1439 (14)00215-4.

Conclusiones

En conclusión, el ÍNDICE G+P mostrado en este documento muestra una oportunidad significativa para cambiar la utilidad clínica. El ÍNDICE G+P tiene la capacidad de descartar mediante triaje a pacientes que presentan hematuria en la consulta médica, que tienen bajas probabilidades de CU con una elevada frecuencia general de prueba negativa, un alto nivel de sensibilidad y alto NPV. Este modelo es adecuado para su uso durante la evaluación primaria de pacientes con hematuria para descartar mediante triaje a pacientes que no requieren pruebas complementarias urológicas completas, reduciendo de este modo potencialmente el número de pacientes con hematuria que son derivados al urólogo especialista para evaluar el CU, ayudando a mantener la QoL del paciente y ayudando a reducir los costes relacionados con el diagnóstico.

Ejemplo 5: Triaje de pacientes con hematuria usando el índice G+P III ("G2")

Un uso adicional del ÍNDICE G+P es categorizar pacientes según el riesgo de CU. Posteriormente se usa la categoría de riesgo para dar prioridad a los pacientes para una investigación de seguimiento.

Este ejemplo proporciona un método alternativo para el triaje de pacientes usando un índice, denominado en el presente documento el "índice G2+P". Es similar al índice G+P descrito anteriormente en el ejemplo 4.

Definición de G_2

Este clasificador calcula la siguiente fórmula:

$$M1 = [IGF] - [HOXA] + [MDK] + [CDC];$$

por tanto,

$$\begin{aligned} R = & - 6{,}9802 \\ & + 0{,}2007 * M1 \\ & + 1{,}7893 * [IL8R] \\ & + 0{,}1552 * M1^2 \\ & - 0{,}2882 * [IL8R]^2 \\ & - 0{,}0720 * M1 * [IL8R] \end{aligned}$$

para obtener la puntuación:

$$\text{PUNTUACIÓN} = e^R/(1+ e^R),$$

donde [IGF], [HOXA], [MDK], [CDC] e [IL8R] son los logaritmos de las concentraciones de la muestra para los genes

IGF, HOXA, MDK, CDC e IL8R, respectivamente; '*' es la multiplicación convencional y e = 2,718282... es la base del logaritmo natural (Neperiano).

Una alta PUNTUACIÓN indica la presencia de una mayor probabilidad de CU. A modo de ejemplo, se puede establecer un umbral de 0,12 y declarar una PUNTUACIÓN >= 0,12 como probabilidad elevada de tener CU y las puntuaciones menores de 0,12 como baja probabilidad de tener CU.

Ejemplo 6: Uso concurrente del índice [G1 = P] y G2 para descartar mediante triaje a pacientes que presentan hematuria que tienen bajas probabilidades de tener carcinoma urotelial

A. Pacientes con microhematuria

Conjunto de datos de pacientes

- Se evaluaron 45 muestras de pacientes
- 5 pacientes no tienen estado de tabaquismo y se excluyeron de este análisis
- En todos los pacientes se ha efectuado una batería completa de pruebas urológicas complementarias y ninguno tiene carcinoma urotelial.
- Variables fenotípicas: Género, Edad, Estado de tabaquismo
- La demografía de pacientes se muestra en la tabla 13 a continuación.

Tabla 13

Género	Edad	No fumador	Exfumador	Fumador	
Femenino	Edad <60	10	1	2	
	Edad >=60	9	1	2	
Masculino	Edad <60	3	2	3	
	Edad >=60	3	2	2	

Frecuencia de prueba negativa observada:

Usando una frecuencia de prueba negativa del 40 % como umbral, se muestran los datos en la tabla 14 a continuación.

Tabla 14

TNR 40 %	LR < 0,12	0,12 <= LR < 0,23	0,23 <= LR
Triaje negativo	28	3	1
Triaje positivo	4	2	2

Usando una frecuencia de prueba negativa del 50 % como umbral, se muestran los datos en la tabla 15 a continuación.

Tabla 15

TNR 50 %	LR < 0,12	0,12 <= LR < 0,23	0,23 <= LR
Triaje negativo	28	3	3
Triaje positivo	4	2	0

Usando una frecuencia de prueba negativa del 60 % como umbral se obtienen resultados idénticos.
El grupo de mayor riesgo (varón, fumador o exfumador, edad >= 60) fue positivo para esta clasificación mediante triaje.

Resumen y conclusiones

1. Conforme a las guías clínicas, los 41 pacientes se habrían sometido a una batería completa de pruebas urológicas complementarias.
2. Los 41 se sometieron a una batería completa de pruebas complementarias y se determinó que los 41 no tenían carcinoma urotelial.
3. Con una frecuencia de prueba negativa (TNR del 50 %) se habría descartado a un 85 % de los pacientes con microhematuria y como consecuencia, no se habrían sometido a una batería completa de pruebas urológicas complementarias.
4. Si se hubiese usado triaje con Cxbladder a una TNR del 50 % (índice de triaje de -3.33) se habría descartado mediante triaje a un 85,4 % de los pacientes y por consiguiente, de manera correcta no se les habría sometido a una batería completa de pruebas urológicas complementarias.
5. Si se hubiese usado triaje con Cxbladder a una TNR del 40 % (índice de triaje de -3.0) se habría descartado

mediante triaje a un 80,5 % de los pacientes y por consiguiente, de manera correcta no se les habría sometido a una batería completa de pruebas urológicas complementarias.

B. Pacientes con macrohematuria

Conjunto de datos de pacientes

Se usaron 587 muestras de datos de un ensayo clínico y los ensayos de producto North Shore y CURT. Este conjunto de datos era un subconjunto que consistía en datos completos que contenían la edad, el sexo, el estado de tabaquismo y la frecuencia de la hematuria, así como las concentraciones génicas para IGF, HOXA, MDK, CDC, IL8R.

Se usaron los mismos datos para desarrollar el modelo de triaje por Cxbladder a continuación;

ÍNDICE G+P = -8,46 + 0,79 IGF - 1,60 HOXA + 2,10 MDK + 0,95 CDC - 0,38 IL8R + 0,98 SNS + 0,56 Hfreq + 1,11 Género + 0,64 Edad

Se representó la puntuación del triaje frente a la puntuación diagnóstica de G2 en la FIG. 24. Los umbrales del triaje de -3,33, -2,99 y -2,71 corresponden a las frecuencias de prueba negativa del 40 %, 50 % y 60 %, respectivamente. Las verticales son los umbrales de Cxbladder de 0,12 y 0,23. Los círculos rellenos corresponden a tumores; en verde son Ta y en rojo son todos los demás estadios tumorales. La tabla 16 muestra todas las observaciones clínicas.

Tabla 16

Estadio	Sin cáncer	T1	T2	T2a	T3	Ta	Tis
Recuento	515	16	10	1	2	40	3

Se tomaron en consideración los recuentos en los 4 cuadrantes de la FIG. 24 y se determinaron mediante diversos cortes para el índice de triaje y G2.

El umbral para el triaje por Cxbladder (índice de triaje) de -3 corresponde a una frecuencia de prueba negativa del 50 %. La tabla 17 muestra estos resultados.

Tabla 17

Cuadrante	Umbrales	Control	CCT
Inferior izquierdo	Triaje < -3 Y G2 < 0,12	272	6
Superior izquierdo	Triaje > -3 Y G2 < 0,12	155	7
Inferior izquierdo	Triaje < -3 Y G2 >= 0,12	16	0
Superior derecho	Triaje > -3 Y G2 >= 0,12	72	59

Usando un umbral de triaje por Cxbladder de -3,33 (tasa de prueba negativa del 40 %) con el mismo umbral de G2 de 0,12, se observaron los datos mostrados en la **tabla 18**.

Tabla 18

Cuadrante	Umbrales	Control	CCT
Inferior izquierdo	Triaje < -3,33 Y G2 < 0,12	222	4
Superior izquierdo	Triaje > -3,33 Y G2 < 0,12	205	9
Inferior derecho	Triaje < -3,33 Y G2 >= 0,12	9	0
Superior derecho	Triaje > -3,33 Y G2 >= 0,12	79	59

Resumen y conclusiones

1. El uso de una combinación en serie de Cxbladder Triage (G+P) y Cxbladder Detect (G2) en los mismos pacientes en el mismo intervalo de tiempo proporcionó una segregación exhaustiva de los pacientes en cuatro grupos clínicos clave.
2. Para una población combinada de pacientes que presentan microhematuria y macrohematuria y usando una tasa de prueba negativa del 40 %, se descartó mediante triaje a un total de 235 de los 587 pacientes (40,0 %). Se concluye que estos pacientes no necesitarán pruebas complementarias completas para CU.
3. Para la misma población, el grupo residual correspondiente de 352 (60 %) pacientes se someterá a una batería completa de pruebas urológicas complementarias.
4. Este grupo residual contenía todos los tumores de grado alto y estadio tardío. Estos pacientes no fueron descartados mediante triaje y por consiguiente, se someten correctamente a una batería completa de pruebas urológicas.
5. Se descartará un total de 4 Ta de bajo grado (5,6 % del número total de pacientes) y no se someterá a una batería completa de pruebas complementarias.

6. En caso de que las reglas del triaje permitan descartar del triaje a todos los pacientes por debajo de la tasa de prueba negativa del 40 % y con puntuaciones de G2 < 0,12, se descartó mediante triaje a un total de 13 Ta de bajo grado que no se someterán a una batería completa de pruebas complementarias y se detectarán todos los tumores en estadio tardío y de alto grado y se someterán a una batería completa de pruebas urológicas complementarias.
7. Estas reglas de triaje modificadas también dieron como resultado el descarte mediante triaje de un total de 440 del total de 587 pacientes (75 %).

Ventajas y conclusiones generales

En conclusión, el ÍNDICE G+P mostrado en este documento muestra una oportunidad significativa para cambiar la utilidad clínica. El ÍNDICE G+P tiene la capacidad de descartar mediante triaje a pacientes que presentan hematuria en la consulta médica, que tienen bajas probabilidades de CU con una elevada frecuencia general de prueba negativa, un alto nivel de sensibilidad y alto NPV. Este modelo es adecuado para su uso por médicos de atención primaria para descartar mediante triaje a pacientes que no requieren una batería completa de pruebas urológicas complementarias, reduciendo de este modo el número de pacientes con hematuria que sean derivados a un especialista para una evaluación urológica para CU, lo que ayuda a mantener la calidad de vida del paciente y reduce los costes relacionados con el diagnóstico. Los métodos divulgados proporcionaron una evaluación inesperadamente precisa de la ausencia de necesidad de efectuar una investigación de seguimiento para aquellos con hematuria. Estos representan unos efectos excelentes que podrían no haberse logrado sin usar la divulgación contenida en el presente documento.

Aplicabilidad industrial

Las realizaciones de la presente invención son útiles en los campos de la atención sanitaria y la medicina.

Especialidades técnicas

Los aspectos descritos en el presente documento proporcionan métodos implementados informáticamente altamente precisos, sensibles y específicos para el triaje de pacientes para determinar qué pacientes no requieren procedimientos o seguimiento sustancial a corto plazo. Los métodos mejoran las operaciones computarizadas proporcionando operaciones computarizadas novedosas y no evidentes basadas en el análisis de información genética y fenotípica específica de pacientes con hematuria que proporciona resultados tangibles, útiles y concretos para mejorar la calidad de la atención sanitaria y reducir los costes.

LISTADO DE SECUENCIAS

<110> Pacific Edge Limited

<120> Triaje de pacientes que tienen hematuria asintomática usando biomarcadores genotípicos y fenotípicos

<130> PEBL 1028 WO1

<160> 17

<170> PatentIn versión 3.5

<210> 1
<211> 360
<212> PRT
<213> Ser humano

<400> 1

```
Met Glu Asp Phe Asn Met Glu Ser Asp Ser Phe Glu Asp Phe Trp Lys
1               5                  10                  15
Gly Glu Asp Leu Ser Asn Tyr Ser Tyr Ser Ser Thr Leu Pro Pro Phe
            20                  25                  30
Leu Leu Asp Ala Ala Pro Cys Glu Pro Glu Ser Leu Glu Ile Asn Lys
        35                  40                  45
Tyr Phe Val Val Ile Ile Tyr Ala Leu Val Phe Leu Leu Ser Leu Leu
    50                  55                  60
Gly Asn Ser Leu Val Met Leu Val Ile Leu Tyr Ser Arg Val Gly Arg
65                  70                  75                  80
Ser Val Thr Asp Val Tyr Leu Leu Asn Leu Ala Leu Ala Asp Leu Leu
                85                  90                  95
Phe Ala Leu Thr Leu Pro Ile Trp Ala Ala Ser Lys Val Asn Gly Trp
            100                 105                 110
Ile Phe Gly Thr Phe Leu Cys Lys Val Val Ser Leu Leu Lys Glu Val
        115                 120                 125
Asn Phe Tyr Ser Gly Ile Leu Leu Leu Ala Cys Ile Ser Val Asp Arg
    130                 135                 140
Tyr Leu Ala Ile Val His Ala Thr Arg Thr Leu Thr Gln Lys Arg Tyr
145                 150                 155                 160
Leu Val Lys Phe Ile Cys Leu Ser Ile Trp Gly Leu Ser Leu Leu Leu
                165                 170                 175
Ala Leu Pro Val Leu Leu Phe Arg Arg Thr Val Tyr Ser Ser Asn Val
            180                 185                 190
Ser Pro Ala Cys Tyr Glu Asp Met Gly Asn Asn Thr Ala Asn Trp Arg
        195                 200                 205
Met Leu Leu Arg Ile Leu Pro Gln Ser Phe Gly Phe Ile Val Pro Leu
    210                 215                 220
Leu Ile Met Leu Phe Cys Tyr Gly Phe Thr Leu Arg Thr Leu Phe Lys
225                 230                 235                 240
Ala His Met Gly Gln Lys His Arg Ala Met Arg Val Ile Phe Ala Val
                245                 250                 255
Val Leu Ile Phe Leu Leu Cys Trp Leu Pro Tyr Asn Leu Val Leu Leu
            260                 265                 270
Ala Asp Thr Leu Met Arg Thr Gln Val Ile Gln Glu Thr Cys Glu Arg
        275                 280                 285
Arg Asn His Ile Asp Arg Ala Leu Asp Ala Thr Glu Ile Leu Gly Ile
    290                 295                 300
Leu His Ser Cys Leu Asn Pro Leu Ile Tyr Ala Phe Ile Gly Gln Lys
305                 310                 315                 320
Phe Arg His Gly Leu Leu Lys Ile Leu Ala Ile His Gly Leu Ile Ser
                325                 330                 335
Lys Asp Ser Leu Pro Lys Asp Ser Arg Pro Ser Phe Val Gly Ser Ser
            340                 345                 350
Ser Gly His Thr Ser Thr Thr Leu
        355                 360
```

<210> 2
<211> 2880
<212> ADN
<213> Ser humano

<400> 2

aggttcaaaa cattcagaga cagaaggtgg atagacaaat ctccaccttc agactggtag 60

gctcctccag aagccatcag acaggaagat gtgaaaatcc ccagcactca tcccagaatc 120
actaagtggc acctgtcctg ggccaaagtc ccaggacaga cctcattgtt cctctgtggg 180
aatacctccc caggagggca tcctggattt ccccccttgca acccaggtca gaagtttcat 240
cgtcaaggtt gtttcatctt ttttttcctg tctaacagct ctgactacca cccaaccttg 300
aggcacagtg aagacatcgg tggccactcc aataacagca ggtcacagct gctcttctgg 360
aggtgtccta caggtgaaaa gcccagcgac ccagtcagga tttaagttta cctcaaaaat 420
ggaagatttt aacatggaga gtgacagctt tgaagatttc tggaaaggtg aagatcttag 480
taattacagt tacagctcta ccctgccccc ttttctacta gatgccgccc catgtgaacc 540
agaatccctg gaaatcaaca agtattttgt ggtcattatc tatgccctgg tattcctgct 600
gagcctgctg ggaaactccc tcgtgatgct ggtcatctta tacagcaggg tcggccgctc 660
cgtcactgat gtctacctgc tgaacctagc cttggccgac ctactctttg ccctgacctt 720
gcccatctgg gccgcctcca aggtgaatgg ctggattttt ggcacattcc tgtgcaaggt 780
ggtctcactc ctgaaggaag tcaacttcta tagtggcatc ctgctactgg cctgcatcag 840
tgtggaccgt tacctggcca ttgtccatgc cacacgcaca ctgacccaga agcgctactt 900
ggtcaaattc atatgtctca gcatctgggg tctgtccttg ctcctggccc tgcctgtctt 960
acttttccga aggaccgtct actcatccaa tgttagccca gcctgctatg aggacatggg 1020
caacaataca gcaaactggc ggatgctgtt acggatcctg ccccagtcct ttggcttcat 1080
cgtgccactg ctgatcatgc tgttctgcta cggattcacc ctgcgtacgc tgtttaaggc 1140
ccacatgggg cagaagcacc gggccatgcg ggtcatcttt gctgtcgtcc tcatcttcct 1200
gctctgctgg ctgccctaca acctggtcct gctggcagac accctcatga ggacccaggt 1260
gatccaggag acctgtgagc gccgcaatca catcgaccgg gctctggatg ccaccgagat 1320
tctgggcatc cttcacagct gcctcaaccc cctcatctac gccttcattg gccagaagtt 1380
tcgccatgga ctcctcaaga ttctagctat acatggcttg atcagcaagg actccctgcc 1440
caaagacagc aggccttcct ttgttggctc ttcttcaggg cacacttcca ctactctcta 1500
agacctcctg cctaagtgca gccccgtggg gttcctccct tctcttcaca gtcacattcc 1560
aagcctcatg tccactggtt cttcttggtc tcagtgtcaa tgcagccccc attgtggtca 1620
caggaagtag aggaggccac gttcttacta gtttcccttg catggtttag aaagcttgcc 1680
ctggtgcctc acccttgcc ataattacta tgtcatttgc tggagctctg cccatcctgc 1740
ccctgagccc atggcactct atgttctaag aagtgaaaat ctacactcca gtgagacagc 1800
tctgcatact cattaggatg gctagtatca aaagaaagaa aatcaggctg gccaacgggg 1860
tgaaaccctg tctctactaa aaatacaaaa aaaaaaaaaa attagccggg cgtggtggtg 1920
agtgcctgta atcacagcta cttgggaggc tgagatggga gaatcacttg aacccgggag 1980
gcagaggttg cagtgagccg agattgtgcc cctgcactcc agcctgagcg acagtgagac 2040
tctgtctcag tccatgaaga tgtagaggag aaactggaac tctcgagcgt tgctgggggg 2100
gattgtaaaa tggtgtgacc actgcagaag acagtatggc agctttcctc aaaacttcag 2160
acatagaatt aacacatgat cctgcaattc cacttatagg aattgaccca caagaaatga 2220
aagcagggac ttgaacccat atttgtacac caatattcat agcagcttat tcacaagacc 2280
caaaaggcag aagcaaccca aatgttcatc aatgaatgaa tgaatggcta agcaaaatgt 2340
gatatgtacc taacgaagta tccttcagcc tgaaagagga atgaagtact catacatgtt 2400
acaacacgga cgaaccttga aaactttatg ctaagtgaaa taagccagac atcaacagat 2460
aaatagttta tgattccacc tacatgaggt actgagagtg aacaaattta cagagacaga 2520
aagcagaaca gtgattacca gggactgagg ggaggggagc atgggaagtg acggtttaat 2580
gggcacaggg tttatgttta ggatgttgaa aaagttctgc agataaacag tagtgatagt 2640
tgtaccgcaa tgtgacttaa tgccactaaa ttgacactta aaaatggttt aaatggtcaa 2700
ttttgttatg tatattttat atcaatttaa aaaaaaacct gagccccaaa aggtatttta 2760
atcaccaagg ctgattaaac caaggctaga accacctgcc tatatttttt gttaaatgat 2820
ttcattcaat atctttttt taataaacca tttttacttg ggtgtttata aaaaaaaaaa 2880

<210> 3
<211> 18
<212> ADN
<213> Sintético

<400> 3
tgcaccccca agaccaaa 18

<210> 4
<211> 26
<212> ADN
<213> Sintético

<400> 4

tgattaaagc taacgagcag acagaa 26

<210> 5
<211> 26
<212> ADN
<213> Sintético

<400> 5
ccttcccttt cttggctttg gccttt 26

<210> 6
<211> 22
<212> ADN
<213> Sintético

<400> 6
cgttgtacct gcccaattgt ga 22

<210> 7
<211> 20
<212> ADN
<213> Sintético

<400> 7
gggacgcatc actcaacgtt 20

<210> 8
<211> 27
<212> ADN
<213> Sintético

<400> 8
aagagaaagc agtgcaaacc ttcccgt 27

<210> 9
<211> 16
<212> ADN
<213> Sintético

<400> 9
gccgccgcgg aataat 16

<210> 10
<211> 28
<212> ADN
<213> Sintético

<400> 10
tgtctaccct tatacacaac tccatagg 28

<210> 11
<211> 31
<212> ADN
<213> Sintético

<400> 11
agccgggatc taccataccc attgactaac t 31

<210> 12
<211> 20
<212> ADN
<213> Sintético

<400> 12
tggaacggcc aaatgtactg 20

<210> 13
<211> 20
<212> ADN
<213> Sintético

<400> 13
tggcgtattc ccgttcaagt 20

<210> 14
<211> 23
<212> ADN
<213> Sintético

<400> 14
actctgcccg acgtggtctc cca 23

<210> 15
<211> 23
<212> ADN
<213> Sintético

<400> 15
ccttgaggca cagtgaagac atc 23

<210> 16
<211> 23
<212> ADN
<213> Sintético

<400> 16
cctgtaggac acctccagaa gag 23

<210> 17
<211> 27
<212> ADN
<213> Sintético

<400> 17
tggccactcc aataacagca ggtcaca 27

REIVINDICACIONES

1. Un método para determinar, en un paciente que presenta hematuria, el nivel de riesgo de tener cáncer urotelial, que comprende:

a. proporcionar una muestra de orina de dicho paciente;
b. cuantificar un valor, MI, que comprende cuantificar los niveles de expresión de MDK, CDC2, HOXA13 e IGFBP5 humano y un valor adicional, IL-8Rb, en dicha muestra;
c. evaluar las variables fenotípicas de Frecuencia de hematuria (HFREQ), Age (EdadMQ50), Género, Tabaquismo (SMK) y Número de glóbulos rojos (RBC) de dicho paciente;
d. calcular un índice genotípico más fenotípico (ÍNDICE G+P) de acuerdo con:

la fórmula (i), ÍNDICE G+P = (1 * HFREQ + 3 * Género + 4 * SMK) + (5 * MI + 2 * IL-8Rb),

la fórmula (ii), ÍNDICE G+P = (w1 * HFREQ + w2 * EdadMQ50 + w3 * Género + w4 * SMK + w5 * RBC) + (w6 * M1 + w7 * IL-8Rb),

o

la fórmula (iii), ÍNDICE G+P = -8,46 + 0,79 IGFBP5 - 1,60 HOXA13 + 2,10 MDK + 0,95 CDC2 - 0,38 IL8-Rb + 0,98 SMK + 0,56 HFREQ + 1,11 Género + 0,64 EdadMQ50;

en donde HFREQ significa la frecuencia de hallazgo de 3 o más glóbulos rojos por campo de alto aumento en un periodo de 6 meses; si la frecuencia es baja, entonces HFREQ = 0 y si es mayor de 3 glóbulos rojos por campo de alto aumento, entonces es 1; EdadMQ50 se refiere a la edad del sujeto; si es mayor que 50 años, entonces EdadMQ50 = 1 y si es menor de 50 años, entonces es 0; al género se le asigna un valor de 1 para masculino y 0 para femenino; SMK significa si el sujeto es fumador o exfumador; en caso de que sea no fumador, entonces SMK = 0 y si es fumador, entonces es 1; RBC significa el número de glóbulos rojos; si es de 25 o más, entonces RBC se establece en 1 y si es menor de 25, entonces es 0; y
e. determinar si el ÍNDICE G+P es mayor que un umbral que indica el nivel de riesgo de que el paciente tenga cáncer urotelial.

2. El método de la reivindicación 1, en donde dicho umbral se selecciona entre el grupo de valores del ÍNDICE G+P de 0 a 5, de 6 a 10 o de 11-15, donde dicho valor de 0 a 5 indica bajo riesgo, de 6 a 10 indica riesgo moderado y de 11-15 indica alto riesgo.

3. El método de la reivindicación 2, donde en caso de que dicho umbral sea un valor del ÍNDICE G+P sea de 6-10, se determina que dicho paciente necesita someterse a pruebas clínicas o de laboratorio adicionales.

4. El método de la reivindicación 2, donde en caso de que dicho umbral sea un valor del ÍNDICE G+P sea de 11-15, se determina que dicho paciente necesita someterse a pruebas clínicas o de laboratorio adicionales.

5. El método de la reivindicación 2, donde en caso de que dicho umbral sea un valor del ÍNDICE G+P sea de 0-5, se asigna al paciente a una lista de observación para pruebas clínicas o de laboratorio adicionales.

6. El método de cualquiera de las reivindicaciones 1 a 5, donde el umbral se establece usando un método estadístico, opcionalmente en donde el método estadístico es uno cualquiera de análisis de discriminantes lineales (LDA), regresión logística (LogReg), máquina de vectores de soporte (SVM), K 5 vecinos más próximos (KN5N) y clasificación de árbol de partición (TREE).

7. El método de cualquiera de las reivindicaciones 1 a 6, que además comprende cuantificar la expresión de un marcador genotípico adicional seleccionado entre la figura 6 o la figura 7.

8. El método de cualquiera de las reivindicaciones 1 a 7, donde dicha etapa de cuantificación de la expresión génica se lleva a cabo detectando los niveles de ARNm.

9. El método de cualquiera de las reivindicaciones 1 a 7, en donde dicha etapa de cuantificar la expresión génica se lleva a cabo detectando los niveles de ADNc, opcionalmente donde dicha etapa de cuantificar la expresión genética se lleva a cabo usando un oligonucleótido complementario a al menos una porción de dicho ADNc.

10. El método de cualquiera de las reivindicaciones 1 a 8, donde dicha etapa de cuantificar la expresión genética se lleva a cabo usando un método de qRT-PCR usando un cebador directo y un cebador inverso.

11. El método de cualquiera de las reivindicaciones 1 a 7, donde dicha etapa de cuantificar la expresión génica se lleva a cabo:

a. detectando los niveles de una proteína; o
b. detectando los niveles de un péptido.

12. El método de una cualquiera de las reivindicaciones 1 a 7 u 11, donde dicha etapa de cuantificar la expresión genética se lleva a cabo usando un anticuerpo dirigido contra dicho marcador.

13. El método de una cualquiera de las reivindicaciones 1 a 7 u 11-12, donde dicha etapa de cuantificar la expresión genética se lleva a cabo usando un método de inmunoensayo de tipo sándwich o usando una microplaca de anticuerpos.

14. El método de una cualquiera de las reivindicaciones 1 a 7 u 11-13, donde dicha cuantificación de la expresión genética se lleva a cabo usando un anticuerpo monoclonal.

15. El método de una cualquiera de las reivindicaciones 1 a 7 u 11-14, donde dicha cuantificación de la expresión genética se lleva a cabo usando un antisuero policlonal.

LOCUS NM_001557 2880 pb mRNA
DEFINICIÓN Receptor 2 (CXCR2) de quimiocinas (motivo C-X-C) de Homo sapiens, variante de transcrito 1, ARNm.
ACCESSION NM_001557
VERSION NM_001557.3 GI:269973857

Secuencia de proteína: "MEDFNMESDSFEDFWKGEDLSNYSYSSTLPPFLLDAAPCEPESL
EINKYFVVIIYALVFLLSLLGNSLVMLVILYSRVGRSVTDVYLLNLALADLLFALTLP
IWAASKVNGWIFGTFLCKVVSLLKEVNFYSGILLLACISVDRYLAIVHATRTLTQKRY
LVKFICLSIWGLSLLLALPVLLFRRTVYSSNVSPACYEDMGNNTANWRMLLRILPQSF
GFIVPLLIMLFCYGFTLRTLFKAHMGQKHRAMRVIFAVVLIFLLCWLPYNLVLLADTL
MRTQVIQETCERRNHIDRALDATEILGILHSCLNPLIYAFIGQKFRHGLLKILAIHGL
ISKDSLPKDSRPSFVGSSSGHTSTTL"

ORIGEN

```
      1 aggttcaaaa cattcagaga cagaaggtgg atagacaaat ctccaccttc agactggtag
     61 gctcctccag aagccatcag acaggaagat gtgaaaatcc ccagcactca tcccagaatc
    121 actaagtggc acctgtcctg ggccaaagtc ccaggacaga cctcattgtt cctctgtggg
    181 aatacctccc caggagggca tcctggattt ccccccttgca acccaggtca gaagtttcat
    241 cgtcaaggtt gtttcatctt ttttttcctg tctaacagct ctgactacca cccaaccttg
    301 aggcacagtg aagacatcgg tggccactcc aataacagca ggtcacagct gctcttctgg
    361 aggtgtccta caggtgaaaa gcccagcgac ccagtcagga tttaagttta cctcaaaaat
    421 ggaagatttt aacatggaga gtgacagctt tgaagatttc tggaaaggtg aagatcttag
    481 taattacagt tacagctcta ccctgccccc ttttctacta gatgccgccc catgtgaacc
    541 agaatccctg gaaatcaaca agtattttgt ggtcattatc tatgccctgg tattcctgct
```

Figura 1

```
    601 gagcctgctg ggaaactccc tcgtgatgct ggtcatctta tacagcaggg tcggccgctc
```

```
 661 cgtcactgat gtctacctgc tgaacctagc cttggccgac ctactcttg ccctgacctt
 721 gcccatctgg gccgcctcca aggtgaatgg ctggatttt ggcacattcc tgtgcaaggt
 781 ggtctcactc ctgaaggaag tcaacttcta tagtggcatc ctgctactgg cctgcatcag
 841 tgtggaccgt tacctggcca ttgtccatgc cacacgcaca ctgacccaga agcgctactt
 901 ggtcaaattc atatgtctca gcatctgggg tctgtcctgg ctcctggccc tgcctgtctt
 961 acttttccga aggaccgtct actcatccaa tgttagccca gcctgctatg aggacatggg
1021 caacaataca gcaaactggc ggatgctgtt acggatcctg ccccagtcct ttggcttcat
1081 cgtgccactg ctgatcatgc tgttctgcta cggattcacc ctgcgtacgc tgtttaaggc
1141 ccacatgggg cagaagcacc gggccatgcg ggtcatcttt gctgtcgtcc tcatcttcct
1201 gctctgctgg ctgccctaca acctggtcct gctggcagac accctcatga ggacccaggt
1261 gatccaggag acctgtgagc gccgcaatca catcgaccgg gctctggatg ccaccgagat
1321 tctgggcatc cttcacagct gcctcaaccc cctcatctac gccttcattg gccagaagtt
1381 tcgccatgga ctcctcaaga ttctagctat acatggcttg atcagcaagg actccctgcc
1441 caaagacagc aggccttcct ttgttggctc ttcttcaggg cacacttcca ctactcttca
1501 agacctcctg cctaagtgca gccccgtggg gttcctccct tctcttcaca gtcacattcc
1561 aagcctcatg tccactggtt cttcttggtc tcagtgtcaa tgcagccccc attgtggtca
1621 caggaagtag aggaggccac gttcttacta gtttcccttg catggtttag aaagcttgcc
1681 ctggtgcctc acccccttgcc ataattacta tgtcatttgc tggagctcctg cccatcctgc
1741 ccctgagccc atggcactct atgttctaag aagtgaaaat ctacactcca gtgagacagc
1801 tctgcatact cattaggatg gctagtatca aaagaaagaa aatcaggctg gccaacgggg
1861 tgaaaccctg tctctactaa aaatacaaaa aaaaaaaaaa attagccggg cgtggtggtg
1921 agtgcctgta atcacagcta cttgggaggc tgagatgggа gaatcacttg aacccgggag
1981 gcagaggttg cagtgagccg agattgtgcc cctgcactcc agcctgagcg acagtgagac
2041 tctgtctcag tccatgaaga tgtagaggag aaactggaac tctcgagcgt tgctgggggg
2101 gattgtaaaa tggtgtgacc actgcagaag acagtatggc agctttcctc aaaacttcag
2161 acatagaatt aacacatgat cctgcaattc cacttatagg aattgaccca caagaaatga
2221 aagcagggac ttgaacccat attgtacac caatattcat agcagcttat tcacaagacc
2281 caaaaggcag aagcaaccca aatgttcatc aatgaatgaa tgaatggcta agcaaaatgt
2341 gatatgtacc taacgaagta tccttcagcc tgaaagagga atgaagtact catacatgtt
```

Figura 1 (cont)

```
2401 acaacacgga cgaaccttga aaactttatg ctaagtgaaa taagccagac atcaacagat
```

```
2461 aaatagttta tgattccacc tacatgaggt actgagagtg aacaaattta cagagacaga
2521 aagcagaaca gtgattacca gggactgagg ggaggggagc atgggaagtg acggtttaat
2581 gggcacaggg tttatgttta ggatgttgaa aaagttctgc agataaacag tagtgatagt
2641 tgtaccgcaa tgtgacttaa tgccactaaa ttgacactta aaaatggttt aaatggtcaa
2701 ttttgttatg tatattttat atcaatttaa aaaaaaacct gagccccaaa aggtatttta
2761 atcaccaagg ctgattaaac caaggctaga accacctgcc tatatttttt gttaaatgat
2821 ttcattcaat atctttttct taataaacca tttttacttg ggtgtttata aaaaaaaaaa
```

Figura 1 (cont)

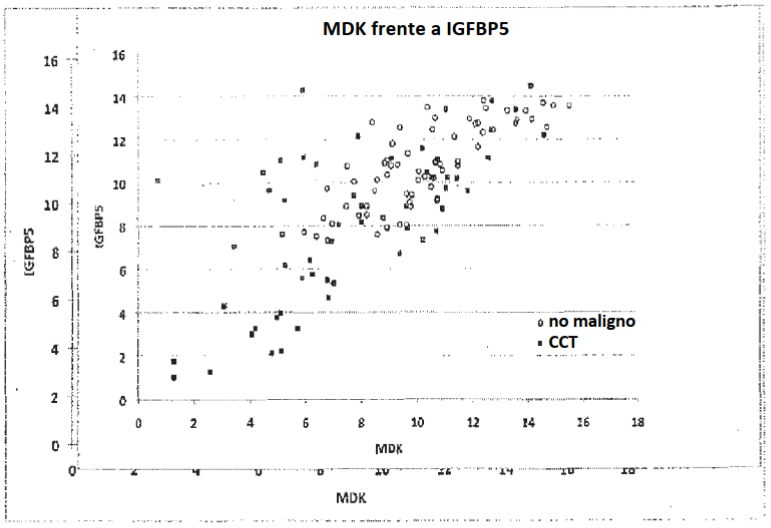


Fig. 2a

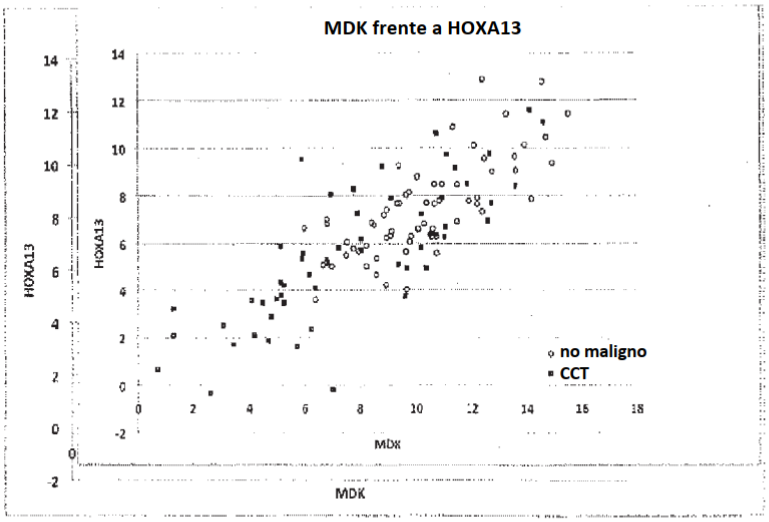


Fig. 2b

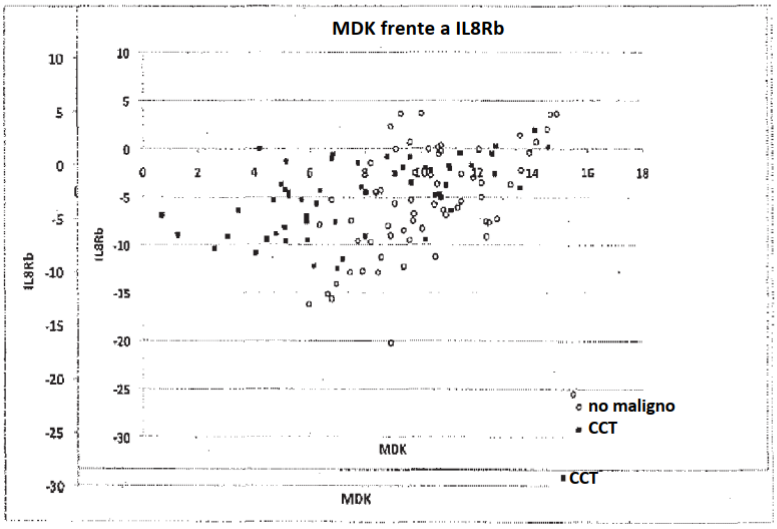


Fig. 2c

Figura 2.

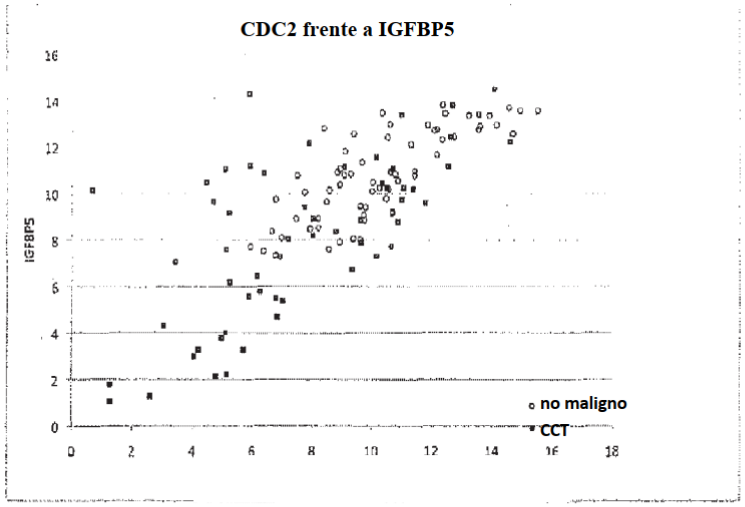


Fig. 2d

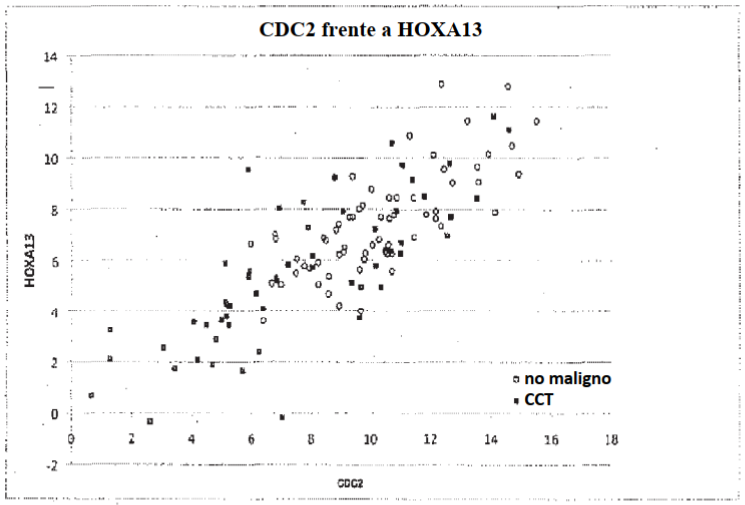


Fig. 2e

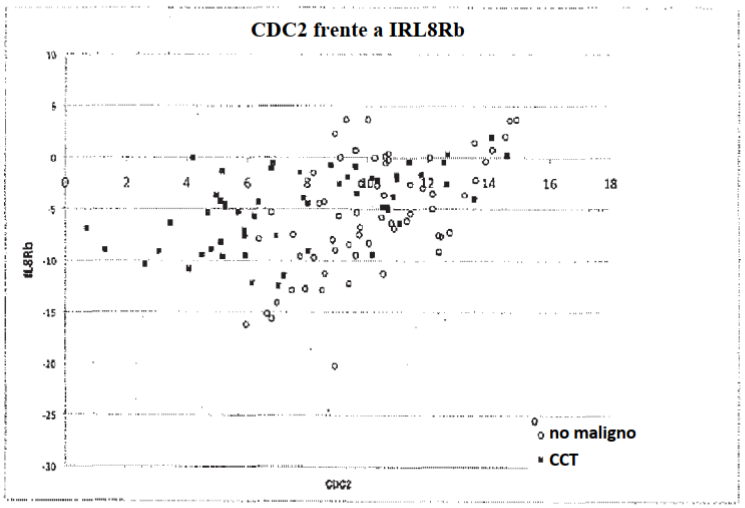


Fig. 2f

Figura 2 cont.

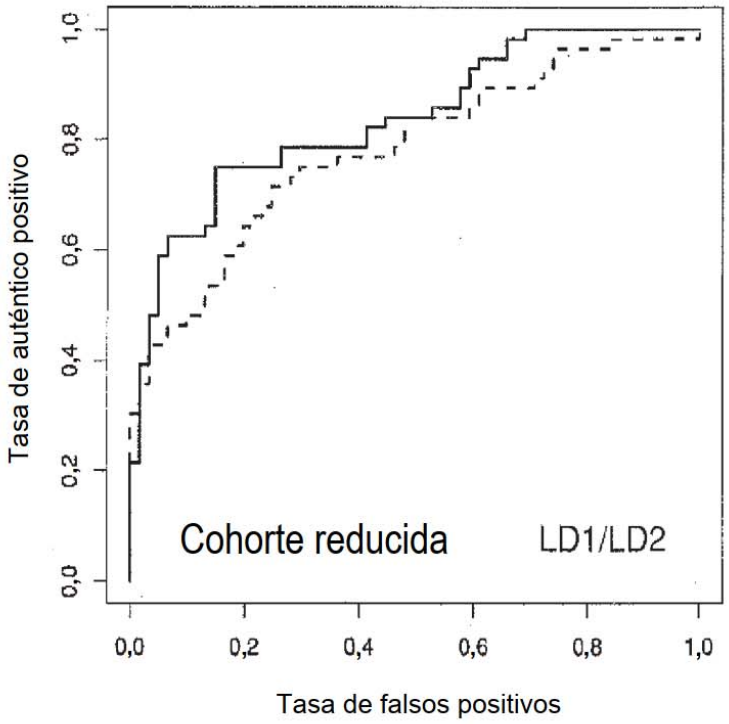


Fig. 3a

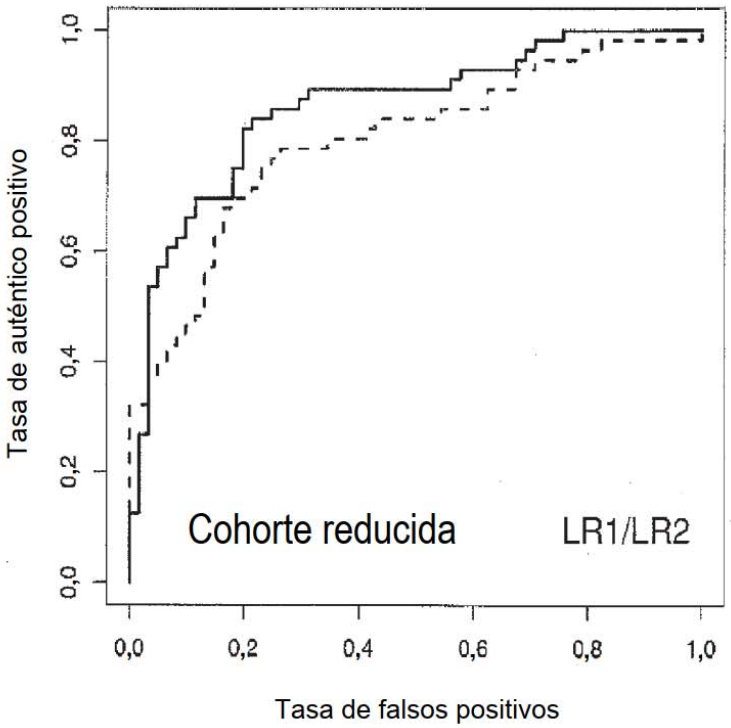


Fig. 3b

Figura 3a-b

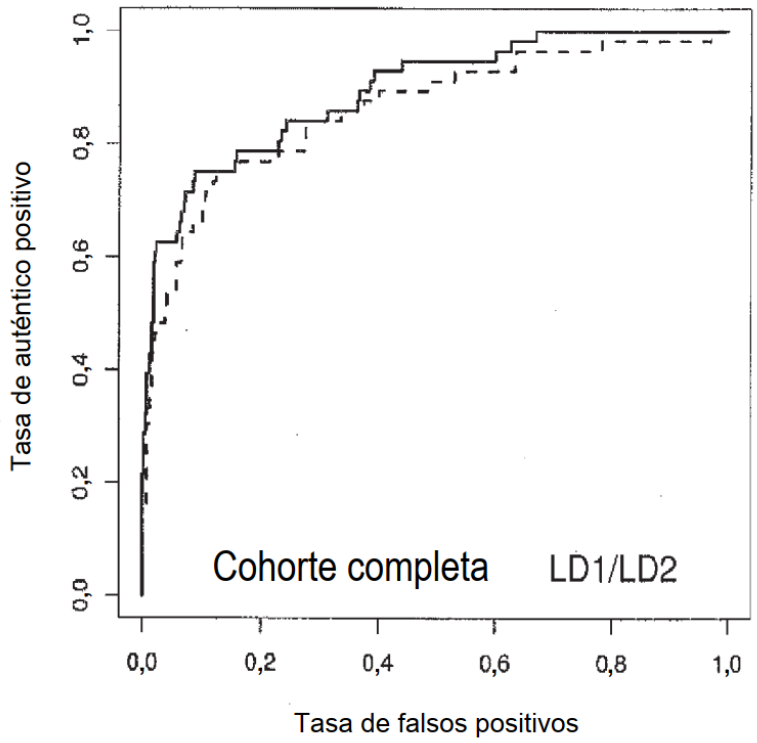


Fig. 4a

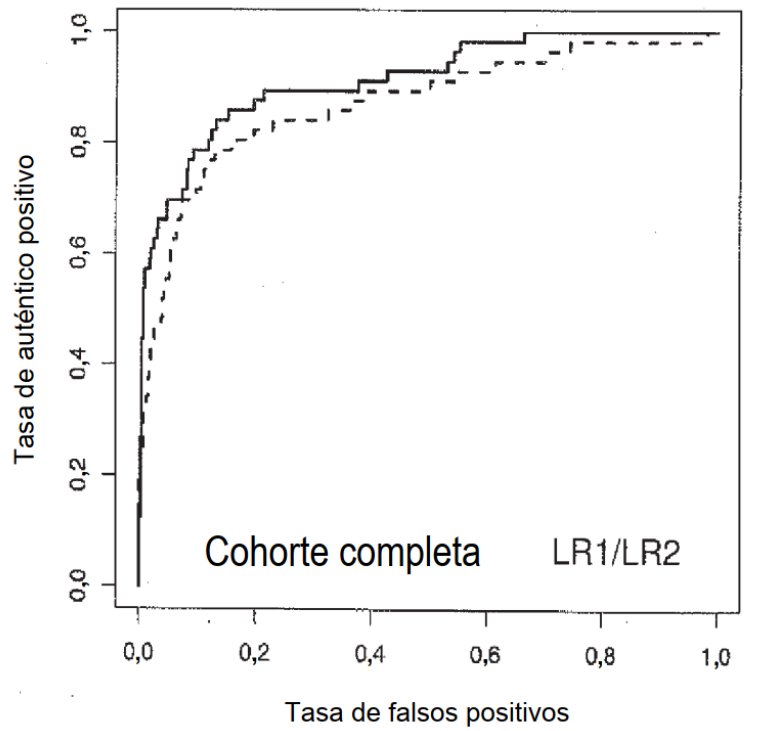


Fig. 4b

Figura 4a-b

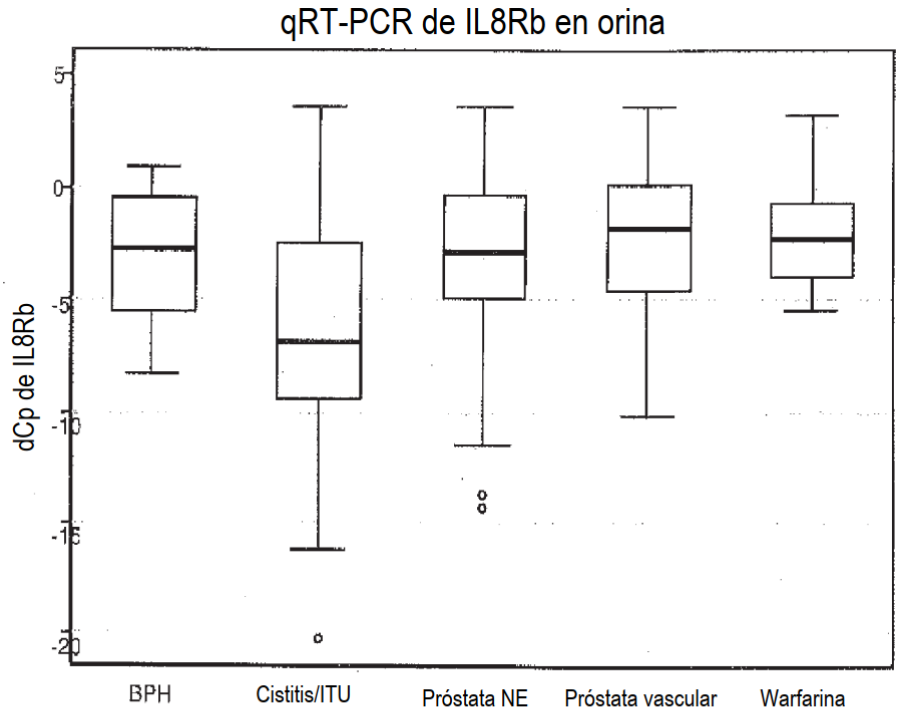

qRT-PCR de IL8Rb en orina
dCp de IL8Rb
5
0
-5
-10
-15
-20
BPH
Cistitis/ITU
Próstata NE
Próstata vascular
Warfarina

Figura 5.

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre	
familia de genes homólogos de ras, miembro B (RHOB), ARNm	NP_004031.1	NM_004040.2	mwghuman30K#A:08114	-2,684592	2,13794E-10	2,25758E-09	-10,53678	38	8789,5	4176	23	
proteína teórica FLJ21511 (FLJ21511), ARNm	NP_079363.1	NM_025087.1	mwghuman30K#A:00410	-2,080908	2,54894E-11	3,22908E-08	-6,925773	68	14455	9272,5	0	
gen de caja emparejada 8 (PAX8), variante de transcrito PAX8A, ARNm	NP_003457.1	NM_003466.3	mwghuman30K#B:3648	-1,961896	4,71674E-11	4,07014E-08	-6,806281	82	19419	12999,5	0	
homólogo A (levadura) de regulador UPF3 de transcritos sin sentido (UPF3A), variante de transcrito 1, ARNm	NP_075387.1	NM_023011.2	mwghuman30K#A:06295	-3,84007	1,03703E-13	3,34027E-08	-7,811653	96,5	5663,5	1825	0	
12-hidroxideshidrogenasa de leucotrienos B4 (LTB4DH), ARNm	NP_036344.1	NM_012212.2	mwghuman30K#B:8686	-5,034613	3,08873E-10	1,41836E-07	-6,436167	122,5	11340,5	3138	0	

Figura 6

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
proteína 2 que interactúa con la familia RAB11 (clase I) (RAB11FIP2), ARNm	NP_055719.1	NM_014904.1	mwghuman30K#A:10479	-1,993629	4,57925E-09	5,83951E-08	-7,711624	156	19636,5	12951	0
familia de dominio TSC22, miembro 1 (TSC22D1), variante de transcrito 2, ARNm	NP_006013.1	NM_006022.2	mwghuman30K#A:00577	-2,133065	1,12033E-07	1,8543E-07	-7,698568	156,5	8415	6130	1
sulfito oxidasa (SUOX), gen nuclear que codifica proteína mitocondrial, variante de transcrito 1, ARNm	NP_000447.2	NM_000456.2	mwghuman30K#A:05512	-1,622803	4,52404E-10	5,29156E-09	-6,728892	223	11312,5	8789,5	0
dominio 21A de repetición de tetratricopéptido (TTC21A), ARNm	NP_665698.1	NM_145755.1	mwghuman30K#B:6976	-2,976809	2,11314E-07	4,28806E-07	-6,286432	225	8375,5	3870,5	1

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
activador de proteína RAS p21 (proteína activadora de GTPasa) 1 (RASA1), variante de transcrito 2, ARNm	NP_072179.1	NM_022650.1	mwghuman30K#A:08292	-1,947454	3,74564E-09	2,68048E-06	-5,99299	235	6895,5	5277,5	7
predicción de ensembl genscan		AC024384.3.107975.119435.1	mwghuman30K#C:6641	-2,471322	4,47058E-08	1,54585E-06	-6,085076	255,5	3327,5	1537	
predicción de ensembl genscan		AL031669.28.1.94224.1	mwghuman30K#C:0930	-1,774123	6,77812E-09	1,13272E-06	-5,56116	288,5	4367	3218	
geminina, inhibidor de la replicación del ADN (GMNN), ARNm	NP_056979.1	NM_015895.3	mwghuman30K#A:03435	-2,046753	1,94846E-07	4,07898E-06	-5,99032	311	11844	8413,75	2
proteína de unión a ácidos grasos 1, hígado (FABP1), ARNm	NP_001434.1	NM_001443.1	mwghuman30K#A:09506	-1,985247	2,20148E-06	6,83453E-07	-6,750782	312,5	25595	19500,75	0
homólogo de oncogén (aviar) del virus 17 del sarcoma v-jun (JUN), ARNm	NP_002219.1	NM_002228.3	mwghuman30K#A:04848	-3,222311	6,2171E-07	4,59274E-06	-6,383849	318,5	12139	5321,5	11

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre	
dominios transmembrana y de hélice superenrollada 4 (TMCO4), ARNm	NP_859070.2	NM_181719.2	mwghuman30K#B:3094	-1,799143	8,41415E-09	3,21175E-06	-5,702577	322,5	4584,5	3158	5	
predicción de ensembl genscan		AP002080.2.37398.74587.1	mwghuman30K#C:4035	-1,860121	5,75299E-08	1,01121E-07	-7,300326	332,5	4972,5	2643		
similar 2 a sinaptotagmina (SYTL2), variante de transcrito b, ARNm	NP_115755.2	NM_032379.3	mwghuman30K#B:9411	-2,343357	4,07526E-10	9,87494E-09	-7,993821	342	19920	12053,5	0	
ciclina G1 (CCNG1), variante de transcrito 1, ARNm	NP_004051.1	NM_004060.3	mwghuman30K#B:5261	-1,83737	2,79204E-08	4,64998E-07	-6,214776	365	9269,5	4453	7	
proteína 34 de caja F (FBXO34), ARNm	NP_060413.2	NM_017943.2	mwghuman30K#B:4885	-1,801846	1,15594E-06	1,13272E-06	-7,165506	365,5	17957,5	13192,75	5	
proteína teórica xp_097916 loc150582		XM_097916	mwghuman30K#B:6228	-2,291304	2,69775E-08	3,84322E-06	-5,299383	375	3591	1276		
contenedor 4 de dominio de ARN pseudouridilato sintasa (RPUSD4), ARNm	NP_116184.1	NM_032795.1	mwghuman30K#B:6888	-1,588438	2,0145E-10	1,14227E-08	-8,404434	407	15747,5	11247	5	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
dominio de repetición de WD 33 (WDR3), variante de transcrito 2, ARNm	NP_0010066 23.1	NM_0010066 22.1	mwghuman3 0K#A:08115	-1,765322	1,47232E-07	1,45307E-06	-6,190057	413	14876	13787,5	8
gen FLJ20859 (FLJ20859), variante de transcrito 1, ARNm	NP_0010251 62.1	NM_0010299 91.1	mwghuman3 0K#A:10470	-1,708985	8,133E-08	1,13272E-06	-6,182853	420,5	3781	2693,75	
factor activador de la transcripción 3 (ATF3), variante de transcrito 2, ARNm	NP_004015.3	NM_004024.3	mwghuman3 0K#A:00568	-3,134291	8,40437E-06	1,51698E-07	-7,446028	421	13832,5	5301,25	3
citocromo p450, familia 3, subfamilia A, polipéptido 5 (CYP3A5), ARNm	NP_000768.1	NM_000777.2	mwghuman3 0K#A:04559	-2,482162	7,0225E-08	1,9772E-06	-4,813425	451,5	20350,5	12190,5	2
predicción de ensembl genscan		AL161725.13.1.181179.3	mwghuman3 0K#C:3443	-1,955495	1,19555E-09	9,24336E-06	-5,681512	455	2738,5	1799	
ligando de KIT (KITLG), variante de transcrito b, ARNm	NP_000890.1	NM_000899.3	mwghuman3 0K#A:07355	-4,28251	3,8867E-06	7,23184E-07	-5,884133	470	19497	10599,75	2

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
factor acelerante de la degradación para el complemento (CD44, sistema de grupo sanguíneo de Cromer) (DAF), ARNm	NP_000565.1	NM_000574.2	mwghuman30K#A:06760	-2,146033	1,44112E-06	6,53282E-06	-5,668164	471,5	9806,5	4753,25	10
familia de transportadores de soluto 25 (transportador mitocondrial; transportador de fosfato), miembro 23 (SLC25A23), ARNm	NP_077008.2	NM_024103.2	mwghuman30K#A:05530	-1,465915	7,79426E-09	8,80903E-07	-5,741772	497	16285,5	12418,75	4
familia de sitio de integración de MMTV de tipo sin alas, miembro 2B (WNT2B), variante de transcrito WNT-2B1, ARNm	NP_004176.2	NM_004185.2	mwghuman30K#A:05696	-10,745275	2,92382E-12	3,19726E-07	-6,175471	501	10716	2016,25	0
subfamilia A de protocadherina gamma, 12 (PCDHGA12), variante de transcrito 1, ARNm	NP_003726.1	NM_003735.2	mwghuman30K#A:03393	-2,070447	2,89557E-07	6,92439E-06	-4,982742	511	2537	1931,5	0

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
predicción de ensembl genscan		AC010828.3.125573.158193.2	mwghuman30K#C:4737	-1,679465	2,34945E-07	6,18322E-07	-5,341693	511	26962	25287,75	
receptor olfatorio, familia 1, subfamilia D, miembro 5 (OR1D5), ARNm	NP_055381.1	NM_014566.1	mwghuman30K#A:08766	-1,640751	5,38884E-08	4,32856E-06	-5,134756	539,5	1644	1134	0
receptor acoplado a proteína G 126 (GPR126), transcrito b1, ARNm	NP_940971.1	NM_198569.1	mwghuman30K#B:2989	-2,181735	3,9002E-07	1,05021E-05	-5,047901	549	15965,5	8040,75	0
similar a nueva proteína ba408e5.4 similar a dmpk cinasa beta de unión a cdc42bpb, cdc42bpb loc144850		XM_090553	mwghuman30K#B:9158	-4,265957	2,3863E-08	5,29006E-07	-6,259229	573	1784	632	
peptidasa 9 específica para ubiquitina, unida a Y (similar a facetas grasas, Drosophila) (USP9Y), ARNm	NP_004645.2	NM_004654.3	mwghuman30K#A:10655	-1,584065	3,40815E-07	8,23838E-06	-5,287284	587	10035,5	7858,25	1

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
predicción de ensembl genscan		AC022554.2.1 1725.14597.1	mwghuman30K#C:7523	-2,800527	1,74508E-07	6,41199E-07	-5,902891	597	3330	1296	
familia de transportador de solutos 35 (transportador de CMP-ácido siálico), miembro A1 (SLC35A1), ARNm	NP_006407.1	NM_006416.2	mwghuman30K#B:6203	-1,879113	1,20877E-06	5,48118E-06	-5,297694	606	20639	15436,75	1
región de unión de cadena v-j de cadena alfa del receptor t de la línea celular sc30, tcr v 21.1 j 11, la secuencia codificante informada abarca desde la posición de nucleótidos		U14083	mwghuman30K#B:9790	-2,442854	9,94216E-08	1,14277E-08	-7,365379	617	6673	2661,5	
marco abierto de lectura 33 del cromosoma 2 (C2orf33), ARNm	NP_064579.3	NM_020194.4	mwghuman30K#A:06988	-1,615311	3,24982E-07	9,24336E-06	-5,392632	624	11555,5	8868,25	4
predicción de ensembl genscan		AC073108.7.1 14033.192530 .1	mwghuman30K#C:4991	-2,137812	1,453E-05	1,06394E-06	-6,345675	650,5	9963,5	5676,75	

Figura 6 (cont)

Nombre/ simbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
UDP-N-acetil-alfa-D-galactosamina: polipéptido N-acetilgalactosaminiltransferasa 7 (GalNAc-T7 (GALNT7)), ARNm	NP_059119.1	NM_017423.1	mwghuman30K#A:01420	-1,401046	7,53042E-08	1,13272E-06	-6,143981	656,5	10573	8997,75	2
marco abierto de lectura 1 del cromosoma 11 (C11orf1), ARNm	NP_073598.1	NM_022761.1	mwghuman30K#A:01834	-1,81726	1,38032E-06	1,22937E-05	-5,104229	673,5	12025,5	9585,25	0
ribonucleoproteína nuclear heterogénea H2 (H') (NHRPH2), variante de transcrito 1, ARNm	NP_062543.1	NM_019597.3	mwghuman30K#A:08547	-2,03901	2,40065E-06	6,16248E-06	-5,465662	684	777	472	9
Síndrome de Werner (WRN), ARNm	NP_000544.1	NM_000553.2	mwghuman30K#A:03163	-1,605462	7,49706E-08	6,16248E-06	-4,681771	696	4585	4351,25	3
predicción de ensembl genscan		AC040160.3.61245.138821.5	mwghuman30K#C:7317	-4,953847	7,54477E-12	1,43666E-08	-6,602968	706,5	4523,5	744,5	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre	
aducina 3 (Gamma) (ADD3), variante de transcrito 1, ARNm	NP_058432.1	NM_016824.3	mwghuman30K#B:5211	-2,092475	1,34331E-05	1,16156E-05	-5,883251	716	7668,5	4808,75	8	
similar 2 a hidroxiesteroide deshidrogenasa (HSDL2), ARNm	NP_115679.2	NM_032303.2	mwghuman30K#B:3888	-2,1317	9,39999E-07	5,81228E-06	-5,58267	732,5	7276,5	5055	2	
transcrito 4 inducible por daño en ADN (DDIT4), ARNm	NP_061931.1	NM_019058.2	mwghuman30K#B:2115	-2,660775	6,68834E-05	5,16818E-06	-5,969252	750	10960,5	5062,75	12	
proteína de unión a ARN, autoantigénica (homólogo de hnRNP asociado con amarillo letal (ratón)) (RALY), variante de transcrito 1, ARNm	NP_057951.1	NM_016732.1	mwghuman30K#A:06376	-1,901576	1,7076E-07	4,32856E-06	-4,392163	750,5	977	386,5	26	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
miembro 3 de la familia de cadena media de acil-CoA sintetasa (ACSM3), variante de transcrito 1, ARNm	NP_005613.2	NM_005622.3	mwghuman30K#B:9076	-1,967123	1,7751E-05	7,33836E-06	-5,050721	765,5	18906,5	12252,25	3
atanogen asociado con BCL2 (BAG1), ARNm	NP_004314.3	NM_004323.3	mwghuman30K#A:09581	-2,525417	2,13163E-08	5,29006E-07	-5,934904	776	9031	4503,5	1
expresado en células precursoras neurales, regulado negativamente en el desarrollo 4 (NEDD4), variante de transcrito 1, ARNm	NP_006145.1	NM_006154.1	mwghuman30K#B:7862	-1,900818	3,7763E-06	8,23334E-06	-5,249282	781	15497	12075	2
marco abierto de lectura 19 del cromosoma 7 (C7orf19), ARNm	NP_116220.1	NM_032831.1	mwghuman30K#B:2324	-1,780735	4,63759E-06	0,000016624	-5,43532	783	11586,5	6704,75	16
calpaína 13 (CAPN13), ARNm	NP_653176.2	NM_144575.2	mwghuman30K#B:6754	-1,542431	4,41704E-06	5,81228E-06	-4,843249	808,5	23318	19953,5	0
caja homeostática B2 (HOXB2), ARNm	NP_002136.1	NM_002145.2	mwghuman30K#A:01639	-2,702649	2,94498E-11	5,08294E-08	-6,774876	814,5	11537,5	6408	3

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre	
contenedor de dominio de hélice superenrollada 28A (CCDC28A), ARNm	NP_056254.1	NM_015439.2	mwghuman30K#B:3128	-1,960606	1,97345E-06	5,48118E-06	-5,330597	826	19022	14223,75	4	
regulador de la miofibrilogénesis 1 (MR-1), variante de transcrito 1, ARNm	NP_056303.2	NM_015488.3	mwghuman30K#B:3179	-1,526799	1,22924E-05	6,92439E-06	-5,874478	832	13007	10755,75	6	
proteína de unión a nucleótidos de guanina (proteína G), similar 1 a polipéptido beta 2 (GNB2L1), ARNm	NP_006089.1	NM_006098.4	mwghuman30K#C:0841	-1,789085	2,16356E-06	4,35419E-05	-4,900058	844,5	1710	1178	106	
contenedor de repetición de espectrina, envuelta nuclear 1 (SYNE1), variante de transcrito beta, ARNm	NP_056108.1	NM_015293.1	mwghuman30K#B:9104	-1,414534	7,52733E-07	3,02433E-06	-5,983107	851	20159,5	17233,75	64	
predicción de ensembl genscan		AL021155.1.1.107603.5	mwghuman30K#C:7334	-1,766409	2,12121E-06	2,01355E-05	-4,468876	853	1264	847,5		

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre	
receptor de proteína morfogenética ósea, tipo IA (BMPR1A), ARNm	NP_004320.2	NM_004329.2	mwghuman30K#A:02484	-1,558632	2,22237E-07	1,22937E-05	-4,618211	874,5	5116	5037,5	4	
contenedor B1 de motivo IQ (IQCB1), variante de transcrito 1, ARNm	NP_0010188 64.1	NM_0010235 70.1	mwghuman30K#B:4020	-1,646698	2,79495E-05	1,62908E-05	-5,169855	882,5	10922	8773,5	2	
predicción de ensembl genscan		AC069502.9.1.18380.1	mwghuman30K#C:3663	-1,464573	6,94397E-07	2,53657E-05	-4,951029	901	21078	19035		
similar a proteína asociada con muerte (LOC92196), ARNm	NP_0010179 20.1	NM_0010179 20.1	mwghuman30K#B:6718	-2,553467	5,76345E-07	7,76152E-07	-5,221107	907	23255	15868	0	
predicción de ensembl genscan		AC012203.4.1 05191.123571.1	mwghuman30K#C:4578	-1,438718	3,21936E-05	4,85862E-07	-5,989899	907	27071,5	25867,5		
predicción de ensembl genscan, el oligo de mwg coincide con estos números de RefSeq NM_00986		AL158153.8.45139.113162.3	mwghuman30K#C:2173	-2,089238	4,80586E-07	4,32856E-06	-5,395407	909,5	1290	1010,25		

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
inducible por detención del crecimiento y daño en el ADN, alfa (GADD45A), ARNm	NP_001915.1	NM_001924.2	mwghuman30K#A:04346	-1,68216	7,53906E-06	2,67916E-05	-4,924921	916,5	11634,5	8596,75	3
chaperona de proteína de estrés 70, asociada a microsomas, 60kDA (STCH), ARNm	NP_008879.3	NM_006948.4	mwghuman30K#A:05064	-1,761334	4,58163E-05	1,72268E-05	-5,784774	916,5	10794	6231,25	5
predicción de ensembl genscan		Z84474.1.1.107526.2	mwghuman30K#C:0911	-1,884499	1,62444E-08	2,23413E-06	-5,434325	933,5	1974	1488,5	
familia con similitud de secuencia 44, miembro B (FAM44B), ARNm	NP_612378.1	NM_138369.1	mwghuman30K#B:4192	-1,357114	7,91633E-07	1,64431E-06	-6,047548	942,5	15308,5	13004,75	1
predicción de ensembl genscan		AC078789.15.1.55074.2	mwghuman30K#C:6283	-1,786484	1,80836E-06	3,70831E-05	-4,861731	948	25553,5	22990,75	
similar a citrato liasa beta (CLYBL), variante de transcrito 1, ARNm	NP_612124.3	NM_138280.3	mwghuman30K#B:8855	-1,444822	1,17336E-06	5,16818E-06	-5,775319	953	21231	19330,25	1

Figura 6 (cont)

Nombre/ simbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre	
predicción de ensembl genscan		AP001015.3.6 6054.115655. 1	mwghuman3 0K#C:3998	-1,709242	1,48259E-07	7,36585E -05	-4,504002	972,5	3374,5	2739,5		
predicción de ensembl genscan		AL159169.14. 1.86155.2	mwghuman3 0K#C:4465	-1,474035	8,96434E-07	3,25437E -06	-5,020854	998,5	28575	28030		
marco abierto de lectura 130 del cromosoma 6 (C6orf130), ARNm	NP_659500.1	NM_145063.2	mwghuman3 0K#B:4652	-1,674296	1,77778E-05	3,32944E -05	-4,846371	1011, 5	14804	10754,75	3	
similar a metiltransferasa 1 (METTL1), variante de transcrito 1, ARNm	NP_005362.1	NM_005371.3	mwghuman3 0K#A:00607	-1,464425	1,73108E-07	5,48118E -06	-5,447819	1026	943	645,5	2	
similar 4 a proteína de ensamblaje de nucleosomas 1 (NAP1L4), ARNm	NP_005960.1	NM_005969.3	mwghuman3 0K#A:00706	-1,902267	1,44614E-05	6,99321E -05	-4,962357	1042, 5	10336,5	8978,5	12	
proteína tumoral D52 (TDP52), variante de transcrito 3, ARNm	NP_005070.1	NM_005079.2	mwghuman3 0K#A:05633	-1,55805	1,79544E-06	3,15409E -05	-4,599069	1070	19033	14646,75	4	
contenedor 4 de dominio EH (EHD4), ARNm	NP_644670.1	NM_139265.2	mwghuman3 0K#B:5911	-1,64935	1,97896E-05	3,70831E -05	-4,903865	1077	12134	9187,5	4	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre	
receptor 4 de factor de crecimiento de fibroblastos (PGFR4), variante de transcrito 2, ARNm	NP_075252.2	NM_022963.2	mwghuman30K#B:8174	-2,638426	2,02402E-06	3,21175E-06	-6,251766	1087,5	17096	9762,5	0	
KIAA0674 (KIAA0674), ARNm	XP_376903.2	XM_376903.2	mwghuman30K#B:0240	-1,362276	3,68739E-06	3,62055E-06	-5,432041	1092,5	10744	9097,25	4	
clon bac de cit987sk-a-363e6 del cromosoma 16		U91321	mwghuman30K#B:5375	-1,568388	5,91844E-05	6,53282E-06	-6,349455	1114,5	10576,5	6083,25		
predicción de ensembl genscan		AC073610.11.63494.102478.3	mwghuman30K#C:1337	-1,828108	1,68631E-06	1,30096E-05	-5,253574	1114,5	1434,5	737		
receptor de andrógenos (AR), variante de transcrito 1, ARNm	NP_000035.2	NM_000044.2	mwghuman30K#A:01003	-1,575215	1,18973E-07	1,54035E-05	-4,291628	1120,5	18067,5	15252,5	0	
factor de transcripción tiroideo 1 (TITF1), ARNm	NP_003308.1	NM_003317.3	mwghuman30K#A:07197	-1,466987	8,57665E-07	1,54585E-06	-6,023277	1124,5	2139	1026,25	0	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
contenedor de repetición de espectrina, envuelta nuclear 1 (SYNE1), variante de transcrito más largo, ARNm	NP_892006.1	NM_182961.1	mwghuman30K#C:2340	-1,440164	6,37872E-05	2,15085E-05	-5,805222	1124,5	26092,5	21681,5	64
predicción de ensembl genscan		AC019084.7.93515.134992.2	mwghuman30K#C:5482	-1,55364	1,48892E-06	7,77593E-06	-5,043927	1126,5	8785,5	8789,5	
predicción de ensembl genscan		AC026235.12.9210.157122.4	mwghuman30K#C:3969	-1,693854	8,74991E-08	0,000122803	-4,393663	1136,5	804	493,5	
predicción de ensembl genscan		AP002840.1.99623.121675.1	mwghuman30K#C:5950	-1,784079	2,96992E-05	7,75722E-05	-4,69759	1156	16693,5	14033,25	
cinasa 1 similar a CDC (CLK1), variante de transcrito 1, ARNm	NP_004062.2	NM_004071.2	mwghuman30K#A:07840	-2,034707	6,15281E-05	6,99321E-05	-5,146308	1157,5	12199	7934,75	4
lipina 1 (LPIN1), ARNm	NP_663731.1	NM_145693.1	mwghuman30K#B:8172	-1,851742	1,40757E-07	2,98753E-05	-4,301729	1159,5	3757	2609	2
lamina A/C (LMNA), variante de transcrito 2, ARNm	NP_005563.1	NM_005572.2	mwghuman30K#A:08943	-1,724223	9,63727E-05	5,67358E-05	-5,305029	1162,5	4447	3295	10
predicción de ensembl genscan		AC079456.15.119608.151490.1	mwghuman30K#C:3309	-1,39662	6,38081E-06	4,32856E-06	-6,439308	1169,5	26755,5	26493,5	

Figura 6 (cont)

Nombre/ simbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
similar a adnc de riken 1500002b03 clon mgc: 12928 imagen: 4298414		BC007768	mwghuman30K#B:4242	-1,518362	5,21322E-06	3,51403E-05	-4,463332	1171	21159	20375	
predicción de ensembl genscan		AC027086.2.23033.31927.1	mwghuman30K#C:6982	-1,659003	1,16113E-11	3,82785E-07	-6,295271	1172,5	2757	2062,75	
hialuronoglucosaminidasa 3 (HYAL3), ARNm	NP_003540.2	NM_003549.2	mwghuman30K#A:09529	-2,01277	6,70905E-08	1,16156E-05	-4,684969	1175	4326	3464,5	4
predicción de ensembl genscan		AC083822.10.126464.129122.1	mwghuman30K#C:9101	-1,479602	6,01319E-06	2,53657E-05	-4,526959	1177	22091,5	20578,5	
proteína VprBP (VprBP), ARNm	NP_055518.1	NM_014703.1	mwghuman30K#B:0332	-1,720326	3,28082E-05	2,40122E-05	-5,246863	1178	17312,5	14031	0
factor de exportación de ARN nuclear 1 (NXF1), ARNm	NP_006353.2	NM_006362.3	mwghuman30K#B:1047	-1,485619	7,48559E-05	1,62908E-05	-5,176445	1195,5	15210	14049,75	20
homólogo 2 de espiráculos vacíos (Drosophila) (EMX2), ARNm	NP_004089.1	NM_004098.2	mwghuman30K#B:9279	-1,745086	9,4996E-09	5,64117E-07	-5,709572	1199,5	19380	10749,25	0
caja homeostática D8 (HOXD8), ARNm	NP_062458.1	NM_019558.2	mwghuman30K#B:3465	-1,41555	6,70306E-08	2,10019E-06	-5,674131	1206	9672,5	7512,25	0

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
proteína predicha pro2591 de hq2591		AF119886	mwghuman30K#B:1001	-1,487408	7,83971E-05	1,30297E-05	-4,929888	1212	27543,5	26699	0
predicción de ensembl genscan		AC016432.3.16946.19371.1	mwghuman30K#C:6058	-3,528975	4,33482E-13	5,08294E-08	-6,731611	1216,5	3962,5	1303	
predicción de ensembl genscan		AF286885.1.76372.115954.3	mwghuman30K#C:7800	-1,83051	1,14443E-05	5,11795E-05	-4,359199	1218	21163,5	21224,5	
proteína de dedo de cinc 25 (KOX 19) (ZNF25), ARNm	NP_659448.1	NM_145011.2	mwghuman30K#C:1648	-1,473477	6,43814E-09	4,35863E-07	-5,84264	1253,5	21026,5	19171	1
proteína de dedo de cinc 626 (ZNF626), ARNm	NP_660340.1	NM_145297.2	mwghuman30K#B:4148	-1,773183	2,31455E-06	6,63845E-05	-4,294361	1266	14806,5	10341	1
inducible por detención del crecimiento y daño en el ADN beta (GADD45B), ARNm	NP_056490.1	NM_015675.1	mwghuman30K#A:00537	-1,950707	0,000200032	2,03519E-05	-5,558029	1270,5	13309	8615,5	5
nebulette (NEBL), variante de transcrito 1, ARNm	NP_006384.1	NM_006393.1	mwghuman30K#A:10459	-1,513391	1,95755E-05	3,15409E-05	-4,789714	1276	22936	20052,5	1

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre	
proteína teórica pro2832 pro2832		NM_018541	mwghuman30K#A:00371	-1,457098	3,06355E-05	6,99321E-05	-5,164495	1289,5	16692,5	14266,5		
proteína de dedo anular 44 (RNF44), ARNm	NP_055716.1	NM_014901.4	mwghuman30K#A:10680	-1,353366	7,76079E-07	2,37433E-06	-5,842265	1292	6288,5	4664,5	1	
REX1, homólogo de ARN exonucleasa 1 (S. cerevisiae) (REXO1), ARNm	NP_065746.2	NM_020695.2	mwghuman30K#B:8218	-1,36867	1,73225E-05	2,38069E-05	-5,111364	1305,5	4933	3309,5	6	
receptor pirimidinérgico P2Y, acoplado a proteína G, 4 (P2RY4), ARNm	NP_002556.1	NM_002565.3	mwghuman30K#A:00433	-1,732075	5,0193E-08	2,68048E-06	-5,510614	1309	18307	12139,5	0	
predicción de ensembl genscan		AL023804.2.1.96460.1	mwghuman30K#C:5405	-1,533348	1,03637E-06	2,806E-05	-4,693772	1313	28610	27980,5		
predicción de ensembl genscan		AL139823.3.49422.51911.1	mwghuman30K#C:8800	-1,894452	4,41409E-05	0,00012914	-4,650026	1320	11214,5	7548		
similar a proteína teórica (LOC440804), ARNm	XP_036936.3	XM_036936.3	mwghuman30K#B:4832	-4,049648	5,38047E-11	1,51698E-07	-6,579174	1332	3647	996,5	0	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre	
marco abierto de lectura 41 del cromosoma X (CXorf41), ARNm	NP_775765.1	NM_173494.1	mwghuman30K#C:1526	-1,448331	1,77872E-05	2,53657E-05	-5,140193	1334	27414	24062,75	0	
marco abierto de lectura 152 del cromosoma 20 (C20orf152), ARNm	NP_543024.1	NM_080834.1	mwghuman30K#B:3251	-1,948612	6,84261E-06	1,82138E-05	-4,987586	1349	10775	8652,5	0	
predicción de ensembl genscan		AL390766.11.35686.142382.1	mwghuman30K#C:4546	-1,516487	5,60986E-05	4,81914E-05	-4,642986	1352	28469,5	27078,25		
fosforilasa, glucógeno; hígado (enfermedad de HER, enfermedad de almacenamiento de glucógeno de tipo VI) (PYGL), ARNm	NP_002854.3	NM_002863.3	mwghuman30K#A:04342	-2,190281	5,14475E-06	0,000269936	-4,316788	1388,5	10289,5	7222	21	
predicción de ensembl genscan		AC013764.3.117501.157349.1	mwghuman30K#C:3627	-1,671949	8,77886E-06	7,36585E-05	-4,28185	1395,5	1205	826,5		
similar 4 a folistatina (FSTL4), ARNm	NP_055897.1	NM_015082.1	mwghuman30K#B:0510	-1,544042	4,79688E-05	4,35419E-05	-4,607654	1405,5	17051	12319	0	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
marco abierto de lectura 168 del cromosoma 14 (C14orf168), ARNm	NP_113615.1	NM_031427.1	mwghuman30K#B:4011	-1,512002	1,56095E-05	5,97939E-05	-4,695177	1414	15738	14705,75	1
familia de aldehído deshidrogenasa 7, miembro A1 (ALDH7A1), ARNm	NP_001173.1	NM_001182.2	mwghuman30K#A:08361	-1,837545	4,80787E-07	0,000105506	-3,903403	1433	20103,5	14345,5	2
predicción de ensembl genscan		AC010904.9.383.204857.4	mwghuman30K#C:7433	-1,40403	4,74755E-05	3,07694E-05	-4,852328	1444	27423,5	25979,25	
calicreína 8 (neuropsina/ ovasina) (KLK8), variante de transcrito 2, ARNm	NP_653088.1	NM_144505.1	mwghuman30K#B:0129	-2,091505	3,56009E-06	2,61655E-06	-5,991679	1449,5	8391	6488	0
cremallera de leucina, similar a regulado negativamente en cáncer 1 (LDOC1L),ARNm	NP_115663.2	NM_032287.2	mwghuman30K#B:8710	-1,501021	2,6566E-06	1,22937E-05	-4,280727	1454,5	16509,5	14157	0
canal 1 de segmento de dos poros (TPCN1), ARNm	NP_060371.2	NM_017901.3	mwghuman30K#B:7759	-1,497443	4,32588E-06	3,32944E-05	-4,567011	1460,5	17545	11635,75	8

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
proteína teórica MGC11242 (MGC11242), ARNm	NP_077296.1	NM_024320.2	mwghuman30K#A:05130	-1,527839	2,70218E-05	1,72268E-05	-4,004461	1463,5	17565	13134,5	0
proteína 1 de clasificación de cúmulos ácidos de fosfofurina (PACS1), ARNm	NP_060496.2	NM_018026.2	mwghuman30K#B:7703	-1,494259	5,23361E-05	7,36585E-05	-4,899266	1482,5	9352	7757,75	19
proteína teórica xp_036406 loc91138		XM_036406	mwghuman30K#B:8080	-1,533764	1,02161E-05	4,59223E-05	-4,93538	1486,5	15372,5	14652	
KIAA1274 (KIAA1274), ARNm	NP_055246.1	NM_014431.1	mwghuman30K#B:8047	-1,41942	7,87044E-05	2,27276E-05	-5,659345	1488,5	8059	5065	0
predicción de ensembl genscan		AC090651.1.9901.36661.2	mwghuman30K#C:6027	-1,525691	2,39913E-05	7,36585E-05	-4,620984	1489,5	21092	20253	
proteína teórica FLJ13111 (FLJ13111), ARNm	NP_079358.1	NM_025082.1	mwghuman30K#A:02690	-1,496988	3,69096E-05	7,08197E-05	-4,489468	1494	4861	7579,25	10
glucoforina C (grupo sanguíneo de Gerbich) (GYPC), variante de transcrito 1, ARNm	NP_002092.1	NM_002101.3	mwghuman30K#A:01278	-1,628829	9,49623E-06	0,000135784	-4,110107	1502	8042	6771	2

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
LOC401510 teórica (LOC401510), ARNm	XP_376843.2	XM_376843.2	mwghuman30K#B:6807	-1,402922	3,03035E-06	0,000105506	-4,471835	1504	13882,5	11233,25	0
predicción de ensembl genscan		AC068282.4.168870.201224.1	mwghuman30K#C:1421	-1,273265	1,4683E-06	4,32856E-06	-5,673395	1505,5	4650	3232,25	
fosfolipasa A2, grupo IVA (citosólica, dependiente de calcio) (PLA2G4A), ARNm	NP_077734.1	NM_024420.1	mwghuman30K#C:2726	-1,656839	1,32347E-05	2,82936E-05	-4,522341	1508,5	17687,5	14140,5	5
plastina 1 (isoforma I) (PLS1), ARNm	NP_002661.1	NM_002670.1	mwghuman30K#A:06551	-1,626206	5,67628E-05	2,98753E-05	-4,952263	1509,5	20943	14040,5	1
similar a neurobeachina 1 (NBEAL1), ARNm	NP_945183.1	NM_198945.1	mwghuman30K#C:8646	-1,528665	7,51176E-06	2,82936E-05	-5,005478	1511,5	23554,5	20060	3
alcohol deshidrogenasa 4 (clase II), polipéptido pi (ADH4), ARNm	NP_000661.2	NM_000670.3	mwghuman30K#A:04002	-1,485933	6,40488E-05	7,36585E-05	-4,608649	1520,5	22288	20240,75	0

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre	
clon rp5-94718 en el cromosoma 1p34.1-31.11. contiene ests, gss y stss. parte de nuevo dominio de repetición de cub-sushi-scr que contiene p		AL355178	mwghuman30K#B:9614	-1,466888	8,43097E-06	3,15409E-05	-4,397922	1529	4947,5	4650,25		
citocromo b5 reductasa 3 (CYB5R3), variante de transcrito M, ARNm	NP_000389.1	NM_000398.4	mwghuman30K#A:01423	-1,497796	0,000121606	9,05266E-05	-5,28459	1531,5	9574,5	6353,5	9	
clon fib5634 proteína predicha pro1477 de hq1477		AF130059	mwghuman30K#B:1074	-1,699661	1,17568E-05	3,32944E-05	-4,802532	1535	5685	3946,75	10	
contenedor 2 de tipo dedo de cinc DHHC (ZDHHC2), ARNm	NP_057437.1	NM_016353.2	mwghuman30K#A:03494	-1,588633	8,03858E-06	0,000105506	-4,741236	1538	13684,5	10087,5	0	
caja homeostática de cono-varilla (CRX), ARNm	NP_000545.1	NM_000554.2	mwghuman30K#A:09219	-3,022299	2,51326E-05	9,24336E-06	-5,849282	1548	16588,5	7500,5	0	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
marco abierto de lectura 154 del cromosoma 14 (C14orf154), variante de transcrito 1, ARNm	NP_115609.2	NM_032233.2	mwghuman30K#B:2665	-1,704609	8,00302E-05	0,000166293	-4,427372	1561,5	16629	15080,75	5
proteína teórica xp_098665 loc155085		XM_098665	mwghuman30K#B:6056	-1,672814	4,27782E-05	0,000142749	-4,312941	1566,5	8795	6847,25	
clon hq0117 pro0117		AF090895	mwghuman30K#B:0914	-1,888405	1,18377E-05	9,05266E-05	-4,735768	1570,5	20158	14933,5	4
sobreexpresada 1 en cáncer de próstata y mama (PBOV1), ARNm	NP_067648.1	NM_021635.1	mwghuman30K#A:05114	-1,610345	5,40979E-06	8,1682E-05	-4,437567	1571,5	22494,5	19845,75	0
esterasa D/formilglutatión hidrolasa (ESD), ARNm	NP_001975.1	NM_001984.1	mwghuman30K#C:2641	-1,535083	6,75303E-06	4,84259E-05	-4,777065	1574	8049	4838,75	6
homólogo B relacionado 4 con autofagia ATG4 (S. cerevisiae) (ATG4B), variante de transcrito 1, ARNm	NP_037457.3	NM_013325.4	mwghuman30K#B:2748	-1,348856	8,59727E-06	1,6224E-05	-4,71932	1575	15166	16712,25	13

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre	
NADPH oxidasa 1 (NOX1), variante de transcrito NOH-1Lv, ARNm	NP_039249.1	NM_013955.1	mwghuman30K#B:1370	-1,903809	5,04367E-06	2,65189E-05	-3,799677	1585,5	18807,5	17371,5	0	
predicción de ensembl genscan		AL445675.9.1.171985.3	mwghuman30K#C:8741	-1,774793	2,60881E-05	0,000182916	-4,133137	1598	22483,5	18071,5		
supuesta proteína nuclear ORF1-FL49 (ORF1-FL49), ARNm	NP_115788.1	NM_032412.2	mwghuman30K#B:7958	-1,499659	6,10749E-05	3,32944E-05	-4,457256	1610,5	8598	4986	8	
contenedor 1 de dominio THUMP (THUMPD1), ARNm	NP_060206.2	NM_017736.3	mwghuman30K#B:2088	-1,313257	4,11592E-07	6,53282E-06	-5,325518	1651,5	14877	15313,5	0	
glutaminil-péptido ciclotransferasa (glutaminil ciclasa) (QPCT), ARNm	NP_036545.1	NM_012413.3	mwghuman30K#A:09422	-1,769571	4,69509E-07	0,000122803	-3,882851	1651,5	11960	8357,75	2	
proteína de dedo PDH 23 (PHF23), ARNm	NP_077273.1	NM_024297.1	mwghuman30K#A:02678	-1,398461	1,43961E-05	0,000109884	-4,356774	1657,5	10656,5	9355,75	0	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
proteína-8 de gen asociado con carcinoma nasofaríngeo (NAG8), ARNm	NP_055226.1	NM_014411.2	mwghuman30K#B:1482	-1,388055	2,38325E-05	5,10584E-05	-4,601658	1660	15731,5	12161	0
predicción de ensembl genscan		AC013487.5.102379.117540.1	mwghuman30K#C:3348	-1,465932	0,000144857	2,37433E-06	-6,225756	1662	14711	9423	
predicción de ensembl genscan		AC079757.5.54037.62672.1	mwghuman30K#C:3844	-1,486502	2,69625E-05	4,12788E-05	-4,765229	1684,5	25678,5	23575,5	
proteína teórica FLJ22313 (FLJ22313), ARNm	NP_071768.2	NM_022373.3	mwghuman30K#B:3974	-1,47429	0,000104725	5,67358E-05	-4,784258	1688,5	8360,5	9165,25	2
holocitocromo c sintasa (citocromo c hemo-liasa) (HCCS), ARNm	NP_005324.2	NM_005333.2	mwghuman30K#A:04510	-2,084412	0,000332632	0,000211886	-5,979649	1692,5	10130,5	5488,25	4
fosfatasa de especificidad dual 5 (DUSP5), ARNm	NP_004410.3	NM_004419.3	mwghuman30K#A:04028	-1,866767	5,52516E-05	2,27276E-05	-5,708882	1694,5	16243,5	8885	2
proteína teórica FLJ20245 (FLJ20245), ARNm	NP_060193.1	NM_017723.1	mwghuman30K#B:2083	-1,525601	6,09068E-05	0,00011676	-4,325617	1697	16122	13101	3

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
proteína cinasa, activada por AMP, subunidad no catalítica gamma 2 (PRKAG2), ARNm	NP_057287.2	NM_016203.2	mwghuman30K#A:00771	-1,408092	1,41299E-05	8,59969E-05	-4,435082	1697,5	16315,5	13437,25	5
clon rp11-314a4 en el cromosoma 20. contiene parte del homólogo 2 de ojos ausentes eya2 de Drosophila ests an sts y gsss comienza con i		AL359434	mwghuman30K#B:9638	-1,591958	1,67257E-05	6,56055E-05	-4,203762	1702	25348,5	23840,75	
proteína teórica pro1598 pro1598		NM_018503	mwghuman30K#B:0983	-1,446989	3,89533E-06	1,30096E-05	-5,339446	1716,5	24112,5	22618,75	
cinasa 9 relacionada con NIMA (gen a de nunca en mitosis) (NEK9), ARNm	NP_149107.3	NM_033116.3	mwghuman30K#C:0238	-1,359292	3,922E-05	4,35419E-05	-4,78725	1722,5	10134,5	7455,5	5
receptor de glicina, alfa 3 (GLRA3), ARNm	NP_006520.1	NM_006529.1	mwghuman30K#A:03690	-1,412224	0,000103763	8,1682E-05	-4,59741	1726	9472,5	8083,25	2

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
predicción de ensembl genscan		AC004408.1.1.113252.2	mwghuman30K#C:9512	-1,54073	2,55537E-06	1,54035E-05	-3,618774	1727	2977	2622,75	
predicción de ensembl genscan		AP001655.2.148977.153525.1	mwghuman30K#C:7868	-1,285997	2,64478E-05	1,96829E-05	-4,601991	1746	28566	28245,5	
asociación 3 con LTR de HERV-H (HHLA3), variante de transcrito 3, ARNm	NP_0010366 22.1	NM_0010366 45.1	mwghuman30K#B:1046	-2,043346	8,67677E-05	9,52067E-05	-3,941702	1759	13925,5	11798	0
predicción de ensembl genscan		AL355497.14.1.215397.9	mwghuman30K#C:7911	-3,056765	1,10766E-06	3,82785E-07	-6,916789	1760	5672,5	2481,25	
proteína 3 transmembrana rica en fibronectina leucina (FLRT3), variante de transcrito 1, ARNm	NP_037413.1	NM_013281.2	mwghuman30K#A:05822	-4,204569	1,4941E-08	9,99201E-07	-6,042443	1770	18156	9231,25	0
cyp3a5 alelo cyp3a5*3 con corte y empalme alternativo		AF355802	mwghuman30K#B:1830	-1,371576	7,33293E-06	1,91497E-05	-4,362299	1784	26198,5	24518,25	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
similar a proteína 3 asociada con microfibrillas (PFAP3L), variante de transcrito 1, ARNm	NP_067679.5	NM_021647.5	mwghuman30K#B:0217	-1,987417	0,000137375	1,98715E-06	-4,959479	1785	22262	18475,25	3
asociada 7 con el ciclo de división celular (CDCA7), variante de transcrito 1, ARNm	NP_114148.2	NM_031942.3	mwghuman30K#B:4810	-1,959598	9,26637E-05	0,00015005	-4,246093	1788	14826	12577,25	1
dominio 61 de repetición de WD (WDR61), ARNm	NP_079510.1	NM_025234.1	mwghuman30K#B:1624	-1,531798	0,000154491	0,000105506	-4,38785	1801,5	9756,5	8041,75	3
proteína cinasa, activada por AMP, subunidad no catalítica beta 1 (PRKAB1), ARNm	NP_006244.2	NM_006253.4	mwghuman30K#A:04996	-1,514877	9,29449E-05	0,000105506	-4,4366	1805,5	16011,5	13954	3
dedo de cinc, contenedor 2 de dominio A20 (ZA20D2), ARNm	NP_005998.1	NM_006007.1	mwghuman30K#A:10213	-1,857923	2,78106E-06	2,67916E-05	-4,781221	1808,5	6856,5	4563,25	13
similar a pro2852 loc93374		XM_050978	mwghuman30K#B:7206	-1,359998	8,48914E-05	6,30077E-05	-5,206744	1817	19166	16438	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
catalasa (CAT), ARNm	NP_001743.1	NM_001752.2	mwghuman30K#C:3530	-2,225675	8,1938E-07	6,92439E-06	-4,777325	1820	7219,5	4208,75	31
similar 2 a peptidilprolil isomerasa (ciclofilina) (PPIL2), variante de transcrito 2, ARNm	NP_680480.1	NM_148175.1	mwghuman30K#B:6262	-1,384263	4,77437E-07	4,59274E-06	-5,396075	1834,5	3806	3514	5
periaxina (PRX), variante de transcrito 2, ARNm	NP_870998.1	NM_181882.1	mwghuman30K#A:01781	-1,399169	1,03356E-05	6,63845E-05	-4,549994	1838	19643,5	17122,5	0
secuencia amplificada en carcinoma de mama 3 (BCAS3), ARNm	NP_060149.2	NM_017679.2	mwghuman30K#A:01576	-1,41274	2,55423E-07	3,64015E-06	-4,03611	1839	15754,5	17960,25	2
HtrA serina peptidasa 2 HTRA2	AAF66598.1	AF141307.1	mwghuman30K#B:1170	-1,284282	2,27894E-05	1,86362E-05	-4,071282	1850	22760,5	20044,75	
predicción de ensembl genscan		AC012095.13.12848.17419.1	mwghuman30K#C:7012	-1,517696	0,000134314	0,000172435	-4,407333	1854	26139,5	25010,5	
proteína de unión a formina 4 (FNBP4), ARNm	NP_056123.1	NM_015308.1	mwghuman30K#B:8897	-1,360042	0,000119422	2,15085E-05	-5,532889	1855	17142	14814,25	11

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
familia de aldehído deshidrogenasa 1, miembro L1 (ALDH1L1), ARNm	NP_036322.2	NM_012190.2	mwghuman30K#A:07479	-2,396851	2,32809E-05	0,000157701	-3,740555	1856,5	13469	8189,5	3
predicción de ensembl genscan		AP001487.3.85282.114407.1	mwghuman30K#C:8733	-1,603991	4,09016E-06	5,97939E-05	-4,353485	1870	3882	3994,75	
predicción de ensembl genscan		AC015651.18.1.191583.4	mwghuman30K#C:2191	-1,392136	5,55423E-07	1,72268E-05	-4,562358	1873,5	2463	1928,25	
contenedor 5 de dominio OUT (OTUD5), ARNm	NP_060072.1	NM_017602.2	mwghuman30K#B:7999	-1,721318	3,42749E-06	1,37651E-05	-5,356883	1894,5	1453	1011	14
predicción de ensembl genscan		AC026668.5.84589.93793.1	mwghuman30K#C:5546	-1,668723	3,74924E-06	2,40122E-05	-5,209366	1899	3201	1676,5	
marco abierto de lectura 121 del cromosoma 20 (C20orf121), ARNm	NP_077307.1	NM_024331.2	mwghuman30K#A:03031	-1,520517	5,93503E-08	1,82138E-05	-4,163768	1906	8788	6894	2
piruvato deshidrogenasa cinasa, isozima 4 (PDK4), ARNm	NP_002603.1	NM_002612.2	mwghuman30K#A:07564	-1,735687	1,72294E-05	1,22937E-05	-5,445156	1911,5	18067	13997,25	1

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre	
proteína KIAA1049 (KIAA1049), ARNm	NP_055787.1	NM_014972.1	mwghuman30K#B:2555	-1,400381	1,30483E-05	7,36585E-05	-4,350759	1913,5	10972	8868,5	12	
homólogo de DnaJ (Hsp40), subfamilia B, miembro 6 (DNAJB6), variante de transcrito 1, ARNm	NP_490647.1	NM_058246.3	mwghuman30K#A:04664	-1,455502	9,60314E-05	0,000150 05	-4,286094	1917	3609	2895,5	14	
lectina, unión a galactósido, soluble, 3 (galectina 3) (LGALS3), ARNm	NP_002297.1	NM_002306.1	mwghuman30K#A:10306	-1,93733	6,69527E-06	2,55121E-05	-4,690215	1923	2753	1803,5	6	
proteína de interacción con PTPRF, proteína de unión 2 (liprina beta 2) (PPFIBP2), ARNm	NP_003612.1	NM_003621.1	mwghuman30K#B:0552	-1,881545	1,28489E-06	2,98753E-05	-4,14602	1937,5	6372,5	4594,5	2	
polimerasa (dirigida por ADN), lambda (POLL), ARNm	NP_037406.1	NM_013274.2	mwghuman30K#B:1279	-1,399567	3,5582E-05	2,98753E-05	-4,793654	1938	15108,5	11461,25	8	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre	
factor de diferenciación de crecimiento 15 (GDF15), ARNm	NP_004855.1	NM_004864.1	mwghuman30K#A:07217	-2,483648	0,000224134	0,00025725	-4,296268	1944	6377	3127,5	2	
proteína teórica xp_089809 loc159765		XM_089809	mwghuman30K#B:6474	-1,381507	9,90157E-07	2,40122E-05	-4,093057	1949	11626,5	11522,5		
contenedor 1 de repetición rica en leucina y motivo estéril alfa (LRSAM1), variante de transcrito 1, ARNm	NP_612370.3	NM_138361.3	mwghuman30K#B:4416	-2,034574	0,000184031	0,000233537	-4,13512	1959,5	7671,5	4421,75	1	
complejo de partículas de tráfico de proteínas 6A (TRAPPC6A), ARNm	NP_077013.1	NM_024108.1	mwghuman30K#A:02117	-1,452954	2,44741E-06	2,53657E-05	-3,509555	1985	1279	1294	8	
predicción de ensembl genscan		AL357146.10.1.82494.1	mwghuman30K#C:3608	-1,287782	1,49736E-05	6,05474E-05	-4,826559	1989	28444,5	26083,25		
proteína teórica MGC14327 (MGC14327), ARNm	NP_444273.1	NM_053045.1	mwghuman30K#B:4427	-1,504495	0,000150258	0,000174117	-4,521635	1992	16805	15440,5	4	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
similar a miembro 5b de la familia de cinesina clon mgc:15265 image:4297793		BC009353	mwghuman30K#B:4438	-1,51682	0,000188214	0,000192131	-4,531075	1992	23071,5	21809,75	
contenedor 3 de dominio de tetramerización de canal de potasio (KCTD3), ARNm	NP_057205.2	NM_016121.3	mwghuman30K#A:04386	-1,358006	1,34593E-05	5,38262E-05	-4,810824	1994,5	17479,5	13203,75	3
predicción de ensembl genscan		AC018425.3.153134.186975.1	mwghuman30K#C:3925	-1,496534	0,000149525	0,000192131	-4,315259	1997	25336	25414,75	
proteína de unión a nucleótidos de guanina (proteína G), polipéptido 1 inhibidor de la actividad alfa (GNAI1), ARNm	NP_002060.4	NM_002069.4	mwghuman30K#A:10201	-1,631361	5,34686E-06	0,000433377	-3,885346	2003	16803,5	14082	1
proteína 4 asociada a microtúbulos (MAP4), variante de transcrito 1, ARNm	NP_002366.2	NM_002375.3	mwghuman30K#B:4355	-1,468752	0,000155187	0,000257 25	-4,423897	2004	10052	9080	11

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
familia 6 de proteína nucleolar (asociada a ARN) (NOL6), variante de transcrito alfa, ARNm	NP_075068.2	NM_022917.4	mwghuman30K#A:07182	-1,30257	1,12604E-06	3,70831E-05	-4,767639	2009	10513	8821	7
molécula 1 de adhesión de células endoteliales cerebrales (CEECAM1), ARNm	NP_057258.2	NM_016174.3	mwghuman30K#A:10231	-1,516413	4,80848E-05	0,000109198	-3,667581	2014	16662,5	13757,75	0
ubiquinol-citocromo c reductasa, complejo III, subunidad VII, 9.5kDa (UQCRQ), gen nuclear que codifica proteína mitocondrial, ARNm	NP_055217.2	NM_014402.3	mwghuman30K#B:8060	-1,71221	4,79681E-05	4,35419E-05	-5,049156	2014	4199,5	2985,25	14
proteína teórica xp_092745 loc146337		XM_092745	mwghuman30K#B:3255	-1,386871	3,26973E-05	0,000165718	-4,513374	2041	19510	15741,25	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
proteína teórica FLJ22965 (FLJ22965), ARNm	NP_071384.1	NM_022101.2	mwghuman30K#A:09776	-1,903469	7,16955E-05	0,000475592	-4,078844	2041,5	2596	2337	0
marco abierto de lectura 102 del cromosoma 9 (C9orf102), variante de transcrito 1, ARNm	NP_064592.1	NM_020207.2	mwghuman30K#B:6956	-1,369819	3,05523E-05	0,000105506	-4,676244	2049	16457,5	13970	0
esfingosina-1-fosfato fosfatasa 1 (SGPP1), ARNm	NP_110418.1	NM_030791.2	mwghuman30K#B:1809	-1,477332	4,90035E-05	0,000142749	-4,328137	2062,5	20094	17541,5	1
contenedor 3 de dominio de plecstrina y Sec7 (PSD3), variante de transcrito 1, ARNm	NP_056125.2	NM_015310.2	mwghuman30K#B:9515	-1,395381	2,70551E-05	0,000142896	-3,953529	2072	14023	13223,25	1
proteína teórica LOC283874 (LOC283874), ARNm	NP_001012749.1	NM_001012731.1	mwghuman30K#B:2101	-1,376341	1,45793E-05	2,67916E-05	-5,079903	2090,5	23511,5	23000	1

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
dedo de cinc, contenedor 21 de dominio FYVE (ZFYVE21), ARNm	NP_076976.1	NM_024071.2	mwghuman30K#A:05642	-1,579562	0,000114253	0,000342648	-4,238027	2094	2994,5	2740,25	0
caja E1 de forkhead (factor de transcripción tiroideo 2) (FOXE1), ARNm	NP_004464.2	NM_004473.3	mwghuman30K#A:07899	-1,678652	1,3104E-05	0,000192131	-3,837954	2107	7269	5293,75	0
predicción de ensembl genscan		AL158168.17.82811.114103.1	mwghuman30K#C:5892	-1,54059	5,72345E-05	8,88591E-06	-4,136599	2114	28578	27753,75	
similar a proteína 5 de repetición de hexámero de per loc149135		XM_089208	mwghuman30K#B:9694	-1,363387	0,000105564	3,77672E-05	-3,9371	2131	27002,5	27249,5	
dedo de cinc, matrina tipo 1 (ZMAT1), variante de transcrito 3, ARNm	NP_115817.1	NM_032441.1	mwghuman30K#C:0106	-1,423696	0,000156267	0,000211886	-4,538065	2134,5	20339,5	17190,25	2
señalización 2 del retículo endoplasmático al núcleo (ERN2), ARNm	NP_150296.2	NM_033266.2	mwghuman30K#B:9444	-1,735187	1,98165E-05	2,70994E-05	-4,509363	2135	24789	18903	0

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre	
proteína de dedo anular 38 (RNF38), variante de transcrito 1, ARNm	NP_073618.3	NM_022781.3	mwghuman30K#B:5588	-1,302965	2,50283E-06	2,15085E-05	-4,986001	2138	24546,5	22148,75	1	
proteína mitocondrial inducida por Myc (mimitina), ARNm	NP_777549.1	NM_174889.2	mwghuman30K#B:3716	-1,474873	8,439E-05	0,000359234	-4,207401	2151,5	9215	7030,25	1	
proteína 1 inducible por TGF beta (TINP1), ARNm	NP_055701.1	NM_014886.2	mwghuman30K#A:06520	-1,655993	0,000318759	0,000571767	-4,796546	2153	7530	5545,5	6	
similar a asociado 5 con pluripotencia en el desarrollo; gen 1 específico de células madre embrionarias (LOC341912), ARNm	XP_292301.3	XM_292301.3	mwghuman30K#B:8134	-1,639786	7,64251E-07	4,32856E-06	-5,437632	2155	18581,5	12815,25	1	
miembro 3 de deshidrogenasa/ reductasa (familia SDR) (DHRS3), ARNm	NP_004744.2	NM_004753.4	mwghuman30K#B:0738	-1,439285	4,39582E-05	0,000222464	-3,974197	2166,5	4626	3570	3	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
monooxigenasa 4 que contiene flavina (FMO4), ARNm	NP_002013.1	NM_002022.1	mwghuman30K#B:2927	-1,804252	0,000195967	0,000222464	-4,159771	2173,5	21428	18241	1
predicción de ensembl genscan		AC022526.4.9087.13827.1	mwghuman30K#C:9081	-1,518222	0,000126409	0,000174117	-4,437239	2174	25713,5	23316	
antígeno 8 de cáncer de colon definido serológicamente (SDCCAG8), ARNm	NP_006633.1	NM_006642.1	mwghuman30K#B:4992	-1,359218	5,81032E-06	1,22937E-05	-4,628131	2177,5	18648	16566	4
homólogo B de ciclo de división celular 14 CDC14 (S. cerevisiae) (CDC14B), variante de transcrito 2, ARNm	NP_201588.1	NM_033331.1	mwghuman30K#B:0753	-1,438486	0,000150076	0,000245124	-4,702655	2178,5	23322	22971	2
leucemia mieloide/linfoide o de linaje mixto (homólogo de tritórax, Drosophila); traslocado a, 3 (MLLT3), ARNm	NP_004520.1	NM_004529.1	mwghuman30K#A:00725	-1,503791	3,87447E-05	0,000311603	-4,282874	2188	20164,5	17462	4

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre	
citocromo p450, familia 27, subfamilia A, polipéptido 1 (CYP27A1), gen nuclear que codifica proteína mitocondrial, ARNm	NP_000775.1	NM_000784.2	mwghuman30K#A:07422	-1,399906	4,05019E-05	3,91276E-05	-4,065883	2197	20714	13781,5	10	
extremo 3' de la proteína asociada con células pre-tnk 1f6		L17326	mwghuman30K#B:9025	-1,643879	5,92984E-08	3,84322E-06	-5,055927	2201	6449,5	4787,5		
péptido antimicrobiano de hepcidina (HAMP), ARNm	NP_066998.1	NM_021175.2	mwghuman30K#A:06397	-1,279742	2,53831E-05	1,72268E-05	-4,893297	2205,5	3722,5	3385,25	0	
predicción de ensembl genscan		AC078819.13.1.41283.1	mwghuman30K#C:7792	-1,349603	2,94705E-06	7,88194E-05	-4,262185	2213	21381,5	17800,75		
proteína teórica BC009862 (LOC90113), ARNm	XP_291077.3	XM_291077.3	mwghuman30K#B:3668	-1,769275	0,000340569	0,000454027	-4,413577	2221,5	6261,5	5176	0	
factor de elongación, ARN polimerasa II, 2 (ELL2), ARNm	NP_036213.1	NM_012081.3	mwghuman30K#A:00356	-1,499648	5,40285E-05	0,000157701	-4,384409	2223,5	14138	11702,5	4	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre	
compañero de oncogén FGFR1 (FGFR1OP), variante de transcrito 1, ARNm	NP_008976.1	NM_007045.2	mwghuman30K#B:4675	-1,513357	9,11409E-05	0,000233537	-4,085108	2228,5	19894	19145,25	0	
predicción de ensembl genscan		AC025953.2.116948.127472.1	mwghuman30K#C:6647	-2,043219	0,000155145	0,000283207	-3,955032	2231	21041	17429,5		
proteína 1 de unión a dominio EH (EHBP1), ARNm	NP_056067.1	NM_015252.2	mwghuman30K#B:0385	-1,35833	2,50342E-06	6,99321E-05	-4,628579	2249,5	13151,5	12131	5	
propionil coenzima A carboxilasa, polipéptido alfa (PCCA), ARNm	NP_000273.2	NM_000282.2	mwghuman30K#B:3551	-1,505038	6,41E-05	0,000110999	-4,042034	2251	16323	13459	5	
predicción de ensembl genscan		AL138499.4.1.185713.1	mwghuman30K#C:4137	-1,655664	1,65577E-05	8,59969E-05	-3,700791	2252,5	6728	6542,5		
predicción de ensembl genscan		AC025937.3.28407.38498.1	mwghuman30K#C:9420	-5,821331	1,29566E-11	3,14688E-07	-6,568115	2255,5	3524	1068,5		

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
respuesta de glucocorticoides inducible por malla trabecular de miocilina (MYOC), ARNm	NP_000252.1	NM_000261.1	mwghuman30K#A:09642	-1,364785	4,61814E-05	2,61402E-05	-4,638227	2263	23815,5	22357,25	0
hspc074		AF161337	mwghuman30K#B:1281	-1,61806	1,54965E-06	1,54035E-05	-4,426706	2264,5	1792	1727	3
epóxido hidrolasa 2, EPHX2 citoplasmática, ARNm	NP_001970.2	NM_001979.4	mwghuman30K#A:08300	-1,416398	5,8666E-06	5,81228E-06	-4,695716	2267,5	19752,5	16857	3
proteína centromérica C1 (CENPC1), ARNm	NP_001803.2	NM_001812.2	mwghuman30K#A:10535	-1,578112	0,000183356	0,000192131	-4,258306	2275	19346,5	14107,25	4
poli(A) polimerasa beta (específica de testículo) (PAPOLB), ARNm	NP_064529.4	NM_020144.4	mwghuman30K#A:00034	-1,533242	0,000179394	0,000311603	-4,111646	2284	24777	23772,75	1
contenedor 1 de pliegue de unión a oligonucleótidos/ oligosacáridos (OBFC1), ARNm	NP_079204.1	NM_024928.3	mwghuman30K#A:07366	-1,597262	0,000151906	0,00015005	-4,339099	2292	15940,5	11655	4

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
contenedor 10 de motivo tripartito (TRIM10), variante de transcrito 1, ARNm	NP_006769.1	NM_006778.2	mwghuman30K#A:10515	-1,750912	5,52205E-05	0,000135784	-4,621742	2296,5	4089,5	1565,5	0
proteína 9 de unión a Rap2 (RPIB9), ARNm	NP_612147.1	NM_138290.1	mwghuman30K#B:7696	-1,451227	1,03873E-06	1,37651E-05	-5,068475	2297	24136,5	22036,5	1
fksg17		NM_032031	mwghuman30K#B:1656	-1,576144	0,000141834	0,000233537	-4,037961	2320,5	2050	1701,75	1
receptor acoplado a proteína G 83 (GPR83), ARNm	NP_057624.2	NM_016540.2	mwghuman30K#A:05359	-1,472466	6,54909E-05	0,000182916	-3,811779	2322,5	18971,5	15879,25	0
predicción de ensembl genscan		AC078987.9.135718.144151.1	mwghuman30K#C:4632	-1,300374	2,50018E-05	1,72268E-05	-4,918105	2322,5	25018,5	21332,5	
caja homeostática de LIM 9 (LHX9), variante de transcrito 1, ARNm	NP_064589.2	NM_020204.2	mwghuman30K#B:8634	-1,374971	4,6951E-06	9,24336E-06	-5,312469	2333,5	16158	13872,25	0
homólogo 1 de pellino (Drosophila) (PELI1), ARNm	NP_065702.2	NM_020651.2	mwghuman30K#A:10770	-1,492094	8,8328E-06	8,23838E-06	-5,23835	2336	22421,5	17524,5	7

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre	
LOC401500 teórica (LOC401500), ARNm	XP_379629.1	XM_379629.1	mwghuman30K#B:6223	-1,325858	2,27785E-05	4,12788E-05	-4,402696	2336,5	26315	25320,25	0	
familia de dominio TSC22, miembro 3 (TSC22D3), variante de transcrito 2, ARNm	NP_004080.2	NM_004089.3	mwghuman30K#B:0462	-1,491298	5,41371E-05	1,72268E-05	-5,010311	2350,5	18693,5	13200,5	43	
receptor acoplado a proteína G 51 (GPR51), ARNm	NP_005449.5	NM_005458.5	mwghuman30K#A:04815	-1,389185	0,000101706	2,98753E-05	-4,845168	2361	3420	2489,25	0	
marco abierto de lectura 61 del cromosoma 10 (C10orf61), variante de transcrito 2, ARNm	NP_056446.1	NM_015631.2	mwghuman30K#B:3105	-1,555923	0,000177095	0,000297088	-4,510054	2366,5	10134	8091,5	3	
contenedor 2C de dominio jumonji (JMJD2C),ARNm	NP_055876.1	NM_015061.1	mwghuman30K#B:8433	-1,317474	6,41124E-05	0,000142749	-4,870495	2366,5	24516,5	22272,75	11	
proteína de dedo de cinc 547 (ZNF547), ARNm	NP_775902.2	NM_173631.2	mwghuman30K#B:6716	-1,63256	0,000251174	0,00025725	-3,984452	2370	19484	19537	2	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
proteína teórica FLJ20032 (FLJ20032), ARNm	NP_060098.2	NM_017628.2	mwghuman30K#B:2068	-1,74414	6,61052E-05	0,000105506	-4,166213	2374	15233,5	9767,75	16
predicción de ensembl genscan		AC064877.3.121131.166993.1	mwghuman30K#C:7466	-1,319284	4,28953E-05	2,03519E-05	-4,802131	2374	27348	26740,25	
predicción de ensembl genscan		AC027755.2.81287.97090.1	mwghuman30K#C:3411	-1,439071	0,000125901	0,000331474	-4,145868	2378,5	14385,5	10781,5	
KIAA0690 (KIAA0690), ARNm	NP_055994.1	NM_015179.2	mwghuman30K#B:7202	-1,642414	0,000173936	0,000433377	-4,08872	2380	14459	11844,25	10
similar 2 a proteína rica en ácido glutámico de unión a dominio SH3 (SH3BGRL2), ARNm	NP_113657.1	NM_031469.1	mwghuman30K#B:1794	-1,995457	0,000279676	0,000376568	-4,033794	2381	9893,5	6985,75	3
cdna:fli21394 fis clon col03536 producto de proteína sin nombre		AK025047	mwghuman30K#B:2580	-2,258109	0,000447376	0,000359234	-4,071961	2387,5	17288	11926,75	0
proteína teórica xp_039231 loc91565		XM_039231	mwghuman30K#B:6431	-1,337232	2,70835E-05	0,000174117	-4,141157	2398	24099,5	20885	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
proteína similar a molécula anclada a GPI (GML), ARNm	NP_002057.1	NM_002066.1	mwghuman30K#A:02108	-1,438045	0,000072493	0,00012914	-4,162956	2405	16818	14265,75	0
complejo de activación de ARN nuclear pequeño, polipéptido 5, 19kDa (SNAPC5), ARNm	NP_006040.1	NM_006049.1	mwghuman30K#B:0945	-1,323754	1,74702E-05	0,000184787	-4,006319	2444	17556,5	17969,25	6
predicción de ensembl genscan		AC027689.10.1.180573.5	mwghuman30K#C:9758	-1,644263	0,000255065	2,98753E-05	-3,674603	2451	16829,5	10684,25	
supresor tumoral 1 mitocondrial (MTUS1), gen nuclear que codifica proteína mitocondrial, variante de transcrito 1, ARNm	NP_0010019 24.1	NM_0010019 24.1	mwghuman30K#B:7775	-1,715161	0,000303663	0,00026993 6	-4,435392	2461,5	17489	15873	0
predicción de ensembl genscan		AL049829.4.1.196292.3	mwghuman30K#C:4548	-1,402914	1,70061E-06	2,49927E-06	-5,01844	2471	28131	26784	
predicción de ensembl genscan		AC084842.1.7903.10639.1	mwghuman30K#C:7638	-1,400793	0,000812094	7,36585E-05	-5,107784	2475,5	14235,5	11061,25	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre
similar a SAFB, modulador de la transcripción (SLTM), variante de transcrito 1, ARNm	NP_079031.2	NM_024755.2	mwghuman30K#B:2154	-1,309539	2,00445E-05	0,000122803	-4,680439	2477,5	12490	12280	12
proteína de unión a nucleótidos de guanina (proteína G), polipéptido de actividad activadora alfa, tipo olfatorio (GNAL), variante de transcrito 2, ARNm	NP_002062.1	NM_002071.1	mwghuman30K#A:06286	-1,836112	0,00030026	0,000498109	-4,10637	2478	19959	16690,75	0
similar a cg14182 producto loc146175		XM_085352	mwghuman30K#B:7365	-1,746066	0,000194923	0,000433377	-3,812815	2487	14791,5	9814,5	
proteína fosfatasa de repetición rica en dominio PH y leucina (PHLPP), ARNm	NP_919431.1	NM_194449.1	mwghuman30K#B:2242	-2,043619	0,000307945	0,000498109	-4,006924	2500	15125	10009,75	7

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre	
candidato 2 de susceptibilidad al autismo (AUTS2), ARNm	NP_056385.1	NM_015570.1	mwghuman30K#B:0110	-1,807633	0,000369299	0,000382814	-4,213775	2503	18404	14412,75	8	
proteína teórica xp_097338 loc147909		XM_097338	mwghuman30K#B:5652	-1,428873	1,32234E-06	2,67916E-05	-4,387947	2513	4388	2991,75		
receptor 4 acoplado a proteína G que contiene repetición rica en leucina (LGR4), ARNm	NP_060960.1	NM_018490.1	mwghuman30K#A:06106	-1,519205	7,81416E-05	0,000122803	-3,810559	2514,5	21417,5	19684	2	
predicción de ensembl		ENSG00000113407	mwghuman30K#C:2560	-1,740805	1,93346E-05	9,52811E-05	-4,388894	2525	8797,5	6359,5		
muerte celular programada 1 (PDCD1), ARNm	NP_005009.1	NM_005018.1	mwghuman30K#B:9147	-1,519488	1,50412E-06	3,41028E-06	-5,713059	2537	8575	7118,5	0	
cadherina 20, tipo 2 (CDH20), ARNm	NP_114097.2	NM_031891.2	mwghuman30K#B:7584	-1,393019	1,91884E-05	2,73302E-05	-4,362445	2538	15588	12275,75	0	

Figura 6 (cont)

Nombre/ símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	nombre de la sonda de MWG	mediana del factor de cambio	Prueba de la t	Prueba de Wilcoxon	SAM	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	sangre	
glicina deshidrogenasa (descarboxilante; glicina descarboxilasa, proteína P de sistema de escisión de glicina) (GLDC), ARNm	NP_000161.1	NM_000170.1	mwghuman30K#B:8959	-1,405054	9,33312E-05	0,00011676	-4,398207	2542,5	6039	4550,5	1	
proteasa, serina, 35 (PRSS35), ARNm	NP_699193.1	NM_153362.1	mwghuman30K#B:3364	-1,451818	0,000237039	0,00025725	-4,319803	2545	16909	14109	0	
predicción de ensembl genscan		AC008134.3.1.175132.1	mwghuman30K#C:3643	-1,601839	0,00037884	0,000286633	-4,042534	2545	27071,5	25825		
casete de unión a ATP, subfamilia A (ABC1), miembro 12 (ABCA12), variante de transcrito 1, ARNm	NP_775099.2	NM_173076.2	mwghuman30K#B:3505	-1,653773	0,000237178	0,000475592	-3,927921	2561,5	19931	18593,5	0	
homólogo 3 de supresor de alas peludas (Drosophila) (SUHW3), ARNm	NP_060136.1	NM_017666.2	mwghuman30K#B:2904	-1,678106	0,000186359	0,000546156	-3,864729	2568	17527	15364,5	1	

Figura 6 (cont)

Nombre/símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	mediana del factor de cambio	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	número EST sangre
proteína teórica FLJ21511 (FLJ21511), ARNm	NP_079363.1	NM_025087.1	-2,080908	68	14455	9272,5	0
gen de caja emparejada 8 (PAX8), variante de transcrito PAX8A, ARNm	NP_003457.1	NM_003466.3	-1,961896	82	19419	12999,5	0
homólogo A (levadura) de regulador UPF3 de transcritos sin sentido (UPF3A), variante de transcrito 1, ARNm	NP_075387.1	NM_023011.2	-3,84007	96,5	5663,5	1825	0
12-hidroxideshidrogenasa de leucotrienos B4 (LTB4DH), ARNm	NP_036344.1	NM_012212.2	-5,034613	122,5	11340,5	3138	0
proteína 2 que interactúa con la familia RAB11 (clase 1) (RA811FIP2), ARNm	NP_055719.1	NM_014904.1	-1,993629	156	19636,5	12951	0
familia de dominio TSC22, miembro 1 (TSC22D1), variante de transcrito 2, ARNm	NP_006013.1	NM_006022.2	-2,133065	156,5	8415	6130	1
sulfito oxidasa (SUOX), gen nuclear que codifica proteína mitocondrial, variante de transcrito 1, ARNm	NP_000447.2	NM_000456.2	-1,622803	223	11312,5	8789,5	0
dominio 21A de repetición de tetratricopéptido (TTC21A), ARNm	NP_665698.1	NM_145755.1	-2,976809	225	8375,5	3870,5	1
similar 2 a sinaptotagmina (SYTL2), variante de transcrito 1, ARNm	NP_115755.2	NM_032379.3	-2,343357	342	19920	12053,5	0
familia de sitio de integración de MMTV de tipo sin alas, miembro 2B (WNT28), variante de transcrito WNT-2B1, ARNm	NP_004176.2	NM_004185.2	-10,745275	501	10716	2016,25	0
subfamilia A de protocadherina gamma, 12 (PCDHGA 12), variante de transcrito 1, ARNm	NP_003726.1	NM_003735.2	-2,070447	511	2537	1931,5	0
receptor olfatorio, familia 1, subfamilia D, miembro 5 (OR1D5), ARNm	NP_055381.1	NM_014566.1	-1,640751	539,5	1644	1134	0
receptor acoplado a proteína G 126 (GPR126), variante de transcrito b1, ARNm	NP_940971.1	NM_198569.1	-2,181735	549	15965,5	8040,75	0
peptidasa 9 específica para ubiquitina, unida a Y (similar a facetas grasas, Drosophila) (USP9Y), ARNm	NP_004645.2	NM_004654.3	-1,584065	587	10035,5	7858,25	1

Figura 7

Nombre/símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	mediana del factor de cambio	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	número EST sangre
marco abierto de lectura 1 del cromosoma 11 (C11orf1), ARNm	NP_073598.1	NM_022761.1	-1,81726	673,5	12025,5	9585,25	0
atanogen asociado con BCL2 (BAG1), ARNm	NP_004314.3	NM_004323.3	-2,525417	776	9031	4503,5	1
familia con similitud de secuencia 44, miembro B (FAM448), ARNm	NP_612378.1	NM_138369.1	-1,357114	942,5	15308,5	13004,75	1
receptor 4 de factor de crecimiento de fibroblastos (FGFR4), variante de transcrito 2, ARNm	NP_075252.2	NM_022963.2	-2,638426	1087,5	17096	9762,5	0
factor de transcripción tiroideo 1 (TITF1), ARNm	NP_003308.1	NM_003317.3	-1,466987	1124,5	2139	1026,25	0
proteína VprBP (VprBP), ARNm	NP_055518.1	NM_014703.1	-1,720326	1178	17312,5	14031	0
homólogo 2 de espiráculos vacíos (Drosophila) (EMX2), ARNm	NP_004089.1	NM_004098.2	-1,745086	1199,5	19380	10749,25	0
caja homeostática D8 (HOXD8), ARNm	NP_062458.1	NM_019558.2	-1,41555	1206	9672,5	7512,25	0
proteína de dedo de cinc 626 (ZNF626), ARNm	NP_660340.1	NM_145297.2	-1,773183	1266	14806,5	10341	1
proteína de dedo anular 44 (RNF44), ARNm	NP_055716.1	NM_014901.4	-1,353366	1292	6288,5	4664,5	1
receptor pirimidinérgico P2Y, acoplado a proteína G, 4 (P2RY4), ARNm	NP_002556.1	NM_002565.3	-1,732075	1309	18307	12139,5	0
similar a proteína teórica (LOC440804), ARNm	XP_036936.3	XM_036936.3	-4,049648	1332	3647	996,5	0
marco abierto de lectura 152 del cromosoma 20 (C20orf152), ARNm	NP_543024.1	NM_080834.1	-1,948612	1349	10775	8652,5	0
similar 4 a folistatina (FSTL4), ARNm	NP_055897.1	NM_015082.1	-1,544042	1405,5	17051	12319	0
marco abierto de lectura 168 del cromosoma 14 (C14orf168), ARNm	NP_113615.1	NM_031427.1	-1,512002	1414	15738	14705,75	1
calicreína 8 (neuropsina/ovasina) (KLK8), variante de transcrito 2, ARNm	NP_653088.1	NM_144505.1	-2,091505	1449,5	8391	6488	0
cremallera de leucina, similar a regulado negativamente en cáncer 1 (LDOC1L), ARNm	NP_115663.2	NM_032287.2	-1,501021	1454,5	16509,5	14157	0
proteína teórica MGC11242 (MGC11242), ARNm	NP_077296.1	NM_024320.2	-1,527839	1463,5	17565	13134,5	0
KIAA1274 (KIAA1274), ARNm	NP_055246.1	NM_014431.1	-1,41942	1488,5	8059	5065	0
LOC401510 teórica (LOC401510), ARNm	XP_376843.2	XM_376843.2	-1,402922	1504	13882,5	11233,25	0

Figura 7 (cont)

Nombre/símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	mediana del factor de cambio	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	número EST sangre
plastina 1 (isoforma I) (PLS1), ARNm	NP_002661.1	NM_002670.1	-1,626206	1509,5	20943	14040,5	1
dedo de cinc, contenedor 2 de tipo DHHC (ZDHHC2), ARNm	NP_057437.1	NM_016353.2	-1,588633	1536	13684,5	10087,5	0
caja homeostática de cono-varilla (CRX), ARNm	NP_000545.1	NM_000554.2	-3,022299	1548	16588,5	7500,5	0
proteína de dedo PHD 23 (PHF23), ARNm	NP_077273.1	NM_024297.1	-1,398461	1657,5	10656,5	9355,75	0
proteína-8 de gen asociado con carcinoma nasofaríngeo (NAG8), ARNm	NP_055226.1	NM_014411.2	-1,388055	1660	15731,5	12161	0
asociación 3 con LTR de HERV-H (HHLA3), variante de transcrito 3, ARNm	NP_001031722.1	NM_001036645.1	-2,043346	1759	13925,5	11798	0
proteína 3 transmembrana rica en fibronectina leucina (FLRT3), variante de transcrito 1, ARNm	NP_037413.1	NM_013281.2	-4,204569	1770	18156	9231,25	0
asociada 7 con el ciclo de división celular (CDCA7), variante de transcrito 1, ARNm	NP_114148.2	NM_031942.3	-1,959598	1788	14826	12577,25	1
piruvato deshidrogenasa cinasa, isozima 4 (PDK4), ARNm	NP_002603.1	NM_002612.2	-1,735687	1911,5	18067	13997,25	1
contenedor 1 de repetición rica en leucina y motivo estéril alfa (LRSAM1), variante de transcrito 1, ARNm	NP_612370.3	NM_138361.3	-2,034574	1959,5	7671,5	4421,75	1
proteína de unión a nucleótidos de guanina (proteína G), polipéptido 1 inhibidor de la actividad alfa (GNAI1), ARNm	NP_002060.4	NM_002069.4	-1,631361	2003	16803,5	14082	1
molécula 1 de adhesión de células endoteliales cerebrales (CEECAM1), ARNm	NP_057258.2	NM_016174.3	-1,516413	2014	16662,5	13757,75	0
proteína teórica FLJ22965 (FLJ22965), ARNm	NP_071384.1	NM_022101.2	-1,903469	2041,5	2596	2337	0
marco abierto de lectura 102 del cromosoma 9 (C9orf102), variante de transcrito 1, ARNm	NP_064592.1	NM_020207.2	-1,369819	2049	16457,5	13970	0
contenedor 3 de dominio de plecstrina y Sec7 (PSD3), variante de transcrito 1, ARNm	NP_056125.2	NM_015310.2	-1,395381	2072	14023	13223,25	1
dedo de cinc, contenedor 21 de dominio FYVE (ZFYVE21), ARNm	NP_076976.1	NM_024071.2	-1,579562	2094	2994,5	2740,25	0
caja E1 de forkhead (factor de transcripción tiroideo 2) (FOXE1), ARNm	NP_004464.2	NM_004473.3	-1,678652	2107	7269	5293,75	0

Figura 7 (cont)

Nombre/símbolo del gen	identificador de la proteína	identificador del ref seq del ARNm	mediana del factor de cambio	puntuación de rango	rango de la mediana de intensidad de tejido tumoral	rango de la mediana de intensidad de tejido no maligno	número EST sangre
proteína mitocondrial inducida por Myc (mimitina), ARNm	NP_777549.1	NM_174889.2	-1,474873	2151,5	9215	7030,25	1
similar a asociado 5 con pluripotencia en el desarrollo; gen 1 específico de células madre embrionarias (LOC341912), ARNm	XP_292301.3	XM_292301.3	-1,639786	2155	18581,5	12815,25	1
péptido antimicrobiano de hepcidina (HAMP), ARNm	NP_066998.1	NM_021175.2	-1,279742	2205,5	3722,5	3385,25	0
proteína teórica BC009862 (LOC90113), ARNm	XP_291077.3	XM_291077.3	-1,769275	2221,5	6261,5	5176	0
contenedor 10 de motivo tripartito (TRIM10), variante de transcrito 1, ARNm	NP_006769.1	NM_006778.2	-1,750912	2296,5	4089,5	1565,5	0
fksg17 fksg 17		NM_032031	-1,576144	2320,5	2050	1701,75	1
caja homeostática de LIM 9 (LHX9), variante de transcrito 1, ARNm	NP_064589.2	NM_020204.2	-1,374971	2333,5	16158	13872,25	0
receptor acoplado a proteína G 51 (GPR51), ARNm	NP_005449.5	NM_005458.5	-1,389185	2361	3420	2489,25	0
cdna: flj21394 fis clon col03536 producto de proteína sin nombre		AK025047	-2,258109	2387,5	17288	11926,75	0
proteína similar a molécula anclada a GPI (GML), ARNm	NP_002057.1	NM_002066.1	-1,438045	2405	16818	14265,75	0
muerte celular programada 1 (PDCD1), ARNm	NP_005009.1	NM_005018.1	-1,519488	2537	8575	7118,5	0
cadherina 20, tipo 2 (CDH20), ARNm	NP_114097.2	NM_031891.2	-1,393019	2538	15588	12275,75	0
glicina deshidrogenasa (descarboxilante; glicina descarboxilasa, proteína P de sistema de escisión de glicina) (GLDC), ARNm	NP_000161.1	NM_000170.1	-1,405054	2542,5	6039	4550,5	1
proteasa, serina, 35 (PRSS35), ARNm	NP_699193.1	NM_153362.1	-1,451818	2545	16909	14109	0

Figura 7 (cont)

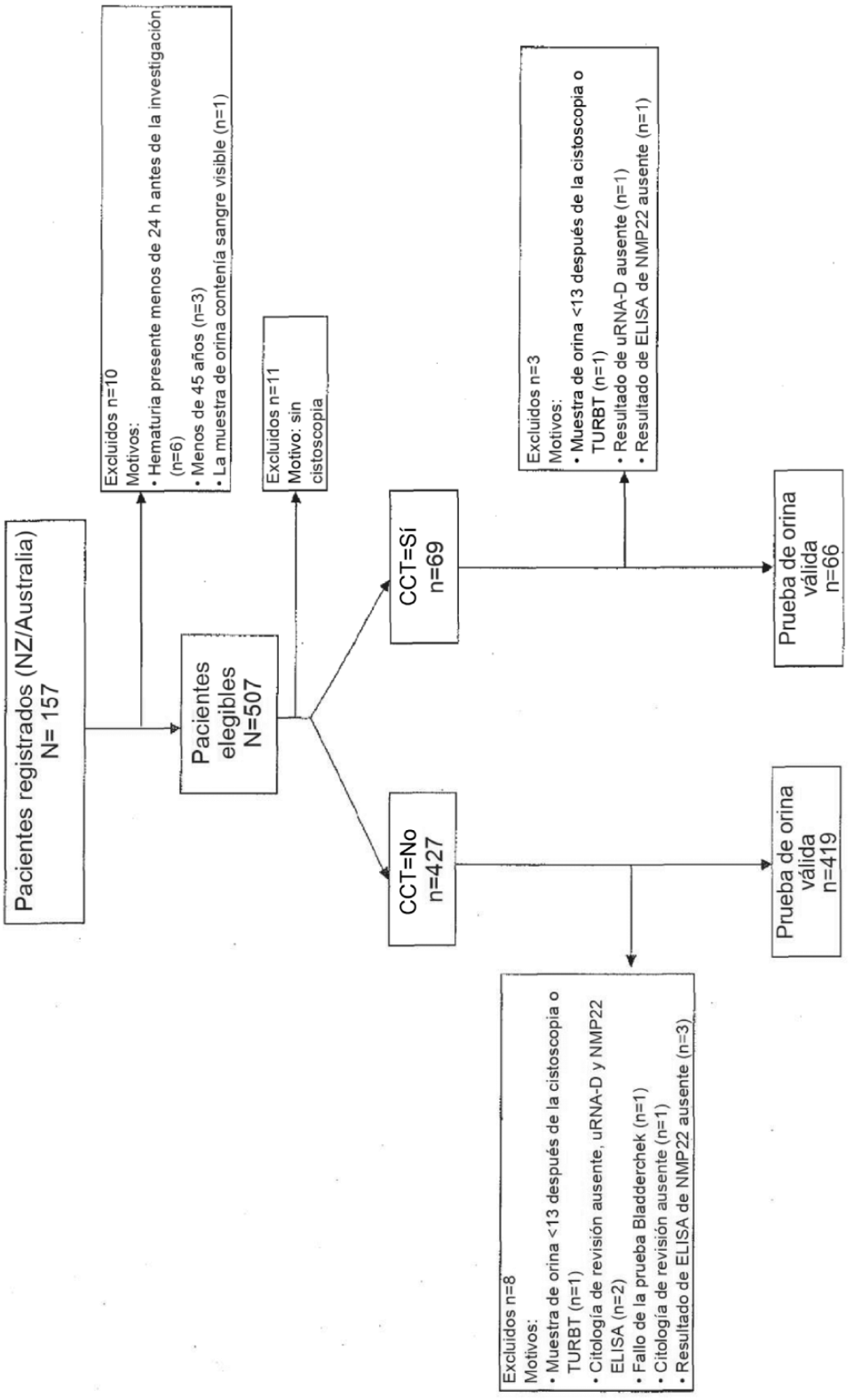

Pacientes registrados (NZ/Australia) N= 157
Pacientes elegibles N=507
Excluidos n=10
Motivos:
• Hematuria presente menos de 24 h antes de la investigación (n=6)
• Menos de 45 años (n=3)
• La muestra de orina contenía sangre visible (n=1)
Excluidos n=11
Motivo: sin cistoscopia
CCT=Sí n=69
CCT=No n=427
Excluidos n=3
Motivos:
• Muestra de orina <13 después de la cistoscopia o TURBT (n=1)
• Resultado de uRNA-D ausente (n=1)
• Resultado de ELISA de NMP22 ausente (n=1)
Prueba de orina válida n=66
Prueba de orina válida n=419
Excluidos n=8
Motivos:
• Muestra de orina <13 después de la cistoscopia o TURBT (n=1)
• Citología de revisión ausente, uRNA-D y NMP22 ELISA (n=2)
• Fallo de la prueba Bladderchek (n=1)
• Citología de revisión ausente (n=1)
• Resultado de ELISA de NMP22 ausente (n=3)

Figura 8

	Sin diagnóstico específico (N=164)	Enfermedad no maligna (N=255)	TCC (N=66)	Global (N=485)
País, n (%)				
Nueva Zelanda	147 (89,6)	239 (93,7)	64 (97,0)	450 (92,8)
Australia	17 (10,4)	16 (6,27)	2 (3,0)	35 (7,2)
Edad, mediana (Intervalo de IQ)	64 (55, 75)	71 (62, 78)	71 (63, 80)	69 (59, 77)
Sexo, n (%)				
Masculino	112 (68,3)	216 (84,7)	61 (92,4)	389 (80,2)
Femenino	52 (31,7)	39 (15,3)	5 (7,5)	96 (19,8)
Etnia, n (%)				
Europea	133 (81,1)	228 (89,4)	62 (93,9)	423 (87,2)
Maorí	15 (9,1)	17 (6,7)	1 (1,5)	33 (6,8)
Otra	16 (9,8)	10 (3,9)	3 (4,5)	29 (6,0)
Hematuria, n (%)				
Años desde aparición				
<6 meses	131 (79,9)	198 (77,6)	57 (86,4)	386 (79,6)
6 meses-1 año	11 (6,7)	25 (9,8)	6 (9,1)	42 (8,7)
1-2 años	11 (6,7)	18 (7,1)	3 (4,5)	32 (6,6)
3-5 años	5 (3,0)	5 (2,0)	0 (0,0)	10 (2,1)
> 5 años	6 (3,7)	9 (3,5)	0 (0,0)	15 (3,1)
Fin del último episodio				
1-7 días	9 (5,5)	26 (10,2)	20 (30,3)	55 (11,3)
8-28 días	14 (8,5)	27 (10,6)	13 (19,7)	54 (11,1)
1-2 meses	60 (36,6)	86 (33,7)	14 (21,2)	160 (33,0)
3-6 meses	70 (42,7)	100 (39,2)	19 (28,8)	189 (39,0)
>6 meses	11 (6,7)	15 (5,9)	0 (0,0)	26 (5,4)
Desconocido	0 (0,0)	1 (0,4)	0 (0,0)	1 (0,2)
Frecuencia				
<=1 al día	104 (63,4)	160 (62,7)	30 (45,5)	294 (60,6)
>1 al día	59 (36,0)	95 (37,3)	36 (54,5)	190 (39,2)
Desconocido	1 (0,6)	0 (0,0)	0 (0,0)	1 (0,2)
Estado de tabaquismo, n (%)				
Fumador actual	34 (20,7)	28 (11,0)	14 (21,2)	76 (15,7)
Exfumador	66 (40,2)	114 (44,7)	35 (53,0)	215 (44,3)
Nunca ha fumado	64 (39,0)	113 (44,3)	17 (25,8)	194 (40,0)

Figura 9

Prueba	Sensibilidad general	Especificidad general
uRNA-D	62,1% (49,3 73,8)	85,2% (corregido)
ELISA de NMP22	50,0% (37,4, 62,6)	88,0% (84,6, 91,0)
BladderChek de NMP22	37,9% (26,2, 50,7)	96,4% (94,2, 98,0)
Citología*	56,1% (43,3, 68,3)	94,5 (91,9, 96,5)
Cinco genes (incluye IL8BRb)†	81,8%	85,1% (corregido)
Cinco genes (incluye IL8BRb)† (90 % de especificidad)	72,7%	89,9% (corregido)

* Incluido en la determinación de CCT para algunos pacientes

† Determinado en los datos del estudio, por lo que no se presentan intervalos de confianza

Figura 10

Fig 11a.

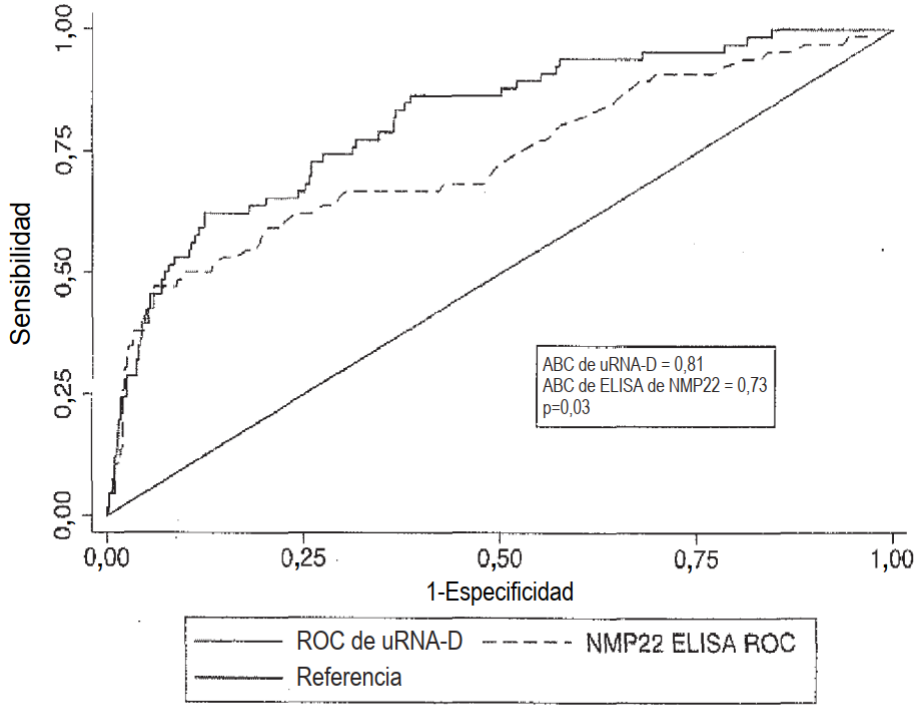


Fig 11b.

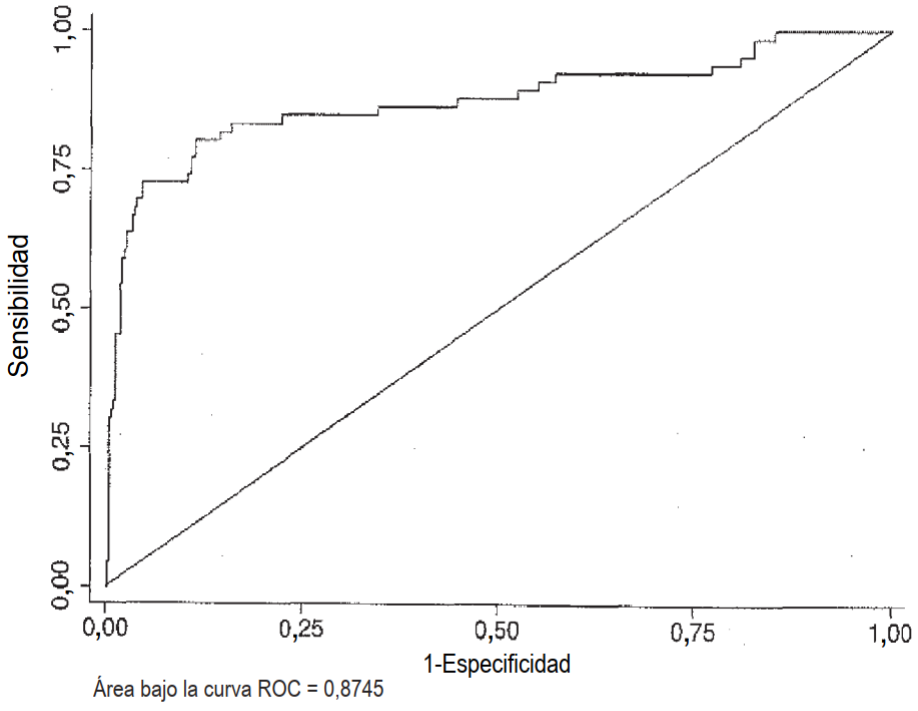


Figura 11

	Citología	ELISA de NMP22	BladderChek	uRNA-D	Cinco marcadores
Estadio (n=66)	**n, % (95% CI)**	**n, % (95% CI)**	**n, % (95% CI)**	**n, % (95% CI)**	**n, %**
Tis (n=2)	2, 100% (16, 100)	0, 0% (0, 84)	0, 0% (0, 84)	1, 50% (1, 99)	2, 100%
Ta (n=37)	13, 35% (20, 53)	13, 35% (20, 53)	14, 38% (22, 55)	15, 41% (25, 58)	25, 68%
T1 (n=16)	11, 69% (41, 89)	12, 75% (48, 93)	8, 50% (25, 75)	15, 94 (70, 100)	16, 100%
T2 (n=9)	9, 100% (66, 100)	6, 67% (30, 93)	2, 22% (3, 60)	9, 100 (66, 100)	9, 100%
≥T3 (n=2)	2, 100% (16, 100)	2, 100% (16, 100)	1, 50% (11, 99)	1, 50% (1.3, 99)	2, 100%
Grado WHO 73 (n=65)					*p*=0,019*
1 (n=3)	1, 33% (1, 91)	1, 33% (1, 91)	1, 33% (1, 91))	1, 33% (1, 91))	1, 33%
2 (n=38)	17, 44% (29, 62)	15, 39% (24, 57)	13, 34% (20, 51)	21, 55% (38, 71)	29, 76%
3 (n=24)	19, 79% (58, 93)	17, 71% (49, 87)	11, 46% (26, 67)	19, 79% (58, 93)	23, 96%
Grado WHO ISUP 1998 (n=65)					*p*=0,016*
Bajo (n=32)	9, 28% (14, 47)	10, 31% (16, 50)	13, 41% (24, 59)	13, 41% (24, 59)	22, 69%
Mixto (n=4)	4, 100% (40, 100)	3, 75% (19, 99)	1, 25% (1, 81)	4, 100 (40,100)	4, 100%
Alto (n=29)	24, 83% (64, 94)	20, 69% (49, 85)	11, 38% (21, 58)	24, 83% (64, 94)	28, 97%
Ubicación del tumor (n=66)					*p*=0,012*
Vejiga (n=62)	35, 56% (43, 69)	30, 48% (35, 61)	25, 40% (28. 54)	39, 63% (50, 75)	50, 81%
Tracto sup. (n=4)	2, 50% (7, 93)	3, 75% (19, 99)	0, 0% (0, 60)	2, 50% (7, 93)	4, 100%
Multiplicidad (n=66)					*p*=0,33*
Única (n=52)	27, 52% (38. 66)	23, 44% (30. 59)	17, 33% (20, 47)	30, 58% (43, 71)	41, 79%
Multifocal (n=13)	10, 77% (46, 95)	10, 77% (46, 95)	8, 62% (32, 86)	11, 85% (55, 98)	12, 92%
Microhematuria (n=65)					*p*=0,43*
Sí (n=43)	27, 63% (47, 77)	27, 63% (47, 77)	20, 47% (31, 62)	31, 72% (56, 85)	41, 95%
No (n=22)	9, 41% (21, 64))	5, 23% (8, 45)	4, 18% (5, 40)	9, 41% (21, 64)	12, 55%
Creatinina (n=55)					*p*<0,0005*
0-4,3 (n=11)	5, 45 (17, 77)	3, 27% (6,61)	4, 36% (11, 69)	4, 36% (11, 69)	7, 64%
4,3-7,1 (n=14)	9, 64% (35, 87)	6, 43% (18,71)	4, 29% (8, 58)	8, 57% (29, 82)	12, 86%
7,1-11,2 (n=22)	11, 50% (28. 72)	14, 64% (41, 83)	11, 50% (28. 72)	16, 73% (50, 89)	18, 82%
≥11,2 (n=8)	6, 75% (35, 97)	3, 38% (9, 76)	2, 25% (3, 65)	6, 75% (35, 97)	8, 100%
Sexo (n=66)					*p*=0,29*
Masculino (n=61)	35, 57% (44, 70)	29, 48% (35, 61)	23, 38% (26, 51)	37, 61% (47, 73)	49, 80%
Femenino (n=5)	2, 40% (5, 85)	4, 80% (28, 99)	2, 40% (5, 85)	4, 80% (28, 99)	5, 100%
					p=0,58*

*valores de p de la prueba exacta de Fisher de asociación entre cada característica de TCC y el resultado de la prueba Cxbladder

Figura 12

	Citología	ELISA de NMP22	BladderChek de NMP22	uRNA-D*	Cinco marcadores*
	n, % 95% CI)	n, % 95% CI)	n, % 95% CI)	n, %	n ,%
Diagnóstico					
Sin diagnóstico (n=164)					
	154, 94% (89, 97)	144, 88% (82, 92)	160, 98% (94, 99)	144, 88%	144, 88%
Diagnóstico no maligno					
Hipertrofia benigna de próstata/prostatitis (n=130)					
	123, 95% (89, 98)	117, 90.% (84, 95)	127, 98% (93, 100)	113, 87%	109/128=85%
Cistitis/infección o inflamación del tracto urinario (n=39)					
	36, 92% (79, 98)	34, 87% (73, 96)	34, 87% (73, 96)	28, 72%	32,,82%
Cálculos (n=28)					
	25, 89% (72, 98)	23, 82% (63, 94)	27, 96% (82, 100)	20. 71%	19, 68%
Hematuria secundaria a Warfarina (n=10)					
	10, 100% (69, 100)	9, 90% (55, 100)	10, 100% (69, 100)	8, 80%	8, 80%
Otro cáncer urológico (n=5)					
	5, 100% (48, 100)	4, 80% (28, 99)	5, 100% (48, 100)	5, 100%	4, 80%
Microhematuria (n=417)					*p†=0,12*
Sí (n=99)	87, 88% (80, 94)	81, 82% (73, 89)	91, 92% (85, 96)	74, 75%	73/97=75%
No (n=318)	307, 97% (94, 98)	287, 90% (86, 93)	311, 98% (96, 99)	281, 88%	279/317=88%
Creatinina (n=386)					*p†=0,002*
0 – 4,3, n=97	95, 98% (93, 100)	90, 93% (86, 97)	95, 98% (93, 100)	87, 90%	83/96=86%
4,3 – 7,1, n=94	92, 98% (93, 100)	81, 86% (78, 92)	90, 96% (89, 99)	81, 86%	84/93=90%
7,1 – 11,2, n=92	79, 86% (77, 92)	83, 90% (82, 95)	90, 98% (92, 100)	76, 83%	76/92=83%
≥11,2, n=103	101 98% (93, 100)	91, 88% (81, 94)	101, 98% (93, 100)	88, 85%	82/102=80%
Sexo (n=419)					*p†=0,32*
Masculino (n=328)	311, 95% (92, 97)	290, 88% (84, 92)	316, 96% (94, 98)	291, 89%	280/325=86%
Femenino (n=91)	85, 93% (86, 98)	79, 87% (78, 93)	88 97% (91, 99)	65, 71%	74/91=81%
					p†=0,25

* especificidad general fijada al 85 %

†valor de p de la prueba de Chi cuadrado de asociación entre cada característica (diagnóstico, microhematuria, creatinina, sexo) y el resultado de la prueba Cxbladder

Figura 13

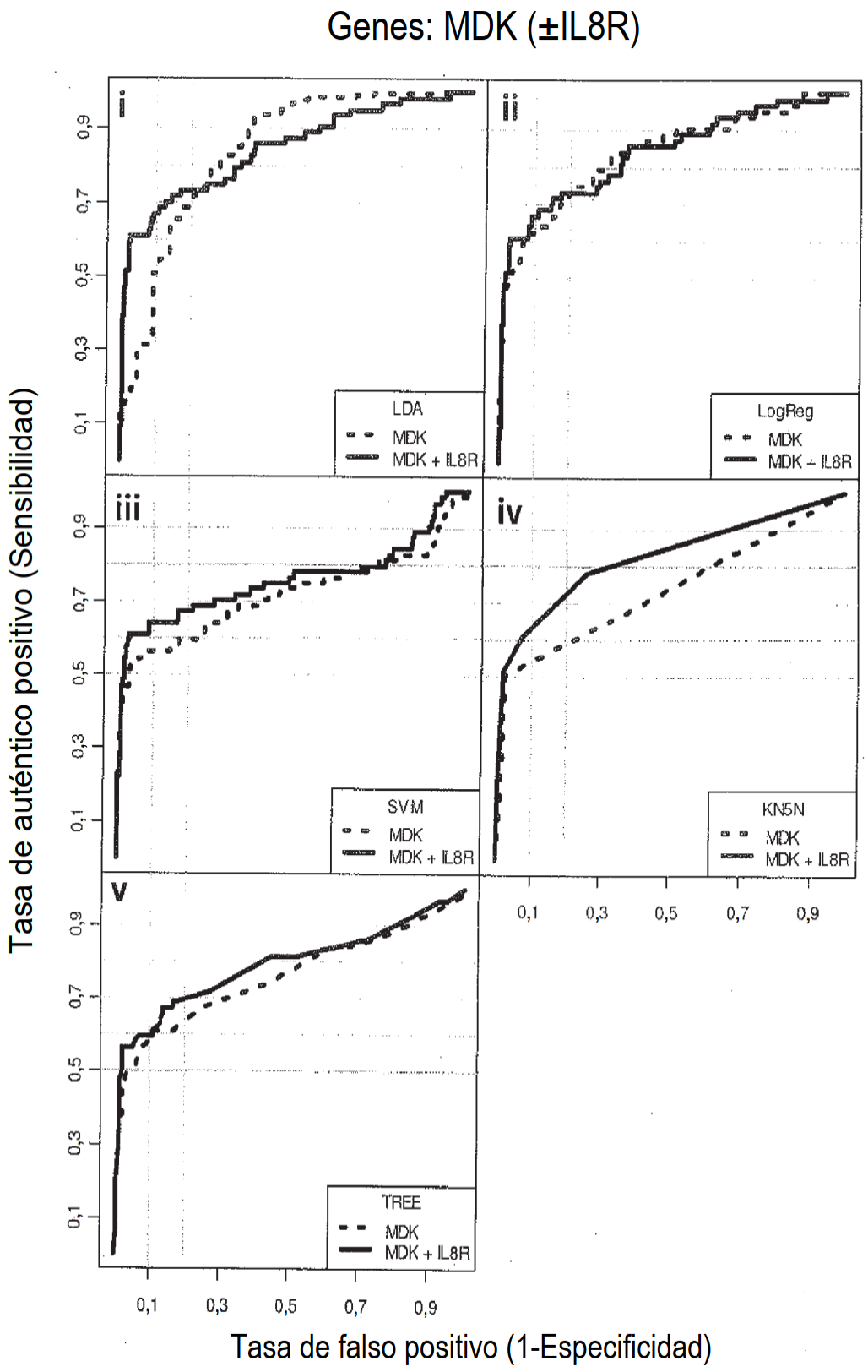

Genes: MDK (±IL8R)
Tasa de auténtico positivo (Sensibilidad)
Tasa de falso positivo (1-Especificidad)
i
LDA
MDK
MDK + IL8R
ii
LogReg
MDK
MDK + IL8R
iii
SVM
MDK
MDK + IL8R
iv
KN5N
MDK
MDK + IL8R
v
TREE
MDK
MDK + IL8R
0,1
0,3
0,5
0,7
0,9

Figura 14a

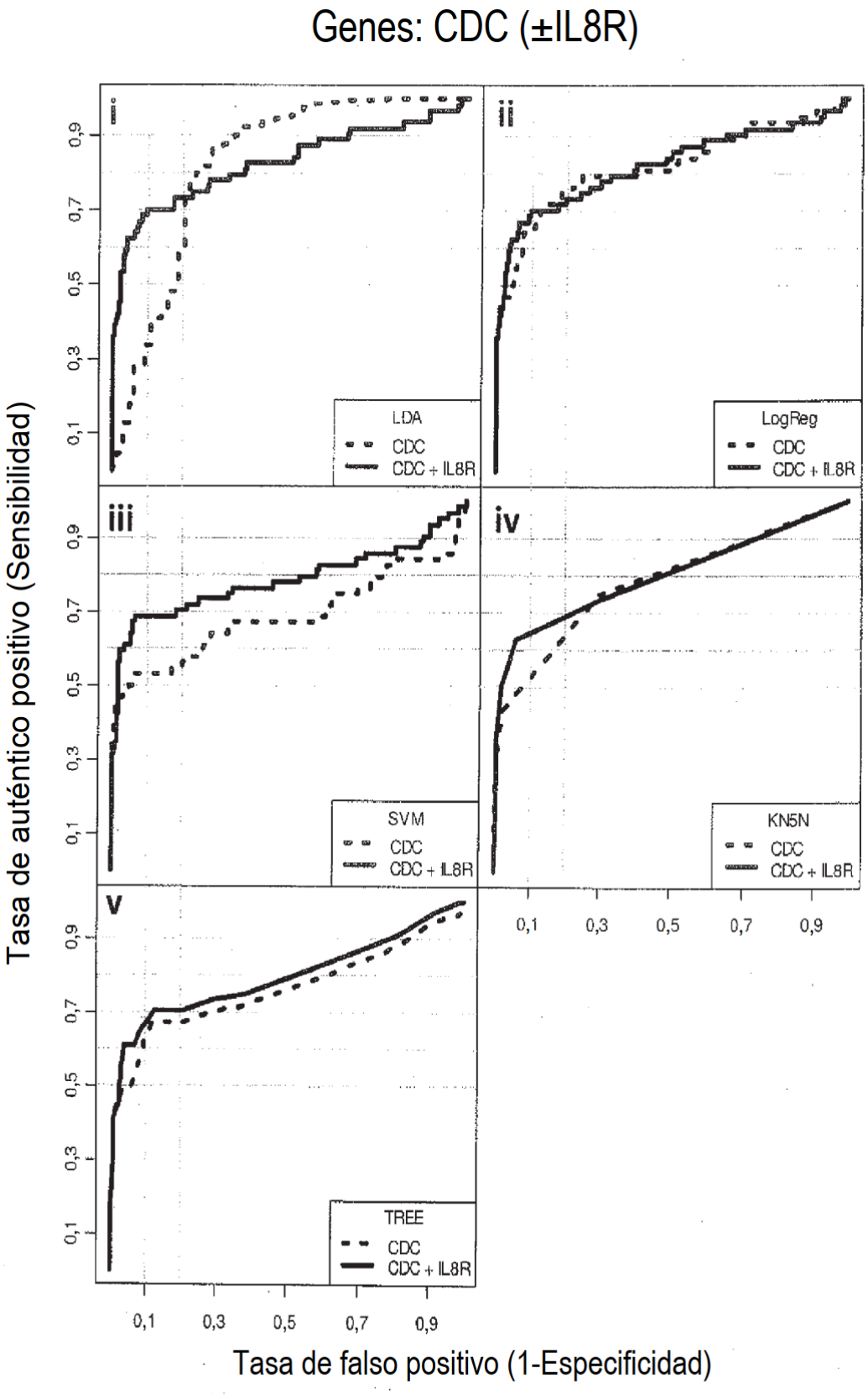

Genes: CDC (±IL8R)
Tasa de auténtico positivo (Sensibilidad)
Tasa de falso positivo (1-Especificidad)
i
ii
iii
iv
v
LDA
LogReg
SVM
KN5N
TREE
CDC
CDC + IL8R
0,1
0,3
0,5
0,7
0,9

Figura 14b

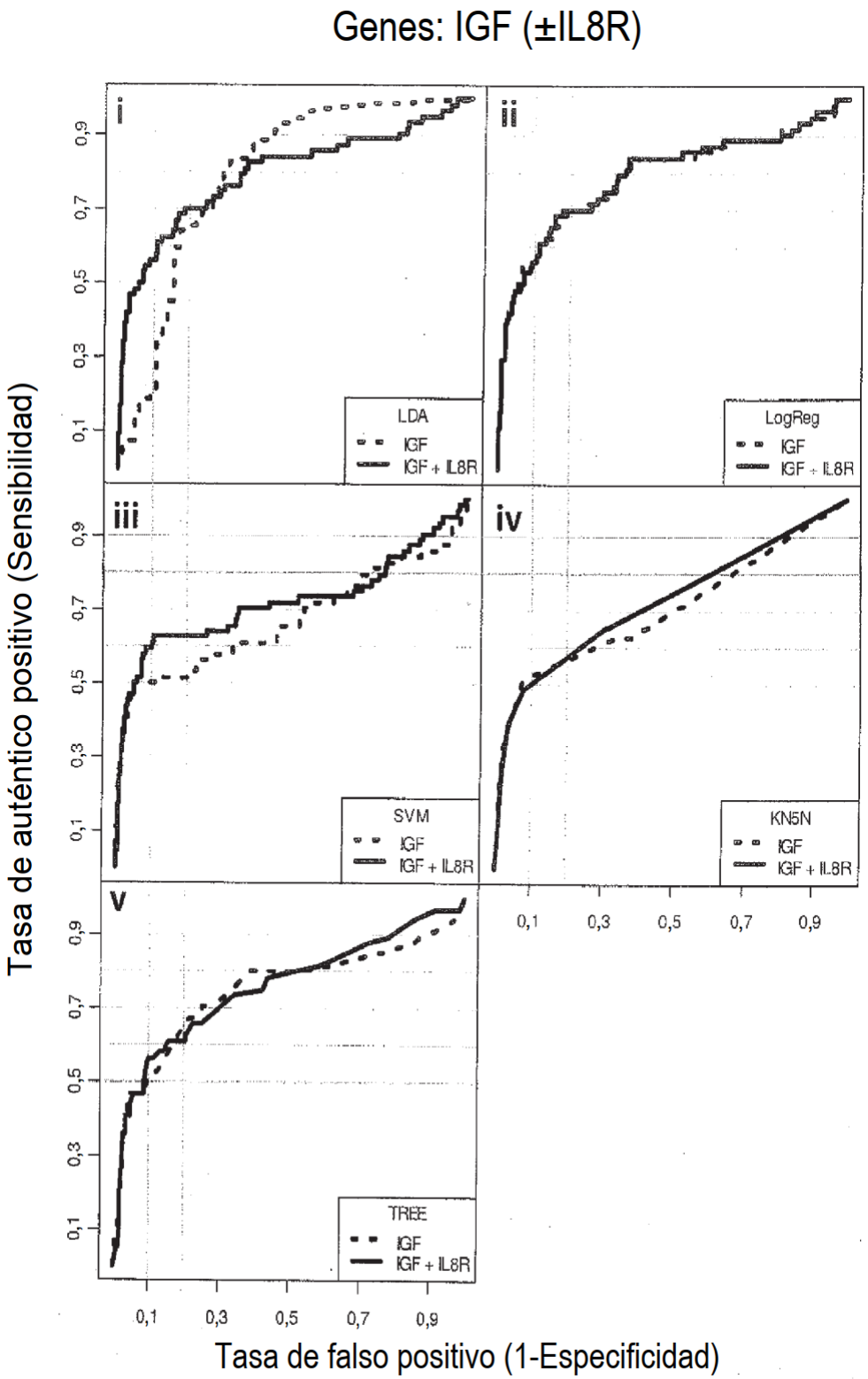

Genes: IGF (±IL8R)
Tasa de auténtico positivo (Sensibilidad)
Tasa de falso positivo (1-Especificidad)
i
LDA
IGF
IGF + IL8R
ii
LogReg
IGF
IGF + IL8R
iii
SVM
IGF
IGF + IL8R
iv
KN5N
IGF
IGF + IL8R
v
TREE
IGF
IGF + IL8R
0,1
0,3
0,5
0,7
0,9

Figura 14c

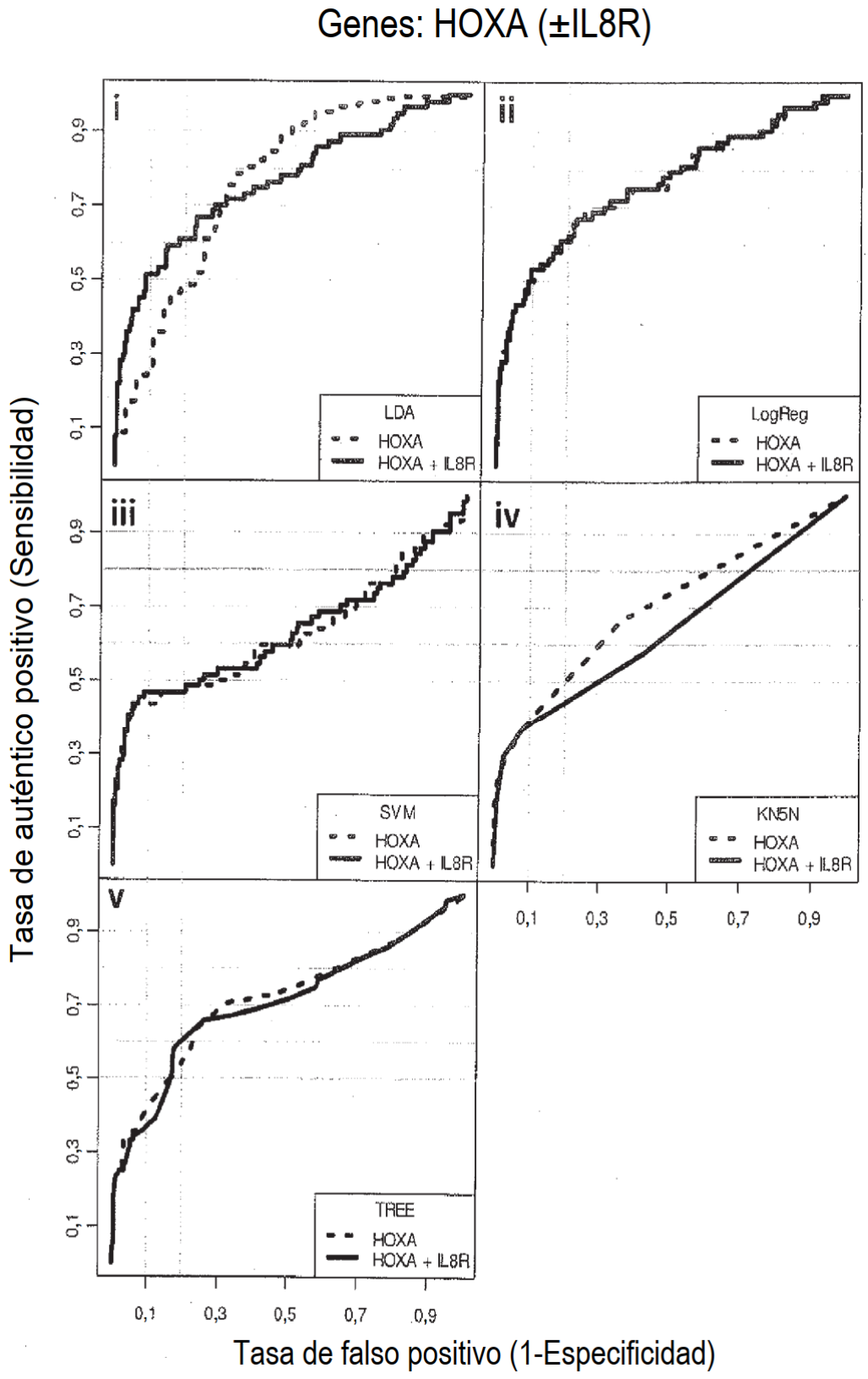

Genes: HOXA (±IL8R)
i
LDA
HOXA
HOXA + IL8R
ii
LogReg
HOXA
HOXA + IL8R
iii
SVM
HOXA
HOXA + IL8R
iv
KN5N
HOXA
HOXA + IL8R
v
TREE
HOXA
HOXA + IL8R
0,1
0,3
0,5
0,7
0,9
Tasa de auténtico positivo (Sensibilidad)
Tasa de falso positivo (1-Especificidad)

Figura 14d

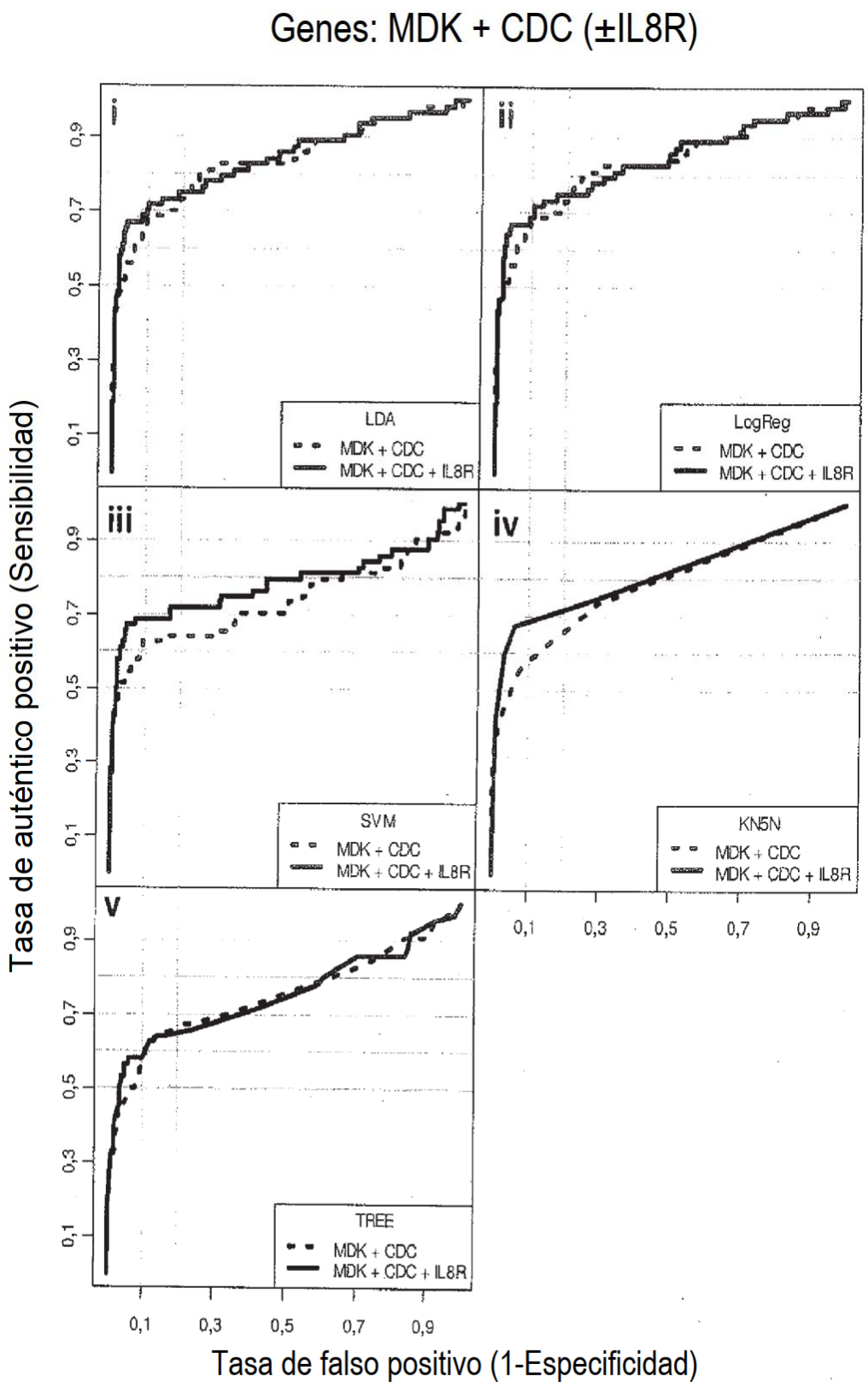

Genes: MDK + CDC (±IL8R)
Tasa de auténtico positivo (Sensibilidad)
Tasa de falso positivo (1-Especificidad)
i
LDA
MDK + CDC
MDK + CDC + IL8R
ii
LogReg
MDK + CDC
MDK + CDC + IL8R
iii
SVM
MDK + CDC
MDK + CDC + IL8R
iv
KN5N
MDK + CDC
MDK + CDC + IL8R
v
TREE
MDK + CDC
MDK + CDC + IL8R
0,1
0,3
0,5
0,7
0,9

Figura 14e

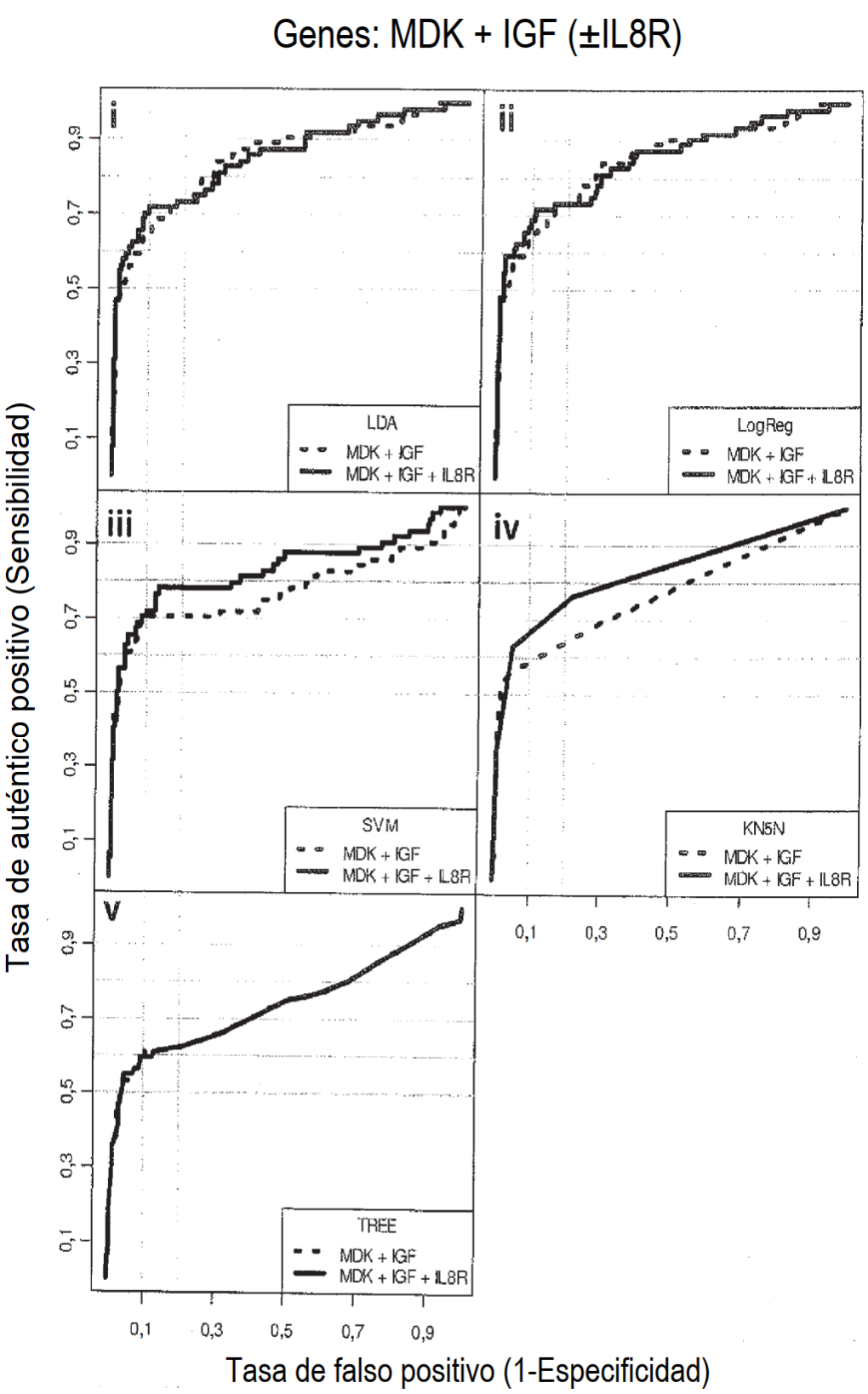

Genes: MDK + IGF (±IL8R)
i
ii
iii
iv
v
LDA
LogReg
SVM
KN5N
TREE
MDK + IGF
MDK + IGF + IL8R
Tasa de auténtico positivo (Sensibilidad)
Tasa de falso positivo (1-Especificidad)
0,1
0,3
0,5
0,7
0,9

Figura 14f

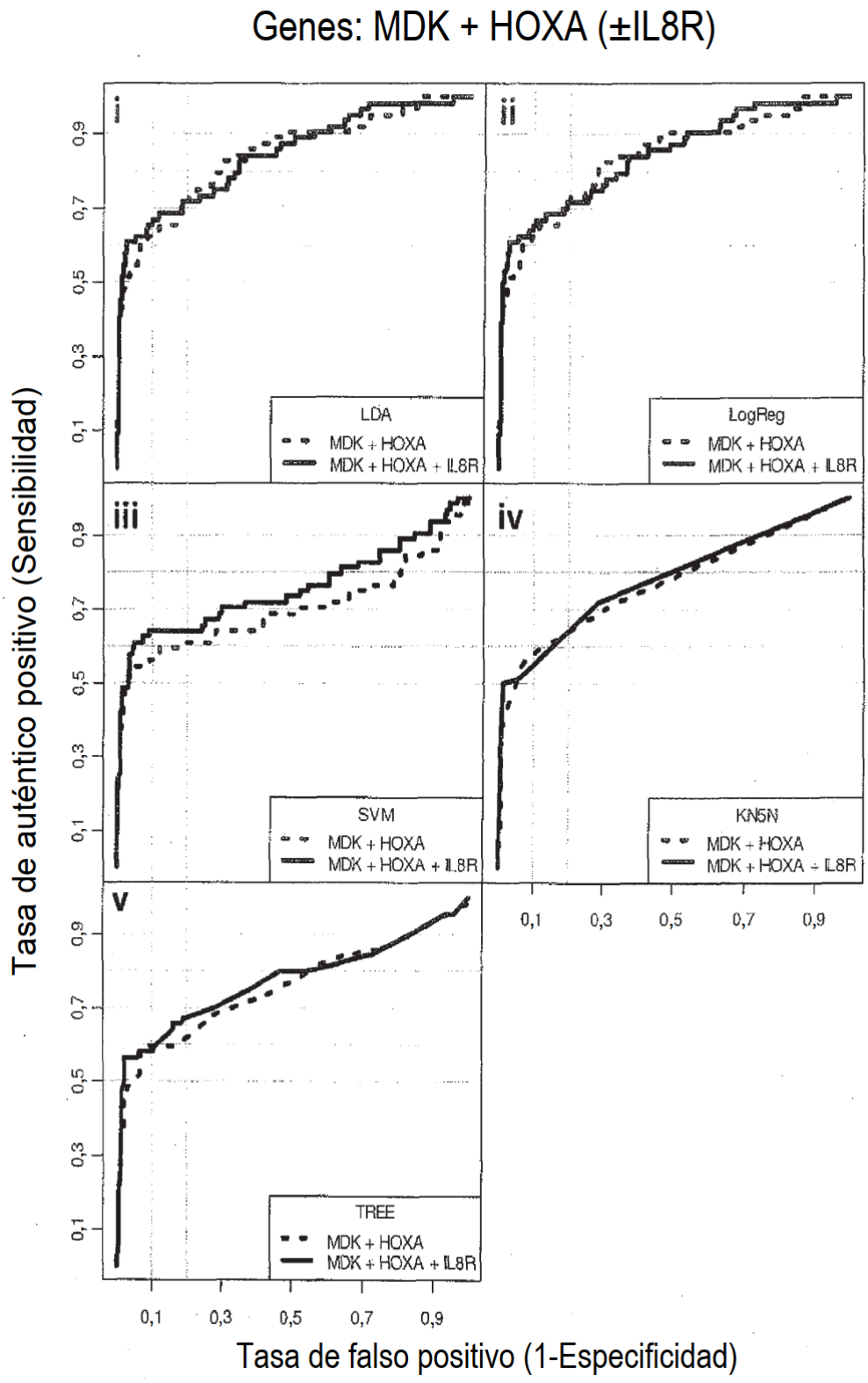

Genes: MDK + HOXA (±IL8R)
Tasa de auténtico positivo (Sensibilidad)
Tasa de falso positivo (1-Especificidad)
i
ii
iii
iv
v
LDA
LogReg
SVM
KNN
TREE
MDK + HOXA
MDK + HOXA + IL8R
0,1
0,3
0,5
0,7
0,9

Figura 14g

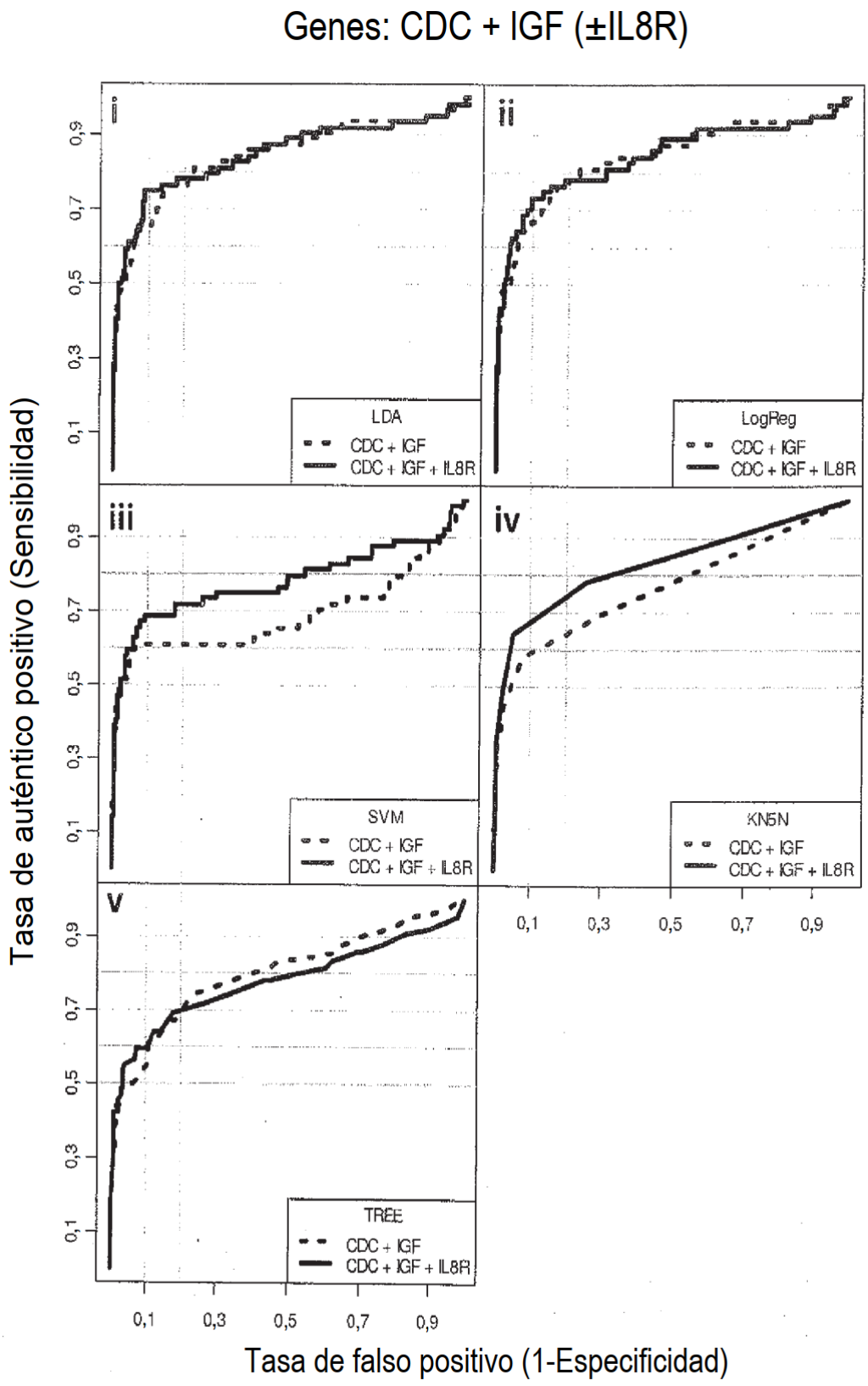

Genes: CDC + IGF (±IL8R)
Tasa de auténtico positivo (Sensibilidad)
Tasa de falso positivo (1-Especificidad)
i
ii
iii
iv
v
LDA
LogReg
SVM
KNSN
TREE
CDC + IGF
CDC + IGF + IL8R
0,1
0,3
0,5
0,7
0,9

Figura 14h

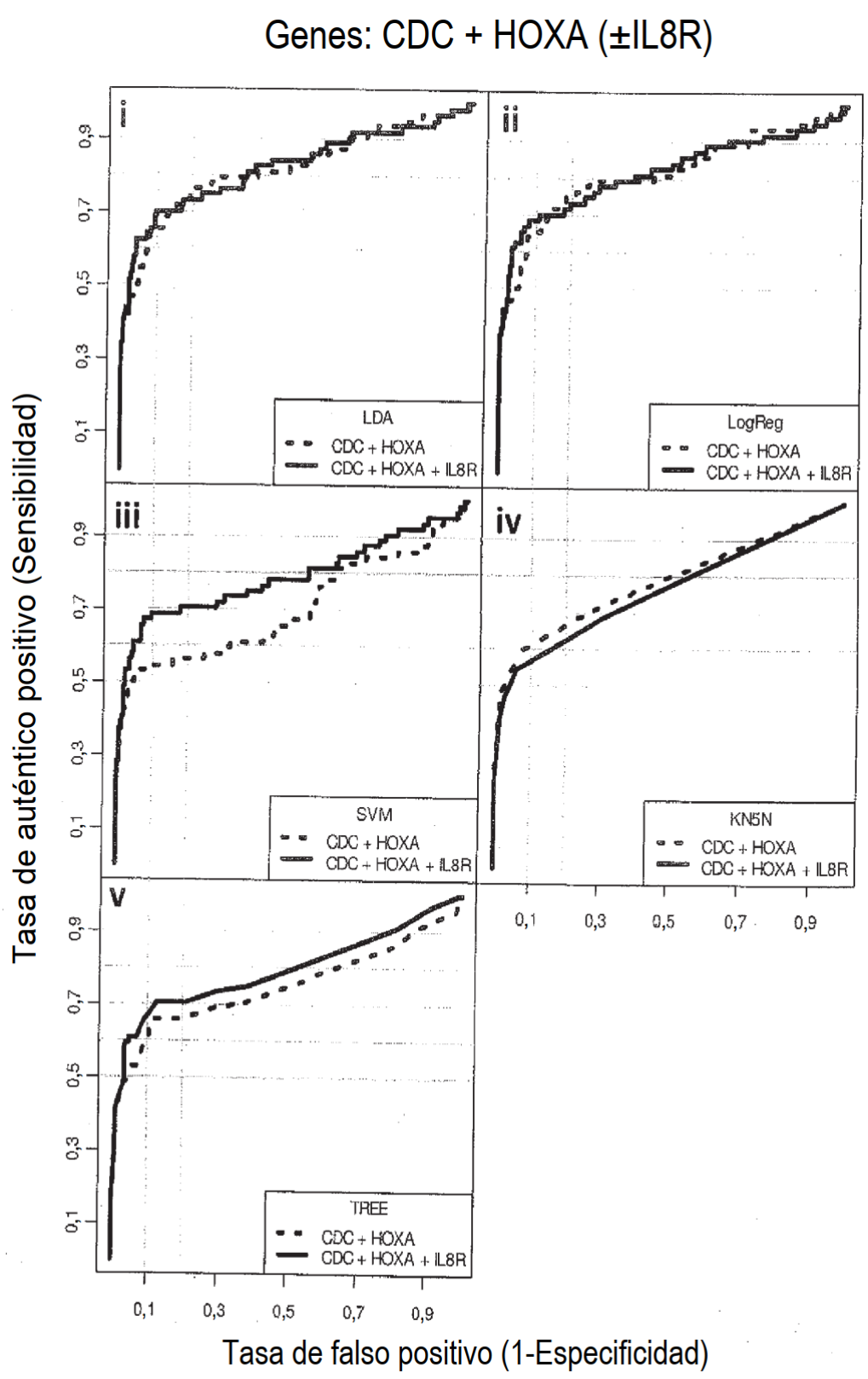

Genes: CDC + HOXA (±IL8R)
Tasa de auténtico positivo (Sensibilidad)
Tasa de falso positivo (1-Especificidad)
i
LDA
CDC + HOXA
CDC + HOXA + IL8R
ii
LogReg
CDC + HOXA
CDC + HOXA + IL8R
iii
SVM
CDC + HOXA
CDC + HOXA + IL8R
iv
KN5N
CDC + HOXA
CDC + HOXA + IL8R
v
TREE
CDC + HOXA
CDC + HOXA + IL8R
0,1
0,3
0,5
0,7
0,9

Figura 14i

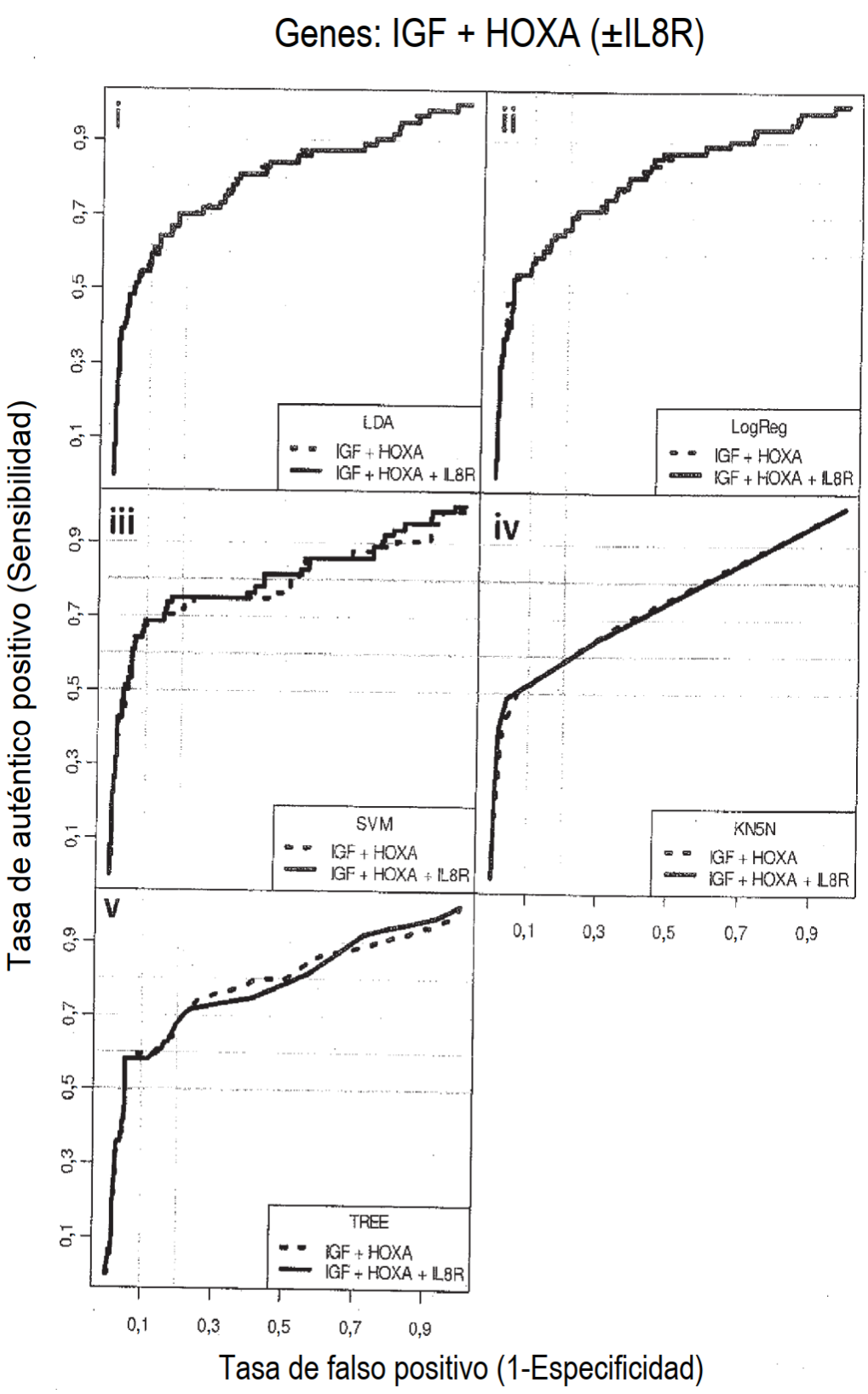

Genes: IGF + HOXA (±IL8R)
Tasa de auténtico positivo (Sensibilidad)
Tasa de falso positivo (1-Especificidad)
i
LDA
IGF + HOXA
IGF + HOXA + IL8R
ii
LogReg
IGF + HOXA
IGF + HOXA + IL8R
iii
SVM
IGF + HOXA
IGF + HOXA + IL8R
iv
KN5N
IGF + HOXA
IGF + HOXA + IL8R
v
TREE
IGF + HOXA
IGF + HOXA + IL8R
0,1
0,3
0,5
0,7
0,9

Figura 14j

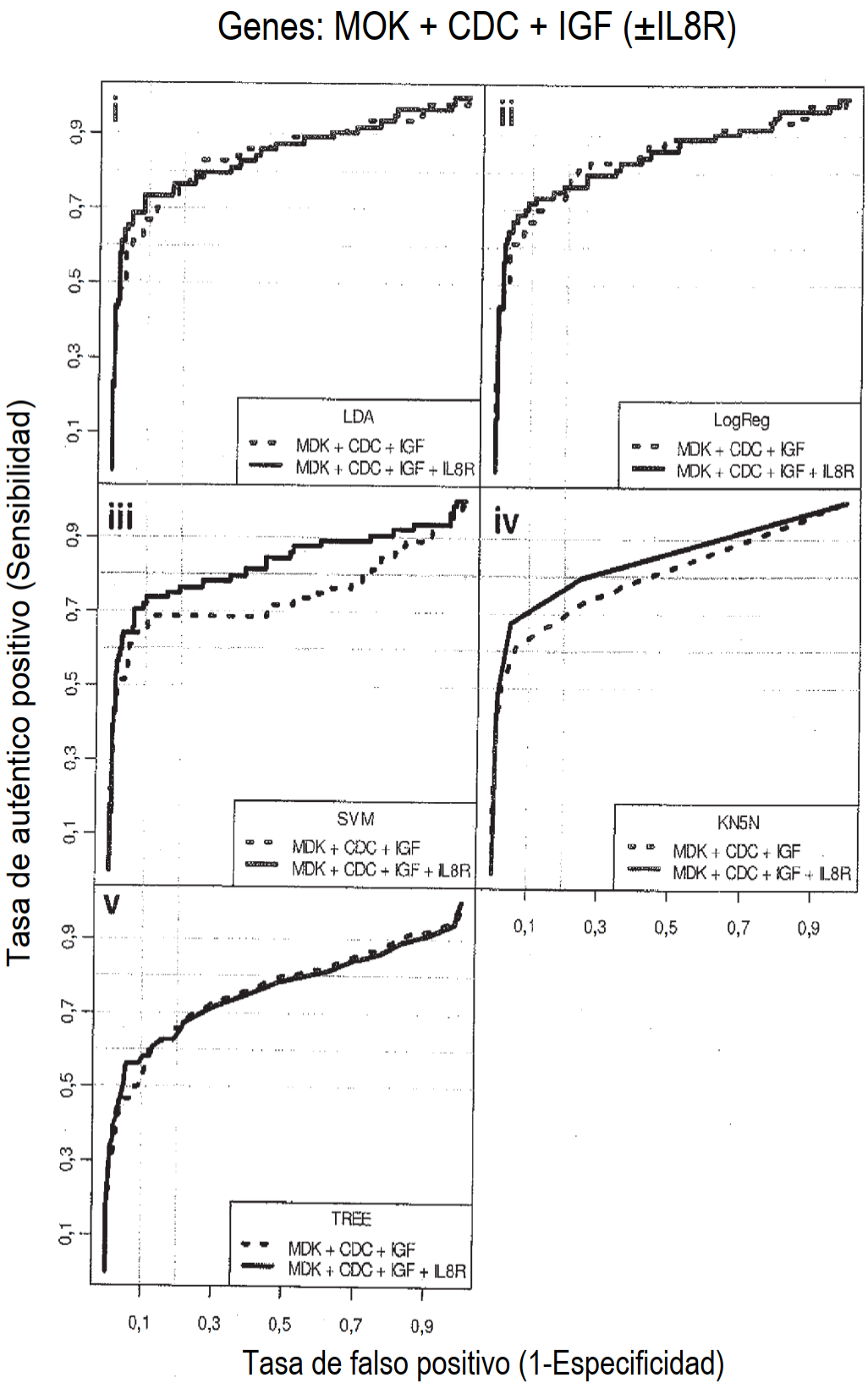

Genes: MOK + CDC + IGF (±IL8R)
Tasa de auténtico positivo (Sensibilidad)
Tasa de falso positivo (1-Especificidad)
i
LDA
MDK + CDC + IGF
MDK + CDC + IGF + IL8R
ii
LogReg
MDK + CDC + IGF
MDK + CDC + IGF + IL8R
iii
SVM
MDK + CDC + IGF
MDK + CDC + IGF + IL8R
iv
KNSN
MDK + CDC + IGF
MDK + CDC + IGF + IL8R
v
TREE
MDK + CDC + IGF
MDK + CDC + IGF + IL8R
0,1
0,3
0,5
0,7
0,9

Figura 14k

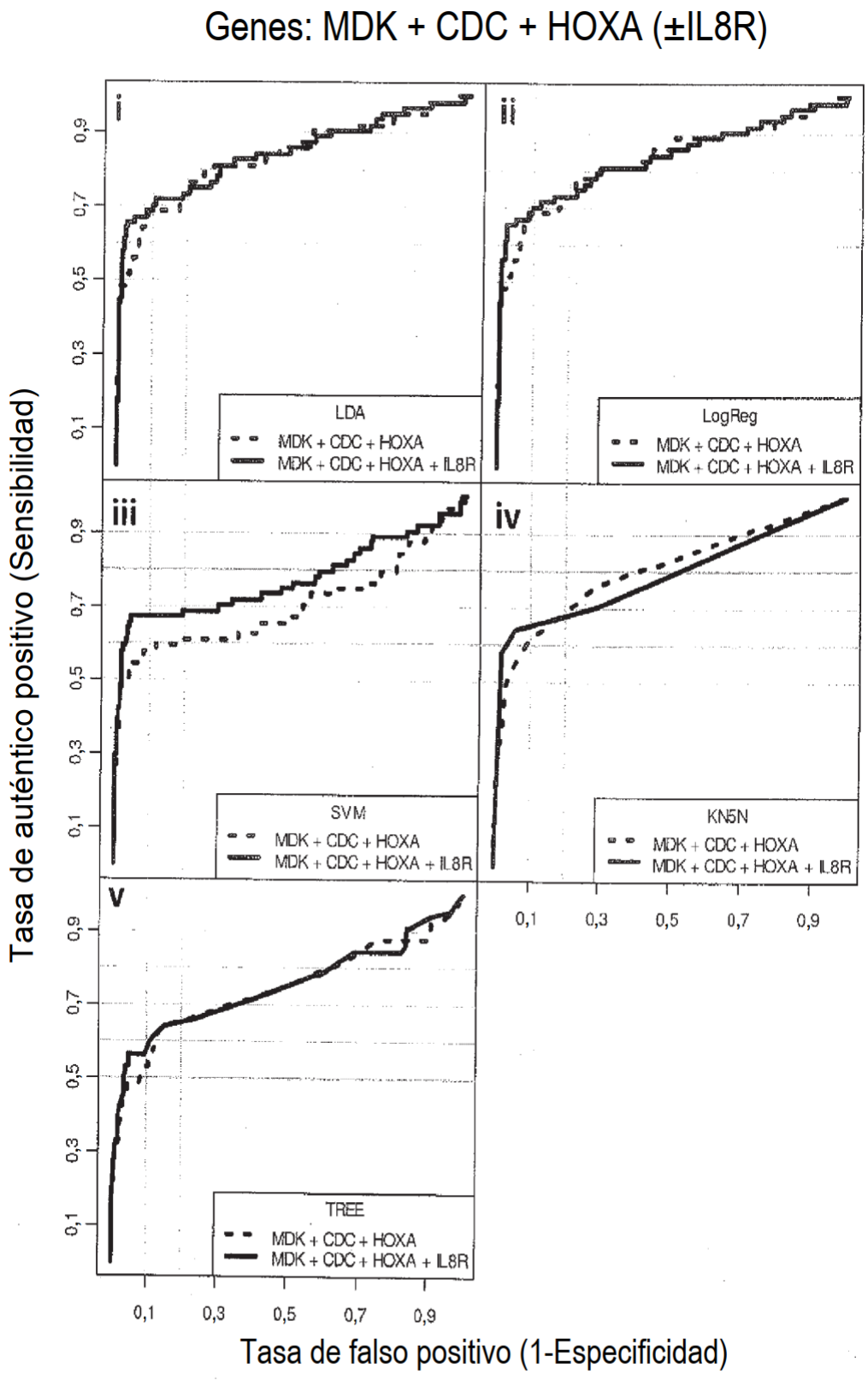

Genes: MDK + CDC + HOXA (±IL8R)
Tasa de auténtico positivo (Sensibilidad)
Tasa de falso positivo (1-Especificidad)
i
LDA
MDK + CDC + HOXA
MDK + CDC + HOXA + IL8R
ii
LogReg
MDK + CDC + HOXA
MDK + CDC + HOXA + IL8R
iii
SVM
MDK + CDC + HOXA
MDK + CDC + HOXA + IL8R
iv
KN5N
MDK + CDC + HOXA
MDK + CDC + HOXA + IL8R
v
TREE
MDK + CDC + HOXA
MDK + CDC + HOXA + IL8R
0,1
0,3
0,5
0,7
0,9

Figura 14l

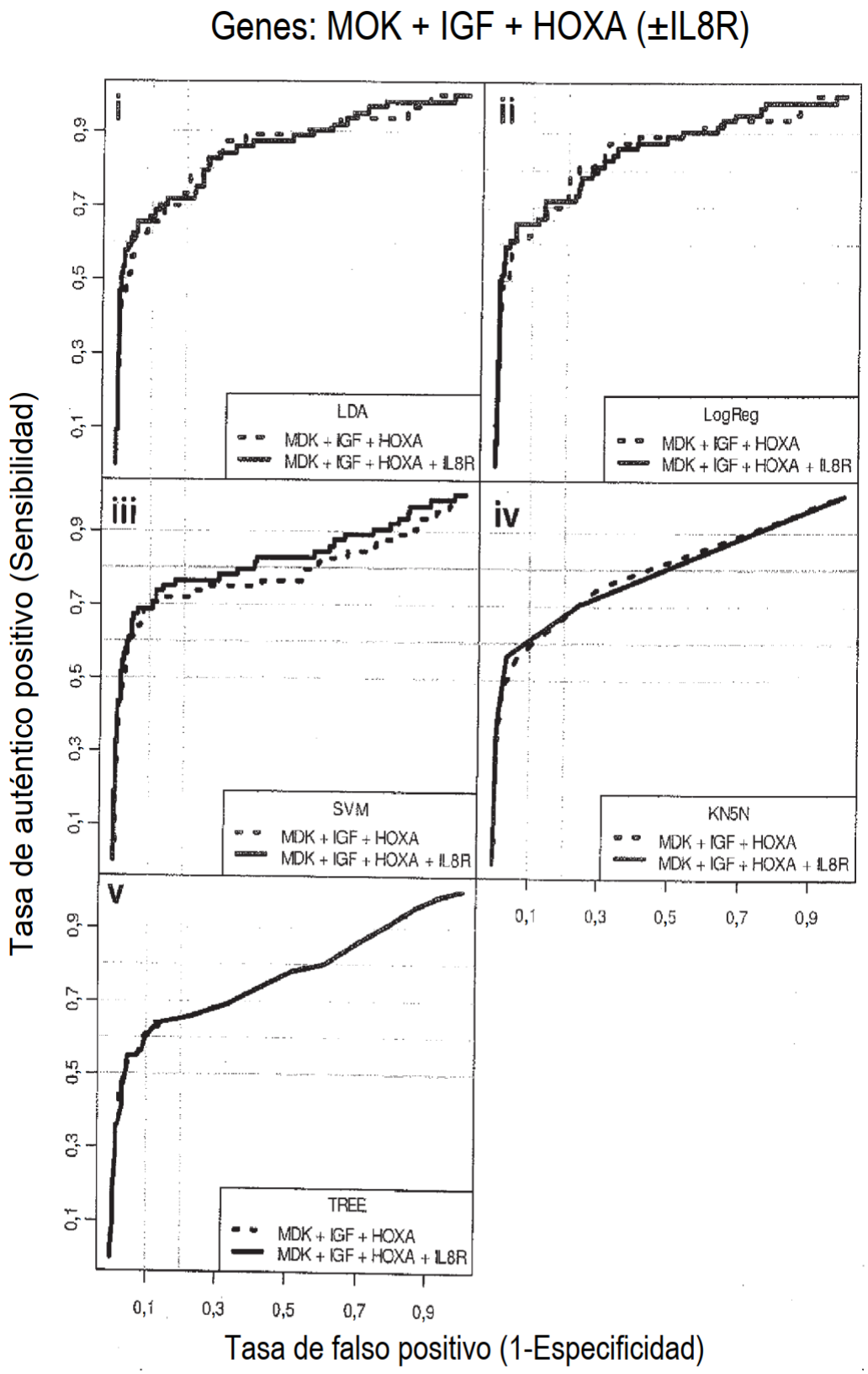

Genes: MOK + IGF + HOXA (±IL8R)
i
ii
iii
iv
v
LDA
LogReg
SVM
KN5N
TREE
MDK + IGF + HOXA
MDK + IGF + HOXA + IL8R
Tasa de auténtico positivo (Sensibilidad)
Tasa de falso positivo (1-Especificidad)
0,1
0,3
0,5
0,7
0,9

Figura 14m

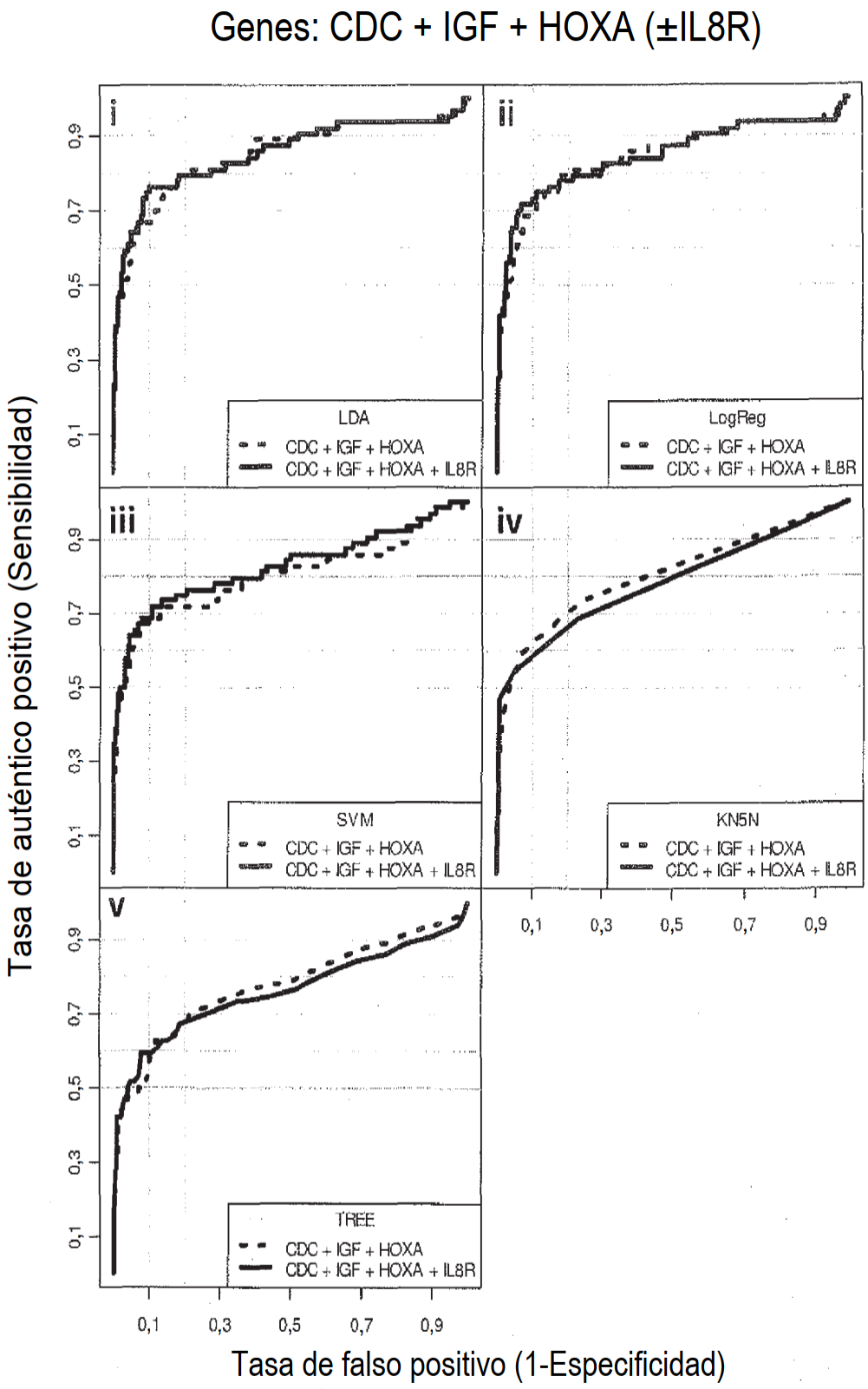

Genes: CDC + IGF + HOXA (±IL8R)
i
ii
iii
iv
v
LDA
LogReg
SVM
KN5N
TREE
CDC + IGF + HOXA
CDC + IGF + HOXA + IL8R
Tasa de auténtico positivo (Sensibilidad)
Tasa de falso positivo (1-Especificidad)
0,1
0,3
0,5
0,7
0,9

Figura 14n

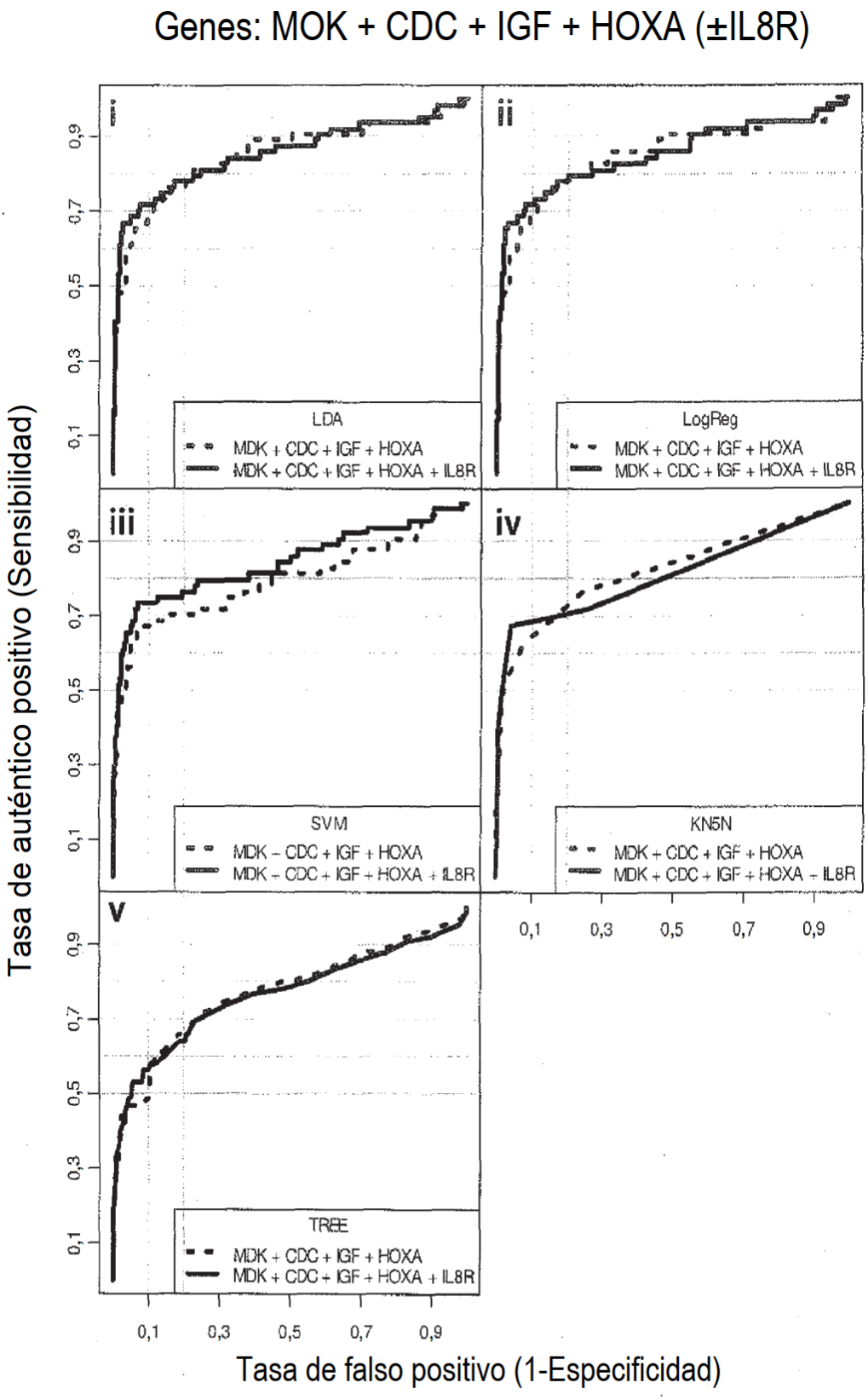

Genes: MOK + CDC + IGF + HOXA (±IL8R)
Tasa de auténtico positivo (Sensibilidad)
Tasa de falso positivo (1-Especificidad)
i
ii
iii
iv
v
LDA
LogReg
SVM
KNN
TREE
MDK + CDC + IGF + HOXA
MDK + CDC + IGF + HOXA + IL8R
0,1
0,3
0,5
0,7
0,9

Figura 14o

Modelo	LDA		LogReg		SVM		KN5N		TREE	
	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb
MDK	0,087	0,127	0,119	0,127	0,107	0,121	0,100	0,120	0,107	0,117
CDC2	0,062	0,129	0,124	0,129	0,102	0,127	0,102	0,121	0,114	0,124
IGFBP5	0,060	0,108	0,109	0,108	0,094	0,107	0,093	0,090	0,093	0,096
HOXA13	0,058	0,094	0,094	0,094	0,082	0,084	0,075	0,071	0,078	0,075
MDK + CDC2	0,124	0,134	0,124	0,133	0,115	0,131	0,109	0,128	0,107	0,111
MDK + IGFBP5	0,122	0,129	0,121	0,129	0,125	0,133	0,111	0,122	0,107	0,107
MDK + HOXA13	0,119	0,126	0,118	0,125	0,108	0,118	0,107	0,107	0,107	0,114
CDC2 + IGFBP5	0,128	0,134	0,127	0,133	0,114	0,125	0,109	0,125	0,109	0,115
CDC + HOXA13	0,122	0,128	0,123	0,129	0,103	0,125	0,116	0,108	0,113	0,123
IGF + HOXA13	0,109	0,109	0,108	0,107	0,119	0,121	0,096	0,098	0,102	0,101
MDK + CDC2 + IGFBP5	0,127	0,135	0,127	0,135	0,123	0,134	0,119	0,132	0,103	0,108
MDK + CDC2 + HOXA13	0,125	0,133	0,125	0,134	0,109	0,127	0,113	0,122	0,104	0,111
MDK + IGFBP5 + HOXA13	0,122	0,128	0,121	0,127	0,126	0,132	0,111	0,115	0,110	0,109
CDC2 + IGFBP5 + HOXA13	0,129	0,136	0,131	0,136	0,126	0,131	0,117	0,114	0,108	0,111
MDK+CDC2+IGFBP5+HOXA13	0,130	0,137	0,130	0,138	0,124	0,137	0,122	0,129	0,103	0,106

Area bajo la curva (AUC) ROC para la tasa de falso positivo de 0,00 a 0,20

Figura 15a

Modelo	LDA	LogReg	SVM	KN5N	TREE
MDK	0,040	0,008	0,014	0,019	0,010
CDC2	0,067	0,005	0,025	0,018	0,009
IGFBP5	0,048	-0,001	0,013	-0,003	0,002
HOXA13	0,036	0,000	0,001	-0,004	-0,003
MDK + CDC2	0,009	0,009	0,015	0,019	0,005
MDK + IGFBP5	0,007	0,007	0,008	0,011	-0,000
MDK + HOXA13	0,007	0,007	0,010	0,000	0,007
CDC2 + IGFBP5	0,007	0,006	0,012	0,016	0,006
CDC + HOXA13	0,006	0,006	0,022	-0,007	0,010
IGF + HOXA13	0,000	-0,001	0,002	0,002	-0,000
MDK + CDC2 + IGFBP5	0,009	0,008	0,012	0,013	0,004
MDK + CDC2 + HOXA13	0,008	0,008	0,018	0,009	0,006
MDK + IGFBP5 + HOXA13	0,006	0,006	0,006	0,004	-0,001
CDC2 + IGFBP5 + HOXA13	0,006	0,006	0,005	-0,003	0,004
MDK+CDC2+IGFBP5+HOXA13	0,007	0,007	0,013	0,007	0,002
Aumento en la ABC (frente a FPR de 0,0 a 0,2)					

Figura 15b

a	Modelo	LDA		LogReg		SVM		KN5N		TREE	
		-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb
	MDK	0,721	0,734	0,734	0,734	0,594	0,672	0,531	0,781	0,625	0,688
	CDC2	0,753	0,734	0,750	0,734	0,562	0,703	0,750	0,734	0,672	0,703
	IGFBP5	0,657	0,703	0,688	0,703	0,516	0,625	0,516	0,656	0,656	0,609
	HOXA13	0,486	0,609	0,625	0,609	0,484	0,469	0,406	0,375	0,547	0,578
	MDK + CDC2	0,734	0,750	0,734	0,750	0,641	0,719	0,734	0,750	0,672	0,641
	MDK + IGFBP5	0,734	0,734	0,734	0,734	0,703	0,781	0,688	0,766	0,625	0,625
	MDK + HOXA13	0,719	0,719	0,734	0,719	0,609	0,641	0,703	0,719	0,609	0,672
	CDC2 + IGFBP5	0,766	0,781	0,766	0,781	0,609	0,719	0,688	0,781	0,672	0,688
	CDC + HOXA13	0,750	0,734	0,750	0,734	0,562	0,703	0,703	0,688	0,656	0,703
	IGF + HOXA13	0,703	0,703	0,672	0,672	0,719	0,750	0,500	0,641	0,688	0,672
	MDK + CDC2 + IGFBP5	0,766	0,766	0,766	0,766	0,688	0,766	0,734	0,797	0,656	0,625
	MDK + CDC2 + HOXA13	0,734	0,734	0,734	0,734	0,609	0,688	0,766	0,703	0,656	0,656
	MDK + IGFBP5 + HOXA13	0,734	0,719	0,734	0,719	0,719	0,766	0,750	0,703	0,656	0,656
	CDC2 + IGFBP5 + HOXA13	0,797	0,797	0,812	0,781	0,719	0,750	0,734	0,688	0,672	0,672
	MDK+CDC2+IGFBP5+HOXA13	0,766	0,781	0,781	0,797	0,703	0,766	0,766	0,719	0,656	0,641
	Sensibilidad a la especificidad: Sp.80										

Figura 16a

b	Modelo	LDA		LogReg		SVM		KN5N		TREE	
		-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb
	MDK	0,657	0,719	0,672	0,719	0,578	0,641	0,531	0,609	0,609	0,672
	CDC2	0,481	0,703	0,719	0,703	0,531	0,688	0,500	0,625	0,672	0,703
	IGFBP5	0,595	0,641	0,625	0,656	0,516	0,625	0,516	0,484	0,562	0,578
	HOXA13	0,462	0,594	0,562	0,562	0,469	0,469	0,406	0,375	0,500	0,406
	MDK + CDC2	0,688	0,734	0,688	0,734	0,641	0,688	0,562	0,672	0,641	0,641
	MDK + IGFBP5	0,703	0,719	0,703	0,719	0,703	0,781	0,562	0,766	0,609	0,609
	MDK + HOXA13	0,656	0,688	0,656	0,688	0,594	0,641	0,562	0,516	0,594	0,625
	CDC2 + IGFBP5	0,750	0,766	0,750	0,766	0,609	0,688	0,578	0,641	0,625	0,641
	CDC + HOXA13	0,719	0,703	0,719	0,703	0,547	0,688	0,594	0,547	0,656	0,703
	IGF + HOXA13	0,641	0,641	0,641	0,641	0,688	0,719	0,500	0,484	0,609	0,609
	MDK + CDC2 + IGFBP5	0,734	0,734	0,734	0,734	0,688	0,734	0,609	0,797	0,609	0,625
	MDK + CDC2 + HOXA13	0,688	0,719	0,688	0,719	0,594	0,672	0,609	0,641	0,641	0,641
	MDK + IGFBP5 + HOXA13	0,703	0,719	0,703	0,719	0,719	0,750	0,562	0,703	0,641	0,641
	CDC2 + IGFBP5 + HOXA13	0,750	0,766	0,750	0,766	0,719	0,734	0,594	0,688	0,625	0,625
	MDK+CDC2+IGFBP5+HOXA13	0,766	0,750	0,750	0,750	0,688	0,750	0,625	0,719	0,609	0,594
	Sensibilidad a la especificidad: Sp.85										

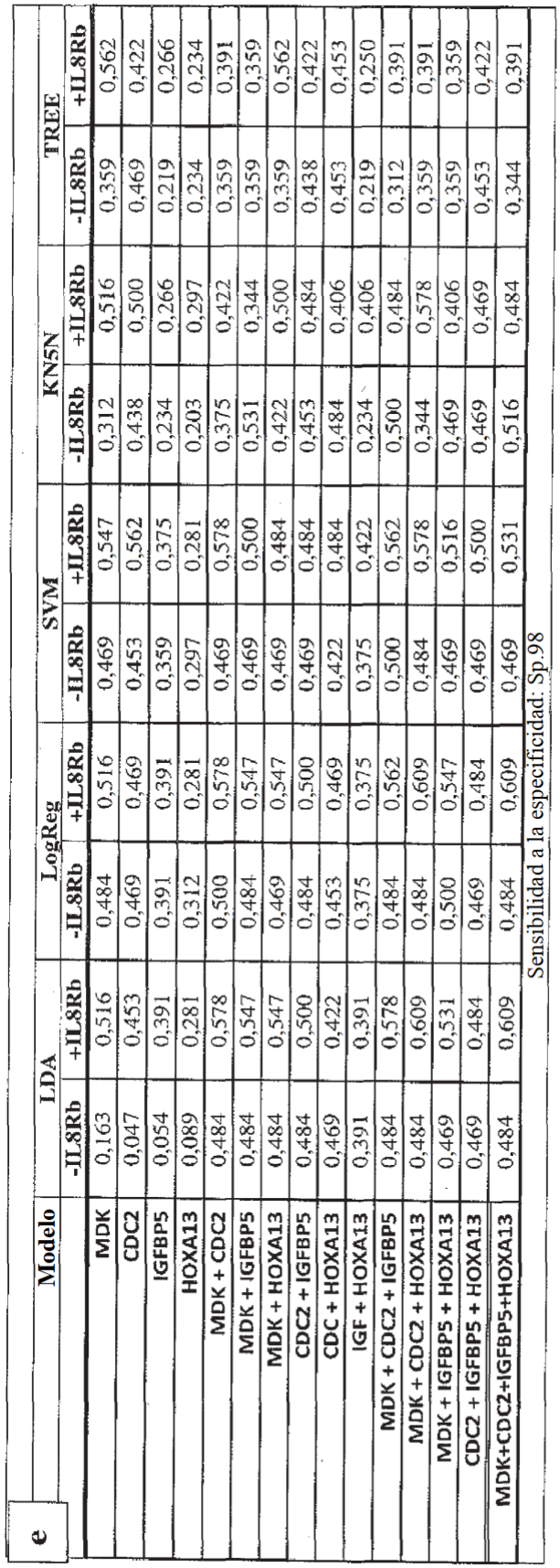
Figura 16b

c	Modelo	LDA		LogReg		SVM		KN5N		TREE	
		-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb
	MDK	0,511	0,672	0,625	0,672	0,562	0,641	0,531	0,609	0,609	0,594
	CDC2	0,338	0,703	0,656	0,703	0,531	0,688	0,500	0,625	0,625	0,672
	IGFBP5	0,212	0,562	0,562	0,562	0,500	0,594	0,516	0,484	0,500	0,562
	HOXA13	0,257	0,516	0,500	0,531	0,438	0,469	0,406	0,375	0,391	0,359
	MDK + CDC2	0,688	0,703	0,688	0,688	0,625	0,688	0,562	0,672	0,562	0,578
	MDK + IGFBP5	0,641	0,719	0,641	0,688	0,703	0,719	0,562	0,625	0,578	0,594
	MDK + HOXA13	0,641	0,672	0,625	0,656	0,562	0,641	0,562	0,516	0,594	0,578
	CDC2 + IGFBP5	0,656	0,750	0,672	0,734	0,609	0,688	0,578	0,641	0,547	0,594
	CDC + HOXA13	0,641	0,703	0,656	0,688	0,547	0,688	0,594	0,547	0,594	0,672
	IGF + HOXA13	0,594	0,594	0,578	0,578	0,672	0,688	0,500	0,484	0,594	0,578
	MDK + CDC2 + IGFBP5	0,672	0,734	0,672	0,719	0,656	0,734	0,609	0,672	0,531	0,578
	MDK + CDC2 + HOXA13	0,688	0,688	0,688	0,703	0,578	0,672	0,609	0,641	0,531	0,562
	MDK + IGFBP5 + HOXA13	0,641	0,672	0,625	0,656	0,688	0,688	0,562	0,562	0,609	0,594
	CDC2 + IGFBP5 + HOXA13	0,672	0,766	0,703	0,734	0,672	0,688	0,594	0,547	0,578	0,594
	MDK+CDC2+IGFBP5+HOXA13	0,688	0,719	0,703	0,719	0,672	0,734	0,625	0,672	0,500	0,562
	Sensibilidad a la especificidad: Sp.90										

Figura 16c

d	Modelo	LDA		LogReg		SVM		KN5N		TREE	
		-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb
	MDK	0,304	0,609	0,547	0,609	0,531	0,609	0,500	0,609	0,500	0,562
	CDC2	0,180	0,625	0,516	0,625	0,500	0,609	0,438	0,625	0,500	0,609
	IGFBP5	0,148	0,484	0,484	0,484	0,500	0,500	0,359	0,391	0,469	0,438
	HOXA13	0,173	0,422	0,438	0,422	0,422	0,406	0,297	0,297	0,344	0,297
	MDK + CDC2	0,562	0,672	0,562	0,672	0,547	0,672	0,453	0,672	0,453	0,562
	MDK + IGFBP5	0,562	0,609	0,578	0,609	0,609	0,656	0,562	0,625	0,531	0,547
	MDK + HOXA13	0,531	0,609	0,531	0,609	0,531	0,609	0,562	0,516	0,500	0,562
	CDC2 + IGFBP5	0,562	0,609	0,562	0,625	0,562	0,594	0,453	0,641	0,500	0,547
	CDC + HOXA13	0,516	0,625	0,500	0,625	0,516	0,609	0,594	0,547	0,516	0,609
	IGF + HOXA13	0,484	0,484	0,531	0,531	0,547	0,531	0,500	0,484	0,406	0,406
	MDK + CDC2 + IGFBP5	0,594	0,656	0,609	0,672	0,578	0,641	0,609	0,672	0,469	0,500
	MDK + CDC2 + HOXA13	0,562	0,672	0,562	0,672	0,547	0,672	0,516	0,641	0,484	0,562
	MDK + IGFBP5 + HOXA13	0,609	0,625	0,609	0,609	0,609	0,656	0,562	0,562	0,531	0,547
	CDC2 + IGFBP5 + HOXA13	0,609	0,641	0,578	0,656	0,594	0,641	0,594	0,547	0,484	0,516
	MDK+CDC2+IGFBP5+HOXA13	0,609	0,688	0,594	0,672	0,625	0,672	0,625	0,672	0,469	0,484
	Sensibilidad a la especificidad: Sp.95										

Figura 16d



e

Modelo	LDA		LogReg		SVM		KNSN		TREE	
	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb	-IL8Rb	+IL8Rb
MDK	0,163	0,516	0,484	0,516	0,469	0,547	0,312	0,516	0,359	0,562
CDC2	0,047	0,453	0,469	0,469	0,453	0,562	0,438	0,500	0,469	0,422
IGFBP5	0,054	0,391	0,391	0,391	0,359	0,375	0,234	0,266	0,219	0,266
HOXA13	0,089	0,281	0,312	0,281	0,297	0,281	0,203	0,297	0,234	0,234
MDK + CDC2	0,484	0,578	0,500	0,578	0,469	0,578	0,375	0,422	0,359	0,391
MDK + IGFBP5	0,484	0,547	0,484	0,547	0,469	0,500	0,531	0,344	0,359	0,359
MDK + HOXA13	0,484	0,547	0,469	0,547	0,469	0,484	0,422	0,500	0,359	0,562
CDC2 + IGFBP5	0,484	0,500	0,484	0,500	0,469	0,484	0,453	0,484	0,438	0,422
CDC + HOXA13	0,469	0,422	0,453	0,469	0,422	0,484	0,484	0,406	0,453	0,453
IGF + HOXA13	0,391	0,391	0,375	0,375	0,375	0,422	0,234	0,406	0,219	0,250
MDK + CDC2 + IGFBP5	0,484	0,578	0,484	0,562	0,500	0,562	0,500	0,484	0,312	0,391
MDK + CDC2 + HOXA13	0,484	0,609	0,484	0,609	0,484	0,578	0,344	0,578	0,359	0,391
MDK + IGFBP5 + HOXA13	0,469	0,531	0,500	0,547	0,469	0,516	0,469	0,406	0,359	0,359
CDC2 + IGFBP5 + HOXA13	0,469	0,484	0,469	0,484	0,469	0,500	0,469	0,469	0,453	0,422
MDK+CDC2+IGFBP5+HOXA13	0,484	0,609	0,484	0,609	0,469	0,531	0,516	0,484	0,344	0,391

Sensibilidad a la especificidad: Sp.98

Figura 16e

Modelo	LDA	LogReg	SVM	KN5N	TREE
MDK	0,013	0,000	0,078	0,250	0,062
CDC2	-0,019	-0,016	0,141	-0,016	0,031
IGFBP5	0,046	0,016	0,109	0,141	-0,047
HOXA13	0,123	-0,016	-0,016	-0,031	0,031
MDK + CDC2	0,016	0,016	0,078	0,016	-0,031
MDK + IGFBP5	0,000	0,000	0,078	0,078	0,000
MDK + HOXA13	0,000	-0,016	0,031	0,016	0,062
CDC2 + IGFBP5	0,016	0,016	0,109	0,094	0,016
CDC + HOXA13	-0,016	-0,016	0,141	-0,016	0,047
IGF + HOXA13	0,000	0,000	0,031	0,141	-0,016
MDK + CDC2 + IGFBP5	0,000	0,000	0,078	0,062	-0,031
MDK + CDC2 + HOXA13	0,000	0,000	0,078	-0,062	0,000
MDK + IGFBP5 + HOXA13	-0,016	-0,016	0,047	-0,047	0,000
CDC2 + IGFBP5 + HOXA13	0,000	-0,031	0,031	-0,047	0,000
MDK+CDC2+IGFBP5+HOXA13	0,016	0,016	0,062	-0,047	-0,016
Aumento de sensibilidad, comenzando a especificidad=0,80					

Figura 17a

Modelo	LDA	LogReg	SVM	KN5N	TREE
MDK	0,062	0,047	0,047	0,078	0,062
CDC2	0,222	-0,016	-0,016	0,125	0,031
IGFBP5	0,046	0,031	0,031	-0,031	0,016
HOXA13	0,132	0,000	0,031	-0,031	-0,094
MDK + CDC2	0,047	0,047	0,016	0,109	0,000
MDK + IGFBP5	0,016	0,016	0,016	0,203	0,000
MDK + HOXA13	0,031	0,031	0,016	-0,047	0,031
CDC2 + IGFBP5	0,016	0,016	0,016	0,062	0,016
CDC + HOXA13	-0,016	-0,016	-0,016	-0,047	0,047
IGF + HOXA13	0,000	0,000	0,062	-0,016	0,000
MDK + CDC2 + IGFBP5	0,000	0,000	0,016	0,188	0,016
MDK + CDC2 + HOXA13	0,031	0,031	0,000	0,031	0,000
MDK + IGFBP5 + HOXA13	0,016	0,016	0,047	0,141	0,000
CDC2 + IGFBP5 + HOXA13	0,016	0,016	0,031	0,094	0,000
MDK+CDC2+IGFBP5+HOXA13	-0,016	0,000	-0,016	0,094	-0,016
Aumento de sensibilidad, comenzando a especificidad=0,85					

Figura 17b

Modelo	LDA	LogReg	SVM	KN5N	TREE
MDK	0,161	0,047	0,078	0,078	-0,016
CDC2	0,365	0,047	0,156	0,125	0,047
IGFBP5	0,350	0,000	0,094	-0,031	0,062
HOXA13	0,259	0,031	0,031	-0,031	-0,031
MDK + CDC2	0,016	0,000	0,062	0,109	0,016
MDK + IGFBP5	0,078	0,047	0,016	0,062	0,016
MDK + HOXA13	0,031	0,031	0,078	-0,047	-0,016
CDC2 + IGFBP5	0,094	0,062	0,078	0,062	0,047
CDC + HOXA13	0,062	0,031	0,141	-0,047	0,078
IGF + HOXA13	0,000	0,000	0,016	-0,016	-0,016
MDK + CDC2 + IGFBP5	0,062	0,047	0,078	0,062	0,047
MDK + CDC2 + HOXA13	0,000	0,016	0,094	0,031	0,031
MDK + IGFBP5 + HOXA13	0,031	0,031	0,000	0,000	-0,016
CDC2 + IGFBP5 + HOXA13	0,094	0,031	0,016	-0,047	0,016
MDK+CDC2+IGFBP5+HOXA13	0,031	0,016	0,062	0,047	0,062

Aumento de sensibilidad, comenzando a especificidad=0,90

Figura 17c

Modelo	LDA	LogReg	SVM	KNSN	TREE
MDK	0,306	0,062	0,078	0,109	0,062
CDC2	0,445	0,109	0,109	0,188	0,109
IGFBP5	0,336	0,000	0,000	0,031	-0,031
HOXA13	0,249	-0,016	-0,016	0,000	-0,047
MDK + CDC2	0,109	0,109	0,125	0,219	0,109
MDK + IGFBP5	0,047	0,031	0,047	0,062	0,016
MDK + HOXA13	0,078	0,078	0,078	-0,047	0,062
CDC2 + IGFBP5	0,047	0,062	0,031	0,188	0,047
CDC + HOXA13	0,109	0,125	0,094	-0,047	0,094
IGF + HOXA13	0,000	0,000	-0,016	-0,016	0,000
MDK + CDC2 + IGFBP5	0,062	0,062	0,062	0,062	0,031
MDK + CDC2 + HOXA13	0,109	0,109	0,125	0,125	0,078
MDK + IGFBP5 + HOXA13	0,016	0,000	0,047	0,000	0,016
CDC2 + IGFBP5 + HOXA13	0,031	0,078	0,047	-0,047	0,031
MDK+CDC2+IGFBP5+HOXA13	0,078	0,078	0,047	0,047	0,016

Aumento de sensibilidad, comenzando a especificidad=0,95

Figura 17d

Modelo	LDA	LogReg	SVM	KN5N	TREE
MDK	0,353	0,031	0,078	0,203	0,203
CDC2	0,406	0,000	0,109	0,062	-0,047
IGFBP5	0,336	0,000	0,016	0,031	0,047
HOXA13	0,192	-0,031	-0,016	0,094	0,000
MDK + CDC2	0,094	0,078	0,109	0,047	0,031
MDK + IGFBP5	0,062	0,062	0,031	-0,188	0,000
MDK + HOXA13	0,062	0,078	0,016	0,078	0,203
CDC2 + IGFBP5	0,016	0,016	0,016	0,031	-0,016
CDC + HOXA13	-0,047	0,016	0,062	-0,078	0,000
IGF + HOXA13	0,000	0,000	0,047	0,172	0,031
MDK + CDC2 + IGFBP5	0,094	0,078	0,062	-0,016	0,078
MDK + CDC2 + HOXA13	0,125	0,125	0,094	0,234	0,031
MDK + IGFBP5 + HOXA13	0,062	0,047	0,047	-0,062	0,000
CDC2 + IGFBP5 + HOXA13	0,016	0,016	0,031	0,000	-0,031
MDK+CDC2+IGFBP5+HOXA13	0,125	0,125	0,062	-0,031	0,047
Aumento de sensibilidad, comenzando a especificidad=0,98					

Figura 17e

Modelo	LDA	LogReg	SVM	KN5N	TREE
MDK	0,030	0,000	0,168	0,180	0,069
CDC2	-0,025	-0,040	0,178	-0,104	0,094
IGFBP5	0,037	0,022	0,131	0,119	-0,025
HOXA13	0,114	-0,007	-0,002	0,116	0,027
MDK + CDC2	0,012	0,030	0,158	-0,101	-0,049
MDK + IGFBP5	0,000	0,000	0,101	0,146	-0,022
MDK + HOXA13	0,000	-0,054	0,126	-0,089	0,072
CDC2 + IGFBP5	0,022	0,012	0,143	0,146	0,035
CDC + HOXA13	-0,025	-0,040	0,158	-0,109	0,106
IGF + HOXA13	0,000	0,000	0,042	0,158	0,005
MDK + CDC2 + IGFBP5	0,000	0,000	0,133	0,151	-0,012
MDK + CDC2 + HOXA13	0,000	0,000	0,165	-0,091	-0,035
MDK + IGFBP5 + HOXA13	-0,022	-0,022	0,081	-0,042	-0,032
CDC2 + IGFBP5 + HOXA13	0,000	-0,094	0,064	-0,032	0,015
MDK+CDC2+IGFBP5+HOXA13	0,027	0,000	0,138	-0,057	-0,010
Aumento de especificidad, comenzando a especificidad=0,80					

Figura 17f

Modelo	LDA	LogReg	SVM	KN5N	TREE
MDK	0,059	0,041	0,123	0,130	0,039
CDC2	0,125	-0,023	0,130	0,130	0,044
IGFBP5	0,041	0,009	0,081	0,069	0,036
HOXA13	0,064	-0,001	-0,001	0,066	-0,015
MDK + CDC2	0,054	0,046	0,108	0,118	-0,001
MDK + IGFBP5	0,051	0,041	0,051	0,096	0,019
MDK + HOXA13	0,051	0,044	0,103	0,086	0,041
CDC2 + IGFBP5	0,017	0,002	0,093	0,096	0,031
CDC + HOXA13	-0,018	-0,028	0,113	0,086	0,056
IGF + HOXA13	-0,001	-0,001	0,002	0,108	-0,008
MDK + CDC2 + IGFBP5	-0,001	-0,001	0,083	0,101	0,014
MDK + CDC2 + HOXA13	0,049	0,054	0,118	0,130	-0,001
MDK + IGFBP5 + HOXA13	0,007	0,014	0,031	0,113	0,009
CDC2 + IGFBP5 + HOXA13	0,051	0,007	0,014	0,101	-0,003
MDK+CDC2+IGFBP5+HOXA13	-0,015	-0,001	0,088	0,110	0,004
Aumento de especificidad, comenzando a especificidad=0,85					

Figura 17g

Modelo	LDA	LogReg	SVM	KN5N	TREE
MDK	0,078	0,021	0,075	0,080	-0,011
CDC2	0,098	0,038	0,080	0,080	0,028
IGFBP5	0,090	0,001	0,043	0,019	0,011
HOXA13	0,088	0,004	0,028	0,016	-0,026
MDK + CDC2	0,004	0,001	0,058	0,068	0,043
MDK + IGFBP5	0,028	0,026	0,001	0,046	0,009
MDK + HOXA13	0,011	0,016	0,063	0,036	-0,009
CDC2 + IGFBP5	0,021	0,028	0,043	0,046	0,058
CDC + HOXA13	0,019	0,033	0,063	0,036	0,053
IGF + HOXA13	0,001	0,001	0,004	0,058	-0,031
MDK + CDC2 + IGFBP5	0,048	0,043	0,036	0,051	0,048
MDK + CDC2 + HOXA13	0,001	0,004	0,075	0,080	0,053
MDK + IGFBP5 + HOXA13	0,041	0,046	0,001	0,063	0,004
CDC2 + IGFBP5 + HOXA13	0,021	0,033	0,021	0,051	-0,001
MDK+CDC2+IGFBP5+HOXA13	0,031	0,021	0,048	0,060	0,046
Aumento de especificidad, comenzando a especificidad=0,90					

Figura 17h

Modelo	LDA	LogReg	SVM	KNSN	TREE
MDK	0,043	0,025	0,030	0,030	0,028
CDC2	0,050	0,025	0,030	0,030	0,023
IGFBP5	0,043	0,001	0,001	0,015	-0,007
HOXA13	0,045	-0,002	-0,004	0,020	-0,012
MDK + CDC2	0,030	0,030	0,033	0,038	0,013
MDK + IGFBP5	0,023	0,025	0,010	-0,004	0,006
MDK + HOXA13	0,030	0,030	0,013	-0,014	0,028
CDC2 + IGFBP5	0,020	0,018	0,015	0,025	0,015
CDC + HOXA13	0,023	0,025	0,025	-0,014	0,015
IGF + HOXA13	0,001	0,001	-0,007	0,008	0,003
MDK + CDC2 + IGFBP5	0,025	0,023	0,023	0,001	0,010
MDK + CDC2 + HOXA13	0,035	0,033	0,033	0,030	0,013
MDK + IGFBP5 + HOXA13	0,006	0,001	0,003	0,013	0,006
CDC2 + IGFBP5 + HOXA13	0,003	0,015	0,008	0,001	0,008
MDK+CDC2+IGFBP5+HOXA13	0,028	0,030	0,015	0,010	0,010
Aumento de especificidad, comenzando a especificidad=0,95					

Figura 17i

Modelo	LDA	LogReg	SVM	KN5N	TREE
MDK	0,015	0,005	0,008	0,005	0,008
CDC2	0,020	0,000	0,000	0,000	-0,007
IGFBP5	0,018	0,000	0,000	0,000	0,003
HOXA13	0,015	-0,007	-0,007	0,010	0,000
MDK + CDC2	0,003	0,003	0,005	0,008	0,003
MDK + IGFBP5	0,000	0,000	0,000	-0,010	0,000
MDK + HOXA13	0,008	0,008	0,008	0,010	0,008
CDC2 + IGFBP5	0,005	0,000	0,003	-0,005	0,000
CDC + HOXA13	-0,002	0,000	0,000	-0,010	-0,002
IGF + HOXA13	0,000	0,000	0,003	0,008	0,003
MDK + CDC2 + IGFBP5	0,003	0,003	0,005	0,005	0,010
MDK + CDC2 + HOXA13	0,008	0,008	0,003	0,005	0,003
MDK + IGFBP5 + HOXA13	0,008	0,005	0,000	0,005	0,000
CDC2 + IGFBP5 + HOXA13	0,000	0,000	0,008	0,013	-0,007
MDK+CDC2+IGFBP5+HOXA13	0,005	0,005	0,005	0,005	0,010
Aumento de especificidad, comenzando a especificidad=0,98					

Figura 17j

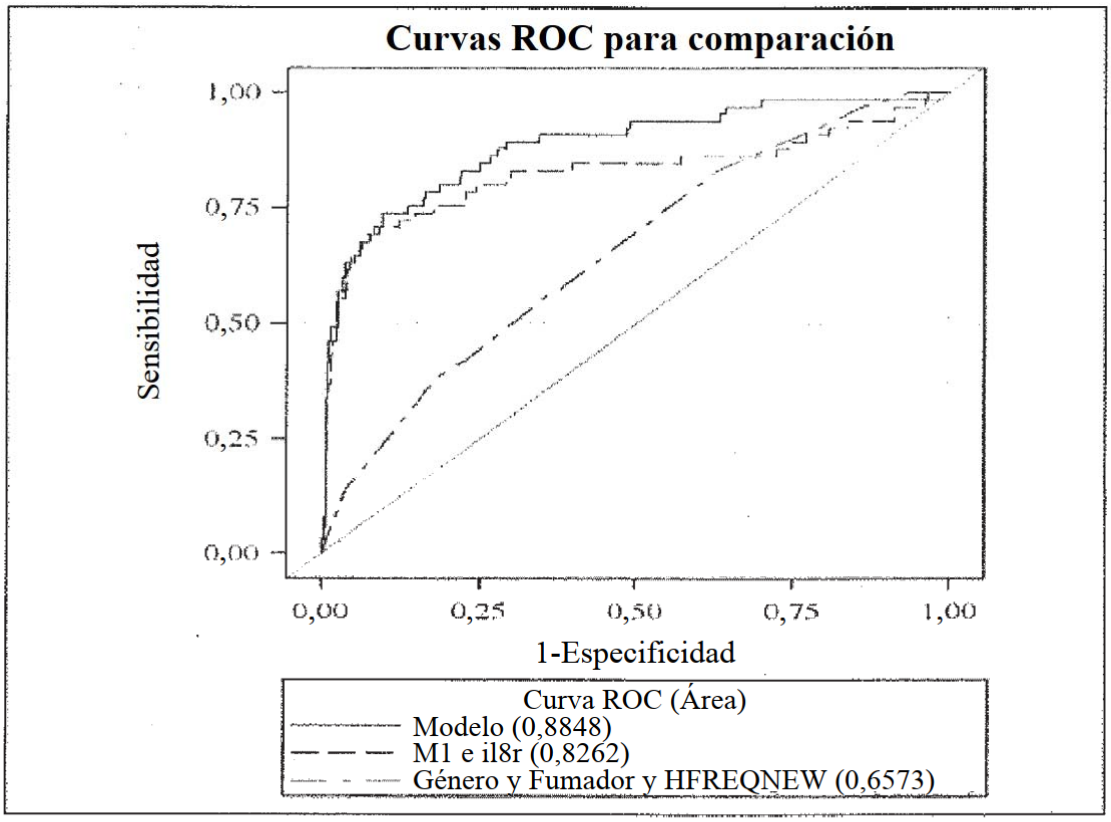

Curvas ROC para comparación
Sensibilidad
1,00
0,75
0,50
0,25
0,00
0,00
0,25
0,50
0,75
1,00
1-Especificidad
Curva ROC (Área)
Modelo (0,8848)
M1 e il8r (0,8262)
Género y Fumador y HFREQNEW (0,6573)

FIG. 18

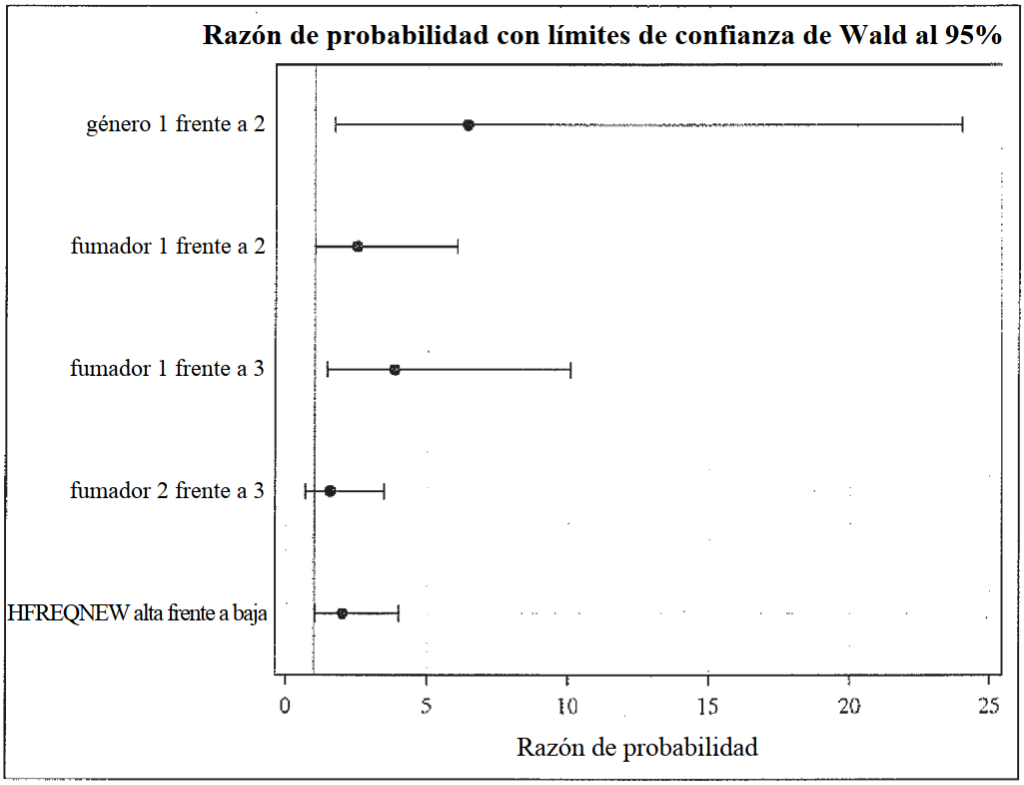

Razón de probabilidad con límites de confianza de Wald al 95%
género 1 frente a 2
fumador 1 frente a 2
fumador 1 frente a 3
fumador 2 frente a 3
HFREQNEW alta frente a baja
0
5
10
15
20
25
Razón de probabilidad

FIG. 19

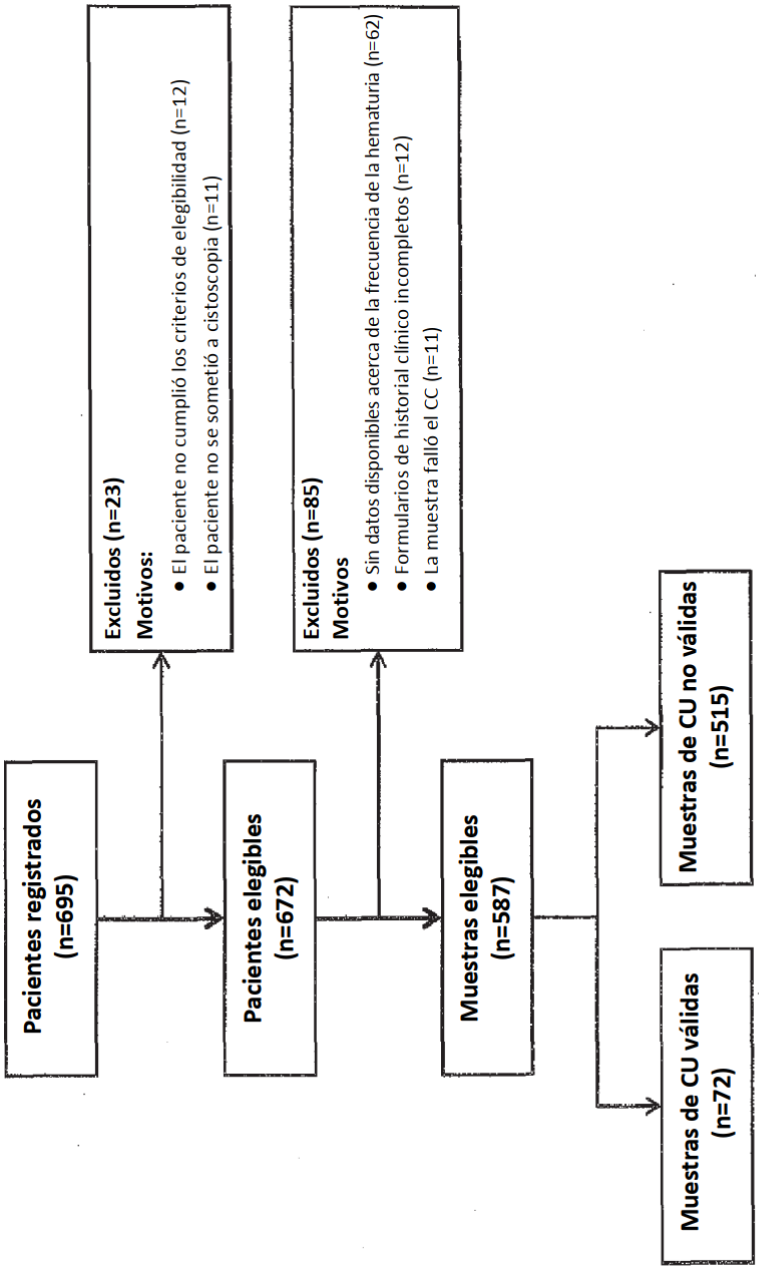

Pacientes registrados (n=695)
Excluidos (n=23)
Motivos:
• El paciente no cumplió los criterios de elegibilidad (n=12)
• El paciente no se sometió a cistoscopia (n=11)
Pacientes elegibles (n=672)
Excluidos (n=85)
Motivos
• Sin datos disponibles acerca de la frecuencia de la hematuria (n=62)
• Formularios de historial clínico incompletos (n=12)
• La muestra falló el CC (n=11)
Muestras elegibles (n=587)
Muestras de CU no válidas (n=515)
Muestras de CU válidas (n=72)

FIG. 20A

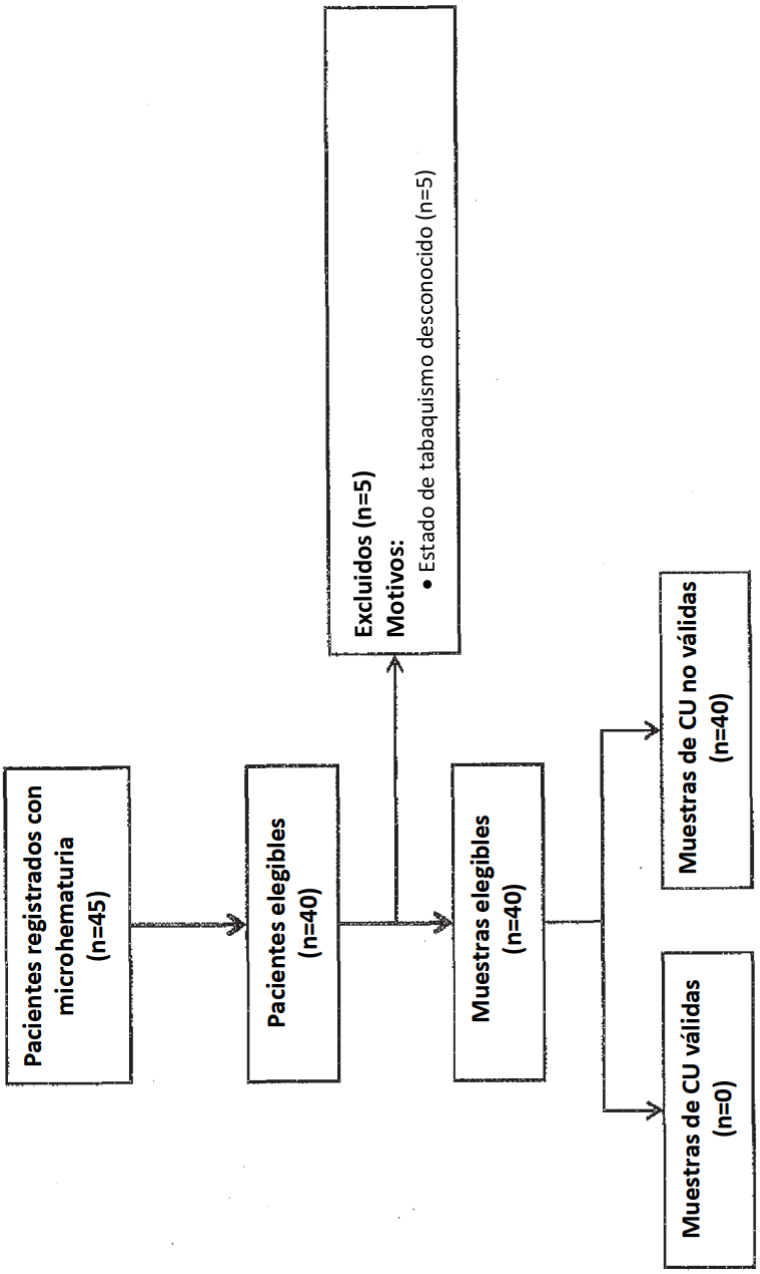

Pacientes registrados con microhematuria (n=45)
Pacientes elegibles (n=40)
Excluidos (n=5)
Motivos:
• Estado de tabaquismo desconocido (n=5)
Muestras elegibles (n=40)
Muestras de CU no válidas (n=40)
Muestras de CU válidas (n=0)

FIG. 20B

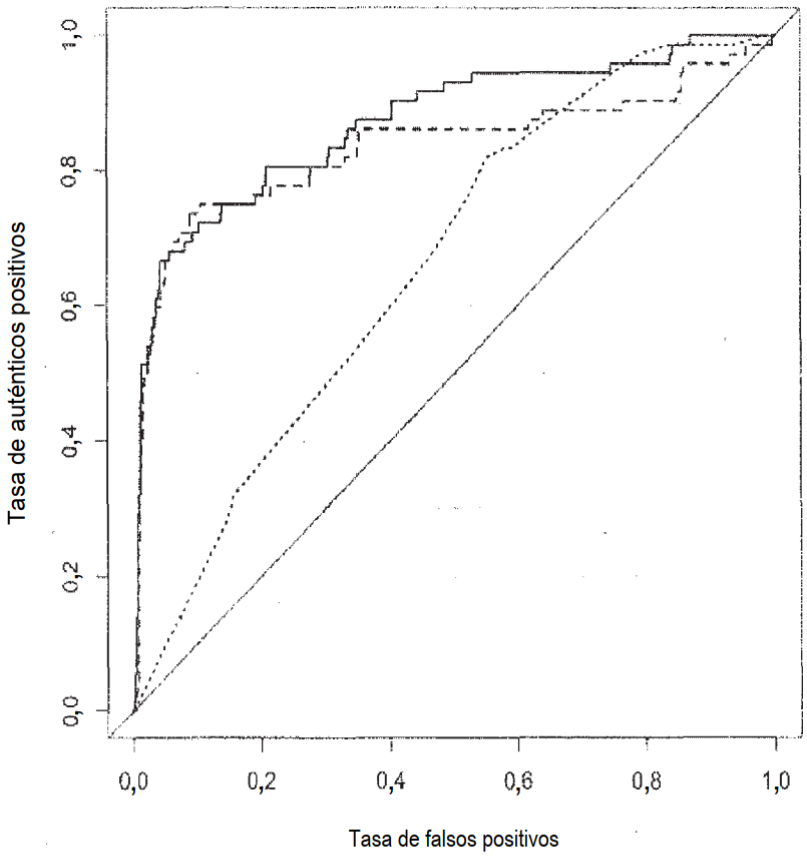

Tasa de auténticos positivos
0,0
0,2
0,4
0,6
0,8
1,0
Tasa de falsos positivos
0,0
0,2
0,4
0,6
0,8
1,0

FIG. 21

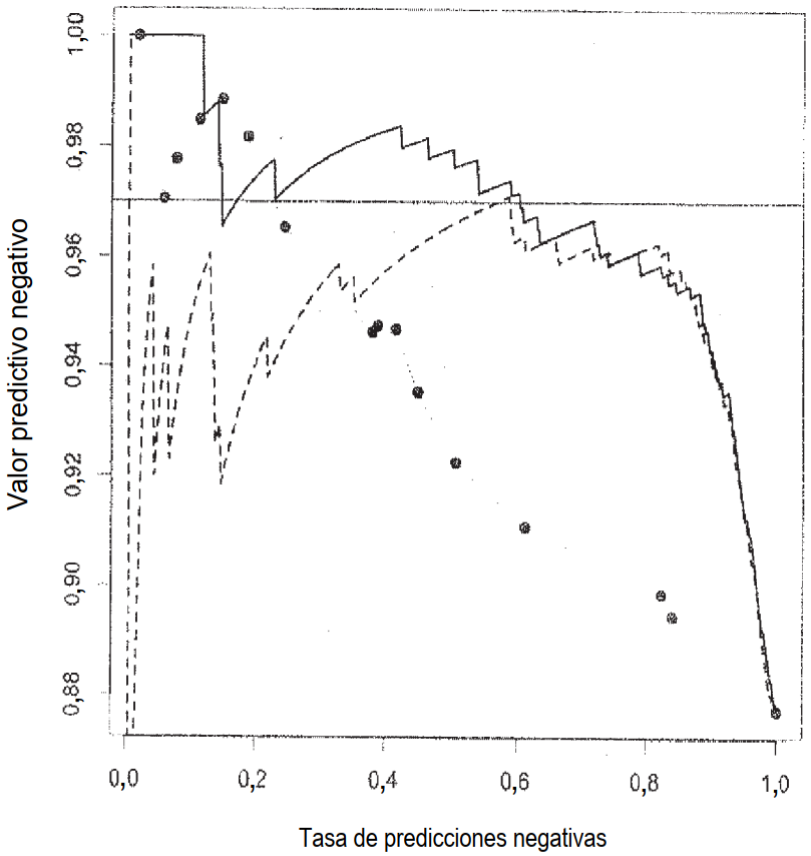

Valor predictivo negativo
1,00
0,98
0,96
0,94
0,92
0,90
0,88
0,0
0,2
0,4
0,6
0,8
1,0
Tasa de predicciones negativas

FIG. 22

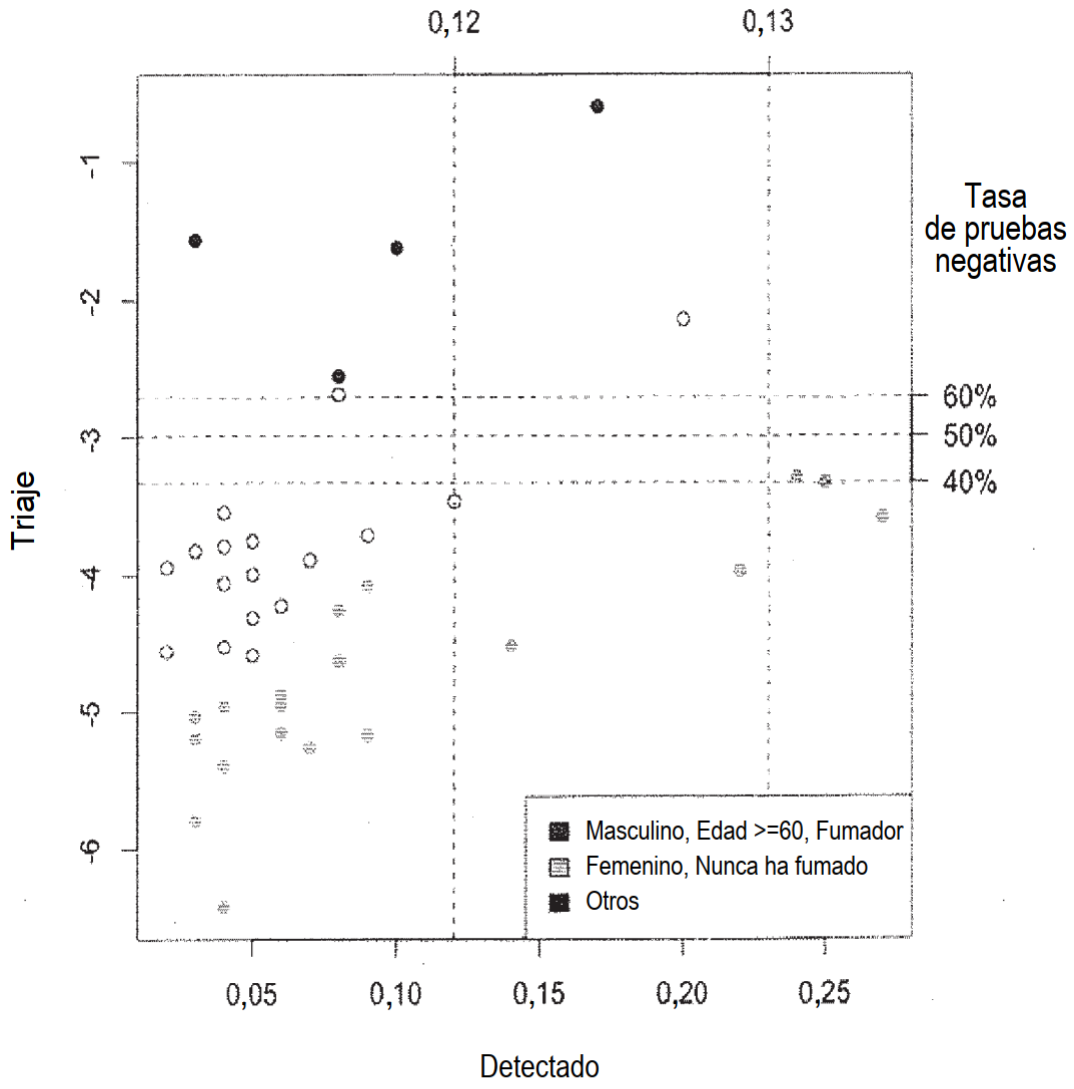

0,12
0,13
Tasa
de pruebas
negativas
60%
50%
40%
Triaje
-1
-2
-3
-4
-5
-6
0,05
0,10
0,15
0,20
0,25
Detectado
Masculino, Edad >=60, Fumador
Femenino, Nunca ha fumado
Otros

FIG. 23

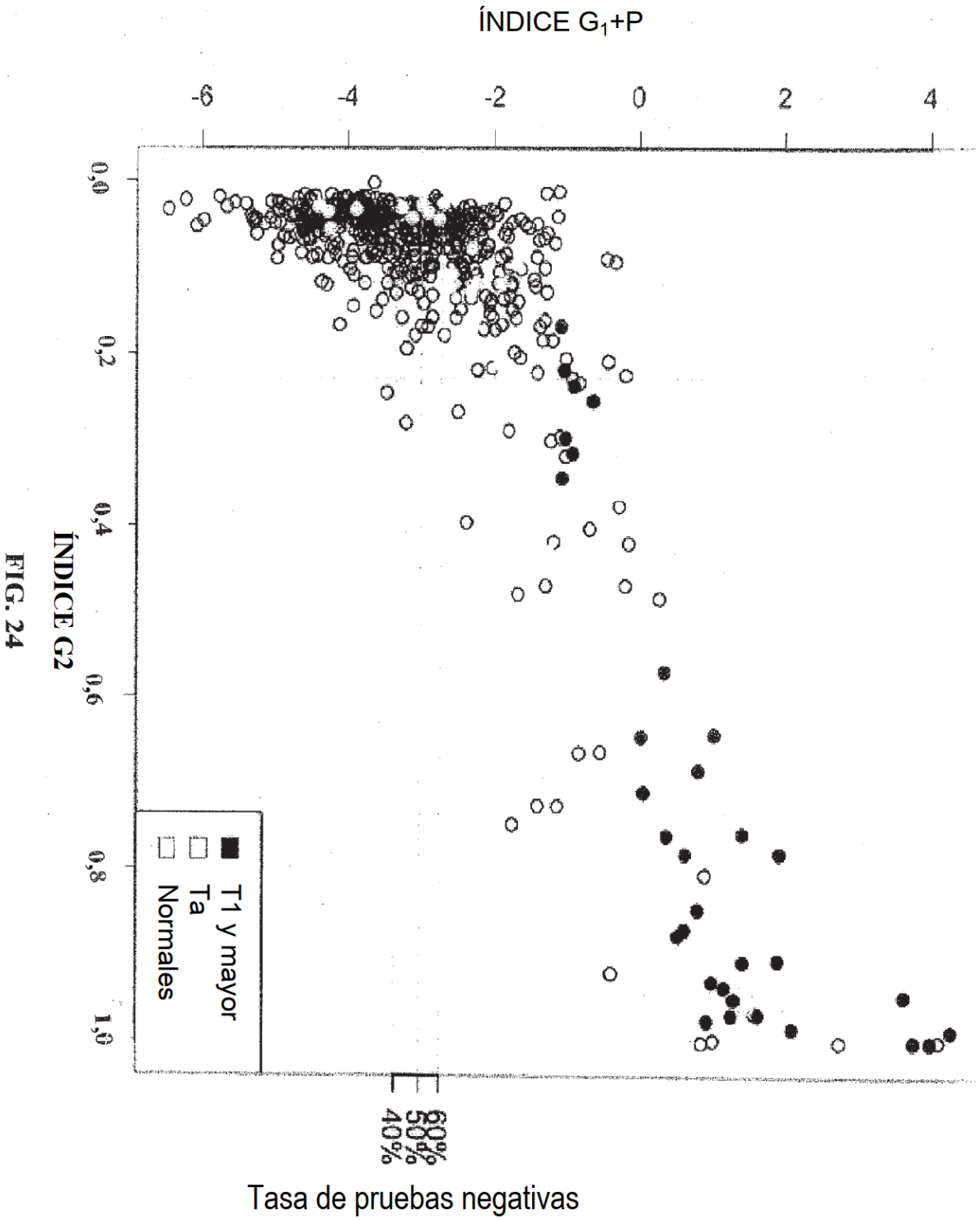

ÍNDICE G1+P
-6
-4
-2
0
2
4
0,0
0,2
0,4
0,6
0,8
1,0
ÍNDICE G2
T1 y mayor
Ta
Normales
60%
50%
40%
Tasa de pruebas negativas

FIG. 24

FIG. 24