(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 337 115**

(21) Número de solicitud: 200702532

(51) Int. Cl.:
C12Q 1/68 (2006.01)

(12) PATENTE DE INVENCIÓN CON EXAMEN PREVIO B2

(22) Fecha de presentación: **26.09.2007**

(43) Fecha de publicación de la solicitud: **20.04.2010**

Fecha de la concesión: **12.01.2011**

(45) Fecha de anuncio de la concesión: **24.01.2011**

(45) Fecha de publicación del folleto de la patente: **24.01.2011**

(73) Titular/es:
Universidade de Santiago de Compostela
Edificio CACTUS-CITT-Campus Sur
15782 Santiago de Compostela, A Coruña, ES
Universidad de Valladolid

(72) Inventor/es: **Brión Martínez, María;**
Pastor Jimeno, José Carlos;
Sanabria Ruiz, María Rosa;
Sobrino Rey, Beatriz;
Fernández Martínez, Itziar;
García Gutiérrez, María Teresa;
Carracedo Álvarez, Ángel María y
Rojas Spano, Jimena

(74) Agente: **Pons Ariño, Ángel**

(54) Título: **Método para la detección del riesgo de desarrollar vitreorretinopatía proliferante (VRP).**

(57) Resumen:
Método para la detección del riesgo de desarrollar vitreorretinopatía proliferante (VRP).
La presente invención se refiere a un método que permite identificar el riesgo de desarrollar vitreorretinopatía proliferante en aquéllas personas que son sometidas a cirugías oculares. Más concretamente, el método se basa en la detección de una serie de polimorfismo de un solo nucleótido (SNP: "single nucleotide polymorphisms") o grupos de SNPs (Haplotipos) que predisponen al desarrollo de la enfermedad.

ES 2 337 115 B2

DESCRIPCIÓN

Método para la detección del riesgo de desarrollar vitreorretinopatía proliferante (VRP).

La presente invención se refiere a un método que permite identificar el riesgo de desarrollar vitreorretinopatía proliferante en aquéllas personas que son sometidas a cirugías oculares. Más concretamente, el método se basa en la detección de una serie de polimorfismo de un solo nucleótido (SNP: "single nucleotide polymorphisms") o grupos de SNPs (Haplotipos) que predisponen al desarrollo de la enfermedad.

Estado de la técnica

La vitreorretinopatía proliferante (VRP) es una respuesta cicatrizal anómala, que constituye la causa más frecuente del fracaso de la cirugía de desprendimiento de retina (DR) y para la que, a pesar de los avances que ha vivido la oftalmología en las últimas décadas, aún siguen sin conocerse sus causas o los factores que predisponen a la misma.

Desde hace más de 20 años se han realizado importantes esfuerzos para conocer factores predictores del riesgo de sufrir esta complicación, en un intento por identificar a los pacientes que se beneficiarían de terapias más específicas o para mejorar la comprensión de los mecanismos intrínsecos que provocan esta complicación.

Hasta el momento, todos los trabajos en esta línea se han desarrollado en base al análisis de las características clínicas de los pacientes con DR con y sin VRP, y los modelos predictivos establecidos en función de éstas variables, aunque explican parcialmente el riesgo de desarrollar esta complicación, son incapaces de aportar un análisis con valores de sensibilidad y especificidad satisfactorios.

En 2006 Sanabria *et al* realizaron un estudio piloto sobre el componente genético en el desarrollo de la VRP, encontrándose una distribución diferente de los (SNP) estudiados en los codones 10 y 25 del Transforming Growth Factor beta 1 (TGF-beta1) entre los pacientes con y sin VRP, lo que les llevó a concluir que éstos podrían conferir una mayor susceptibilidad a presentar esta complicación (Sanabria Ruiz-Colmenares *et al*. Acta Ophthalmol Scand. Jun; 84(3):309-13 (2006)). Sin embargo, estos resultados no consiguen demostrar en absoluto existencia de factores genéticos que predispongan al desarrollo de VRP, puesto que el escaso tamaño muestral de ese trabajo no permite utilizar las herramientas estadísticas adecuadas para el análisis de datos genéticos.

La confirmación de la contribución genética de un individuo a presentar una mayor susceptibilidad para el desarrollo de la VRP y, así como, la identificación del perfil genético que confiere ese incremento del riesgo, permitiría la apertura de nuevas posibilidades de tratamiento enfocados hacia el desarrollo de terapias más personalizadas.

Breve descripción de la invención

Los autores de la presente invención han conseguido demostrar, sorprendentemente, la presencia de factores genéticos que predisponen al desarrollo de VRP. El análisis de estos factores genéticos, vendrán a completar los estudios de aquéllos factores clínicos que en la actualidad ya se emplean para determinar el riesgo que un determinado sujeto tiene a padecer VRP tras una intervención ocular.

Tal y como informa el articulo publicado en 2005 de Nature Genetics (Nat. Genet 37:1299-1300), aunque muchos de los 7 millones de SNPs que se conocen tienen unas frecuencias superiores al 5%, en la actualidad no existen herramientas que permitan llevar a cabo el análisis en bloque de todos ellos. Este hecho, que supone la dificultad de hacer un estudio genético completo, unido al desconocimiento acerca de qué genes podrían estar relacionados con la VRP llevó a la necesidad de llevar a cabo la selección de todos aquellos genes que *a priori* pudiesen, de algún modo, estar relacionados con la complicación.

Para la selección de los marcadores, fue necesaria la formulación de una hipótesis de partida que planteó la posibilidad de que algunos marcadores de genes de citoquinas, factores nucleares y factores de crecimiento involucrados en la inflamación y en procesos fibróticos, así como los intermediarios correspondientes de las vías de señalización, pudieran estar presentes de forma significativa en pacientes que hubiesen desarrollado VRP tras un DR.

Así, se seleccionaron para el análisis 30 genes candidatos involucrados en los diferentes mecanismos implicados en la inflamación y en procesos profibróticos en general, que se detallan a continuación:

El TNF, TNFR2, TGFB1, TGFB2, SMAD3, SMAD7, IFNG, IL1A, IL1B, IL1RN, IL6, IL8, IL10, NFKB1, NFKBIA, NFKBIB, HGF, CTGF, PDGF, PDGFRA, PI3K, EGF, FGF2, MIF, MMP2, MMP9, MCP1, IGF-11, IGF2, IGF-IR.

El gen del TNFα está ubicado en el cromosoma 6, dentro de una región en medio del complejo mayor de histocompatibilidad. Los genes ubicados en esta región presentan un alto desequilibrio de ligamiento, de forma tal que los marcadores estudiados en un gen señalan a otros marcadores en otros genes dentro de la región. Se considera además que la mayoría regula la expresión del TNFα. De esta forma, se ha descrito esta región como TNF block (bloque del TNF). El bloque del TNF está compuesto por los siguientes genes: los llamados genes del TNF (TNFα; LTA; LTB), el gen del AIF1, el gen del NCR3, el gen del NFKBIL1, el gen del ATP6P1G, y el gen del BAT1. (Allcock, R. J. N.;

Windsor, L.; Gut, I. G.; Kucharzak, R.; Sobre, L.; Lechner, D.; Garnier, J.-G.; Baltic, S.; Christiansen, F. T.; Price, P. High-density SNP genotyping defines 17 distinct haplotypes of the TNF block in the Caucasian population: implications for haplotype tagging. Hum. Mutat. 24: 517-525, 2004). En el presente trabajo se han incluido marcadores ubicados en los genes que componen este bloque y se han denominado indistintamente TNF.

A partir de estos genes, se seleccionaron un total de 230 de SNPs, que se encontraban distribuidos entre los 30 genes candidatos mencionados anteriormente y que se analizaron para un total de 450 muestras. De todos los SNPs analizados únicamente aquéllos localizados en los genes SMAD7, TNF, PIK3CG y TNFR2 presentaron una asociación significativa con la VRP, una vez analizados un total de 88.650 polimorfismos para el total de muestras. Estos resultados demuestran la existencia de factores genéticos que predisponen al desarrollo de VRP y, más concretamente, la asociación de determinados genes con la complicación.

En una segunda fase, y a partir de los modelos estadísticos se identificaron combinaciones de SNPs que permiten evaluar el riesgo de desarrollar una VRP en un paciente nuevo. Se identificaron modelos predictivos de 2,10 y 42 SNP.

Así, un primer aspecto de la presente invención se refiere a un método para la determinación del riesgo de desarrollar VRP (en adelante, método de la invención) que comprende identificar en una muestra la presencia de factores de riesgo o la ausencia de factores de protección de cualquiera de los marcadores seleccionados del siguiente grupo de marcadores genéticos:

i) Marcadores ubicados en el gen SMAD7,

ii) Marcadores ubicados en el gen TNF,

iii) Marcadores ubicados en el gen PIK3CG,

iv) Marcadores ubicados en el gen TNFR2,

v) marcadores ligados a cualquiera de los marcadores anteriores, preferentemente con un desequilibrio de ligamiento (r^2) mayor o igual 0.70, 0.80 ó 0.90 y más preferentemente ubicados en el mismo gen. Los marcadores ubicados en los mencionados genes son, preferentemente, SNPs que pueden ser seleccionados, sin ninguna limitación, de las bases de datos dbSNP, HapMap, etc.

donde la presencia de al menos uno de los de los marcadores i-v es indicativa de la existencia de un mayor o menor riesgo a desarrollar VRP.

En una realización preferida de este aspecto de la invención, los marcadores ubicados en los genes SMAD7, TNF, PIK3CG, TNFR2 pueden ser seleccionados del grupo que comprende, sin ningún tipo de limitación, SNPs, microsatélites, minisatélites, inserciones, deleciones, variaciones del numero de copias, inversiones y traslocaciones, donde dichos marcadores pueden ser analizados mediante técnicas sobradamente conocidas en el estado de la técnica (Nat Rev Genet. 2001 Dec; 2(12):930-42, Forensic Sci. Int. (2005) 154: 181-194, Nature. 2006, 444(7118):444-54, PCR Methods Appl. 3:13-22, 1993, etc).

En una realización preferida de la presente invención los marcadores genéticos ubicados en los genes SMAD7, TNF, PIK3CG, TNFR2 son seleccionados respectivamente del grupo que comprende:

i) rs7226855 (tabla 2),

ii) rs2229094, rs2256974, rs2857706 (tabla 2) y haplotipos de TNF seleccionados de la tabla 3.

iii) El haplotipo seleccionado de la tabla 3 (PIK3CG).

iv) El haplotipo seleccionado de la tabla 3 (TNFR2), ó

v) Marcadores ligados a cualquiera de los marcadores anteriores, preferentemente con un desequilibrio de ligamiento (r^2) entre marcadores de r^2 mayor o igual 0.70, y más preferentemente SNPs que pueden ser seleccionados, sin ninguna limitación, de las bases de datos dbSNP, HapMap, etc.

En una realización aun más preferida de este aspecto de la invención el método comprende las etapas de:

i) Extracción y purificación de material genético (genómico o mitocondrial) a partir de una muestra,

ii) Amplificación de al menos uno de los marcadores mencionados anteriormente,

iii) Identificación de los marcadores mediante técnicas sobradamente conocidas en el estado de la técnica.

iv) Determinación del riesgo a desarrollar VRP.

Los oligonucleótidos necesarios para llevar a cabo la amplificación de los marcadores genéticos pueden ser desarrollados, sin ningún tipo de limitación, a partir la información disponible en bases de datos tales como EMBL-EBI (www.ebi.ac.uk) o NCBI (www.ncbi.nlm.nih.gov) y programas informáticos tales como Primer 3, Assay Design 3.0, entre otros.

En una realización preferida de la invención el método de la invención puede estar automatizado empleando cualquiera de las técnicas seleccionadas del grupo de sistemas que comprende, sin ningún tipo de limitación: MassArray system de SequenomTM, SNPlex genotipyng system de Applied Biosystems, GenomeLab SNPstream system, entre otros.

Un segundo aspecto de la invención se refiere a un método para la determinación del riesgo de desarrollar VRP que comprende los siguientes pasos:

a. Extraer el material genético de una muestra biológica aislada.

b. Genotipar al menos un SNP ubicado en al menos uno de los genes seleccionados del grupo que comprende: TNF, TNFR2, TGFB1, TGFB2, SMAD3, SMAD7, IFNG, IL1A, IL1B, IL1RN, IL6, IL8, IL10, NFKB1, NFKBIA, NFKBIB, HGF, CTGF, PDGF, PDGFRA, PI3K, EGF, FGF2, MIF, MMP2, MMP9, MCP1, IGF-11, IGF2, IGF-IR, preferentemente seleccionados de cualquiera de las tablas 1 y 8.

c. Seleccionar al menos un SNP del paso 2 a partir de procedimientos basados en la minimización del error de predicción.

d. i) Introducir los datos de genotipado de los SNPs seleccionados en el paso c) de una muestra problema en un modelo ajustado mediante la aplicación de la técnica Random Forest, ó

ii) Introducir los datos de genotipado de los SNPs seleccionados en el paso c) de una muestra problema en un modelo ajustado mediante la aplicación de cualquiera de las técnicas Support Vector Machines con kernel lineal ó Support Vector Machines con Kernel radial.

En una realización más preferida de este aspecto de la invención, los SNPs genotipados de cada uno de los genes del paso b deben formar un conjunto mínimo de SNPs representativos, localizados en un determinado gen o región capaces de caracterizar a dicho gen/región. La selección de este conjunto implica la eliminación de aquellos SNPs que proporcionan información redundante debido al alto desequilibrio de ligamiento respecto del conjunto de SNPs elegido.

En una realización preferida de este segundo aspecto de la invención en el paso b) se genotipa al menos un SNP o conjunto mínimo de SNPs donde dicho SNPs son TagSNPs y/o están ubicado en exones o regiones promotoras del gen.

En una más realización también preferida de este aspecto de la invención en el paso b) se genotipan al menos dos SNPs que se encuentran en bloques de desequilibrio diferentes, y más preferentemente en genes distintos.

En una realización todavía más preferida de este aspecto de la invención en el paso c) la selección del los SNPs del paso 2 (SNPs más informativos) se realiza por cualquiera por cualquiera de los siguientes procedimientos:

a. un procedimiento basado en la búsqueda exhaustiva del mejor subconjunto de SNPs

b. un procedimiento basado en la búsqueda secuencial partiendo del subconjunto formado por todos los SNPs del paso 2 y desestimando progresivamente aquellos SNPs cuya eliminación del modelo no proporciona un menor error de predicción (Díaz-Uriarte R. y Álvarez de Andrés S. (2006). BMC Bioinformatics 2006, 7:3 doi:10.1186/1471-2105-7-3)

c. un procedimiento basado en la búsqueda secuencial partiendo del subconjunto formado por un único SNP y seleccionando progresivamente aquellos SNPs del paso 2 cuya incorporación al modelo proporciona un menor error de predicción.

En una realización aun más preferida de este aspecto de la invención los SNPs de los genes TNF, TNFR2, TGFB1, TGFB2, SMAD3, SMAD7, IFNG, IL1A, IL1B, IL1RN, IL6, IL8, IL10, NFKB1, NFKBIA, NFKBIB, HGF, CTGF,

PDGF, PDGFRA, PI3K, EGF, FGF2, MIF, MMP2, MMP9, MCP1, IGF-11, IGF2, IGF-IR son seleccionados de la tabla 8 y, en una realización preferida los genes seleccionados del paso c) son aquellos presentados en cualquiera de las tablas 4, 5 o 6.

Un tercer aspecto de la invención proporciona un Kit (en adelante kit de la invención) para determinar el riesgo de desarrollar VRP que comprende medios capaces de llevar a cabo la reacción de detección de los marcadores genéticos ubicados en los genes SMAD7, TNF, PIK3CG, TNFR2. Dichos medios de detección pueden comprender, sin ningún tipo de limitación, oligonucleótidos, enzimas (polimerasas, enzimas de restricción, etc.), tampones, dNTPs, y cualquier otro reactivo requerido para la puesta a punto del kit.

Breve descripción de las figuras

Figura 1.- Cada punto representa el número de SNPs frente a la estimación de error que se comete al clasificar a los individuos utilizando cada SNP para cada modelo. El error de predicción mejora a medida añadimos SNPs al modelo, hasta alcanzar un punto en el que, mientras más SNPs se utilizan, peor es la predicción.

Figura 2. Cada punto representa el número de SNPs frente a la estimación de error que se comete al clasificar a los individuos utilizando cada SNP para cada modelo. El error de predicción mejora a medida añadimos SNPs al modelo, hasta alcanzar un punto en el que, mientras más SNPs se utilizan, peor es la predicción.

Figura 3. Curvas ROC. Establece la capacidad de predicción del modelo. Cuanto más se aleje de la diagonal (hacia arriba), mejor capacidad de predicción tiene el modelo.

Descripción detallada de la invención

A continuación se presentan una serie de ensayos realizados por los inventores, cuya finalidad es ilustrar y poner de manifiesto la especificidad y efectividad de la invención, no suponiendo en ningún caso una limitación a la misma.

1.- *Selección de genes candidatos*

La selección de genes candidatos se realizó en función de la información disponible sobre patologías inflamatorias o profibróticas. Para ello se realizó una extensa busqueda bibliográfica en las bases de datos públicas PubMed y OMIM (www.ncbi.nlm.nih.gov/entrez/), que llevó a la selección de los siguientes 30 genes:

TNF, TNFR2, TGFB1, TGFB2, SMAD3, SMAD7, IFNG, IL1A, IL1B, IL1RN, IL6, IL8, IL10, NFKB1, NFKBIA, NFKBIB, HGF, CTGF, PDGF, PDGFRA, PI3K, EGF, FGF2, MIF, MMP2, MMP9, MCP1, IGF-11, IGF2, IGF-IR

2.- *Selección de SNPs*

La búsqueda y selección de los polimorfismos se realizó utilizando las bases de datos públicas, como dbSNP o HapMap. Se seleccionaron aquellos polimorfismos que fueran bialélicos, como la mayoría de los SNPs, y aquellas deleciones que no superaran las 6 pares de bases.

Para la selección se tuvieron en cuenta los siguientes criterios:

- La localización en el gen: se priorizaron los SNPs en regiones promotoras y exónicas que provoquen un cambio no sinónimo en la transcripción del gen.
- Los bloques de desequilibrio de ligamiento: en función de la información disponible en el proyecto HapMap (www.hapmap.org), se priorizaron los TagSNPs o polimorfismos en diferentes bloques de desequilibrio.
- La frecuencia alélica: se priorizaron los polimorfismos con una frecuencia del alelo menor superior al 10% en Caucásicos.
- los SNPs descritos en la literatura en relación a patologías inflamatorias o profibróticas.

En función de los criterios utilizados se preselecionaron 230 polimorfismos, distribuidos entre los 30 genes candidatos, que de algún modo podría estar relacionados con la enfermedad.

3.- *Selección de pacientes, materiales y métodos*

Los pacientes, seleccionados a partir de centros de referencia ubicados en 5 comunidades autónomas españolas diferentes, se clasificaron según las características clínicas que presentaban en el momento de su inclusión en el estudio, siendo asignados en los grupos caso o control:

Para el *grupo VRP (casos)* se incluyeron pacientes con DR primario que tuviesen una o más de las siguientes características:

- Pliegues fijos en uno o más cuadrantes causados por membranas epirretinianas.
- Cuadros de VRP anterior según la modificación propuesta por Machemer y col sobre la establecida por el Retina Society Committee en 1983 (2).
- Áreas de acortamiento retiniano que precisen retinotomías de descarga y/o retinectomías para poder reaplicar la retina durante la vitrectomía por pars plana (VPP).

Para el *grupo DR (controles)* se incluyeron pacientes con DR primario que tras 3 meses de seguimiento no hubiesen desarrollado una VRP.

Esta clasificación fenotípica de los pacientes fue efectuada por cada investigador de acuerdo a los criterios clínicos mencionados.

4.- *Selección del tamaño muestral*

Se calculó el tamaño muestral adecuado teniendo en cuenta la incidencia de la VRP en el entorno. Según la fórmula de tamaño muestral para muestras de distinto tamaño, considerando como casos a los pacientes con VRP y controles a los pacientes con DR, se estimó necesario recoger muestras de 150 casos de VRP y 300 DR. Con ese tamaño muestral se tendría una potencia del 90% para detectar diferencias significativas entre los grupos caso y control.

5.- *Recolección de muestras y procesado de las muestras*

La toma de muestra se realizó a partir de sangre periférica, extrayendo 6 mi de cada uno de los pacientes incluido en el estudio que se conservaron en tubos de plástico con EDTA a 4ºC. Los envíos de las muestras desde los diferentes centros implicados en el estudio se enviaron al Laboratorio de Biología Molecular del IOBA a temperatura ambiente, una vez por semana.

A medida que se recibieron las muestras de sangre periférica, se realizó la extracción de ADN según el protocolo que se detalla en el kit comercial *REAL Kit* para extracción de DNA de sangre periférica SSS ref: RBM02.

Las muestras de ADN se almacenaron en tubos Eppendorf a -20ºC, identificados con el código de barras correspondiente. La cuantificación del ADN se realizó con la técnica de real time-PCR. En aquellos casos en los que hubo alguna duda sobre la pureza de la muestra se llevó a cabo la verificación mediante espectrofotometría. Para ello se utilizó el espectrofotómetro BioPhotometer, de la casa Eppendorf y se tomó como punto de corte de índice de pureza de la muestra 1,6.

Una vez extraído el ADN de todas las muestras se enviaron al Centro Nacional de Genotipado de Santiago de Compostela, donde se realizó el genotipado de las mismas.

Las muestras de ADN se remitieron al centro de genotipado en placas de 96 pocilios debidamente numeradas y selladas, en concentraciones de 100 ng/ul y con un volumen de 50 ul cada uno.

6.- *Genotipado de las muestras*

El genotipado de las muestras se llevó a cabo de forma enmascarada y tanto lo casos como los controles se analizaron simultáneamente y bajo las mismas condiciones. Se utilizó para este proceso una plataforma de alto rendimiento (SNPlex genotyping System de Applied Biosystems).

7.- *Análisis estadístico de los resultados*

7.1. *Establecimiento de asociaciones significativas*

Se estimaron las frecuencias alélicas y genotípicas, y se verificó el equilibrio de Hardy-Weinberg tanto en la muestra global como en los controles.

Se establecieron las asociaciones nominales de los diferentes SNPs con los grupos control y enfermedad utilizando los modelos de chi-cuadrado y el Test exacto de Fisher y se ajustaron modelos de regresión logística para elegir el mejor modelo de herencia para cada SNP. Como cada individuo posee 2 alelos y el riesgo puede depender de cada uno de los ellos, se definieron 5 modelos de herencia: co-dominante, en donde cada alelo aporta un riesgo independiente;

dominante, en donde una copia del alelo en cuestión es suficiente para modificar el riesgo (homocigoto para ese alelo y heterocigoto tienen el mismo riesgo); recesivo, son necesarias 2 copias del alelo para modificar el riesgo; sobredominante, es decir, ser heterocigoto es un riesgo, y aditivo, en donde cada copia del alelo involucrado modifica el riesgo, de manera que ser homocigoto para ese alelo implica el doble de riesgo.

Se identificaron los posibles haplotipos consistentes con los datos observados en la muestra, considerando conjuntos formados por todos los SNPs estudiados por gen y subconjuntos de alelos consecutivos de tamaño 2 hasta tantos marcadores estudiados en cada gen menos 1. Se estimaron las frecuencias haplotípicas en la muestra total y en cada grupo por separado utilizando el algoritmo EM. Se utilizó un estadístico score para contrastar la asociación entre los haplotipos y el estado caso/control de la muestra. Para este análisis también se ajustaron modelos de herencia: en este caso aditivo, recesivo y dominante.

Estos resultados se corrigieron utilizando el método corrección para comparaciones múltiples en dos pasos propuesto por Rosemberg (Rosenberg P.S. *et al*. (2006) Che A., Chen B.E. Statist Med, 25: 3134-49), en el que en un primer paso cada asociación de un gen con la enfermedad se resume utilizando un único p-valor que combina los análisis de SNPs simples y de bloques haplotípicos y en una segunda etapa estos p-valores por gen, se ajustan controlando la tasa de falsos descubrimientos (False Discovery Rate, FDR) utilizando el q-valor (Storey J.D. *et al* (2003). Annals of Statistics, 31: 2013-35).

Resultados

En la toma de muestras se recogieron finalmente un total de 450 muestras, 312 del grupo control y 138 del grupo casos, entre los diferentes centros incluidos en el estudio. Una vez reunidas todas las muestras y hecha la extracción y cuantificación del DNA genómico, se realizó el genotipado de las mismas. Todas las muestras pudieron ser analizadas.

De los SNPs seleccionados para su estudio, 6 no pudieron ser estudiados por problemas en el diseño de los oligonucleótidos y 27 no pudieron ser estudiados por fallos en el procedimiento de genotipado. Se estudiaron pues 197 SNPs por muestra, lo que supuso un total de 88.650 polimorfismos.

Para todos los polimorfismos se verificó que cumplían con el equilibrio de Hardy-Weinberg tanto en la muestra global como en el grupo control y casos, excepto el rs4916944 del PDGFA, por lo que éste también fue eliminado del análisis. Además, también se estimaron las frecuencias genotípicas para cada polimorfismo.

Asociaciones significativas

• *Asociaciones simples*

Se analizaron las asociaciones de cada SNP por separado con el estado caso- control y de acuerdo al modelo de herencia que mejor explica la distribución de genotipos que existen en la muestra. Se encontraron un total de 22 asociaciones significativas distribuidas en 15 genes diferentes.

• *Asociaciones múltiples*

Todos los bloques haplotípicos analizados verificaron el desequilibrio, es decir, los alelos no se comportaron con independencia entre si. Del análisis global de los haplotipos completos de cada gen consistentes con la muestra (formados por todos los alelos estudiados en cada gen) se observó que existía una asociación significativa con el estado caso/control para los 4 de los 30 genes estudiados.

Además se analizó qué ocurre con los haplotipos formados por subconjuntos de alelos consecutivos de tamaño 2 hasta tantos marcadores como se hubiesen estudiado por gen menos 1. Así, se encontraron también asociaciones significativas de haplotipos consistentes con la muestra en 11 genes.

Corrección para comparaciones múltiples

Fijando el nivel de FDR (False Discovery Rate) (*Benjamini*, *et al*. (*1995*). *Controlling the False Discovery Rate: A Practical and Powerful Approach to Múltiple Testing. J Roy Stat. Soc. (Ser B) 57: p.289-300*), es decir, la proporción de falsos positivos esperados entre los contrastes que resultaron significativos, en 5%, se encontraron 4 genes significativamente asociados con el estado caso/control: el PIK3CG (p-valor: 0,009; q-valor: 0,03), el SMAD7 (p-valor: 0,004; q-valor:0,0250), el TNF (p-valor: 0,005; q-valor: 0,0250), y el TNFR2 (p-valor: 0,019; q-valor: 0,0475).

Este paso de corrección de los resultados obtenidos previamente es imprescindible para establecer la asociación significativa de un gen o un marcador genético con una determinada enfermedad, ya que de no aplicarse las asociaciones obtenidas *a priori* podrían ser erróneas. En el presente estudio la aplicación de esta corrección supuso el descarte de más del 70% de los genes candidatos iniciales.

GEN	p-valor	q-valor
PIK3CG	0,009	0,0300
SMAD7	0,004	0,0250
TNF	0,005	0,0250
TNFR2	0,019	0,0475

Tabla de genes que resultan significativos con el estado caso/control tras la corrección FDR para comparaciones múltiples.

En las tablas 2 a 3 se muestran las asociaciones simples y múltiples de genes de la tabla 1 que resultaron estar significativamente asociados con la enfermedad.

7.2. *Modelos predictivos*

En esta segunda fase y a partir de los modelos estadísticos, se identificaron combinaciones de SNPs que permiten evaluar el riesgo de desarrollar una VRP en un paciente nuevo. Se identificaron modelos predictivos de 2, 10 y 42 SNP.

Determinación de los SNPs predictivos

Se mide la correlación del SNP con la presencia o no de VRP usando lo que se conoce como "ganancia de información" (IG, Information Gain) ((Cover T.M., Thomas J.A. (1991). Elements of information theory. New York: John Wiley): técnica de aprendizaje automatizado (machine learning) que consiste en conseguir reducir la entropía como consecuencia de realizar una división en los datos. La entropía, que denotamos por H, es una medida del grado de incertidumbre que existe sobre un conjunto de datos.

En nuestro contexto, dado un SNP concreto consideramos dos variables, $X = \{AA,aa,Aa\}$ que representa el genotipo observado e $Y = \{0,1\}$ que representa el estado caso/control. Los pasos para determinar si este SNP es o no informativo son los siguientes:

1) Calcular la entropía de la variable Y como $H(Y) = \sum_{k=\{0,1\}} p_k \log\left(\frac{1}{p_k}\right)$, donde p_k es la probabilidad de pertenecer a cada uno de los grupos, k=0 para los controles y k=1 para los casos. Esta probabilidad se estima como la proporción de individuos de la muestra que pertenecen al grupo concreto.

2) Calcular la entropía de la variable Y condicionada a cada uno de los valores de X como $H(Y/X) = \sum_{i\in\{1,2,3\}} p(X = x_i) H(Y/X = x_i)$, donde $x_i \in \{AA, aa, Aa\}$ y $H(Y/X = x_i)$ es la entropía de la variable Y entre los pacientes cuyo genotipo observado es x_i

3) Calcular la IG como $IG(Y/X) = H(Y)-H(Y/X)$, de forma que un valor de IG pequeño significaría que el SNP no es informativo.

4) Determinar la significación estadística de IG utilizando un test de permutaciones (Welch W.J., 1990. Construction of permutation test. J Am Statist Assoc, 85: 693-8). La hipótesis que se contrasta es que el SNP no es buen predictor del estado caso/control. El p-valor para cada uno de los SNP se determina permutando aleatoriamente el estado caso/control observado y calculando en cada caso el valor de IG en esa permutación. El p-valor será la proporción de permutaciones para las que el valor de IG es mayor o igual que el valor original. Se contrasta el valor informativo de cada SNP con 10.000 permutaciones. Puesto que el número de contrastes es alto, utilizamos el nivel de significación basado en el método False-Discovery rate (Benjamini Y., Hochberg Y. 1995).

Modelos predictivos aplicados

El objetivo es ajustar modelos que clasifiquen bien a nuevos individuos una vez conocido su perfil genotípico, es decir que minimicen el error de predicción. Debido a las dimensiones de los datos, se hace necesario elegir entre todas las variables (SNPs) disponibles, conjuntos que predigan el estado de un individuo de la manera más precisa posible.

Se han utilizado técnicas de aprendizaje automatizado (Machine Learning Techniques). Este tipo de técnicas tratan de construir de forma semi-automática modelos estadísticos predictivos. Hay muchos tipos de modelos, pero es impo-

sible, *a priori*, determinar que clase de modelo es más apropiado para un conjunto de datos dado. Por ese motivo, se prueban diferentes tipos de modelos y posteriormente se validan, para finalmente seleccionar el modelo más fiable en cuanto a su capacidad predictiva.

A) *Modelos basados el clasificador de Naïve Bayes*

El modelo de Naïve Bayes utiliza las frecuencias de los diferentes valores de cada SNP en pacientes cuyo estado caso/control es conocido para predecir el estado de nuevos pacientes, cuyo perfil genotípico es conocido pero no su estado. Es el modelo más simple, pero asume que las distintas variables predictoras (SNPs) son independientes entre sí dado el estado caso/control. A pesar de que esta hipótesis es típicamente falsa, en la práctica funciona bien.

Con el método Naïve Bayes simplemente se tabulan el número de veces que un SNP particular ocurre como homocigoto o heterocigoto en una población (caso o control). Esta tabulación proporciona directamente una probabilidad condicionada de la forma. Para clasificar a un nuevo paciente, se utilizó la regla de Bayes.

La clase con la mayor probabilidad es la que se elige para el paciente no clasificado.

Para seleccionar los SNPs más relevantes, teniendo en cuenta que el objetivo es conseguir la máxima exactitud en la discriminación, se ha utilizado un método de eliminación recursiva (backward). Se parte del modelo estimado a partir de todos los SNPs y se calcula su exactitud a través del error de cross-validación. En el segundo paso se clasifican los datos utilizando N-1 SNPs, donde N es el número total de SNPs, N veces, quitando cada vez un SNP. Se comprueba si al eliminar cada uno de los SNPs aumenta la exactitud y en caso afirmativo, se descarta dicho SNP. Se repite el proceso hasta que el hecho de eliminar un SNP conlleve una pérdida de exactitud significativa.

Para calcular el error de cross-validación, utilizamos la técnica de validación cruzada del tipo n-fold (n-fold cross validation). Esta técnica es un proceso iterativo, de forma que en cada paso se dividen los datos en n-1 muestras de entrenamiento (training set) y 1 muestra de contraste (test set). Con las muestras de entrenamiento se ajusta el mejor modelo posible y con la muestra de contraste se comprueba la validez de dicho modelo. El proceso continúa hasta que todos los grupos se han utilizado como muestra de contraste. El error de cross-validación, será la media de los errores cometidos con cada una de las muestras de contraste. En este caso se ha utilizado n=20.

B) *Modelos basados en SVM (Support Vector Machine)*

La técnica de SVM tiene como objetivo obtener un hiperplano óptimo capaz de separar lo mejor posible dos clases. El caso más sencillo es en el que un hiperplano (líneas de más de 2 dimensiones) es capaz de separar correctamente todos los puntos, pero esto no ocurre siempre. Es posible aplicar funciones kernel cuya misión es transformar el espacio de partida en otro espacio, donde sea posible resolver el problema. Por ejemplo, supongamos que tenemos sólo 2 variables (SNPs) y que los datos no pueden ser correctamente separados en los dos grupos, caso/control, por una línea recta (kernel lineal). SVM creará un modelo apropiado, modificando el espacio de partida. Así por ejemplo, un kernel cuadrático podría convertir puntos 2-dimensionales en puntos 3-dimensionales: {SNP1, SNP2}->{SNP1 x SNP1, SNP1 x SNP2, SNP2 x SNP2}, en donde el problema pueda ser resuelto a través de hiperplanos. El SVM podría llevar a cabo la partición de los puntos en este nuevo espacio a través de hiperplanos. Claramente la elección del kernel es muy importante en este tipo de modelos, los más relevantes son el lineal, el radial y el polinómico. Un tutorial sobre esta técnica puede encontrarse en (Burges C.J.C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery, 2:121-67).

Como en el caso del modelo Naïve Bayes, para seleccionar los SNPs más relevantes se ha utilizado un método de eliminación recursiva (backward) basado en minimizar el error de cross-validación, calculado a partir de validación cruzada de tipo n-fold, con n=20.

C) *Modelos basados en árboles de decisión*

Los modelos estadísticos basados en árboles de decisión trabajan de forma recursiva creando particiones de un conjunto de datos dado para conseguir la mejor clasificación de los individuos en relación a una variable respuesta, en nuestro caso el estado caso/control.

Hay varios algoritmos para construir modelos predictivos basados en árboles de decisión. En el desarrollo de la invención se utilizó el método conocido como CART (Classification and Regression Tress) (Breiman L.F., *et al* (1984). Classification and regression trees. Chapman & Hall, New York). In Machine Learning, 45, 5-32 (2001)), un método de árbol binario, que se construye según el siguiente proceso: en primer lugar se encuentra la variable predictora que mejor divide los datos en los 2 grupos (caso/control). Los datos se separan en estos dos grupos y el proceso se aplica separadamente a cada uno de ellos. El proceso continúa hasta que se alcance un criterio de parada relacionado, habitualmente, bien con un mínimo número de datos en cada subgrupo, bien con un máximo número de variables predictoras que puedan intervenir en el modelo o bien con ambos parámetros a la vez. En cada paso se añade un nuevo "nodo" al árbol, que en este contexto se corresponde con un SNP.

En este tipo de modelos es muy frecuente el problema de sobreajuste (overfitting) que surge al utilizar demasiadas variables predictoras en relación a la cantidad de datos disponibles y se traduce en modelos poco generalizables. Para asegurar que el árbol de decisión ajustado no está sobreajustado, se utiliza lo que se conoce como la "poda" (pruning) del árbol que consiste en eliminar subárboles que podrían ser erróneos. Hay muchas técnicas de poda, aquí se ha utilizado la regla "1 desviación estándar" (1 Standard Error rule) que se basa en el error de cross-validación. Según esta regla el árbol con mejores propiedades de generalización es el modelo más simple con un error de cross-validación menor que menor error de cross-validación + 1 error estándar.

D) *Modelos basados en Random Forest*

Básicamente, los Random Forest (Breiman, L. Random Forests. In Machine Learning, 45, 5-32 (2001)) son modelos consistentes en una colección de árboles de decisión, de forma que cada árbol se construye a partir de una serie de variables seleccionadas aleatoriamente. Una vez construido el modelo para clasificar a un nuevo individuo, cada uno de los árboles emite un "voto", asignándose a la clase que tenga un mayor número de votos.

Un resultado de este tipo de modelo son varias medidas de la importancia de cada una de las variables implicadas. La más importante está basada en el decrecimiento del error de precisión cuando los valores de una variable en un nodo de un árbol se permutan aleatoriamente. Esta es la medida que se utiliza en el procedimiento de selección de los SNPs más predictivos.

Para seleccionar SNPs se utiliza el proceso iterativo propuesto por Díaz-Uriarte y Álvarez de Ándres (Díaz-Uriarte R. y Álvarez de Andrés S. (2006). BMC Bioinformatics 2006, 7:3 doi:10.1186/1471-2105-7-3). En cada iteración, se construye un nuevo modelo después de descartar aquellas variables con menor medida de importancia en la iteración anterior. La diferencia con otros métodos que existen de eliminación progresiva de variables radica en que no se recalculan las medidas de importancia en cada paso del algoritmo, mejorando así el posible problema de sobreajuste. Se elige el modelo más simple (con menor número de variables) con una tasa de error menor que la mínima tasa de error + 1 error estándar (1 Estándar Error rule).

E) *Validación de los modelos*

Para validar el modelo se utiliza la técnica de validación cruzada del tipo n-fold (n-fold cross validation). Es un proceso iterativo, de forma que en cada paso se dividen los datos en n-1 muestras de entrenamiento (training set) y 1 muestra de contraste (test set). Con las muestras de entrenamiento se ajusta el mejor modelo posible y con la muestra de contraste se comprueba la validez del modelo. El proceso continúa hasta que todos los grupos se han utilizado como muestra de contraste.

Utilizando esta técnica se obtendrá un valor predicho por el modelo para cada una de las observaciones, lo que permite utilizar las siguientes herramientas de evaluación:

Sensibilidad. Proporción de casos que son clasificados como tal por el modelo predictivo, por tanto da una idea de la utilidad del modelo para identificar a los casos. Es una característica intrínseca de la prueba no viéndose afectada por la prevalencia de la enfermedad.

Especificidad. Proporción de controles que son clasificados como tal por el modelo predictivo, por tanto da una idea de la utilidad del modelo para identificar a los controles. Al igual que la sensibilidad es una característica intrínseca de la prueba no viéndose afectada por la prevalencia de la enfermedad.

Proporción de falsos positivos. Proporción de pacientes controles clasificados por el modelo como casos.

Proporción de falsos negativos. Proporción de pacientes casos clasificados por el modelo como controles.

Exactitud. Proporción de pacientes clasificados correctamente por el modelo.

Índice J de Youden o seguridad diagnóstica. Se calcula como Sensibilidad + Especificidad -1. Cuanto más se aproxima a 1, mayor es la calidad del resultado predicho por el modelo.

Odds ratio diagnóstica. Indica cuántas veces es más frecuente que el modelo clasifique bien a un paciente. Es la misma interpretación que cualquier otra odds-ratio, si es >1 la relación que existe entre lo que predice el modelo y lo realmente observado es positiva, es decir, es más probable acertar; si es <1 es más probable no acertar con la clasificación y si es 1 indica que no hay relación entre lo que predice el modelo y lo realmente observado.

Resultado de la aplicación de los modelos predictivos

En función de los valores obtenidos para cada modelo, se seleccionaron los siguientes modelos que se ajustaron con los SNPs de la tabla 1:

1.- *Modelos basados en SVM (Support Vector Machine)*

En este tipo de modelos no pueden existir valores pérdida en las variables predictoras (SNP). Hay dos opciones: (1) imputar los valores pérdida, completando los genotipos que faltan en función del resto de genotipos observados en el mismo paciente, y (2) sólo considerar los pacientes para los que hay los datos completos. En este último caso se seleccionaron 166 pacientes, de los cuales 114 (68.7%) eran controles y 52 (31.3%) casos.

Se ajustaron modelos utilizando los kernel lineal, radial y polinómico de grado 2 y 3 (cuadrático y cúbico). A continuación se presentan únicamente aquéllos modelos para los que se obtuvieron resultados satisfactorios.

kernel lineal (Muestra sin tener en cuenta los casos con valores perdidos)

Se seleccionaron, utilizando el método de eliminación backward, 42 SNPs (tabla 4) con un error de predicción estimado del 16.1%. Fig. 1. La presencia de estos 42 SNPs es indicativa de un mayor riesgo a desarrollar VRP.

TABLA 4

SNP	Gen	SNP	Gen
rs718768	EGF	rs4820571	MIF
rs9999824		rs738806	
rs1476217	FGF2	rs2192853	MMP2
rs167428		rs243840	
rs3804158		rs9928731	
rs9990554		rs3787268	MMP9
rs5745687	HGF	rs11722146	NFKB1
rs2098395	IFNG	rs230540	
rs6214	IGF1	rs17103274	NFKBIA
rs7136446		rs3138045	
rs3213221	IGF2	rs3136640	NFKBIB
rs8038015	IGF-IR	rs3136646	
rs17015767	IL10	rs4289498	PDGFRA
rs1800871		rs6961244	PIK3CG
rs3024498		rs849380	
rs4390174		rs3743343	SMAD3
rs3917368	IL1B	rs2241715	TGFB1
rs315946	IL1RN	rs2796821	TGFB2
rs973635		rs2229094	TNF
rs11766273	IL6	rs1061622	TNFR2
		rs1061628	
		rs3397	

kernel radial (Muestra imputando los valores perdidos).

Se seleccionaron, utilizando el método de eliminación backward, 10 SNPs (tabla 5) con un error de predicción estimado del 22,4% (Fig. 2). La presencia de estos 21 SNPs es indicativa de un menor riesgo de desarrollar VRP.

TABLA 5

SNP	Gen
rs17015767	IL10
rs1688072	IL1RN
rs243840	MMP2
rs997476	NFKB1
rs17103274	NFKBIA
rs3138056	
rs3136646	NFKBIB
rs2241715	TGFB1
rs2796821	TGFB2
rs2229094	TNF

2.- *Modelo basado en Random Forest* (Muestra sin tener en cuenta los casos con valores perdidos)

En este tipo de modelos no pueden existir pérdidas en las variables predictoras (SNP). Hay dos opciones: (1) imputar los valores pérdidas, completando los genotipos que faltan en función del resto de genotipos observados en el mismo paciente, y (2) sólo considerar los pacientes para los que tenemos los datos completos. En este último caso se seleccionaron 166 pacientes, de los cuales 114 (68.7%) son controles y 52 (31.3%) casos, resultado que la presencia los SNPs presentados en la tabla 2 es indicativa de un mayor riesgo a desarrollar VRP, según se presenta en la tabla de riesgo siguiente.

TABLA 6

SNPs	Gen	Importancia del SNP
rs7136446	IGF1	2.138141
rs2229094	TNF	2.101151

TABLA DE RIESGO		
rs71364446	rs2229094	Probabilidad de desarrollar VRP
TT	TT	0.084
TT	CC	0.000
TT	CT	0.000
CC	TT	0.202
CC	CC	0.356
CC	CT	0.620
CT	TT	0.648

Validación de los modelos

Las medidas de la capacidad predictiva de cada uno de los modelos los modelos seleccionados se muestra en la siguiente tabla (tabla 7). Comentar brevemente los resultados obtenidos.

TABLA 7

Modelo	Medida	N = 116		
		Valor	IC al 95%	
			Inferior	Superior
SVM kernel lineal	Sensibilidad (%)	69.2	55.7	80.1
	Especificidad (%)	83.3	74.6	89.5
	Falsos positivos (%)	16.7	10.5	25.4
	Falsos negativos (%)	30.8	19.9	44.3
	Exactitud (%)	78.4	71.1	84.2
	Seguridad diagnóstica	0.5	-	-
	Odds-ratio diagnóstica	11.25	5.07	24.96
SVM kernel radial		70.4	57.2	80.9
	Sensibilidad (%)	70.2	61.2	77.8
	Especificidad (%)	29.8	22.2	38.8
	Falsos positivos (%)	29.6	19.1	42.8
	Falsos negativos (%)	70.2	62.9	76.6
	Exactitud (%)	0.4	-	-
	Seguridad diagnóstica	5.59	2.75	11.35
Random Forest	Odds-ratio diagnóstica	65.4	51.8	76.8
	Especificidad (%)	71.1	62.1	78.6
	Falsos positivos (%)	28.9	21.4	37.9
	Falsos negativos (%)	34.6	23.2	48.2
	Exactitud (%)	69.3	61.9	75.8
	Seguridad diagnóstica	0.4	-	-
	Odds-ratio diagnóstica	4.64	2.30	9.34

Para los modelos basados en SVM y en Random Forest se imputan los valores pérdidas. El clasificador de Naïve Bayes y los árboles de decisión, permiten trabajar con este tipo de valores. Fig. 10 (Curvas ROC).

(Tabla pasa a página siguiente)

TABLA 1

Gen	SNP	IG	p-valor	p-valor ajustado FDR
CTGF	rs4897554	0.0763	0.3523	0.6224
	rs6917644	0.0908	0.3295	0.6224
	rs9399005	0.0543	0.0583	0.3498
	rs928501	0.0862	0.9984	0.9984
	rs1931002	0.2914	0.4149	0.6224
	rs9483364	0.0784	0.8452	0.9984
EGF	rs718768	0.0876	0.9985	0.9985
	rs17238095	0.0349	0.8318	0.9985
	rs1024600	0.0763	0.9727	0.9985
	rs11568943	0.2944	**0.0238**	0.2142
	rs1860129	0.0998	0.8954	0.9985
	rs2237051	0.0998	0.9603	0.9985
	rs4698803	0.0301	0.3774	0.9985
	rs6533485	0.1021	0.9510	0.9985
	rs9999824	0.0776	0.9316	0.9985
FGF2	rs308417	0.3459	0.0799	0.1798
	rs308435	0.0025	0.0769	0.1798
	rs167428	0.0742	0.9316	0.9944
	rs9990554	0.0610	**0.0036**	**0.0324**
	rs1048201	0.1433	0.2334	0.4201
	rs3804158	0.0731	0.7028	0.9036
	rs7683093	0.1396	0.0590	0.1798
	rs1476217	0.0908	0.9944	0.9944
	rs1982569	0.0782	0.6526	0.9036
HGF	rs2074724	0.0147	0.3433	0.5663
	rs5745687	0.3294	**0.0135**	0.0540
	rs917183	0.0843	0.9237	0.9237
	rs1558001	0.0701	0.4247	0.5663
IFNG	rs2098395	0.0200	0.7865	0.9517
	rs2193049	0.0346	0.9425	0.9517
	rs2069727	0.0848	0.9195	0.9517
	rs2069718	0.0879	0.9517	0.9517
	rs12306852	0.0725	0.0828	0.4140
IGF1	rs2971575	0.0945	0.8439	0.9717
	rs6214	0.0759	0.7450	0.9717
	rs7136446	0.1025	0.9717	0.9717
	rs9989002	0.0416	0.8866	0.9717
	rs2195240	0.0078	0.1589	0.4237
	rs5742629	0.0219	0.0605	0.4216
	rs1019731	0.2035	0.2994	0.5988
	rs35767	0.0232	0.1054	0.4216
IGF2	rs2585	0.0690	0.2439	0.9967
	rs3213221	0.0934	0.9967	0.9967
	rs3741212	0.0927	0.9575	0.9967
	rs1003483	0.0793	0.8576	0.9967
	rs3741208	0.1000	0.7938	0.9967
IGF-IR	rs12899533	0.0863	0.9441	0.9996
	rs4305005	0.1569	0.8652	0.9996

IL10	rs8038015	0.0825	0.8532	0.9996
	rs7166287	0.0846	0.8274	0.9996
	rs10794486	0.0814	0.9996	0.9996
	rs4966035	0.0848	0.9638	0.9996
	rs1879613	0.0472	0.1982	0.9910
	rs2229765	0.0749	0.6820	0.9996
	rs1568501	0.0873	0.9262	0.9996
	rs2684787	0.0008	0.0534	0.5340
	rs4390174	0.0824	0.8579	0.9963
	rs3024498	0.0115	0.4438	0.9963
	rs3024493	0.1599	0.2163	0.9963
	rs2222202	0.0830	0.9963	0.9963
	rs1800872	0.0659	0.8553	0.9963
	rs1800871	0.0598	0.8016	0.9963
	rs1800890	0.0762	0.9033	0.9963
	rs17015767	0.0370	0.3480	0.9963
	rs10494879	0.0853	0.9943	0.9963
IL1A	rs2048874	0.1558	0.2328	0.9902
	rs2856836	0.0677	0.6869	0.9902
	rs1304037	0.0757	0.9902	0.9902
	rs3783550	0.0675	0.9461	0.9902
	rs17561	0.0731	0.9741	0.9902
	rs3783516	0.0853	0.9899	0.9902
IL1B	rs7596684	0.0421	0.9399	0.9813
	rs3917368	0.0925	0.8864	0.9813
	rs1143634	0.0229	0.6231	0.9813
	rs16944	0.0926	0.9813	0.9813
	rs4848306	0.0930	0.8241	0.9813
IL1RN	rs1688072	0.0877	**0.0108**	**0.0436**
	rs1794066	0.0723	0.4668	0.6909
	rs446433	0.0512	0.6813	0.7786
	rs973635	0.1112	**0.0109**	**0.0436**
	rs315952	0.0502	0.9712	0.9712
	rs315949	0.0720	0.4296	0.6909
	rs3087270	0.0639	0.3488	0.6909
	rs315946	0.1689	0.5182	0.6909
IL6	rs4719713	0.0917	0.3837	0.9926
	rs12700386	0.0303	0.7401	0.9926
	rs1800795	0.0800	0.5987	0.9926
	rs1474347	0.0856	0.9697	0.9926
	rs2069840	0.0848	0.9926	0.9926
	rs11766273	0.4387	0.5225	0.9926
	rs2140543	0.1017	0.9449	0.9926
IL8	rs2227306	0.0953	0.9997	0.9997
	rs2227543	0.0799	0.8647	0.9997
	rs4694178	0.0732	0.2899	0.8697
MCP1	rs4795893	0.0772	0.6211	0.6231
	rs2857653	0.0321	0.4942	0.6231
	rs1024611	0.0224	0.3396	0.6231
	rs3760396	0.0799	**0.0033**	**0.0165**
	rs2530797	0.0759	0.6231	0.6231

MIF	rs9612503	0.0784	0.9990	0.9990
	rs738806	0.0586	0.8494	0.9990
	rs755622	0.0760	0.5729	0.9990
	rs1007888	0.0505	0.2456	0.9990
	rs4820571	0.0866	0.9787	0.9990
MMP2	rs1561220	0.0361	**0.0242**	0.1936
	rs243866	0.0361	0.8412	0.8412
	rs243864	0.0356	0.6630	0.7577
	rs9928731	0.0738	0.5356	0.7141
	rs243845	0.0807	0.2926	0.4682
	rs243840	0.0233	0.2017	0.4034
	rs7201	0.0645	0.1974	0.4034
	rs2192853	0.0608	0.1140	0.4034
MMP9	rs4810482	0.0941	0.9351	0.9351
	rs3918241	0.1627	0.3489	0.6978
	rs3918253	0.0964	0.8926	0.9351
	rs3787268	0.0360	0.0825	0.4950
	rs2250889	0.3963	0.2454	0.6978
	rs2274756	0.1587	0.4741	0.7112
NFKB1	rs3774932	0.0701	0.9574	0.9817
	rs230540	0.0812	0.9793	0.9817
	rs4698858	0.0687	0.2879	0.9817
	rs11722146	0.0470	0.8928	0.9817
	rs4648110	0.0148	0.4593	0.9817
	rs7674640	0.0922	0.9817	0.9817
	rs997476	0.3747	0.3586	0.9817
NFKBIA	rs7152826	0.0826	0.9846	0.9943
	rs3138056	0.0865	0.9691	0.9943
	rs696	0.0901	0.9155	0.9943
	rs3138045	0.0520	0.9943	0.9943
	rs2007960	0.0802	0.7415	0.9943
	rs17103274	0.1948	**0.0039**	**0.0234**
NFKBIB	rs10410544	0.0831	0.9339	0.9947
	rs2053071	0.0981	0.9662	0.9947
	rs11879872	0.0595	0.9457	0.9947
	rs3136640	0.0863	0.9254	0.9947
	rs3136646	0.0404	0.9947	0.9947
PDGFA	rs7777705	0.0847	0.9722	0.9722
	rs11764261	0.1169	0.4503	0.9006
	rs11768030	0.0774	0.1954	0.7816
	rs7806249	0.0843	0.7993	0.9722
PDGFRA	rs7691129	0.1220	0.1005	0.1675
	rs7656613	0.0433	**0.0181**	0.0905
	rs4289498	0.0611	0.0991	0.1675
	rs17739921	0.0910	0.8875	0.8875
	rs4864877	0.0693	0.5742	0.7178
PIK3CG	rs849380	0.0040	0.1249	0.2810
	rs849384	0.0859	0.9186	0.9186
	rs757902	0.0827	0.8869	0.9186
	rs849387	0.0945	0.7982	0.9186
	rs4730204	0.0389	0.8779	0.9186

	rs4727666	0.0328	0.8149	0.9186
	rs4460309	0.0013	0.0747	0.2676
	rs6961244	0.2269	**0.0453**	0.2676
	rs3173908	0.0023	0.0892	0.2676
SMAD3	rs7177795	0.0895	0.7203	0.9993
	rs4776881	0.0918	0.9535	0.9993
	rs6494634	0.0759	0.8182	0.9993
	rs2033785	0.0363	0.6311	0.9993
	rs731874	0.0570	0.9993	0.9993
	rs8032802	0.2336	**0.0242**	0.1936
	rs8031627	0.0455	0.9799	0.9993
	rs3743343	0.0375	0.5346	0.9993
SMAD7	rs4939826	0.3533	0.2059	0.3089
	rs7226855	0.0493	**0.0036**	**0.0216**
	rs9946510	0.0865	0.8799	0.8799
	rs6507877	0.0791	0.8621	0.8799
	rs2878889	0.0660	0.1930	0.3089
	rs2337143	0.0281	0.1817	0.3089
TGFB1	rs8179181	0.0546	0.7354	0.8632
	rs4803455	0.0682	0.5346	0.8632
	rs2241715	0.0811	0.8632	0.8632
	rs2241713	0.0824	0.7369	0.8632
TGFB2	rs4846476	0.0014	0.0966	0.3864
	rs4846267	0.0786	0.9973	0.9987
	rs947712	0.0636	0.9987	0.9987
	rs1891467	0.0756	0.9858	0.9987
	rs2796821	0.0425	**0.0060**	**0.0480**
	rs2000220	0.0877	0.9877	0.9987
	rs9308385	0.0241	0.8961	0.9987
	rs900	0.0783	0.9979	0.9987
TNF	rs2857602	0.0913	0.9781	0.9781
	rs2857706	0.0321	**0.0006**	**0.0054**
	rs909253	0.0582	0.5435	0.9781
	rs2229094	0.0737	0.8881	0.9781
	rs1799964	0.0609	0.6860	0.9781
	rs1800629	0.1099	0.1087	0.3261
	rs673	0.5872	0.0837	0.3261
	rs2256965	0.0771	0.4221	0.9497
	rs2256974	0.0711	0.9630	0.9781
TNFR2	rs652284	0.0932	0.9756	0.9990
	rs542282	0.0334	0.9527	0.9990
	rs1061622	0.0192	0.9077	0.9990
	rs590977	0.0611	0.4635	0.9990
	rs3397	0.0936	0.9990	0.9990
	rs1061628	0.0959	0.7846	0.9990

Tabla 2

GEN	POLIMORFISMO	MODELO	GENOTIPO	CONTROLES		CASOS		OR	IC95%	p-valor	Observaciones
				N	%	N	%				
SMAD7	rs7226855	Sobre-dom	A/G	132	43,7	81	63,8	2,27	1,48-3,48	0,0001369	**F Riesgo**
TNF	rs2229094	Dominante	C/T- C/C	144	48	80	65	2,02	1,31-3,11	0,001326	**F Riesgo**
	rs2256974	Aditivo	T	268	68,7	122	31,3	0,64	0,42-0,97	0,03153	**F Protección**
	rs2857706	Sobre-dom	A/G	74	24,8	48	37,8	1,84	1,18-2,87	0,007672	**F Riesgo**

Tabla 3.

GEN	SNPs ESTUDIADOS	nº SNPs	BLOQUE HAPLOTIPICO	p-score	HAPLOTIPO	FRECUENCIAS %		MOD. HERENCIA	OR	IC95%	RIESGO/ PROTECCIÓN
						CONTROL	CASO				
PIK3CG	rs849380 rs849384 rs757902 rs849387 rs4730204 rs4727666 rs4460309 rs6961244 rs3173908	3	SNP7 al SNP9	0,00291	TCC	0,01	1,14	Dominante	$15,8.10^4$	$15,8.10^4$ -$15,8.10^4$	Factor de riesgo

Tabla 3 (continuación).

GEN	SNPs ESTUDIADOS	nº SNPs	BLOQUE HAPLOTIPICO	p-score	HAPLOTIPO	FRECUENCIAS %		MOD. HERENCIA	OR	IC95%	RIESGO/ PROTECCIÓN
						CONTROL	CASO				
TNF	1- rs2857602 2- rs2857706 3- rs909253 4- rs2229094 5- rs1799964 6- rs1800629 7- rs673 8- rs2256965 9- rs2256974	2	SNP3 al SNP4	0,01764	TC	27,14	37,18	Dominante	1,9775	1,2690- 3,0815	F Riesgo
			SNP4 al SNP5	0,00045	CT	2,98	9,49	Dominante	3,6533	1,8254- 7,3117	F Riesgo
					TC	0,18	0		0	0	F Protección
		3	SNP2 al SNP4	0,03087	ATC	15,16	20,52	Dominante	1,8672	1,1776- 2,9607	F Riesgo
			SNP3 al SNP5	0,00095	CTC	0,18	0	Dominante	0	0	F Protección
					TCT	2,97	9,57		3,6428	1,8003- 7,3708	F Riesgo
			SNP4 al SNP6	0,0001	CCG	24,67	26,6	Dominante	1,5698	1,0140- 2,4304	F Riesgo
					CTG	2,97	9,53		4,0825	2,0114- 8,2861	F Riesgo
					TCG	0,18	0		0	0	F Protección
		4	SNP1 al SNP4	0,037	TATC	15,23	20,52	Dominante	1,8423	1,1625- 2,9195	F Riesgo
			SNP2 al SNP5	0,00145	ATCC	15,29	20,2	Dominante	1,7374	1,0935- 2,7605	F Riesgo
					GCTC	0,18	0		0	0	F Protección
					GTCT	3,07	9,54		3,3584	1,6612- 6,7899	F Riesgo
			SNP3 al SNP6	0,00045	CTCG	0,18	0	Dominante	0	0	F Protección
					TCTG	29,7	9,57		3,9151	1,9192- 7,9866	F Riesgo
			SNP4 al SNP7	0,00055	CCGG	24,34	26,6	Dominante	1,6124	1,0390- 2,5023	F Riesgo
					CTGG	2,98	9,53		4,1539	2,0421- 8,4495	F Riesgo
		5	SNP1 al SNP5	0,00085	TATCC	15,33	20,2	Dominante	1,7284	1,0883- 2,7450	F Riesgo
					TGCTC	0,18	0		0	0	F Protección
					TGTCT	3,06	9,5		3,3495	1,6568- 6,7714	F Riesgo
			SNP2 al SNP6	0,00045	GCTCG	0,18	0	Aditivo	0	0	F Protección
					GTCTG	3,07	9,54		3,4744	1,7616- 6,8529	F Riesgo
			SNP3 al SNP7	0,0009	TCTGG	2,98	9,57	Dominante	3,972	1,9424- 8,1225	F Riesgo
			SNP4 al SNP8	0,0029	CCGGC	24,01	26,36	Dominante	1,5993	1,0215- 2,5039	F Riesgo
					CTGGT	3,2	9,47		3,6146	1,7875- 7,3091	F Riesgo

Tabla 3 (continuación).

GEN	SNPs ESTUDIADOS	nº SNPs	BLOQUE HAPLOTIPICO	p-score	HAPLOTIPO	FRECUENCIAS %		MOD. HERENCIA	OR	IC95%	RIESGO/ PROTECCIÓN
						CONTROL	CASO				
TNF	1- rs2857602 2- rs2857706 3- rs909253 4- rs2229094 5- rs1799964 6- rs1800629 7- rs673 8- rs2256965 9- rs2256974	6	SNP1 al SNP6	0,0001	TGCTCG	0,18	0	Aditivo	0	0	F Protección
					TGTCTG	3,06	9,5		3,4547	1,7510- 6,8161	F Riesgo
			SNP2 al SNP7	0,0005	ATCCGG	14,95	20,2	Dominante	1,7684	1,1056- 2,8286	F Riesgo
					GTCTGG	3,07	9,54		3,653	1,7886- 7,4609	F Riesgo
			SNP3 al SNP8	0,00332	TCCGGC	24,09	26,7	Dominante	1,631	1,0356- 2,5687	F Riesgo
					TCTGGT	3,05	9,48		3,8959	1,9002- 7,9876	F Riesgo
			SNP4 al SNP9	0,0089	CCGGCG	16,85	21,25	Dominante	1,8344	1,1437- 2,9423	F Riesgo
					CTGGTG	3,22	9,47		3,4399	1,6941- 6,9850	F Riesgo
		7	SNP1 al SNP7	0,0003	TATCCGG	15,33	20,2	Dominante	1,7131	1,0713- 2,7394	F Riesgo
					TGTCTGG	3,06	9,5		3,6146	1,7701- 7,3811	F Riesgo
			SNP2 al SNP8	0,00166	ATCCGGC	14,64	20,26	Dominante	1,8938	1,1730- 3,0575	F Riesgo
					GTCTGGT	3,07	9,48		3,5416	1,7452- 7,1873	F Riesgo
			SNP3 al SNP9	0,01135	CTTGGCG	1,24	0	Dominante	0	0	F Protección
					TCCGGCG	16,79	21,11		1,8045	1,1173- 2,9143	F Riesgo
					TCTGGTG	2,95	9,48		3,813	1,8352- 7,9221	F Riesgo
		8	SNP1 al SNP8	0,0022	TATCCGAC	14,94	20,26	Dominante	1,8385	1,1395- 2,9661	F Riesgo
					TGTCTGGT	3,18	9,34		3,4819	1,7174- 7,0592	F Riesgo
			SNP2 al SNP9	0,00563	ATCCGGCG	14,67	20,31	Dominante	1,8278	1,1368- 2,9388	F Riesgo
					CTTGGCG	1,3	0		0	0	F Protección
					CTTGGTG	0,96	0,42		0,421	0,2984- 0,5941	F Riesgo
					GTCTGGTG	2,97	9,48		3,682	1,8032- 7,5181	F Riesgo
		9	SNP1 al SNP9	0,03965	ATCCGGCG	15	20,31	Dominante	1,7532	1,0883- 2,8244	F Riesgo
					TGCTTGGCG	1,36	0		0	0	F Protección
					TGTCTGGTG	3,19	9,34		3,6296	1,7805- 7,3992	F Riesgo

Tabla 3 (continuación).

GEN	SNPs ESTUDIADOS	nº SNPs	BLOQUE HAPLOTIPICO	p-score	HAPLOTIPO	FRECUENCIAS %		MOD. HERENCIA	OR	IC95%	RIESGO/ PROTECCIÓN
						CONTROL	CASO				
TNFR2	1- rs652284	6	SNP1 al SNP6	0,00797	CGGATC	0,13	2,82	Aditivo	16,9589	15,8670- 18,1259	F Riesgo
	2- rs542282				CGGATT	1,2	0		0	0	F Protección
	3- rs1061622				CGGCTT	0,54	0,87		3,0015	2,2336- 4,0333	F riesgo
	4- rs590977				TGGATC	0,76	0		0	0	F Protección
	5- rs3397				TGTATC	0	1,37		$71,6.10^{7}$	$71,6.10^{7}$	F Riesgo
	6- rs1061628										

Tabla 8			
SNP	**Gen**	**SNP**	**Gen**
rs4897554	CTGF	rs2229765	IGF-IR
rs6917644	CTGF	rs1568501	IGF-IR
rs9399005	CTGF	rs2684787	IGF-IR
rs928501	CTGF	rs4390174	IL10
rs1931002	CTGF	rs3024498	IL10
rs9483364	CTGF	rs3024493	IL10
rs718768	EGF	rs2222202	IL10
rs17238095	EGF	rs1800872	IL10
rs1024600	EGF	rs1800871	IL10
rs11568943	EGF	rs1800890	IL10
rs1860129	EGF	rs17015767	IL10
rs2237051	EGF	rs10494879	IL10
rs4698803	EGF	rs2048874	IL1A
rs6533485	EGF	rs2856836	IL1A
rs9999824	EGF	rs1304037	IL1A
rs308417	FGF2	rs3783550	IL1A
rs308435	FGF2	rs17561	IL1A
rs167428	FGF2	rs3783516	IL1A
rs9990554	FGF2	rs7596684	IL1B
rs1048201	FGF2	rs3917368	IL1B
rs3804158	FGF2	rs1143634	IL1B
rs7683093	FGF2	rs16944	IL1B
rs1476217	FGF2	rs4848306	IL1B
rs1982569	FGF2	rs1688072	IL1RN
rs2074724	HGF	rs1794066	IL1RN
rs5745687	HGF	rs446433	IL1RN
rs917183	HGF	rs973635	IL1RN
rs1558001	HGF	rs315952	IL1RN
rs2098395	IFNG	rs315949	IL1RN
rs2193049	IFNG	rs3087270	IL1RN
rs2069727	IFNG	rs315946	IL1RN
rs2069718	IFNG	rs4719713	IL6
rs12306852	IFNG	rs2056576	IL6
rs2971575	IGF1	rs12700386	IL6
rs6214	IGF1	rs1800795	IL6
rs7136446	IGF1	rs1474347	IL6
rs9989002	IGF1	rs2069840	IL6
rs2195240	IGF1	rs11766273	IL6
rs5742629	IGF1	rs2140543	IL6
rs1019731	IGF1	rs2227306	IL8
rs35767	IGF1	rs2227543	IL8
rs2585	IGF2	rs4694178	IL8
rs3213221	IGF2	rs4795893	MCP1
rs3741212	IGF2	rs2857653	MCP1
rs1003483	IGF2	rs1024611	MCP1
rs3741208	IGF2	rs3760396	MCP1
rs12899533	IGF-IR	rs2530797	MCP1
rs4305005	IGF-IR	rs9612503	MIF
rs8038015	IGF-IR	rs738806	MIF
rs7166287	IGF-IR	rs755622	MIF
rs10794486	IGF-IR	rs1007888	MIF
rs4966035	IGF-IR	rs4820571	MIF
rs1879613	IGF-IR	rs1561220	MMP2

Tabla 8 (continuación)			
SNP	Gen	SNP	Gen
rs243866	MMP2	rs7177795	SMAD3
rs243864	MMP2	rs4776881	SMAD3
rs9928731	MMP2	rs6494634	SMAD3
rs243845	MMP2	rs2033785	SMAD3
rs243840	MMP2	rs731874	SMAD3
rs7201	MMP2	rs8032802	SMAD3
rs2192853	MMP2	rs8031627	SMAD3
rs4810482	MMP9	rs3743343	SMAD3
rs3918241	MMP9	rs4939826	SMAD7
rs3918253	MMP9	rs7226855	SMAD7
rs3787268	MMP9	rs9946510	SMAD7
rs2250889	MMP9	rs6507877	SMAD7
rs2274756	MMP9	rs2878889	SMAD7
rs3774932	NFKB1	rs2337143	SMAD7
rs230540	NFKB1	rs8179181	TGFB1
rs4698858	NFKB1	rs4803455	TGFB1
rs11722146	NFKB1	rs2241715	TGFB1
rs4648110	NFKB1	rs2241713	TGFB1
rs7674640	NFKB1	rs4846476	TGFB2
rs997476	NFKB1	rs4846267	TGFB2
rs7152826	NFKBIA	rs947712	TGFB2
rs3138056	NFKBIA	rs1891467	TGFB2
rs696	NFKBIA	rs2796821	TGFB2
rs3138045	NFKBIA	rs2000220	TGFB2
rs2007960	NFKBIA	rs9308385	TGFB2
rs17103274	NFKBIA	rs900	TGFB2
rs10410544	NFKBIB	rs2857602	TNF
rs2053071	NFKBIB	rs2857706	TNF
rs11879872	NFKBIB	rs909253	TNF
rs3136640	NFKBIB	rs2229094	TNF
rs3136646	NFKBIB	rs1799964	TNF
rs7777705	PDGFA	rs1800629	TNF
rs11764261	PDGFA	rs673	TNF
rs11768030	PDGFA	rs2256965	TNF
rs7806249	PDGFA	rs2256974	TNF
rs7691129	PDGFRA	rs652284	TNFR2
rs7656613	PDGFRA	rs542282	TNFR2
rs4289498	PDGFRA	rs1061622	TNFR2
rs17739921	PDGFRA	rs590977	TNFR2
rs4864877	PDGFRA	rs3397	TNFR2
rs849380	PIK3CG	rs1061628	TNFR2
rs849384	PIK3CG		
rs757902	PIK3CG		
rs849387	PIK3CG		
rs4730204	PIK3CG		
rs4727666	PIK3CG		
rs4460309	PIK3CG		
rs6961244	PIK3CG		
rs3173908	PIK3CG		

REIVINDICACIONES

1. Método para la obtención de datos útiles para la determinación del riesgo de desarrollar Vitreo Retinopatía Proliferante (VRP) que comprende identificar en una muestra biológica aislada las variantes polimórficas del gen SMAD 7 asociadas al SNP rs7226855.

2. Método para la obtención de datos útiles para la determinación del riesgo de desarrollar Vitreo Retinopatía Proliferante (VRP) que comprende el método de obtención de datos según la reivindicación 1, donde la variante polimórfica del gen SMAD 7 identificada es el SNP rs7226855.

3. Método para pronosticar el riesgo de desarrollar Vitreo Retinopatía Proliferante (VRP) que comprende llevar a cabo el método para la obtención de datos según cualquiera de las reivindicaciones 1-2, donde la presencia de variantes polimórficas del gen SMAD 7 asociadas al SNP rs7226855, o la presencia del SNP rs7226855 se identifica como un factor de riesgo a desarrollar VRP.

FIG. 1

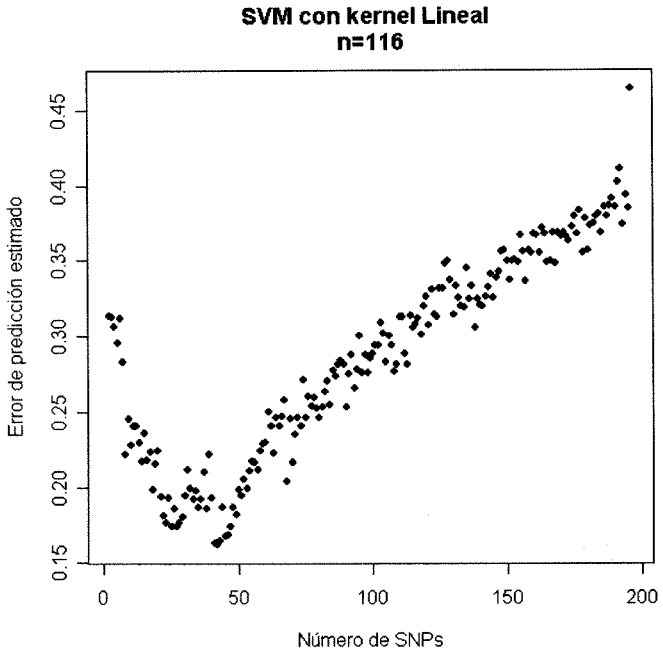

SVM con kernel Lineal
n=116
Error de predicción estimado
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0
50
100
150
200
Número de SNPs

FIG. 2

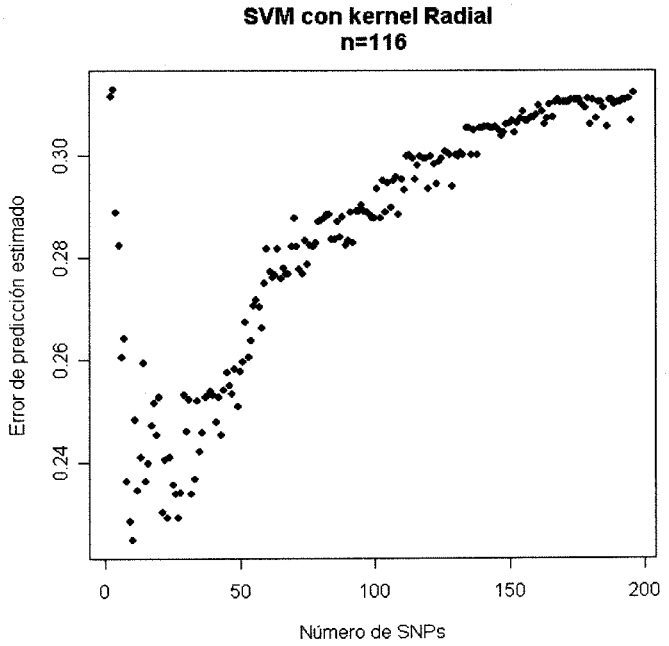
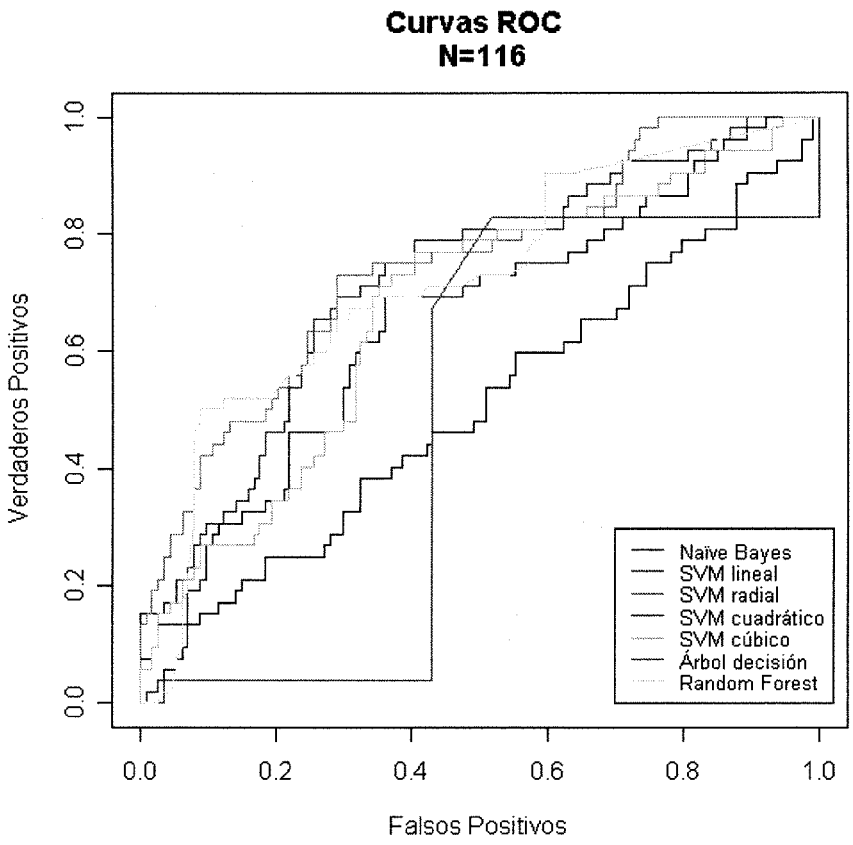

SVM con kernel Radial
n=116
Error de predicción estimado
0.24
0.26
0.28
0.30
0
50
100
150
200
Número de SNPs

FIG. 3


Curvas ROC
N=116
Verdaderos Positivos
Falsos Positivos
0.0
0.2
0.4
0.6
0.8
1.0
Naïve Bayes
SVM lineal
SVM radial
SVM cuadrático
SVM cúbico
Árbol decisión
Random Forest

OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

(11) ES 2 337 115

(21) Nº de solicitud: 200702532

(22) Fecha de presentación de la solicitud: **26.09.2007**

(32) Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TÉCNICA

(51) **Int. Cl.:** ***C12Q 1/68*** (2006.01)

DOCUMENTOS RELEVANTES

Categoría	(56) Documentos citados	Reivindicaciones afectadas
A	Base de datos NCBI [en línea] [recuperado el 22.12.2008] Recuperado de: 25.05.2006, http://www.ncbi.nlm.nih.gov/SNP/ snp_ref.cgi?rs=rs7226855	1-3
A	SAIKA S. et al. Effect of Smad7 gene overexpression on Transforming Growth Factor beta-induced Retinal Pigment Fibrosis in a Proliferative Vitreoretinopathy mouse model. Arch. Ophthalmol. Mayo 2007, Vol. 125, páginas 647-654, todo el documento.	1-3
A	SANABRIA RUIZ-COLMENARES M.R. et al. Cytokine gene polymorphisms in retinal detachment patients with and without proliferative vitreoretinopathy: a preliminary study. Acta Ophthalmol. Scand. 2006, Vol. 84, páginas 309-313, todo el documento.	1-3
A	ES 2251977 T3 (HORMOS MEDICAL CORPORATION) 16.05.2006, todo el documento.	1-3
A	ES 2186926 T3 (INTERLEUKIN GENETICS, INC.) 16.05.2003, todo el documento.	1-3

Categoría de los documentos citados

X: de particular relevancia
Y: de particular relevancia combinado con otro/s de la misma categoría
A: refleja el estado de la técnica
O: referido a divulgación no escrita
P: publicado entre la fecha de prioridad y la de presentación de la solicitud
E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

☒ para todas las reivindicaciones ☐ para las reivindicaciones nº:

Fecha de realización del informe	Examinador	Página
05.04.2010	M. Cumbreño Galindo	1/4

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

C12Q

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

INVENES, EPODOC, WPI, TXTE, XPESP, BIOSIS, MEDLINE, NPL, EMBASE, EBI

Fecha de Realización de la Opinión Escrita: 05.04.2010

Declaración

Novedad (Art. 6.1 LP 11/1986)	Reivindicaciones	1-3	**SÍ**
	Reivindicaciones		**NO**
Actividad inventiva (Art. 8.1 LP 11/1986)	Reivindicaciones	1-3	**SÍ**
	Reivindicaciones		**NO**

Se considera que la solicitud cumple con el requisito de **aplicación industrial.** Este requisito fue evaluado durante la fase de examen formal y técnico de la solicitud (Artículo 31.2 Ley 11/1986).

Base de la Opinión:

La presente opinión se ha realizado sobre la base de la solicitud de patente tal y como ha sido publicada.

1. Documentos considerados:

A continuación se relacionan los documentos pertenecientes al estado de la técnica tomados en consideración para la realización de esta opinión.

Documento	Número Publicación o Identificación	Fecha Publicación
D01	Base de datos NCBI [en línea] [recuperado el 25-03-2009] Recuperado de: http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=rs7226855	25-5-2006
D02	Saika S. Et al. Arch. Ophthalmol. Vol. 125, páginas 647-654	05-2007
D03	Sanabria Ruiz-Colmenares M.R. et al. Acta Ophthalmol. Scand. Vol. 84, páginas 309- 313	2006
D04	ES 2251977 T3	16-5-2006
D05	ES 2186926 T3	16-5-2003

2. Declaración motivada según los artículos 29.6 y 29.7 del Reglamento de ejecución de la Ley 11/1986, de 20 de marzo, de patentes sobre la novedad y la actividad inventiva; citas y explicaciones en apoyo de esta declaración

La presente invención tiene por objeto un método para la determinación del riesgo de desarrollar Vitreo Retinopatía Proliferante (VRP) que comprende identificar, en una muestra biológica aislada, las variantes polimórficas del gen SMAD 7 asociadas al SNP rs7226855 (reivindicaciones de la 1 a la 3).

D01 divulga el SNP rs7226855.

D02 determina los efectos de la transferencia del gen Smad7 en la prevención de las respuestas fibrogénicas del epitelio pigmentario de la retina, principal causa de vitreorretinopatía proliferante, en ratones, pudiendo constituir una nueva estrategia para prevenir y tratar dicha enfermedad.

D03 tiene como propósito analizar la distribución de variantes genéticas de algunas citocinas, como el factor de necrosis tumoral alfa (TNF-alfa), en pacientes con desprendimiento de retina regmatógeno con y sin vitreorretinopatía proliferante.

D04 anticipa métodos para diagnosticar la susceptibilidad de una persona diabética a tener un riesgo aumentado de desarrollo de una retinopatía diabética, determinando si dicho sujeto presenta un polimorfismo en la parte peptídica señal del preproNPY humano.

D05 anticipa un método para detectar predisposición a, y determinar el riesgo de, padecer retinopatía diabética amenazante de la visión mediante la identificación de un modelo de polimorfismo genético de un paciente en IL-1A, IL-1B (-511) e IL-1RN.

NOVEDAD Y ACTIVIDAD INVENTIVA

D02 determina los efectos de la transferencia del gen Smad7 en la prevención de las respuestas fibrogénicas del epitelio pigmentario de la retina, principal causa de vitreorretinopatía proliferante, en ratones, pudiendo constituir una nueva estrategia para prevenir y tratar dicha enfermedad.

Sin embargo, en la documentación consultada, constituida por documentos de patentes y por publicaciones científicas, no se ha encontrado la identificación de la presencia del SNP rs7226855 como método para la determinación del riesgo de desarrollar Vitreo-Retinopatía Proliferante.

Por consiguiente, las reivindicaciones de la 1 a la 3 cumplen con los requisitos de novedad y actividad inventiva.