

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 419 183**

51 Int. Cl.:

**C12Q 1/68**

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **20.08.2007 E 07808542 (0)**

97 Fecha y número de publicación de la concesión europea: **13.03.2013 EP 2195448**

54 Título: **Procedimiento para predecir el color del iris**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**19.08.2013**

73 Titular/es:

**ERASMUS UNIVERSITY MEDICAL CENTER  
ROTTERDAM (100.0%)  
Dr. Molewaterplein 50  
3015 GE Rotterdam , NL**

72 Inventor/es:

**KAYSER, MANFRED, HEINZ;  
VAN DUIJN, CORNELIA, MARJA y  
OOSTRA, BERNARD, ANNE**

74 Agente/Representante:

**CARPINTERO LÓPEZ, Mario**

**ES 2 419 183 T3**

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

## DESCRIPCIÓN

Procedimiento para predecir el color del iris

La presente invención se refiere al campo de la genética humana, más particularmente a la predicción del color del iris basándose en marcadores de polimorfismo.

5 El color del iris humano (también denominado el color de los ojos) fue uno de los primeros rasgos humanos usados para investigar la herencia mendeliana (Davenport y Davenport, 1907), pero ahora se considera un rasgo poligenético (Frudakis y col., 2003; Sturm y Frudakis, 2004). La base física de la variación del color del iris es la cantidad de pigmento de melanina y el número de melanosomas en la capa externa del iris (estroma iridiano anterior). Los iris marrones tienen más pigmento de melanina y más melanosomas que los azules, mientras que el número de melanocitos es similar en iris de  
10 ambos tipos de color (Imesch y col., 1996; Sturm y Frudakis, 2004; Wielgus y Sarna, 2005). El pigmento de melanina en los melanosomas puede producirse en dos formas: eumelanina, una forma de color negra amarronada responsable de colores del iris oscuros, y feomelanina, una forma amarilla rojiza de melanina (Sturm y Frudakis, 2004). El color del iris humano permanece casi constante pasada la infancia temprana, pero puede someterse a regulación adrenérgica pasada la infancia y también puede cambiar debido a medicación (Bito y col., 1997; Imesch y col., 1997). La mayoría de las poblaciones de seres humanos en todo el mundo tienen el color del iris marrón, mientras que los colores azules y verdes se encuentran adicionalmente en personas de descendencia europea. Esto hace que los europeos sean la población objetivo para investigar la base genética de la variación del color del iris.

Estudios genéticos tempranos de la variación del color del iris humano han elegido como diana dos regiones genómicas, una sobre el cromosoma 19 (Eiberg y Mohr, 1987) y la otra sobre el cromosoma 15 (Eiberg y Mohr, 1996). La región del cromosoma 15 también se confirmó en dos barridos del genoma basados en microsatélites independientes recientes en gemelos (Zhu y col., 2004; Posthuma y col., 2006). Esta región aloja el gen *OCA2*, que representa el homólogo humano del sitio de dilución de la conjuntivitis de ratón, codifica la proteína O y participa en albinismo oculocutáneo tipo II (Ramsay y col., 1992; Rinchik y col., 1993; Spritz y col., 1995; Brilliant, 2001). Varios polimorfismos de un solo nucleótido (SNP) dentro de *OCA2* se asociaron al color del iris en seres humanos, incluyendo dos variantes codificantes (Rebbeck y col.,  
25 2002; Frudakis y col., 2003; Duffy y col., 2004). Recientemente, el gen *OCA2* se caracterizó en más detalle. Duffy y col. (Duffy y col., 2007) estudiaron 71 SNP de *OCA2*, la mayoría de los cuales elegían como diana SNP derivados del proyecto HapMap, en aproximadamente 3.800 gemelos, sus hermanos y padres. La asociación más fuerte se observó para tres SNP del intrón 1 con valores de p de 10<sup>-54</sup> a 10<sup>-96</sup> mientras que otros SNP de *OCA2* mostraron menos asociación (<10<sup>-20</sup>). Sin embargo, la señal de alta asociación en el intrón 1 pareció extenderse en la región proximal de 5' fuera del gen  
30 *OCA2*, sugiriendo variantes desconocidas localizadas fuera del gen *OCA2* que determinan la variación del color del iris en seres humanos que no se investigaron por Duffy y col. (2007). Frudakis y col. (Frudakis y col., 2007) investigaron un total de 395 SNP de *OCA2* en dos conjuntos parcialmente solapantes de individuos e informaron de 33 SNP en asociación con el color del iris de los cuales seis se habían identificado previamente (Frudakis y col., 2003) y dos coincidían con SNP del intrón 1 identificado por Duffy y col. (Duffy y col., 2007). Otros genes que participan en el color del iris incluyen *MATP*,  
35 *ASIP*, *TYRP1*, *CYP1A2*, *CYP2C8* y *CYP2C9* (documento WO 2002/097047; Kanetsky y col., 2002; Frudakis y col., 2003; Graf y col., 2005).

Sin embargo, todavía existe la necesidad de mejores marcadores para el color del iris. Tales marcadores pueden usarse ventajosamente en investigaciones forenses y otra aplicación de identificación humana en la que las muestras de ácido nucleico, por ejemplo, de una escena del crimen se usan para obtener información sobre los sujetos que han dejado atrás  
40 este ácido nucleico en el lugar del crimen. Actualmente, la identificación humana usando marcadores de ácido nucleico es completamente comparativa y compara perfiles de marcador (huellas dactilares de ADN, perfiles de ADN) obtenidos de muestras de la escena del crimen con aquellos obtenidos de sospechosos conocidos. Si la policía no sabe de ningún sospechoso, no puede obtenerse el perfil y compararse con el recogido de la escena del crimen. Por consiguiente, en tales casos, la persona que deja la muestra en la escena del crimen y que podría haber cometido el crimen no puede  
45 identificarse usando pruebas genéticas (ADN). Similarmente, las personas desaparecidas se identifican actualmente comparando un perfil de ADN obtenido de sus restos con el obtenido de un pariente conocido. Si no se sabe nada sobre la persona desaparecida, no pueden identificarse parientes para la prueba genética y no está disponible perfil de ADN para comparación. Si se tuvieran marcadores de ácido nucleico que pudieran predecir de forma fidedigna el color de los ojos (iris), esto ayudaría a encontrar personas desconocidas (sospechosos / personas desaparecidas) de una forma directa y  
50 sin comparar perfiles de ADN.

## RESUMEN DE LA INVENCION

Los inventores han encontrado ahora varios marcadores en un área hasta ahora no investigada, la región genómica 5' proximal (fuera) del gen *OCA2* y que incluye el gen *HERC2* vecino directo. Esta región está localizada sobre el cromosoma 15 entre los pares de bases 26018062 y 26240890 (según la versión 36 de NCBI o la versión 46.36h de *Homo sapiens* de Ensembl). Por consiguiente, la invención se refiere a un procedimiento para predecir el color de los ojos de un ser humano, por ejemplo, una persona desconocida a partir de una muestra de ácido nucleico / proteína ensayando

uno o más polimorfismos en el gen *HERC2* en el que el uno o más polimorfismos se seleccionan del grupo constituido por rs916977, rs8028689, rs6497287, rs8041209, rs6497292, rs2240202, rs2346050, rs12592730, rs7183877, rs2240204, rs8039195, rs16950979, rs16950987, rs1667394 y rs1635168 y cualquier marcador en el gen *HERC2* en desequilibrio de ligamiento genético con dichos polimorfismos.

5 Preferentemente, el polimorfismo es rs916977, siendo la predicción ojos azules si el polimorfismo tiene la secuencia de nucleótidos CC (o GG considerando las bases complementarias), u ojos marrones si la secuencia de nucleótidos es TT (o AA considerando las bases complementarias). Para los otros marcadores y sus alelos, la predicción es la siguiente, proporcionando ambos alelos de ambos cromosomas homólogos y las bases complementarias indicadas entre paréntesis:  
 10 rs8028689: CC (GG) marrón y TT (AA) azul, rs6497287: AA (TT) azul y GG (CC) marrón, rs8041209: AA (TT) marrón y CC (GG) azul, rs6497292: AA (TT) azul y GG (CC) marrón, rs2240202: AA (TT) marrón y GG (CC) azul, rs2346050: AA (TT) azul y GG (CC) marrón, rs12592730: AA (TT) marrón y GG (CC) azul, rs7183877: TT (AA) azul y GG (CC) marrón, rs2240204: TT (AA) marrón y CC (GG) azul, rs8039195: TT (AA) azul y CC (GG) marrón, rs16950979: AA (TT) azul y CC (GG) marrón, rs16950987: TT (AA) marrón y CC (GG) azul, rs1667394: TT (AA) azul y CC (GG) marrón y rs1635168: TT (AA) marrón y GG (CC) azul.

15 **LEYENDAS DE LAS FIGURAS**

**Figura 1.** Asociación del genoma completo para el color del iris humano en dos muestras de poblaciones independientes de Los Países Bajos: A) el estudio de ERF (Familia de Erasmus Rucphen) (N=192) usando Affymetrix 250K SNPs, y B) el estudio de Rotterdam (N=481) usando Affymetrix 500K SNPs teniendo en cuenta múltiples pruebas. Se nombran SNP asociados al color del iris al nivel de significancia estadística del genoma completo de <0,001, marcadores correspondientes entre ambos estudios están marcados en negrita.  
 20

**Figura 2.** Asociación de SNP al color del iris de la región 25-26,8 Mb del cromosoma 15 de dos estudios de GWA independientes de Los Países Bajos: A) el estudio de ERF (N=192), y B) el estudio de Rotterdam (N=481). Las estadísticas de la prueba de la chi al cuadrado (eje Y) se representan frente a la distancia del mapa físico, alineada con SNP conocidos y genes en la región. Los genes *OCA2* y *HERC2* están marcados con recuadros verdes y azules respectivamente. La puntuación de  $-\log_{10}$  valores de p para cada asociación de SNP con el color de los ojos azul/no azul se muestra como puntos azules sin conectar, y las asociaciones de haplotipos basadas en ventana deslizante de dos y tres SNP se representan como líneas rosas y naranjas continuas respectivamente. Los  $-\log_{10}$  valores de p están truncados en 20. Se nombran SNP que alcanzaron el nivel de significancia del genoma completo de 0,05.  
 25

30 Marcadores correspondientes entre ambos estudios están marcados en negrita.

**Figura 3.** Resultados de análisis del ligamiento del genoma completo para color del iris humano. Las puntuaciones de LOD se muestran por cromosoma.

**Figura 4.** Oportunidades relativas (OR) para el color del iris marrón para cuatro SNP: rs11855019, rs6497268, rs7495174 en *OCA2* y rs916977 en el gen *HERC2* en las poblaciones del estudio de ERF (N=2217) y de Rotterdam (N=6056). O se derivaron usando regresión logística (color del iris marrón y no marrón binario contra el número de alelos secundarios de cada SNP) en la que los cuatro SNP se ajustaron en el mismo modelo.  
 35

**Figura 5.** Exactitud discriminatoria y probabilidades de predicción del color del iris marrón usando genotipos de cuatro SNP: rs11855019, rs6497268, rs7495174 en el *OCA2* y rs916977 en el gen *HERC2* en la población del estudio de ERF (N=2217) y de Rotterdam (N=6056): A) Exactitud discriminatoria para la predicción del color del iris marrón basándose en los cuatro SNP en diferentes conjuntos de muestras, B) Comparación de las probabilidades de predicción para el color del iris marrón de los cuatro SNP con los basados en rs916977 solo.  
 40

**Figura 6.** Distribución de la frecuencia de alelos de rs916977 en toda Europa: A) Frecuencias de los alelos rs916977 para 23 poblaciones europeas (incluyendo el estudio de Rotterdam) superpuestas sobre un mapa de Europa que indica las clases de color del iris humano. Los tamaños de los gráficos de sectores indican el tamaño de muestra. B) Análisis de autocorrelación espacial de rs916977 en Europa considerando 23 muestras de población. La forma del autocorrelograma describe el patrón geográfico de las frecuencias alélicas rs916977. Valores de I de Moran positivos son distancias cortas y negativos para distancias geográficas grandes indican un patrón clinal de los datos genéticos de un lado del mapa al opuesto. La Figura 6A (sin datos sobre rs916977) es de Beals y Hoijer "An Introducción To Anthropology", 3ª edición, publicado por Almenten y Bacon, Boston, MA. Copyright © 1965 por Pearson Education.  
 45  
 50

**DESCRIPCIÓN DETALLADA DE LA INVENCION**

Hasta la fecha, todas las búsquedas del genoma completo para genes que participan en el color del iris humano se han basado en estudios del ligamiento que eligen como diana genes relativamente raros con grandes efectos y que revelan

grandes regiones genómicas, y estudios de genes candidatos que eligen como diana genes que, basándose en su función, se espera que participen en la pigmentación del color del iris. Se ha demostrado que la asociación del genoma completo (GWA) es un enfoque poderoso para estudiar variantes comunes que participan en los rasgos. Hasta la fecha no se ha realizado ninguna GWA para el color del iris en seres humanos. Por tanto, los presentes inventores aplicaron tanto una GWA como un enfoque del ligamiento a la variación del color del iris humano en dos estudios independientes. Una GWA se realizó en una población humana genéticamente aislada reciente de Los Países Bajos, el estudio de la familia de Erasmus Rucphen (ERF), y la otra en una muestra de poblaciones no consanguíneas de Los Países Bajos, el estudio de Rotterdam.

El objetivo del estudio de los presentes inventores era desenmarañar los principales genes que contribuyen a la variación del color del iris humano en europeos junto con la evaluación de su valor de predecir los fenotipos del color del iris de personas desconocidas, ya que esto puede ser relevante en futuras aplicaciones forenses u otras aplicaciones de identificación humana. Las actuales formas de identificación humana basadas en ADN siguen un enfoque estrictamente comparativo de perfiles de ADN no codificantes (o no) coincidentes obtenidos de la escena del crimen (o una persona desaparecida) con los de un sospechoso conocido (o parientes conocidos de la persona desaparecida). Si los sospechosos (o parientes) son desconocidos, un caso forense no puede resolverse con la ayuda de las pruebas de ADN actualmente usadas. La predicción de rasgos visibles tales como el color del iris de una persona desconocida usando marcadores genéticos informativos, si se hace de forma fidedigna, ayudaría a rastrear personas desconocidas.

De los estudios de los presentes inventores resultó que la región 5' proximal del gen *OCA2* hasta e incluyendo el gen *HERC2* no es la región más importante que participa en el color del iris de dos estudios genéticos de poblaciones independientes. Esta región está localizada sobre el cromosoma 15 entre los pares de bases 26018062 y 26240890 según la versión 36 de NCBI o la versión 46.36h de *Homo sapiens* de Ensembl. En esta región se encontraron varios SNP que discriminan el color del iris entre ojos azules, ojos marrones y ojos de color intermedio. De estos SNP, rs916977 en el intrón 11 del gen *HERC2* es el SNP más fuerte asociado a la variación del color del iris humano. Éste es el único marcador que alcanza la significancia del genoma completo en ambos estudios de GWA; no hubo otras regiones que mostraran significancia del genoma completo coherente en ambos estudios de GWA; no hubo otras regiones que mostraran pruebas del genoma completo coherentes para la asociación del color del iris. El marcador del intrón 11 de *HERC2* está localizado 169 kb proximal de 5' del gen *OCA2*, que se informó previamente que se asociaba al color del iris humano (Rebbeck y col., 2002; Duffy y col., 2004; Jannot y col., 2005; Duffy y col., 2007; Frudakis y col., 2007). Si se analiza *HERC2* y los tres SNP de *OCA2* más importantes informados en un estudio previo (rs11855019, rs6497268 y rs7495174), rs916977 del gen *HERC2* mostró el menor valor de p en ambas poblaciones ampliadas. Además, de estos tres SNP de *OCA2*, solo dos (rs11855019 y rs7495174) permanecieron significativamente asociados al color del iris cuando rs916977 de *HERC2* se incluyó en el modelo. En los modelos de predicción de los presentes inventores, rs916977 de *HERC2* fue el único de los dos SNP más importantes en predecir el color del iris. Los SNP en genes sugeridos anteriormente con participación en el color del iris no contribuyeron (*MATP*, *ASIP*, *CYP1A2*, *CYP2C8* y *CYP2C9*) o contribuyeron marginalmente (*TYRP1*). Los SNP de *OCA2* identificados en la GWA del estudio de Rotterdam también contribuyeron solo marginalmente a la predicción del color del iris, cuando su valor predictivo se evaluó junto con los otros marcadores. En general, la asociación genética al color del iris humano fue mayor para marcadores localizados en el gen *HERC2* que lo que fue para marcadores de cualquier otro gen, que incluye *OCA2*. Pareció que el color del iris azul se asoció a la presencia de los SNP CC (bases complementarias GG) en rs916977, mientras que el color del iris marrón se asoció a la presencia de los nucleótidos TT (bases complementarias AA). Además, los presentes inventores encontraron siete SNP del gen *HERC2* con asociación significativa al color del iris al nivel del genoma completo en la población del estudio de Rotterdam con la siguiente asociación de genotipo - color del iris (considerando tanto cromosomas homólogos como las bases complementarias entre paréntesis): rs8028689: CC (GG) marrón y TT (AA) azul, rs6497287: AA (TT) azul y GG (CC) marrón, rs8041209: AA (TT) marrón y CC (GG) azul, rs6497292: AA (TT) azul y GG (CC) marrón, rs2240202: AA (TT) marrón y GG (CC) azul, rs2346050: AA (TT) azul y GG (CC) marrón, rs12592730: AA (TT) marrón y GG (CC) azul. Además, los presentes inventores encontraron otros siete SNP del gen *HERC2* cuando aplicaron las matrices Illumina 300 Duo SNP a los 733 participantes de ERF y uno solapándose (rs8028689) con los resultados de la matriz Affymetrix 500K en el estudio de Rotterdam con la siguiente asociación del color del iris al genotipo (considerando tanto cromosomas homólogos como las bases complementarias entre paréntesis): rs7183877: TT (AA) azul y GG (CC) marrón, rs2240204: TT (AA) marrón y CC (GG) azul, rs8039195: TT (AA) azul y CC (GG) marrón, rs16950979: AA (TT) azul y CC (GG) marrón, rs16950987: TT (AA) marrón y CC (GG) azul, rs1667394: TT (AA) azul y CC (GG) marrón, y rs1635168: TT (AA) marrón y GG (CC) azul.

“Polimorfismos” son variantes alélicas que se producen en una población. El polimorfismo puede ser una diferencia de un solo nucleótido presente en un sitio, o puede ser una inserción o delección de uno o algunos nucleótidos. Como tal, un polimorfismo de un solo nucleótido (“SNP”) se caracteriza por la presencia en una población de uno o dos, o cuatro nucleótidos (es decir, adenosina, citosina, guanosina o timidina) en un sitio particular en un genoma tal como el genoma humano. Por consiguiente, se reconoció que, mientras que los procedimientos de la invención se ejemplifican principalmente por la detección de SNP, los procedimientos desvelados u otros conocidos en la técnica pueden usarse similarmente para identificar otros polimorfismos en la región genética ejemplificada del *HERC2* en desequilibrio de

ligamiento genético con dichos SNP.

Los haplotipos pueden deducirse de datos del genotipo correspondientes a ciertos SNP usando el algoritmo de Stephens y Donnelly (Am. J. Hum. Genet. 68: 978-989,2001). Las fases del haplotipo (es decir, los alelos del haplotipo particular en un individuo) también pueden determinarse usando el algoritmo de Stephens y Donnelly. Están disponibles programas de software que realizan este algoritmo (por ejemplo, The PHASE program, Departamento de estadística, Universidad de Oxford).

En un ejemplo, llamado el procedimiento Haploscope (véase el documento WO 02/048318), una combinación de SNP candidatos se selecciona de una pluralidad de combinaciones de SNP candidatos para un gen asociado a un rasgo genético. Los datos de haplotipos asociados a esta combinación de SNP candidatos son leídos para una pluralidad de individuos y se agrupan en un grupo de respuesta positiva y un grupo de respuesta negativa basándose en si se cumplen los criterios de rasgos predeterminados para un individuo. Se realiza un análisis estadístico (como se trata más adelante) sobre los datos de haplotipos agrupados para obtener una medición estadística asociada a la combinación de SNP candidatos. Los actos de selección, lectura, agrupación y realización se repiten según sea necesario para identificar la combinación de SNP candidatos que tienen la medición estadística óptima. En un enfoque, todas las posibles combinaciones de SNP se seleccionan y se analizan estadísticamente. En otro enfoque, una búsqueda directa basada en los resultados de análisis estadísticos previos de combinaciones de SNP se realiza hasta que se obtiene la medición estadística óptima. Además, el número de combinaciones de SNP seleccionadas y analizadas puede reducirse basándose en un procedimiento de prueba simultáneo.

Como se usa en el presente documento, el término “deducir” o “deduciendo”, cuando se usa en referencia a una predicción del color del iris, significa sacar una conclusión sobre el color de los ojos de un sujeto usando un procedimiento de analizar individualmente o en combinación la(s) existencia(s) de nucleótidos de uno o más SNP de la invención en una muestra de ácido nucleico del sujeto y comparar la(s) existencia(s) de nucleótidos individual(es) o la combinación de existencia(s) de nucleótidos de los SNP(s) con relaciones conocidas de existencia(s) de nucleótidos del color del iris. Como se ha desvelado en el presente documento, la(s) existencia(s) de nucleótidos puede(n) identificarse directamente examinando moléculas de ácidos nucleicos, o indirectamente examinando un polipéptido codificado por un gen particular, por ejemplo, el gen *HERC2*, cuando el polimorfismo está asociado a un cambio de aminoácidos en el polipéptido codificado o tiene efectos reguladores.

Procedimientos de realización de una comparación tal y de alcanzar una conclusión basándose en esa comparación se ejemplifican en el presente documento. La deducción normalmente implica usar un modelo complejo que implica usar relaciones conocidas de alelos conocidos o existencias de nucleótidos como clasificadores. La comparación puede realizarse aplicando los datos referentes al (a los) alelo(s) del haplotipo del sujeto a un modelo complejo que hace una clasificación por discriminación cuadrática ciega usando una matriz varianza-covarianza. Un modelo tal se ilustra en los ejemplos.

Para determinar si los haplotipos son útiles en una deducción del color de los ojos pueden realizarse numerosos análisis estadísticos. Las frecuencias de alelos pueden calcularse para haplotipos y las frecuencias de haplotipos por emparejamiento estimarse usando un algoritmo EM (Excoffier y Slatkin, Mol Biol Evol. 1995 Sep; 12 (5): 921-7). Entonces pueden calcularse los coeficientes de desequilibrio de ligamiento. Además de diversos parámetros tales como los coeficientes de desequilibrio de ligamiento, frecuencias de alelos y haplotipos (dentro de grupos étnicos, de control y de casos), puede calcularse la estadística de la chi al cuadrado y otros parámetros genéticos de la población tales como índices panmíticos para controlar la variación étnica, ancestral u otra sistemática entre los grupos de casos y de control. Los marcadores/haplotipos con valor para diferenciar la matriz de casos del control, si los hay, pueden presentarse en forma matemática que describe cualquier relación y acompañarse de estadística de asociación (prueba y efecto). Un resultado de análisis estadístico que muestra una asociación de un marcador de SNP o un haplotipo al color de los ojos con al menos una probabilidad de insignificancia inferior a 0,05 (o 5%) puede usarse para identificar marcadores / haplotipos más informativos. Estas herramientas estadísticas pueden probar la significancia relacionada con una hipótesis nula de que un alelo de SNP o alelo del haplotipo no sea diferente entre los grupos. Si la significancia de esta diferencia es baja, sugiere que el alelo no está relacionado con el color de los ojos. Si la significancia es alta (valor de p al menos <0,05, preferible <0,01), sugiere que el alelo / haplotipo está relacionado con el color de los ojos.

El término “cebador” como se usa en el presente documento se refiere a un oligonucleótido que puede hibridarse con la diana de amplificación permitiendo que una ADN polimerasa se una, sirviendo así de punto de iniciación de la síntesis de ADN cuando se dispone en condiciones en las que se induce la síntesis del producto de extensión del cebador que es complementario a una hebra de ácido nucleico, es decir, en presencia de nucleótidos y un agente para polimerización tal como ADN polimerasa y a una temperatura adecuada y pH. El cebador (de amplificación) es preferentemente monocatenario para máxima eficiencia en la amplificación. Preferentemente, el cebador es un oligodesoxirribonucleótido. El cebador debe ser suficientemente largo para cebar la síntesis de productos de extensión en presencia del agente para la polimerización. Las longitudes exactas de los cebadores dependerán de muchos factores, que incluyen temperatura y fuente del cebador. Un “par de cebadores bidireccionales” como se usa en el presente documento se refiere a un cebador

directo y uno inverso como comúnmente se usan en la materia de la amplificación de ADN tal como en amplificación por PCR.

El término “sonda” se refiere a una secuencia de oligonucleótidos monocatenaria que reconocerá y formará un dúplex unido por hidrógeno con una secuencia complementaria en un analito de secuencia de ácido nucleico diana o su derivado de ADNc.

5 Los términos “rigurosidad” o “condiciones de hibridación rigurosas” se refieren a condiciones de hibridación que afectan la estabilidad de los híbridos, por ejemplo, temperatura, concentración de sales, pH, concentración de formamida y similares. Estas condiciones se optimizan empíricamente para maximizar la unión específica y minimizar la unión no específica del cebador o sonda con su secuencia de ácidos nucleicos diana. Los términos como se usan incluyen referencia a  
10 condiciones bajo las que una sonda o cebador se hibridará con su secuencia diana, a un grado detectablemente mayor que otras secuencias (por ejemplo, al menos 2 veces con respecto a la referencia). Las condiciones rigurosas son dependientes de la secuencia y serán diferentes en diferentes circunstancias. Secuencias más largas se hibridan específicamente a mayores temperaturas. Generalmente, las condiciones rigurosas están seleccionadas para ser aproximadamente 5 °C inferior al punto de fusión térmico ( $T_m$ ) para la secuencia específica a una fuerza iónica y pH  
15 definidos. La  $T_m$  es la temperatura (bajo fuerza iónica y pH definidos) a la que el 50% de una secuencia diana complementaria se hibrida con una sonda o cebador perfectamente apareado. Normalmente, condiciones rigurosas serán aquellas en las que la concentración de sales es inferior a aproximadamente  $1,0 \text{ M}$   $\text{Na}^+$ , normalmente aproximadamente concentración de  $0,01$  a  $1,0 \text{ M}$  (u otras sales) a pH 7,0 a 8,3 y la temperatura es al menos aproximadamente 30 °C para sondas o cebadores cortos (por ejemplo, 10 a 50 nucleótidos) y al menos aproximadamente  
20 60 °C para sondas o cebadores largos (por ejemplo, superior a 50 nucleótidos). También pueden lograrse condiciones rigurosas con la adición de agentes desestabilizantes tales como formamida. Condiciones de baja rigurosidad o “condiciones de rigurosidad reducida” a modo de ejemplo incluyen hibridación con una disolución de tampón de 30% de formamida, NaCl 1 M, 1% de SDS a 37 °C y un lavado en 2x SSC a 40 °C. Condiciones de alta rigurosidad a modo de ejemplo incluyen hibridación en 50% de formamida, NaCl 1 M, 1% de SDS a 37 °C, y un lavado en 0,1x SSC a 60 °C. Los  
25 procedimientos de hibridación son muy conocidos en la técnica y se describen, por ejemplo, en Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. eds. (1998) *Current protocols in molecular biology*. V.B. Chanda, series ed. New York: John Wiley & Sons.

Los procedimientos para detectar un cambio de nucleótidos pueden utilizar una o más sondas o cebadores de oligonucleótidos que incluyen, por ejemplo, un par de cebadores de amplificación que se hibridan selectivamente con un  
30 polinucleótido diana, que contiene una o más posiciones de SNP. Sondas de oligonucleótidos útiles en poner en práctica un procedimiento de la invención pueden incluir, por ejemplo, un oligonucleótido que es complementario a y extiende una parte del polinucleótido diana, que incluye la posición del SNP, en el que la presencia de un nucleótido específico en la posición (es decir, el SNP) se detecta por la presencia o ausencia de hibridación selectiva de la sonda. Un procedimiento tal puede incluir adicionalmente poner en contacto el polinucleótido diana y el oligonucleótido hibridado con una  
35 endonucleasa, y detectar la presencia o ausencia de un producto de escisión de la sonda, dependiendo de si la aparición del nucleótido en el sitio de SNP es complementario al nucleótido correspondiente de la sonda. Un par de sondas que se hibridan específicamente en la dirección 5' y adyacente y en la dirección 3' y adyacente al sitio del SNP, en la que una de las sondas incluye un nucleótido complementario a una existencia de nucleótidos del SNP, también puede usarse en un ensayo de ligación de oligonucleótidos en el que la presencia o ausencia de un producto de ligación es indicativo de la  
40 manifestación de nucleótidos en el sitio de SNP. Un oligonucleótido también puede ser útil como cebador, por ejemplo, para una reacción de extensión de cebadores, en la que el producto (o ausencia de un producto) de la reacción de extensión es indicativa de la existencia del nucleótido. Además, puede ser útil un par de cebadores útil para amplificar una parte del polinucleótido diana que incluye el sitio de SNP, examinándose el producto de amplificación para determinar la existencia del nucleótido en el sitio de SNP.

45 Si la existencia de nucleótidos particulares de un SNP, o existencias de nucleótidos de un haplotipo, es tal que la existencia de nucleótidos produzca un cambio de aminoácidos en un polipéptido codificado, la existencia de nucleótidos puede identificarse indirectamente detectando el aminoácido particular en el polipéptido. El procedimiento para determinar el aminoácido dependerá, por ejemplo, de la estructura del polipéptido o de la posición del aminoácido en el polipéptido. Si el polipéptido contiene solo una única existencia de un aminoácido codificado por el SNP particular, el polipéptido puede  
50 examinarse para la presencia o ausencia del aminoácido. Por ejemplo, si el aminoácido está en o próximo al extremo amino o el extremo carboxi del polipéptido, puede realizarse la simple secuenciación de los aminoácidos terminales. Alternativamente, el polipéptido puede tratarse con una o más enzimas y puede examinarse un fragmento de péptido que contiene la posición de aminoácidos de interés, por ejemplo, por secuenciación del péptido, o detectando una migración particular del péptido tras la electroforesis. Si el aminoácido particular comprende un epítipo del polipéptido, puede  
55 detectarse la unión específica, o ausencia de la misma, de un anticuerpo específico para el epítipo. Otros procedimientos para detectar un aminoácido particular en un polipéptido o fragmento de péptido del mismo son muy conocidos y pueden seleccionarse basándose, por ejemplo, en la conveniencia o disponibilidad de equipo tal como un espectrómetro de masas, sistema de electroforesis capilar, equipo de resonancia magnética nuclear y similares.

Un procedimiento para deducir el rasgo del color del iris de un sujeto humano a partir de una muestra de ácido nucleico del sujeto humano que identifica una existencia de nucleótidos en la muestra para un SNP según la invención puede incluir adicionalmente agrupar las existencias de nucleótidos de los SNP para un gen o un área particular del genoma, tal como la región del gen *HERC2*, en un alelo del marcador o más alelos de haplotipo. Los alelos de marcador / de haplotipo individuales identificados pueden entonces compararse con alelos de marcador / de haplotipo conocidos de forma que, si se conoce la relación de los alelos de marcador / de haplotipo conocidos con respecto al color de los ojos, puede hacerse una deducción en cuanto al color del iris del sujeto que proporciona la muestra de ácido nucleico. La identificación de la existencia de nucleótidos puede realizarse usando cualquier procedimiento adecuado para examinar la muestra particular. Por ejemplo, si la muestra contiene moléculas de ácidos nucleicos, la identificación puede realizarse poniendo en contacto polinucleótidos en (o derivados de) la muestra con un miembro del par de unión específico que se hibrida selectivamente con una región del polinucleótido que incluye el SNP o los SNP, en condiciones en las que el miembro del par de unión se une específicamente en o próximo al (a los) SNP. El miembro del par de unión puede ser cualquier molécula que se una específicamente a o se asocie con el polinucleótido diana, que incluye, por ejemplo, un anticuerpo o un oligonucleótido.

En principio, los procedimientos de la invención pueden realizarse usando cualquier procedimiento de amplificación de ácidos nucleicos tal como la reacción en cadena de la polimerasa (PCR; Mullis 1987, patentes de EE.UU. n° 4.683.195, 4.683.202 y 4.800.159) o usando reacciones de amplificación tales como reacción en cadena de la ligasa (LCR; Barany 1991, Proc. Natl. Acad. Sci. USA 88:189-193; solicitud EP n° 320.308), replicación de secuencia autosostenida (3SR; Guatelli y col., 1990, Proc. Natl. Acad. Sci. USA 87:1874-1878), amplificación por desplazamiento de cadenas (SDA; patentes de EE.UU. n° 5.270.184 y 5.455.166), sistema de amplificación transcripcional (TAS; Kwoh y col., Proc. Natl. Acad. Sci. USA 86:1173-1177), Q-beta replicasa (Lizardi y col., 1988, Bio/Technology 6:1197), amplificación por círculo rodante (RCA; patente de EE.UU. n° 5.871.921), amplificación basada en secuencias de ácidos nucleicos (NASBA), polimorfismo de longitud de fragmentos de escisión (patente de EE.UU. n° 5.719.028), amplificación isotérmica y quimérica de ácidos nucleicos iniciada por cebadores (ICAN), procedimiento de amplificación por extensión de ramificaciones (RAM; patentes de EE.UU. n° 5.719.028 y 5.942.391) u otros procedimientos adecuados para la amplificación de ADN.

Con el fin de amplificar ADN con un número pequeño de desapareamientos con uno o más de los cebadores de amplificación, puede realizarse una reacción de amplificación en condiciones de rigurosidad reducida (por ejemplo, una amplificación por PCR usando una temperatura de hibridación de 38 °C, o la presencia de MgCl<sub>2</sub> 3,5 mM). El experto en la materia podrá seleccionar condiciones de rigurosidad adecuada.

Los cebadores en el presente documento están seleccionados para ser “sustancialmente” complementarios (es decir, al menos el 65%, más preferentemente al menos el 80% perfectamente complementarios) a sus regiones diana presentes sobre las diferentes cadenas de cada secuencia específica que va a amplificarse. Es posible usar secuencias de cebadores que contienen, por ejemplo, residuos de inositol o bases ambiguas, o incluso cebadores que contienen uno o más desapareamientos cuando se comparan con la secuencia diana. En general, las secuencias que presentan al menos el 65%, más preferentemente al menos el 80%, de homología con las secuencias de oligonucleótidos de ADN diana se consideran adecuadas para su uso en un procedimiento de la presente invención. Los desapareamientos de secuencias tampoco son críticos si se usan condiciones de hibridación de baja rigurosidad.

La detección de los productos de amplificación puede llevarse a cabo en principio por cualquier procedimiento adecuado conocido en la técnica. Los fragmentos de detección pueden teñirse directamente o marcarse con marcas radiactivas, anticuerpos, colorantes luminiscentes, colorantes fluorescentes o reactivos de enzima. Las tinciones de ADN directas incluyen, por ejemplo, intercalar colorantes tales como naranja de acridina, bromuro de etidio, monoazida de etidio o colorantes de Hoechst.

Alternativamente, los fragmentos de ADN pueden detectarse por la incorporación de bases de dNTP marcadas en los fragmentos de ADN sintetizados. Marcas de detección que pueden asociarse a bases de nucleótido incluyen, por ejemplo, fluoresceína, colorante de cianina o BrdUrd.

Si se usa un sistema de detección basado en sondas, un procedimiento de detección adecuado para su uso en la presente invención puede comprender, por ejemplo, un formato de inmunoensayo enzimático (EIA) (Jacobs y col., 1997, J. Clin. Microbiol. 35, 791-795). Para realizar una detección a manera de un procedimiento de EIA, tanto el cebador directo como el inverso usado en la reacción de amplificación puede comprender un grupo de captura, tal como un grupo de biotina para la inmovilización de amplicones de PCR de ADN diana sobre, por ejemplo, pocillos de una placa de microtitulación recubierta con estreptavidina para la posterior detección por EIA de amplicones de ADN diana (véase más adelante). El experto entenderá que pueden emplearse otros grupos para la inmovilización de amplicones de PCR de ADN diana en un formato de EIA.

Sondas útiles para la detección del ADN diana como se desvelan en el presente documento se unen preferentemente solo a al menos una parte de la región de la secuencia de ADN como se amplifica por el procedimiento de amplificación de ADN. Aquellos expertos en la materia pueden preparar sondas adecuadas para la detección basándose en la secuencia

de nucleótidos del ADN diana sin excesiva experimentación como se explica en el presente documento. También pueden usarse adecuadamente secuencias complementarias del ADN diana como sondas de detección en un procedimiento de la invención, a condición de que una hebra complementaria tal se amplifique en la reacción de amplificación empleada.

5 Procedimientos de detección adecuados para su uso en el presente documento pueden comprender, por ejemplo, inmovilización de los amplicones y sondaje de las secuencias de ADN de los mismos por, por ejemplo, transferencia Southern. Otros formatos pueden comprender un formato de EIA como se ha descrito anteriormente. Para facilitar la detección de la unión, las sondas de detección de amplicones específicos pueden comprender un resto de marca tal como un fluoróforo, un cromóforo, una enzima o una radiomarca, de manera que se facilite el control de la unión de las sondas al producto de reacción de la reacción de amplificación. Tales marcas son muy conocidas para aquellos expertos en la materia e incluyen, por ejemplo, isotiocianato de fluoresceína (FITC), β-galactosidasa, peroxidasa de rábano picante, estreptavidina, biotina, digoxigenina, <sup>35</sup>S o <sup>125</sup>I. Otros ejemplos serán evidentes para aquellos expertos en la materia.

10 La detección también puede realizarse por un llamado ensayo de transferencia de línea inversa (RLB) tal como se ha descrito, por ejemplo, por Van den Brule y col. (2002, J. Clin. Microbiol. 40, 779-787). Sondas de RLB para este fin se sintetizan preferentemente con un grupo amino de 5' para la posterior inmovilización sobre, por ejemplo, membranas de nailon recubiertas con carboxilo. La ventaja de un formato de RLB es la facilidad del sistema y su velocidad, permitiéndose así procesamiento de muestras de alta resolución.

15 Cualquier procedimiento adecuado para cribar los ácidos nucleicos para la presencia o ausencia de polimorfismos se considera que es parte de la presente invención. Tales procedimientos incluyen, pero no se limitan a: secuenciación de ADN, análisis de polimorfismos de longitud de fragmentos de restricción (RFLP), análisis de polimorfismos de longitud de fragmentos amplificados (AFLP); análisis de heterodúplex, análisis de polimorfismos de conformaciones monocatenarias (SSCP), electroforesis en gel en gradiente desnaturizante (DGGE), análisis por PCR en tiempo real (por ejemplo, Taqman®), electroforesis en gel en gradiente de temperatura (TGGE), extensión de cebadores, hibridación específica para alelos y ensayos de análisis genéticos INVADER®, análisis de polimorfismos de longitud de fragmentos de escisión (CFLP), análisis de regiones amplificadas caracterizadas por secuencias (SCAR), análisis de secuencia polimórficas amplificados escindidas (CAPS)

20 El uso de sondas de ácido nucleico para la detección de secuencias de ADN específicas es muy conocido en la técnica. Generalmente, estos procedimientos comprenden la hibridación del ADN diana con la sonda, seguido de lavados post-hibridación. La especificidad es normalmente función de los lavados post-hibridación, siendo los factores críticos la fuerza iónica y la temperatura de la disolución de lavado final. Para híbridos de ADN-ADN, la T<sub>m</sub> puede aproximarse a partir de la ecuación de Meinkoth y Wahl, Anal. Biochem., 138: 267-284 (1984):  $T_m = 81,5 \text{ }^\circ\text{C} + 16,6 (\log M) + 0,41 (\% \text{ de GC}) - 0,61 (\% \text{ de forma}) - 500/L$ ; en la que M es la molaridad de cationes monovalentes, % de GC es el porcentaje de nucleótidos guanosina y citosina en el ADN, % de forma es el porcentaje de formamida en la disolución de hibridación y L es la longitud del híbrido en pares de bases. La T<sub>m</sub> es la temperatura (bajo fuerza iónica y pH definidos) a la que el 50% de una secuencia diana complementaria se hibrida con una sonda perfectamente coincidente. La T<sub>m</sub> se reduce aproximadamente 1 °C por cada 1% de desapareamiento; así, las condiciones de hibridación y/o lavado pueden ajustarse para hibridarse con secuencias de la identidad deseada. Por ejemplo, si se buscan secuencias con > 90% de identidad, la T<sub>m</sub> puede disminuirse 10 °C. Generalmente, las condiciones rigurosas se seleccionan para ser aproximadamente 5 °C inferior al punto de fusión térmico (T<sub>m</sub>) para la secuencia específica y su complemento a una fuerza iónica y pH definidos. Sin embargo, condiciones duramente rigurosas pueden utilizar una hibridación y/o lavado a 1, 2, 3 ó 4 °C inferior al punto de fusión térmico (T<sub>m</sub>); condiciones moderadamente rigurosas pueden utilizar una hibridación y/o lavado a 6, 7, 8, 9 ó 10 °C inferior al punto de fusión térmico (T<sub>m</sub>); condiciones de baja rigurosidad pueden utilizar una hibridación y/o lavado a 11,12,13,14, 15 ó 20 °C inferior al punto de fusión térmico (T<sub>m</sub>). Usando la ecuación, composiciones de hibridación y lavado, y T<sub>m</sub> deseada, aquellos expertos habituales entenderán que las variaciones en la rigurosidad de hibridación y/o disoluciones de lavado están inherentemente descritas. Si el grado deseado de desapareamiento produce una T<sub>m</sub> inferior a 45 °C (disolución acuosa) o 32 °C (disolución de formamida), se prefiere aumentar la concentración de SSC de manera que pueda usarse una mayor temperatura. Una extensa guía para la hibridación de ácidos nucleicos se encuentra en Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes, Parte I, Capítulo 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier. Nueva York (1993); y Current Protocols in Molecular Biology, Capítulo 2, Ausubel, y col., Eds., arriba.

25 El desarrollo de cebadores y sondas útiles para la detección de posiciones polimórficas en un ácido nucleico está dentro del campo de los expertos habituales (véase, por ejemplo, Sambrook, J., Russell D.W., Sambrook, J. (2001) Molecular Cloning: a Laboratory Manual. Cold Spring Harbor Laboratory Press, Plainview, N.Y.).

30 Usando tecnología de ADN convencional es posible producir sondas y cebadores que se hibriden directamente o indirectamente con las muestras de ADN que van a probarse o ADNc producido a partir de ARN por transcripción inversa, y que pueden usarse en ensayos para la detección de los SNP. Las técnicas de amplificación de ácidos nucleicos permiten la amplificación de fragmentos de ácidos nucleicos, que pueden estar presentes en cantidades muy bajas.

Con el fin de desarrollar procedimientos de detección basados en ácidos nucleicos deben determinarse las secuencias específicas para SNP para las que entonces pueden desarrollarse cebadores o sondas. Para detectar los SNP por amplificación de ácidos nucleicos y/o hibridación de sondas, el ácido nucleico puede aislarse de cualquier material de muestra en bruto, opcionalmente transcribirse en forma inversa en ADNc y clonarse y/o secuenciarse directamente. Los kits de aislamiento de ADN y ARN están comercialmente disponibles de, por ejemplo, QIAGEN GmbH, Hilden, Alemania, o Roche Diagnostics, una división de F. Hoffmann-La Roche Ltd, Basilea, Suiza.

Una muestra útil para poner en práctica un procedimiento de la invención puede ser cualquier muestra biológica de un sujeto que contiene moléculas de ácidos nucleicos, que incluyen porciones de las secuencias de genes que van a examinarse, o polipéptidos codificados correspondientes, dependiendo del procedimiento particular. Como tal, la muestra puede ser una muestra de célula, tejido u órgano, o puede ser una muestra de un fluido biológico tal como semen, saliva, sangre y similares. Una muestra de ácido nucleico útil para poner en práctica un procedimiento de la invención dependerá, en parte, de si los SNP del haplotipo que va a identificarse están en regiones codificantes o en regiones no codificantes. Así, si al menos uno de los SNP que va a identificarse está en una región no codificante, la muestra de ácido nucleico es generalmente una muestra de ácido desoxirribonucleico (ADN), particularmente ADN genómico o un producto de amplificación del mismo. Sin embargo, si está disponible ácido ribonucleico (ARN) heteronuclear, que incluye moléculas de ARN precursoras de ARNm sin cortar ni empalmar, puede usarse un ADNc o producto de amplificación del mismo. Si cada uno de los SNP del haplotipo está presente en una región codificante del (de los) gen(es) de pigmentación, la muestra de ácido nucleico puede ser ADN o ARN, o productos derivados de los mismos, por ejemplo, productos de amplificación. Además, aunque los procedimientos de la invención generalmente se ejemplifican con respecto a una muestra de ácido nucleico, se reconocerá que alelos de haplotipos particulares pueden estar en regiones codificantes de un gen y pueden producir polipéptidos que contienen diferentes aminoácidos en las posiciones correspondientes a los SNP debido a cambios de codones no degenerados. Como tal, en otro aspecto, los procedimientos de la invención pueden ponerse en práctica usando una muestra que contiene polipéptidos del sujeto.

Los procedimientos de la invención pueden ponerse en práctica con respecto a sujetos humanos y, por tanto, pueden ser particularmente útiles para análisis forense. En una aplicación forense o un procedimiento de la invención, la muestra de ácido nucleico humana puede obtenerse de una escena del crimen, usando procedimientos de muestreo bien establecidos. Así, la muestra puede ser muestra fluida o una muestra de hisopo, por ejemplo, mancha de sangre, mancha de semen, folículo piloso u otro espécimen biológico tomado de una escena del crimen, o puede ser una muestra de tierra que se sospecha que contiene material biológico de una posible víctima del crimen o perpetrador, puede ser material recuperado de debajo de las uñas de los dedos de una posible víctima del crimen, o similares, en el que los ácidos nucleicos (o polipéptidos) en la muestra pueden usarse como base para sacar una deducción en cuanto al color del iris según un procedimiento de la invención. Otra aplicación de la invención está en identificar personas desaparecidas analizando los marcadores identificados en el presente documento a partir de ácidos nucleicos / proteínas en partes del cuerpo o muestras del cuerpo completo de la persona desconocida que va a identificarse.

Usando tanto el ácido nucleico clonado como una sonda de hibridación, usando información de secuencias derivada del clon, o diseñando cebadores degenerativos basados en la secuencia del SNP y sus secuencias flanqueantes, las sondas de hibridación de ácidos nucleicos y/o cebadores de amplificación de ácidos nucleicos pueden diseñarse y usarse en un ensayo de detección para detectar las características del uno o más SNP en una muestra como se define en el presente documento.

El ADN, o alternativamente, el ADNc puede amplificarse por PCR usando, por ejemplo, Pfu y Taq ADN polimerasas y cebadores de amplificación específicos para las secuencias de ADN de SNP. También pueden usarse sistemas comercialmente disponibles completos para PCR (por ejemplo, disponibles de diversos proveedores tales como Roche Diagnostics). Un procedimiento adecuado puede incluir, por ejemplo, mezclar en un sistema de tamponamiento acuoso adecuado (por ejemplo, un tampón de PCR comercialmente disponible) una cantidad adecuada de ADN total como molde (por ejemplo, 1 a 5 µg), una cantidad adecuada de dNTP y la ADN polimerasa, desnaturalizar los ácidos nucleicos hirviéndolos durante 1 min y realizar una reacción de ciclos de aproximadamente 10-50 ciclos alternos de hibridación de cebadores rigurosa, alargamiento de cadenas y desnaturalización a temperatura adecuada para obtener copias de ADN del molde de ADN como producto de amplificación. La cantidad de copias producidas tras un cierto número de ciclos se correlaciona directamente con la cantidad de ADN diana en el molde de ADN.

El experto conoce muy bien los procedimientos de PCR cuantitativa disponibles presentemente disponibles de proveedores comerciales para cuantificar la cantidad de ADN diana en el molde. El término "señal de hibridación" como se usa en el presente documento se refiere, entre otros, a la cantidad de producto de amplificación producido tras un cierto número de ciclos y así a la cantidad de ADN diana disponible como molde en la reacción.

Con el fin de amplificar un ácido nucleico con un pequeño número de desapareamientos con uno o más de los cebadores de amplificación, una reacción de amplificación puede realizarse en condiciones de rigurosidad reducida (por ejemplo, una amplificación por PCR usando una temperatura de hibridación de 38 °C, o la presencia de MgCl<sub>2</sub> 3,5 mM). El experto en la

materia podrá seleccionar condiciones de rigurosidad adecuada.

Los cebadores en el presente documento se seleccionan por ser “sustancialmente” complementarios (es decir, al menos el 65%, más preferentemente al menos el 80% perfectamente complementarios) a sus regiones diana presentes sobre las diferentes cadenas de cada secuencia específica que va a amplificarse. Es posible usar secuencias de cebadores que contienen, por ejemplo, residuos de inositol o bases ambiguas, o incluso cebadores que contienen uno o más desapareamientos cuando se comparan con la secuencia diana. En general, las secuencias que presentan al menos el 65%, más preferentemente al menos el 80%, de homología con las secuencias de oligonucleótidos de ADN o ARN diana se consideran adecuadas para su uso en un procedimiento de la presente invención. Los desapareamientos de secuencias tampoco son críticos si se usan condiciones de hibridación de baja rigurosidad.

La detección de los productos de amplificación puede llevarse a cabo en principio por cualquier procedimiento adecuado conocido en la técnica. Los fragmentos amplificados pueden teñirse directamente o marcarse con marcas radiactivas, anticuerpos, colorantes luminiscentes, colorantes fluorescentes o reactivos de enzima. Las tinciones de ADN directas incluyen, por ejemplo, intercalar colorantes tales como naranja de acridina, bromuro de etidio, monoazida de etidio o colorantes de Hoechst.

Alternativamente, los fragmentos de ADN o ARN pueden detectarse por la incorporación de bases de dNTP marcadas en los fragmentos sintetizados. Marcas de detección que pueden asociarse a bases de nucleótido incluyen, por ejemplo, fluoresceína, colorante de cianina, digoxigenina (DIG) o bromodesoxiuridina (BrdUrd).

En un procedimiento de PCR cuantitativo, la reacción se realiza preferentemente usando un cebador de oligonucleótidos que contiene uno o más monómeros de ácidos nucleicos 'bloqueados' (LNA®), o usando sondas fluorescentes de LNA®. La tecnología de LNA® implica un oligonucleótido (sonda o cebador que contiene uno o más monómeros de LNA® [modificado con 2'-O,4'-C-metilen-B-D-ribofuranosilo] (por ejemplo, Petersen y Wengel, 2003. TRENDS in Biotechnology, vol. 21(2):74-81). En un monómero de LNA, el resto de azúcar ribosa del nucleótido se modifica, mientras que la propia base sigue sin alterar. El resultado es un puente covalente que 'bloquea' la ribosa en la conformación de tipo N (3'-endo), que potencia el apilamiento de bases y la pre-organización del esqueleto de fosfato. Esto provee al oligonucleótido de afinidad mejorada por secuencias de ADN o ARN complementarias y, por tanto, una mayor T<sub>m</sub>. Si se usan cebadores de LNA®, la detección de los productos de amplificación bicatenarios puede realizarse, por ejemplo, usando una tinción de ADN bicatenaria tal como SYBR Green® [Molecular Probes, Inc.] (véase, por ejemplo, Ponchel y col. 2003, BMC Biotechnology 3:18).

Otros procedimientos de analizar el ácido nucleico comprenden adecuadamente el uso de un ensayo de extensión de cebadores; una PCR Taqman®; un ensayo de hibridación diferencial; un ensayo que detecta la escisión de secuencias específicas para alelo; y/o PCR específica para alelo.

Si se usa un sistema de detección basado en sondas, un procedimiento de detección adecuado para su uso en la presente invención puede comprender, por ejemplo, un formato de inmunoensayo enzimático (EIA) (Jacobs y col., 1997, J. Clin. Microbiol. 35, 791-795). Para realizar una detección a manera de un procedimiento de EIA, tanto el cebador directo como el inverso usado en la reacción de amplificación puede comprender un grupo de captura, tal como un grupo de biotina para la inmovilización de amplicones de PCR de ADN diana sobre, por ejemplo, pocillos de una placa de microtitulación recubierta con estreptavidina o Dynabeads® recubiertas con estreptavidina (DynaL Biotech, Oslo, Noruega) para la posterior detección por EIA de amplicones de ADN diana. El experto entenderá que pueden emplearse otros grupos para la inmovilización de amplicones de PCR de ADN diana en un formato de EIA.

Sondas útiles para la detección de las secuencias de ácidos nucleicos diana como se desvela en el presente documento se unen preferentemente solo a al menos una parte de la región de la secuencia de ácidos nucleicos como se amplifica por el procedimiento de amplificación de ácidos nucleicos. Aquellos expertos en la materia pueden preparar sondas adecuadas para la detección basándose en la secuencia de nucleótidos del ácido nucleico diana sin excesiva experimentación como se explica en el presente documento. También pueden usarse adecuadamente secuencias de nucleótidos complementarias, tanto si es ADN o ARN como análogos químicamente sintetizados, del ácido nucleico diana como sondas de detección específicas para tipo en un procedimiento de la invención, a condición de que una hebra complementaria tal se amplifique en la reacción de amplificación empleada.

Procedimientos de detección adecuados para su uso en el presente documento pueden comprender, por ejemplo, inmovilización de los amplicones y sondaje de las secuencias de ácidos nucleicos de los mismos por, por ejemplo, transferencia Northern y Southern. Otros formatos pueden comprender un formato de EIA como se ha descrito anteriormente. Para facilitar la detección de la unión, las sondas de detección de amplicones específicos pueden comprender un resto de marca tal como un fluoróforo, un cromóforo, una enzima o una radiomarca, de manera que se facilite la monitorización de la unión de las sondas al producto de reacción de la reacción de amplificación. Tales marcas son muy conocidas para aquellos expertos en la materia e incluyen, por ejemplo, isotiocianato de fluoresceína (FITC), β-galactosidasa, peroxidasa de rábano picante, estreptavidina, biotina, digoxigenina, <sup>35</sup>S, <sup>14</sup>C, <sup>32</sup>P o <sup>125</sup>I. Otros ejemplos serán evidentes para aquellos expertos en la materia.

La detección también puede realizarse por un llamado ensayo de transferencia de línea inversa (RLB) tal como se ha descrito, por ejemplo, por Van den Brule y col. (2002). Sondas de RLB para este fin se sintetizan preferentemente con un grupo amino de 5' para la posterior inmovilización sobre, por ejemplo, membranas de nailon recubiertas con carboxilo. La ventaja de un formato de RLB es la facilidad del sistema y su velocidad, permitiéndose así procesamiento de muestras de alta resolución.

El uso de sondas de ácido nucleico para la detección de ARN o fragmentos de ADN es muy conocido en la técnica. Generalmente, estos procedimientos comprenden la hibridación del ácido nucleico diana con la sonda, seguido de lavados post-hibridación. La especificidad es normalmente función de lavados post-hibridación, siendo los factores críticos la fuerza iónica y la temperatura de la disolución de lavado final. Para híbridos de ácidos nucleicos, la  $T_m$  puede aproximarse a partir de la ecuación de Meinkoth y Wahl (1984):  $T_m = 81,5 \text{ °C} + 16,6 (\log M) + 0,41 (\% \text{ de GC}) - 0,61 (\% \text{ de forma}) - 500/L$ ; en la que M es la molaridad de cationes monovalentes, % de GC es el porcentaje de nucleótidos guanosina y citosina en el ácido nucleico, % de forma es el porcentaje de formamida en la disolución de hibridación y L es la longitud del híbrido en pares de bases. La  $T_m$  es la temperatura (bajo fuerza iónica y pH definidos) a la que el 50% de una secuencia diana complementaria se hibrida con una sonda perfectamente coincidente. La  $T_m$  se reduce aproximadamente 1 °C por cada 1% de desapareamiento; así, las condiciones de hibridación y/o lavado pueden ajustarse para hibridarse con secuencias de la identidad deseada. Por ejemplo, si se buscan secuencias con > 90% de identidad, la  $T_m$  puede disminuirse 10 °C. Generalmente, las condiciones rigurosas se seleccionan para ser aproximadamente 5 °C inferior al punto de fusión térmico ( $T_m$ ) para la secuencia específica y su complemento a una fuerza iónica y pH definidos. Sin embargo, condiciones duramente rigurosas pueden utilizar una hibridación y/o lavado a 1, 2, 3 ó 4 °C inferior al punto de fusión térmico ( $T_m$ ); condiciones moderadamente rigurosas pueden utilizar una hibridación y/o lavado a 6, 7, 8, 9 ó 10 °C inferior al punto de fusión térmico ( $T_m$ ); condiciones de baja rigurosidad pueden utilizar una hibridación y/o lavado a 11, 12, 13, 14, 15 ó 20 °C inferior al punto de fusión térmico ( $T_m$ ). Usando la ecuación, composiciones de hibridación y lavado, y  $T_m$  deseada, aquellos expertos habituales entenderán que las variaciones en la rigurosidad de hibridación y/o disoluciones de lavado están inherentemente descritas. Si el grado deseado de desapareamiento produce una  $T_m$  inferior a 45 °C (disolución acuosa) o 32 °C (disolución de formamida), se prefiere aumentar la concentración de SSC de manera que pueda usarse una mayor temperatura. Una extensa guía para la hibridación de ácidos nucleicos se encuentra en Tijssen, 1993, arriba; Ausubel y col., 1998, arriba.

En otro aspecto, la invención proporciona sondas de oligonucleótidos para la detección del SNP. Las sondas de detección en el presente documento se seleccionan por ser "sustancialmente" complementarias a una molécula de ácido nucleico monocatenario, o a una de las hebras de los ácidos nucleicos bicatenarios generadas por una reacción de amplificación de la invención. Preferentemente, las sondas son sustancialmente complementarias a las hebras no codificantes opcionalmente inmovilizadas (por ejemplo, marcadas con biotina) de los amplicones generados a partir de ARN o ADN diana.

Es admisible que las sondas de detección de la presente invención contengan uno o más desapareamientos con respecto a su secuencia diana. En general, las secuencias que presentan al menos el 65%, más preferentemente al menos el 80%, de homología con las secuencias de oligonucleótidos diana se consideran adecuadas para su uso en un procedimiento de la presente invención.

La invención se ilustrará ahora mediante los siguientes ejemplos no limitantes.

## EJEMPLOS

### 40 MATERIAL Y PROCEDIMIENTOS

#### Poblaciones de los estudios

##### *La población del estudio de ERF*

Para el estudio de ERF se invitó a todos los descendientes y cónyuges vivos de 22 parejas que vivieron en la región Rucphen del sudeste de Los Países Bajos en el siglo XIX. Estas parejas tuvieron un mínimo de seis hijos y sus relaciones genealógicas son conocidas hasta mediados del siglo XVIII cuando la población fue fundada por aproximadamente 150 individuos. La población incluye ahora aproximadamente 20.000 residentes que muestran un elevado desequilibrio de ligamiento en comparación con poblaciones no consanguíneas (Pardo y col., 2005; Service y col., 2006).

El Comité de ética médica del Centro médico Erasmus autorizó el protocolo del estudio, y todos los participantes proporcionaron consentimiento informado por escrito. La información sobre el color del iris se recogió para todos los participantes y se extrajo ADN genómico de sangre venosa periférica utilizando el procedimiento de precipitación por sales (Miller y col. 1988). Para el cribado de GWA inicial, los presentes inventores seleccionaron 192 parientes lejanos ( $\geq 5$  generaciones) de la población de ERF (Tabla 1). Los individuos se seleccionaron inicialmente para un estudio de altura, pero manteniendo la suposición de que los genes que participan en la altura humana y el color del iris se transmiten independientemente, los presentes inventores usaron los datos de GWA para el presente proyecto. Los datos de GWA y

el color del iris estuvieron disponibles para 192 participantes y se realizaron estudios de replicación usando 2217 participantes de ERF para los que estaban disponibles los datos sobre el color del iris (Tabla 1). Para el análisis del ligamiento del genoma completo, los presentes inventores usaron 1292 individuos de ERF que tenían datos del color del iris y se genotiparon para el panel de ligamiento de Illumina 6K. Además, los presentes inventores usaron 733 participantes de ERF que tuvieron datos del color del iris y se genotiparon para el panel 300 Duo SNP.

#### *La población del estudio de Rotterdam*

El estudio de Rotterdam es un estudio de cohorte prospectivo basado en población en 7.983 sujetos de 55 años de edad y superior que residían en Ommoord, un suburbio de Rotterdam, que tuvo como objeto evaluar la aparición y determinantes de enfermedades crónicas en la etapa de madurez (Hofman y col. 1991). Ésta es una población no consanguínea, prácticamente de origen europeo holandés. En 1990-1993, los participantes fueron invitados a visitar el centro de investigación para un examen clínico. El Comité de ética médica del Centro médico Erasmus autorizó el protocolo del estudio, y todos los participantes proporcionaron consentimiento informado por escrito. La información sobre el color del iris se recogió para todos los participantes y se extrajo ADN genómico de sangre venosa periférica utilizando el procedimiento de precipitación por sales (Miller y col. 1988). Los presentes inventores seleccionaron 509 mujeres sin parentesco de edades entre 60 y 75 años para un cribado de GWA independiente (Tabla 1). Se excluyeron las mujeres que usaban medicación para hipertensión, diabetes o fármacos hipolipemiantes y terapia hormonal sustitutiva durante el seguimiento, además de cualquier mujer restante con historia de infarto de miocardio nuevo/ya diagnosticado, accidente cerebrovascular, cáncer o fracturas de cadera. Aunque estas mujeres se seleccionaron para un estudio de GWA piloto para osteoporosis y otros trastornos, los presentes inventores asumieron que el color del iris se estaba segregando independientemente de estos rasgos. Para 481 mujeres estuvieron disponibles GWA y datos sobre el color del iris (Tabla 1) y para estudios de replicación estuvo disponible ADN para 6056 participantes con fenotipos del color del iris conocidos (Tabla 1).

#### **Recogida de fenotipos**

En ambos estudios, cada ojo se examinó por examen con lámpara de hendidura y el color del iris se clasificó usando imágenes patrón que mostraban diversos grados de pigmentación del iris. Se distinguieron tres categorías del color del iris (azul, intermedio y marrón) basándose en el color predominante y la cantidad de pigmento amarillo o marrón presente en el iris. Los fenotipos del color del iris se resumieron en la Tabla 1. Las diferencias entre las frecuencias de color del iris azul y marrón entre el estudio de ERF y el estudio de Rotterdam pueden explicarse lo más probablemente por la influencia de la ocupación española de la región en el siglo XVI y XVII, que duró durante casi cien años en las partes de más al sur de Los Países Bajos.

#### **Recogida de genotipos**

##### *Genotipado de micromatrices*

Para el estudio de Rotterdam se utilizó el conjunto de matrices GeneChip® Human Mapping 500K (Affymetrix). Debido al marcado desequilibrio de ligamiento observado previamente en ERF (Aulchenko y col., 2004; Service y col., 2006), los presentes inventores aplicaron la matriz 250K NSP del conjunto de matrices GeneChip® Human Mapping 500K (Affymetrix) en el estudio de ERF. El genotipado basado en micromatrices se realizó en el CM Erasmus según las instrucciones del fabricante. Los marcadores se excluyeron si se desviaron significativamente del equilibrio de Hardy-Weinberg ( $P < 0,001$ ), si tuvieron baja frecuencia de alelos secundarios ( $MAF < 0,025$ ) o si tuvieron una tasa de retirada  $< 95\%$  en todas las muestras. Adicionalmente, los presentes inventores excluyeron 10 mujeres del estudio piloto de Rotterdam que no se agruparon con la muestra de otro modo homogénea que contenía el 98% de los participantes (prueba de concordancia de la población por emparejamiento  $< 0,0001$ ) (Purcell y col. 2007). Para el análisis del ligamiento en la cohorte de ERF, los presentes inventores usaron el ensayo de ligamiento Illumina Infinio. Este panel incluye 5.861 marcadores distribuidos uniformemente en todo el genoma humano (mediana de la distancia entre el marcador 301 kb). Los presentes inventores usaron adicionalmente para GWA el producto Illumina 300 Duo que contenía ~317000 SNP que eligen diana en todo el genoma humano. El genotipado Illumina se realizó en el Centro Nacional de Genotipado en Francia según la instrucción del fabricante.

##### *Genotipado TaqMan*

Para los estudios de conformación, cuatro SNP se genotiparon en las poblaciones del estudio de ERF total ( $N=2217$ ) y de Rotterdam total ( $N=6056$ ) usando tecnología TaqMan con rs11855019, rs7495174 y rs6497268 tipados usando ensayos Custom Taqman (Applied Biosystems) y rs916977 usando el ensayo de genotipado Taqman C\_2567831\_10 (Applied Biosystems). Las secuencias de cebador y sonda de los tres primeros SNP están disponibles bajo petición de los autores, mientras que el ensayo para rs916977 está comercialmente disponible del fabricante. Se dispensaron 1-2 ng de ADN genómico en placas de 384 pocillos usando un robot de pipeteado Caliper Sciclone ALH3000 (Caliper LS). Todos los ensayos se ejecutaron en un volumen total de 2  $\mu$ l usando 2-5 ng de ADN genómico, 0,025-0,05  $\mu$ l de 40x mezcla de ensayo y 1  $\mu$ l de ABSOLUTE QPCR Mix (ABgene) o Taqman Universal PCR Master Mix (Applied Biosystems). Los

reactivos se dispensaron en una placa de 384 pocillos usando Deereac Equator NS808 (Deereac Fluidics). Los programas de PCR fueron 95 °C 15 min, 95 °C 15 s y 60 °C 1 min durante 40 ciclos (ABSOLUTE QPCR Mix) o 94 °C 10 min, 94 °C 15 s, 60 °C 1 min durante 40 ciclos (Taqman Universal PCR Master Mix) en Dual 384-well GeneAmp PCR System 9700 (Applied Biosystems) con posterior lectura del punto final en un sistema de PCR en tiempo real ABI 7900HT (Applied Biosystems).

## **Análisis estadísticos**

### *Análisis de asociación y ligamiento*

Para probar las asociaciones usando grandes cantidades de SNP, los presentes inventores usaron funciones rápidas implementadas en R library GenABEL versión 1.1-8 (Aulchenko y col. 2007). En el análisis de GWA, los valores de p se estimaron usando la prueba de tendencias de Armitage en tablas de contingencia de 2 por 3 (color del iris azul y no azul contra tres genotipos posibles) con 1 grado de libertad. En los análisis de replicación de SNP de interés y basándose en las poblaciones ampliadas, las oportunidades relativas (OR) se derivaron usando regresiones logísticas múltiples usando un resultado binario (color marrón y no marrón). Para ajustar para múltiples pruebas con un gran número de marcadores correlacionados, la significancia también se estimó frente a la distribución empírica de la estadística de la chi al cuadrado después de 1000 permutaciones del genoma completo. Los presentes inventores usaron el procedimiento de control genómico (Devlin y Roeder 1999) para ajustar la relación entre participantes de ERF (Steinthorsdottir y col., 2007). Basándose en el ensayo de ligamiento Illumina Infinio, el factor de inflación se estimó que era 1,12 para el estudio de ERF. No se observó inflación de la estadística de pruebas en el estudio de Rotterdam ( $\lambda=1,00$ ), usando la matriz Affymetrix GeneChip® Human Mapping 500K. Las regiones que flanqueaban SNP significativos en cualquier cribado de GWA se investigaron adicionalmente usando análisis de haplotipos. Se usaron dos procedimientos: (1) ventana deslizante de 2 y 3 SNP continuos usando R library haplo.stats versión 1.3 (Schaid y col. 2002) y (2) análisis de bloques de haplotipos usando el paquete de software Haploview versión 3.32, en el que los bloques se definieron usando umbrales de confianza del 95% de D' (Gabriel y col., 2002). Los cálculos del poder estadístico usando una tasa de error tipo I del genoma completo del 5% muestran que la probabilidad es del 80% para identificar como parte de GWA un SNP que explica el 18% de la varianza fenotípica para el estudio de ERF para detectar un SNP, mientras que para el estudio de Rotterdam el poder estadístico fue del 80% para detectar un SNP que explicaba el 8% de la varianza como parte del GWA. Para el estudio de ERF, usando una tasa de error tipo I del genoma completo del 5%, la probabilidad es del 80% para detectar ligamiento con una variante que explica el 2-3% de la varianza fenotípica. Para el análisis del ligamiento en ERF, los presentes inventores usaron los modelos de componentes de la varianza como se han implementado en el paquete informático SOLAR (rutinas de análisis del ligamiento oligogénico secuenciales) (Almasy y Blangero 1998). Antes del análisis, la genealogía de ERF se fraccionó en linajes de 18 bits usando el software PedSTR. La identidad por descendencia (IBD) genómica se estimó usando el paquete de software MERLIN (Abecasis y col. 2002). El corte de linajes complejos puede conducir a ligamientos positivos falsos debido a que el verdadero parentesco está subestimado. Por tanto, en lugar del parentesco del linaje estimado a partir de los linajes de 18 bits fraccionados, los presentes inventores estimaron parentesco nulo como un promedio de IBD de marcadores a través del genoma.

### *Análisis de valor predictivo*

Para investigar el valor de los SNP identificados para la predicción del color del iris, los presentes inventores construyeron un modelo de predicción en la muestra piloto de GWA de Rotterdam y en una muestra al 50% al azar del estudio de Rotterdam (conjunto de datos de derivación) usando análisis de regresión logística.

Se construyeron modelos separados para color del iris marrón y azul (sí/no). El modelo fue validado en el otro 50% del estudio de Rotterdam (validación interna) y en el estudio de ERF (validación externa). Para cada individuo, los presentes inventores calcularon las probabilidades de iris marrones y azules. El valor predictivo se evaluó por el área bajo la curva de las características de operación del receptor (ABC), que es una medida de la exactitud discriminatoria que indica el grado con el que las probabilidades predichas pueden discriminar entre individuos con iris marrones (o azules) y aquellos sin ellos. ABC oscila de 0,5 que representa ausencia total de discriminación a 1,0 que representa discriminación perfecta (Hanley y McNeil 1982).

### *Autocorrelación espacial y análisis de correlación de Pearson*

Las frecuencias de alelos de rs916977 se obtuvieron de datos de SNP del genoma completo de 23 poblaciones europeas (de las cuales el estudio de Rotterdam es una) como parte de un proyecto diferente. El patrón espacial de las frecuencias alélicas del SNP rs916977 se analizó por medio de autocorrelación espacial (Sokal y Oden, 1978) usando el programa PASSAGE. Este procedimiento representa la cantidad de autocorrelación (expresada como índice I de Moran) entre pares de poblaciones contra su distancia geográfica. Para el análisis de correlación, la distribución de frecuencias de color del iris europeo se obtuvo de un mapa publicado en cualquier parte (Beals y Hoijer 1965) y los valores medios de las clases de frecuencia del color del iris se asignaron a cada una de las 23 poblaciones estudiadas basándose en su origen geográfico. Se realizó una correlación de Pearson entre este valor deducido de los fenotipos del color del iris y la frecuencia del alelo C de rs916977 en cada población.

## RESULTADOS

### *Hallazgos de los estudios de asociación y ligamiento*

En el análisis de GWA (250K SNP) de los 192 individuos sin parentesco del estudio de ERF, rs916977 localizado a 26,2 Mb del cromosoma 15 alcanzó significancia del genoma completo (valor de p nominal =  $1,18 \times 10^{-9}$ , valor de p del genoma completo empírico ajustado para múltiples pruebas  $< 0,001$ , Figuras 1A, 2A). Este marcador se localizó en el intrón 11 del gen *HERC2*.

Siete SNP adicionales que flanquean rs916977 alcanzaron significancia nominal de  $< 0,01$  (rs16950821 y rs8024968 en el gen *OCA2*; rs6497287, rs8041209, rs8028689, rs2346050 en el gen *HERC2*; rs4966231 5' proximal de *HERC2*; véase la Figura 2A para el SNP individual y análisis de haplotipos de ventana deslizante). Esta región cubrió la parte proximal de 5' del gen *OCA2* y el gen *HERC2* completo hasta 5' proximal de *HERC2*. En los datos de Illumina 300 Duo adicionalmente obtenidos de 733 participantes de ERF, los presentes inventores encontraron 18 SNP con asociación del color del iris significativa al nivel del genoma completo ( $P < 0,01$ ) de los que 8 están en el gen *HERC2* (rs7183877, rs2240204, rs8039195, rs16950979, rs16950987, rs1667394, rs1635168, y rs8028689) y 10 están en el gen *OCA2* (rs2594935, rs728405, rs3794604, rs4778232, rs8024968, rs1597196, rs7179994, rs4778138, rs4778241 y rs7495174) (Tabla 5). Por tanto, el análisis del ligamiento del genoma completo en 1457 participantes del estudio de ERF usando el panel de ligamiento 6K Illumina dio pruebas de evidencia muy convincentes para ligamiento del color del iris con el cromosoma 15q13.1. La mayor puntuación de LOD obtenida en todo el genoma fue 29,47 en rs4778137 del intrón 1 del gen *OCA2*; éste fue el único pico en el análisis del ligamiento que alcanzó una puntuación de LOD superior a 3 (véase la Figura 3).

También el segundo GWA independiente (500 K SNP) en 481 mujeres del estudio de Rotterdam identificó la misma región 15q13.1 como la más fuertemente asociada al color del iris. De nuevo, rs916977 en el gen *HERC2* mostró la asociación más fuerte con el color del iris (valor de p nominal =  $2,08 \times 10^{-40}$ , valor de p empírico del genoma completo  $< 0,001$ ). Además, siete otros SNP de *HERC2* que flanqueaban rs916977 en la región de 26,1-26,2 Mb (rs6497287, rs8041209, rs8028689, rs2346050, rs6497292, rs2240202, rs12592730 de los que los tres últimos no se incluyeron en el cribado de 250K en el estudio de ERF) alcanzaron significancia del genoma completo (valor de p  $< 0,001$ , valores de p nominales de  $10^{-13}$  a  $10^{-19}$ , Figura 1b y 2b). Además, cuatro SNP intrónicos (rs3930739, rs1448485, rs16950821, rs8024968) de *OCA2* mostraron valores de p nominales  $< 10^{-5}$  en el estudio de Rotterdam (Figura 2B). Ningún otro marcador fuera del gen *HERC2* reveló asociación significativa del genoma completo al color del iris en el estudio de Rotterdam usando el nivel de  $P < 0,001$ . Una región sobre el cromosoma 9 mostró significancia del genoma completo límite al nivel del 0,5%, pero no se confirmó el estudio de asociación de ERF o del ligamiento.

Para investigar adicionalmente la región de 15q13.1, los presentes inventores examinaron la estructura de bloques de haplotipos y el patrón de desequilibrio de ligamiento en el estudio de ERF y de Rotterdam (Gabriel y col. 2002). El patrón del desequilibrio de ligamiento fue similar en el estudio de ERF y de Rotterdam. Se identificaron tres bloques en ambas poblaciones en las que el 3º bloque contenía el SNP de *HERC2* rs916977 que mostraba la asociación más fuerte en ambas poblaciones (estudio de ERF: valor de p para el bloque 3 =  $5,94 \text{ ó } 3 \times 10^{-6}$ , estudio de Rotterdam: valor de P = 8,93 ó  $8,39 \times 10^{-20}$  (Tabla 2). El 2º y 3º bloque no estuvieron en fuerte desequilibrio de ligamiento en ambas poblaciones (estudio de ERF:  $r^2 = 0,40$ ; estudio de Rotterdam:  $r^2 = 0,39$ ). El 2º bloque está físicamente más próximo al intrón 1 del gen *OCA2*, la región que se informó previamente que era la región más importante para el color del iris (Duffy y col., 2007). Este bloque también mostró algunas pruebas de asociación al color del iris en el estudio de Rotterdam, pero en particular a un nivel de significancia mucho menor en comparación con el 3º bloque (valor de p nominal para el bloque =  $7,61 \times 10^{-3}$ ). El segundo bloque fue solo marginalmente significativo en el estudio de ERF (valor de p nominal del bloque 2 = 0,01) (Tabla 2). El bloque 1 también mostró algunas pruebas de asociación al color del iris en el estudio de Rotterdam, pero esto no está confirmado en el estudio de ERF. Ni el bloque 1 ni el bloque 2 alcanzaron significancia del genoma completo cuando se ajustaron para múltiples pruebas. Un haplotipo común de 5 SNP del 3º bloque en el estudio de Rotterdam, CTCGA, se encontró más frecuentemente en personas con el color del iris azul (97%) que en aquellas con el color del iris marrón (58,7%, color del iris intermedio: 76,5%) (Tabla 2). Se encuentra que el mismo haplotipo se asocia al color del iris en ERF (faltando rs4932620 que no estaba presente en la matriz de 250k) con una frecuencia del 96,8% en aquellos con el color del iris azul en comparación con el 65,7% en marrones (intermedios: 91,9%). Se encuentran dos haplotipos adicionales en ERF y ERGO y están asociados al color del iris no azul (Tabla 2).

### *Tipado de SNP de la región de OCA2-HERC2 en los estudios de ERF y de Rotterdam totales*

Para confirmar el hallazgo de GWA, los presentes inventores tiparon rs916977 en *HERC2* en los 2217 participantes del estudio de ERF y en los 6056 participantes del estudio de Rotterdam disponibles (Tabla 1). Además, los presentes inventores genotiparon los tres SNP en el intrón 1 del gen *OCA2* (rs11855019, rs6497268 y rs7495174) que informaron recientemente que están fuertemente asociados al color del iris (Duffy y col., 2007), pero no fueron parte de las matrices de SNP de Affymetrix que usaron los presentes inventores. El SNP rs916977 está localizado aproximadamente 168 kb proximal de 5' de los tres SNP del intrón 1 del gen *OCA2*. La última región también mostró la mayor puntuación de LOD en el análisis del ligamiento de los presentes inventores del estudio de ERF. Cuando se analizaron por separado e ignorando

el desequilibrio de ligamiento, los cuatro SNP se asociaron fuertemente al color del iris en ambas poblaciones ampliadas ( $p < 5,0 \times 10^{-76}$ ; véase la Tabla 3). Sin embargo, el SNP rs916977 del gen *HERC2* mostró la mayor significancia estadística ( $1,9 \times 10^{-113}$  para ERF y  $< 1,0 \times 10^{-300}$  para el estudio de Rotterdam). Si se incluyen los cuatro SNP juntos en el modelo de regresión, rs916977 (ERF:  $2,02 \times 10^{-33}$ ; estudio de Rotterdam:  $p < 5,84 \times 10^{-113}$ ), rs11855019 (ERF:  $2,63 \times 10^{-9}$ ; estudio de Rotterdam:  $6,75 \times 10^{-6}$ ) y rs7495174 (ERF:  $9,71 \times 10^{-9}$ ; estudio de Rotterdam:  $5,09 \times 10^{-9}$ ), siguieron estando significativamente asociados al color del iris, siendo rs916977 claramente el marcador más informativo de los cuatro probados. El análisis de haplotipos basado en estos cuatro SNP en las dos poblaciones ampliadas mostró que el haplotipo de TGTC (rs11855019, rs6497268, rs7495174 y rs916977) se encontró en el 75,5% de los participantes del estudio de ERF y en el 81,2% del estudio de Rotterdam. Este haplotipo se encontró generalmente en individuos con iris azules, y en frecuencia significativamente menor en aquellos con color del iris marrón e intermedio (estudio de ERF: 92,0%, 55,6% y 79,4%; estudio de Rotterdam: 91,2%, 55,3% y 72,1%) (Tabla 4). Los presentes inventores observaron 10 haplotipos adicionales en el estudio de ERF y de Rotterdam. Todos estos se encontraron en una mayor frecuencia en individuos con el color del iris marrón en comparación con aquellos con el color del iris azul (excepto para el haplotipo CTTC).

#### *Predicción del color del iris basándose en genotipos de SNP*

Primero, los presentes inventores evaluaron el valor predictivo del rs916977 de *HERC2*, el SNP repetido en ERF (GWA y ligamiento) y el estudio de Rotterdam (GWA), junto con los tres SNP en el intrón 1 del gen *OCA2* (rs11855019, rs6497268 y rs7495174) identificados por Duffy y col. (Duffy y col., 2007) y posteriormente lo confirmaron en el estudio de ERF y de Rotterdam. La Figura 5a muestra la exactitud discriminatoria para la predicción del color del iris marrón en los datos estudio de Rotterdam y de ERF basándose en los cuatro SNP. El ABC para la predicción del color del iris marrón fue 0,81 en el conjunto de datos de derivación, 0,81 en el conjunto de datos de validación interna del estudio de Rotterdam y 0,79 en la población de ERF total. Para el color del iris azul, estos valores fueron ligeramente inferiores, pero comparables (0,80, 0,80 y 0,76; datos no mostrados). El ángulo agudo en las curvas de ROC sugiere una función dominante para uno de los SNP, que fue rs916977. La Figura 5b compara la predicción del color del iris marrón basándose en los cuatro SNP con la basada para el SNP más fuertemente asociado, rs916977, por sí misma. Notoriamente, el ABC para la predicción del color del iris marrón basada únicamente en rs916977 fue 0,78 en el estudio de Rotterdam y 0,71 en el estudio de ERF. Basándose en el estudio de Rotterdam, la probabilidad predicha del color del iris marrón fue 10,3% para portadores homocigóticos del principal alelo C de rs916977, 63,3% para heterocigóticos y 84,7% para no portadores. Aunque ninguno de los genes previamente implicados en el color del iris tales como *MATP*, *ASIP*, *TYRP1*, *CYP1A2*, *CYP2C8* y *CYP2C9* (Kanetsky y col., 2002; Frudakis y col., 2003; Graf y col., 2005) alcanzó significancia del genoma completo en dos estudios de GWA independientes de los presentes inventores, ni mostraron pruebas convincentes para ligamiento en el estudio de ERF, los presentes inventores evaluaron todos los SNP que cubrían aquellos genes en el conjunto de datos de los presentes inventores del estudio de ERF y de Rotterdam. El único gen que mostró pruebas coherentes de asociación al color del iris en el estudio de ERF y de Rotterdam fue *TYRP1*. Hubo 8 SNP consecutivos (rs668603, rs497500, rs619678, rs650283, rs678535, rs500021, rs965209 y rs7857363) que mostraron pruebas significativas nominales de asociación ( $0,01 < p < 0,02$ ). La adición de estos SNP al modelo de predicción en la muestra de Rotterdam ( $n=491$ ) mejoró el valor marginalmente predictivo (aumento del ABC de 0,81 a 0,84).

En el modelo final de los presentes inventores, los presentes inventores añadieron cuatro SNP de *OCA2* adicionales (rs768547, rs3751651, rs1448488 y rs730502), que los presentes inventores obtuvieron de análisis de regresión retrospectiva múltiple ( $P < 0,02$ ) cuando se ajusta para el efecto de los 41 SNP disponibles del gen *OCA2* en el estudio de Rotterdam (después del control de calidad). La adición de estos cuatro SNP al modelo que incluye rs916977 de *HERC2*, los tres SNP del intrón 1 en *OCA2* y los 8 SNP de *TYRP1* mejoraron el ABC a 0,87. En todos los análisis, rs916977 de *HERC2* siguió siendo el SNP que dirigía el valor predictivo.

#### *Distribución espacial de alelos de rs916977 y correlación con el color del iris en toda Europa*

Una inspección de las frecuencias de alelos de rs916977 en 23 poblaciones de distribución de ámbito europeo (de las cuales el estudio de Rotterdam era una) mostraron que el alelo C, asociado al color del iris azul en los dos estudios de GWA, fue más frecuente en el norte de Europa (la región en la que el color del iris azul es el rasgo principal), y el alelo T, asociado al color del iris marrón en los dos estudios de GWA, fue más frecuente en el sur de Europa (la región en la que el color del iris marrón es el rasgo principal) (Figura 6A). Un análisis de autocorrelación espacial reveló que las frecuencias de alelos de rs916977 en las 23 poblaciones europeas estudiadas siguieron una distribución clinal estadísticamente significativa en toda Europa (Figura 6B). En la Figura 6A, los presentes inventores superpusieron las frecuencias de alelos de rs916977 en las 23 poblaciones europeas estudiadas sobre la distribución de fenotipos del color del iris en toda Europa. Como es evidente, hay una clara relación entre la distribución del alelo C con la del color de iris claro e inversa la distribución del alelo T con la del color del iris oscuro en Europa. Obsérvese que la mayor frecuencia de rs916977C en Finlandia y Suecia también mostró la mayor frecuencia de color del iris claro ( $> 80\%$ ). Para probar formalmente esta relación, los presentes inventores dedujeron los fenotipos del color del iris para las 23 poblaciones estudiadas a partir de las clases del color del iris representadas en la figura (desafortunadamente, en la medida que saben los presentes inventores, los datos en bruto para esta cifra nunca se publicaron) y revelaron una correlación altamente positiva entre la distribución de alelos de rs916977 y la distribución de fenotipos del color del iris deducidos ( $r$  de Pearson=0,7).

5 Para concluir, los presentes inventores identificaron el SNP intrónico rs916977 a partir del gen *HERC2* como el marcador más significativamente asociado a la variación del color del iris humano a un nivel del genoma completo en dos muestras de poblaciones independientes de Los Países Bajos y como marcador con mayor valor de predicción para el color del iris. Los hallazgos de los inventores sugieren que las variantes genéticas que regulan la expresión del gen *OCA2* existen en el gen *HERC2* o, alternativamente, dentro de la secuencia corta entre los genes *OCA2* y *HERC2*, que causa la variación del color del iris en seres humanos.

10 Los datos de los presentes inventores también sugieren que los marcadores genéticos propuestos aquí, rs916977 (y cualquier marcador estrechamente ligado en particular), son de valor para la predicción del color de los ojos en personas desconocidas para aplicaciones forenses de identificación humana. Esto es ciertamente aplicable a individuos de origen europeo. Finalmente, necesita enfatizarse que cualquier aplicación de marcadores genéticos para la predicción del color del iris u otros rasgos visibles para fines forenses requiere una clara base legal, que no existe en Los Países Bajos.

Tabla 1 Características de las poblaciones de los estudios

	Tamaño de la población, N (%)	Frecuencia del color del iris		
		Azul	Intermedio	Marrón
<i>Cribados de GWA independientes</i>				
Muestra basada en la población del estudio de ERF	192	4D.6	16,2	43,2
Mujeres	95 (49,5)	39,0	16,8	44,2
Hombres	97 (50,5)	42,3	15,4	42,3
Edad media, años (DE)		31,7 (6,6)	32,5 (4,9)	30,2 (6,9)
<hr/>				
Muestra basada en la población del estudio de Rotterdam	481	67,3	10,0	22,7
Mujeres	481 (100)	67,3	10,0	22,7
Hombres	0	NA	NA	NA
Edad media, años (DE)		68,9 (3,7)	68,1 (4,1)	67,8 (3,8)
<hr/>				
<i>Replicación de regiones candidatas</i>				
Estudio de ERF basado en familias totales	2217	41,1	20,9	38,0
Mujeres	1248 (56,3)	38,4	21,6	40,0
Hombres	969 (43,7)	44,6	19,9	35,5
Edad media, años (DE)		49,7 (14,5)	49,5 (13,9)	47,7 (14,1)

(continuación)

Estudio de Rotterdam basado en la población total	6056	67,6	9,7	22,7
Mujeres	3565 (58,9)	66,6	10,9	22,6
Hombres	2491 (41,1)	69,3	7,8	22,9
Edad media, años (DE)		69,3 (8,9)	68,6 (8,8)	68,2 (8,7)

Los valores son porcentajes a menos que se indique lo contrario. NA = no aplicable, el cribado de GWA en el estudio de Rotterdam se realizó en mujeres solo.

Tabla 2. Asociación de bloques de haplotipos con el color del iris humano en la región 25,86-26,19 Mb del cromosoma 15 de dos estudios de GWA independientes de Los Países Bajos.

Conjunto de muestras	de	Bloque de haplotipos <sup>†</sup>	Haplotipo	Frecuencia (%)				Chi al cuadrado	Valores de p		
				Todos	Azul	Intermedio	Marrón				
Estudio de ERF N=192	1*		GATA	70,1	75,0	74,2	65,1	3,73	0,44		
			GGCG	23,4	17,3	19,4	28,9				
			AGCG	6,1	7,7	6,4	4,8				
	2*		CGA	69,5	72,1	72,4	49,2			14,3	0,01
			CGG	22,2	20,2	19,5	24,9				
			TAG	15,9	7,6	8,1	26,7				
	3**		CTCA	82,5	96,8	91,9	65,7			29,6	5,93x10 <sup>-5</sup>
			CTTA	11,2		8,1	20,5				
			ACTG	6,1	1,3	0	13,2				
Estudio de Rotterdam N=481	1*		GATA	77,7	81,9	75,5	68,2	13,7	8,27x10 <sup>-3</sup>		
			GGCG	16,0	11,8	17,0	26,6				
			AGCG	6,0	6,2	5,3	5,1				
	2*		CGA	75,6	79,5	78,6	64,6			13,9	7,61x10 <sup>-3</sup>
			CGG	13,4	12,9	9,7	15,3				
			TAG	10,3	7,4	9,5	18,9				
	3**		CTCGA	85,8	97,0	76,5	58,7			95,6	8,39x10 <sup>-20</sup>
			CTTGA	7,4	1,6	14,9	19,8				
			ACTGG	5,1	1,0	6,4	16,1				

<sup>†</sup>bloque 1 incluye: rs17567007, rs1800411, rs1448488, rs12910433, bloque 2 incluye: rs16950821, rs8024968, rs6497254, bloque 3 incluye: rs8041209, rs8028689, rs916977, rs4932620, rs2346050; \* localizado dentro del gen *OCA2*, \*\* localizado dentro del gen *HERC2*

Tabla 3. Asociación de cuatro SNP de la región de OCA2-HERC2 con el color del iris humano en las poblaciones ampliadas del estudio de ERF y de Rotterdam (N=8272) junto con frecuencias de genotipos en las poblaciones de HapMap de Europa, Asia oriental y África.

SNP	Genotipo	Estudio de ERF (N=2217)						Estudio de Rotterdam (N=6056)						Estudio Hi HapMap					
		N	Azul	Intermedio	Marrón	Valor de P	N	Azul	Intermedio	Marrón	Valores de P	N	Frec. (%)	N	Frec. (%)	N	Frec. (%)	N	Frec. (%)
rs11855019*	TT	1440	54,5	22,9	22,6	$4,4 \times 10^{-100}$	4580	77,3	8,7	14,0	$5,4 \times 10^{-211}$	49	81,7	5	5,5	3	5,0		
	TC	655	15,1	18,8	66,1		1250	36,8	13,3	49,9		11	18,3	32	35,6	26	43,3		
	CC	64	6,3	6,2	87,5		97	15,5	9,2	75,3		0	0	53	58,9	31	51,7		
rs6497268*	GG	144	153,6	21,2	25,2	$5,0 \times 10^{-75}$	4233	80,3	7,7	12,0	$2,8 \times 10^{-267}$	45	75,0	2	2,2	8	13,3		
	GT	652	17,2	20,4	62,4		1486	36,2	15,0	48,8		15	25,0	24	26,7	30	50,0		
	TT	67	7,5	14,9	77,6		151	17,9	11,2	70,9		0	0	64	71,1	22	36,7		
rs7495174*	TT	174	949,7	21,2	29,1	$4,3 \times 10^{-79}$	5273	74,6	8,9	16,5	$1,4 \times 10^{-239}$	54	90,0	7	7,8	43	71,7		
	TC	382	3,4	20,2	76,4		659	16,2	15,2	68,6		6	10,0	38	42,2	16	26,7		
	CC	17	5,9	0	94,1		34	2,9	5,9	91,2		0	0	45	50,0	1	1,6		
rs916977**	CC	1543	54,8	21,3	23,9	$1,9 \times 10^{-113}$	4572	82,4	7,3	10,3	$< 1,0 \times 10^{-300}$	44	73,3	2	2,2	0	0		
	CT	573	6,4	22,1	69,5		1229	18,3	18,4	63,3		16	26,7	28	31,1	6	10,0		
	TT	60	3,3	3,3	93,4		105	6,6	8,6	84,8		0	0	60	66,7	54	90,0		

\* localizados en el intrón 1 del gen OCA2, \*\* localizados en el intrón 11 del gen HERC2

Tabla 4. Asociación de haplotipos de cuatro SNP de la región de OCA2-HERC2 con el color del iris humano en las poblaciones ampliadas del estudio de ERF y de Rotterdam (N=8273) junto con frecuencias de haplotipos en las poblaciones de HapMap de Europa, Asia oriental y África.

Haplotipo	Estudio de ERF (N=2217)						Estudio de Rotterdam (N=6056)						HapMap (N=30)		
	Frecuencia del color del iris (%)			Frecuencia del color del iris (%)			Frecuencia del color del iris (%)			Frecuencia (%) en			Frecuencia (%) en		
	Todos	Azul	Intermedio	Marrón	Todos	Azul	Intermedio	Marrón	Europeos	Asiáticos orientales	Africanos	Europeos	Asiáticos orientales	Africanos	
TGTC	75,5	92,0	79,4	55,6	81,2	91,2	72,1	55,3	82,5	11,6	5,0	82,5	11,6	5,0	
CTCT	5,5	0,6	6,0	10,5	4,5	0,9	6,7	14,5	4,2	65,8	13,3	4,2	65,8	13,3	
TTTT	4,6	1,6	5,2	7,6	4,3	1,5	8,7	10,5	5,8	10,1	18,0	5,8	10,1	18,0	
CTCC	4,0	0,2	2,7	9,0	0,3	0,3	1,6	3,5	0	3,4	0	0	3,4	0	
TGCT	0,1	0	0	0,3	0,3	0	0,5	1,1	0,8	1,2	1,7	0,8	1,2	1,7	
TGTT	1,3	0,3	0,7	2,7	0,9	0,3	1,5	2,6	0	0,4	2,0	0	0,4	2,0	
CGTT	3,9	0,3	2,2	8,7	1,7	0,3	2,7	5,5	2,5	1,7	29,7	2,5	1,7	29,7	
CGTC	1,0	0,5	0,6	1,7	0,8	0,7	0,8	0,8	1,7	0	0	1,7	0	0	
TTTC	0,3	0,1	0,3	0,4	1,0	0,8	0,8	1,9	1,7	0	0	1,7	0	0	
CTTC	3,3	4,2	2,8	2,5	3,5	3,9	3,2	2,6	0,8	2,7	0	0,8	2,7	0	
CTTT	0,5	0,1	0,3	1,0	0,6	0,1	1,4	1,7	0	2,4	30,3	0	2,4	30,3	
CGCT	0	0	0	0	0	0	0	0	0	0,7	0	0	0,7	0	

Los haplotipos consisten en SNP en el siguiente orden: rs11855019, rs6497268, rs7495174 – todos en el gen OCA2, y rs916977 en el gen HERC2. Se muestra todo haplotipo observado en cualquiera de las poblaciones, \*\* combina datos de japoneses y chinos.

ES 2 419 183 T3

Tabla 5. Todos los SNP asociados al color del iris al nivel de significancia del genoma completo del conjunto de datos de Illumina 300 Duo en 733 participantes de ERF

Marcador de SNP	Gen	Valor de p del genoma completo	Valor de p nominal	Genotipo	Frecuencia del color del iris		
					Azul	Intermed.	Marrón
rs8039195	HERC2	<0,01	<10 <sup>-16</sup>	TT	0,5479	0,2035	0,2485
				TC	0,0855	0,2303	0,6842
				CC	0	0	1
rs1667394	HERC2	<0,01	<10 <sup>-16</sup>	TT	0,58	0,2058	0,2141
				TC	0,0791	0,226	0,6949
				CC	0	0	1
rs7183877	HERC2	<0,01	1,73E <sup>-13</sup>	TT	0,4863	0,2072	0,3065
				GT	0,1098	0,2195	0,6707
				GG	0	0	1
rs16950979	HERC2	<0,01	3,35E <sup>-12</sup>	TT	0,4824	0,2044	0,3132
				TC	0,058	0,2319	0,7101
				CC	0	0	1
rs8028689	HERC2	<0,01	3,67E <sup>-12</sup>	TT	0,4817	0,205	0,3133
				TC	0,058	0,2319	0,7101
				CC	0	0	1
rs2240204	HERC2	<0,01	3,67E <sup>-12</sup>	TT	0	0	1
				CT	0,058	0,2319	0,7101
				CC	0,4817	0,205	0,3133
rs16950987	HERC2	<0,01	4,39E <sup>-12</sup>	TT	0	0	1
				CT	0,058	0,2319	0,7101
				CC	0,4799	0,2057	0,3144
rs1597196	OCA2	<0,01	<10 <sup>-16</sup>	TT	0,122	0,2195	0,6585
				GT	0,25	0,2069	0,5431
				GG	0,5779	0,206	0,2161
rs4778138	OCA2	<0,01	<10 <sup>-16</sup>	TT	0,5756	0,2144	0,2099
				TC	0,1779	0,2115	0,6106
				CC	0,05	0	0,95
rs4778241	OCA2	<0,01	<10 <sup>-16</sup>	TT	0,5578	0,2133	0,2289
				GT	0,1961	0,201	0,6029

(continuación)

				GG	0,1176	0,1176	0,7647
rs7495174	OCA2	<0,01	$<10^{-16}$	TT	0,5226	0,2134	0,264
				TC	0,0357	0,1875	0,7768
				CC	0	0	1
rs4778232	OCA2	<0,01	$5,00E^{-14}$	TT	0,1636	0,1818	0,6545
				CT	0,3404	0,1879	0,4716
				CC	0,5629	0,2275	0,2096
rs7179994	OCA2	<0,01	$5,46E^{-14}$	TT	0,5315	0,2017	0,2668
				TC	0,2034	0,226	0,5706
				CC	0,1765	0,1765	0,6471
rs3794604	OCA2	<0,01	$1,99E^{-12}$	TT	0,125	0,1667	0,7083
				CT	0,2556	0,2056	0,5389
				CC	0,5226	0,2108	0,2667
rs1635168	HERC2	<0,01	$2,74E^{-13}$	TT	0	0	1
			GT	GT	0,0548	0,2329	0,7123
				GG	0,484	0,205	0,3109
rs8024968	OCA2	<0,01	$4,12E^{-12}$	TT	0,125	0,1667	0,7083
				CT	0,2623	0,2022	0,5355
				CC	0,5216	0,2112	0,2672
rs2594935	OCA2	<0,01	$1,00E^{-11}$	TT	0,5737	0,2038	0,2226
				CT	0,33	0,1987	0,4714
				CC	0,2182	0,2727	0,5091
rs728405	OCA2	<0,01	$1,95E^{-10}$	TT	0,5314	0,1816	0,287
				TG	0,2624	0,2525	0,4851
				GG	0,1304	0,3043	0,5652

**Referencias**

5 Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 30:97-101

Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 62:1198-1211

10 Aulchenko YS, Heutink P, Mackay I, Bertoli-Avella AM, Pullen J, Vaessen N, Rademaker TA, Sandkuijl LA, Cardon L, Oostra B, van Duijn CM (2004) Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet* 12:527-534

Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 23:1294-1296

- Beals RL, Hoijer H (1965) An introduction to anthropology. Allyn and Bacon, Boston, MA. Bito LZ, Matheny A, Cruickshanks KJ, Nondahl DM, Carino OB (1997) Eye color changes past early childhood. The Louisville Twin Study. *Arch Ophthalmol* 115:659-663
- 5 Brilliant MH (2001) The mouse p (pink-eyed dilution) and human P genes, oculocutaneous albinism type 2 (OCA2), and melanosomal pH. *Pigment Cell Res* 14:86-93
- Butler MG (1989) Hypopigmentation: a common feature of Prader-Labhart-Willi syndrome. *Am J Hum Genet* 45:140-146
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton
- 10 Chai JH, Locke DP, Grealis JM, Knoll JH, Ohta T, Dunai J, Yavor A, Eichler EE, Nicholls RD (2003) Identification of four highly conserved genes between breakpoint hotspots BP1 and BP2 of the Prader-Willi/Angelman syndromes deletion region that have undergone evolutionary transposition mediated by flanking duplicons. *Am J Hum Genet* 73:898-925
- Davenport GC, Davenport CB (1907) Heredity of eye-color in man. *Science* 26:590-592
- 15 Devlin B, Roeder K (1999) Genomic control for association studies. *Biometrics* 55:997-1004
- Duffy DL, Box NF, Chen W, Palmer JS, Montgomery GW, James MR, Hayward NK, Martin NG, Sturm RA (2004) Interactive effects of MC1R and OCA2 on melanoma risk phenotypes. *Hum Mol Genet* 13:447-461
- Duffy DL, Montgomery GW, Chen W, Zhao ZZ, Le L, James MR, Hayward NK, Martin NG, Sturm RA (2007) A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. *Am J Hum Genet* 80:241-252
- 20 Eiberg H, Mohr J (1987) Major genes of eye color and hair color linked to LU and SE. *Clin Genet* 31:186-191
- Eiberg H, Mohr J (1996) Assignment of genes coding for brown eye colour (BEY2) and brown hair colour (HCL3) on chromosome 15q. *Eur J Hum Genet* 4:237-241
- Frost P (2006) European hair and eye color. A case of frequency-dependent sexual selection? *Evolution and Human Behavior* 27:85-103
- 25 Frudakis T, Terravainen T, Thomas M (2007) Multilocus OCA2 genotypes specify human iris color. *Hum Genet Online* published ahead of print:DOI: 10.1007/s00439-00007-00401-00438
- Frudakis T, Thomas M, Gaskin Z, Venkateswarlu K, Chandra KS, Ginjupalli S, Gunturi S, Natrajan S, Ponnuswamy VK, Ponnuswamy KN (2003) Sequences associated with human iris pigmentation. *Genetics* 165:2071-2083
- 30 Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. *Science* 296:2225-2229
- Graf J, Hodgson R, van Daal A (2005) Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation. *Hum Mutat* 25:278-284
- 35 Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29-36
- Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA (1991) Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol* 7:403-422
- 40 Imesch PD, Bindley CD, Khademian Z, Ladd B, Gangnon R, Albert DM, Wallow IH (1996) Melanocytes and iris color. Electron microscopic findings. *Arch Ophthalmol* 114:443-447
- Imesch PD, Wallow IH, Albert DM (1997) The color of the human eye: a review of morphologic correlates and of some conditions that affect iridial pigmentation. *Surv Ophthalmol* 41 Suppl 2:S117-123
- Jannot AS, Meziani R, Bertrand G, Gerard B, Descamps V, Archimbaud A, Picard C, Ollivaud L, Basset-Seguain N, Kerob D, Lanternier G, Lebbe C, Saiag P, Crickx B, Clerget-Darpoux F, Grandchamp B, Soufir N, Melan C (2005) Allele variations in the OCA2 gene (pink-eyed-dilution locus) are associated with genetic susceptibility to melanoma. *Eur J Hum Genet* 13:913-920
- 45

- Ji Y, Rebert NA, Joslin JM, Higgins MJ, Schultz RA, Nicholls RD (2000) Structure of the highly conserved HERC2 gene and of multiple partially duplicated paralogs in human. *Genome Res* 10:319-329
- 5 Ji Y, Walkowicz MJ, Buiting K, Johnson DK, Tarvin RE, Rinchik EM, Horsthemke B, Stubbs L, Nicholls RD (1999) The ancestral gene for transcribed, low-copy repeats in the Prader-Willi/Angelman region encodes a large protein implicated in protein trafficking, which is deficient in mice with neuromuscular and spermiogenic abnormalities. *Hum Mol Genet* 8:533-542
- Kanetsky PA, Swoyer J, Panossian S, Holmes R, Guerry D, Rebbeck TR (2002) A polymorphism in the agouti signaling protein gene is associated with human pigmentation. *Am J Hum Genet* 70:770-775
- 10 King RA, Wiesner GL, Townsend D, White JG (1993) Hypopigmentation in Angelman syndrome. *Am J Med Genet* 46:40-44
- Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E (2003) A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* 12:1725-1735
- 15 Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215
- Pardo LM, MacKay I, Oostra B, van Duijn CM, Aulchenko YS (2005) The effect of genetic drift in a young genetically isolated population. *Ann Hum Genet* 69:288-295
- Posthuma D, Visscher PM, Willemsen G, Zhu G, Martin NG, Slagboom PE, de Geus EJ, Boomsma DI (2006) Replicated linkage for eye color on 15q using comparative ratings of sibling pairs. *Behav Genet* 36:12-17
- 20 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: a toolset for whole-genome association and population-based linkage analyses. *Am J Hum Genet*:in press (preimpresión en línea publicada en mayo de 2007)
- Ramsay M, Colman MA, Stevens G, Zwane E, Kromberg J, Farrall M, Jenkins T (1992) The tyrosinase-positive oculocutaneous albinism locus maps to chromosome 15q11.2-q12. *Am J Hum Genet* 51:879-884
- 25 Rebbeck TR, Kanetsky PA, Walker AH, Holmes R, Halpern AC, Schuchter LM, Elder DE, Guerry D (2002) P gene as an inherited biomarker of human eye color. *Cancer Epidemiol Biomarkers Prev* 11:782-784
- Rinchik EM, Bultman SJ, Horsthemke B, Lee ST, Strunk KM, Spritz RA, Avidano KM, Jong MT, Nicholls RD (1993) A gene for the mouse pink-eyed dilution locus and for human type II oculocutaneous albinism. *Nature* 361:72-76
- 30 Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA (2002) Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 70:425-434
- Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorius H, Bedoya G, Ospina J, et al. (2006) Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nat Genet* 38:556-560
- 35 Sokal RR, Oden NL (1978) Spatial autocorrelation in biology 1. Methodology. *Biological Journal of the Linnean Society* 10:199-228
- Spritz RA, Fukai K, Holmes SA, Luande J (1995) Frequent intragenic deletion of the P gene in Tanzanian patients with type II oculocutaneous albinism (OCA2). *Am J Hum Genet* 56:1320-1323
- 40 Staehling-Hampton K, Proll S, Paeper BW, Zhao L, Charmley P, Brown A, Gardner JC, Galas D, Schatzman RC, Beighton P, Papapoulos S, Hamersma H, Brunkow ME (2002) A 52-kb deletion in the SOST-MEOX1 intergenic region on 17q12-q21 is associated with van Buchem disease in the Dutch population. *Am J Med Genet* 110:144-152
- Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, Styrkarsdottir U, et al. (2007) A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* 39:770-775
- Sturm RA, Frudakis TN (2004) Eye colour: portals into pigmentation genes and ancestry. *Trends Genet* 20:327-332
- 45 Wielgus AR, Sarna T (2005) Melanin in human irides of different color and age of donors. *Pigment Cell Res* 18:454-464

Wiesner GL, Bendel CM, Olds DP, White JG, Arthur DC, Ball DW, King RA (1987) Hypopigmentation in the Prader-Willi syndrome. *Am J Hum Genet* 40:431-442

Zhu G, Evans DM, Duffy DL, Montgomery GW, Medland SE, Gillespie NA, Ewen KR, Jewell M, Liew YW, Hayward NK, Sturm RA, Trent JM, Martin NG (2004) A genome scan for eye color in 502 twin families: most variation is due to a QTL on chromosome 15q. *Twin Res* 7:197-210

5

**REIVINDICACIONES**

1. Un procedimiento para predecir el color del iris de un ser humano a partir de una muestra de ácido nucleico que comprende ensayar uno o más polimorfismos en el gen *HERC2* y basándose en los resultados del ensayo predecir el color de los ojos, en el que el uno o más polimorfismos se seleccionan del grupo constituido por rs916977, rs8028689, rs6497287, rs8041209, rs6497292, rs2240202, rs2346050, rs12592730, rs7183877, rs2240204, rs8039195, rs16950979, rs16950987, rs1667394 y rs1635168 y cualquier marcador en el gen *HERC2* en desequilibrio de ligamiento genético con dichos polimorfismos.
2. Un procedimiento según la reivindicación 1, en el que el polimorfismo es rs916977.
3. Un procedimiento según la reivindicación 2, en el que la predicción es ojos azules si el polimorfismo rs916977 tiene la secuencia CC que considera ambos cromosomas homólogos (o GG si se consideran las bases complementarias), ojos marrones si la secuencia de nucleótidos es TT (o AA si se consideran las bases complementarias).
4. Un procedimiento según la reivindicación 1, en el que los polimorfismos predicen los colores de los ojos de la siguiente manera: rs8028689: CC marrón y TT azul, rs6497287: AA azul y GG marrón, rs8041209: AA marrón y CC azul, rs6497292: AA azul y GG marrón, rs2240202: AA marrón y GG azul, rs2346050: AA azul y GG marrón, rs12592730: AA marrón y GG azul, rs7183877: TT azul y GG marrón, rs2240204: TT marrón y CC azul, rs8039195: TT azul y CC marrón, rs16950979: AA azul y CC marrón, rs16950987: TT marrón y CC azul, rs1667394: TT azul y CC marrón y rs1635168: TT marrón y GG azul.
5. Un procedimiento de indicación del color del iris humano detectando una existencia de nucleótidos para un polimorfismo de un solo nucleótido (SNP), que comprende:
  - (i) incubar una muestra que comprende un polinucleótido con un miembro del par de unión específico, en el que el miembro del par de unión específico se une específicamente a o próximo a un polinucleótido que se sospecha que es polimórfico, comprendiendo el polinucleótido una de las existencias de nucleótidos correspondientes a al menos uno de los polimorfismos rs916977, rs8028689, rs6497287, rs8041209, rs6497292, rs2240202, rs2346050, rs12592730, rs7183877, rs2240204, rs8039195, rs16950979, rs16950987, rs1667394 y rs1635168 o cualquier combinación de los mismos o cualquier marcador en el gen *HERC2* en desequilibrio de ligamiento genético con dichos polimorfismos; y ii) detectar la unión selectiva del miembro del par de unión específico, en el que la unión selectiva es indicativa de la presencia de la existencia de nucleótidos, detectando así la existencia de nucleótidos para el polimorfismo.

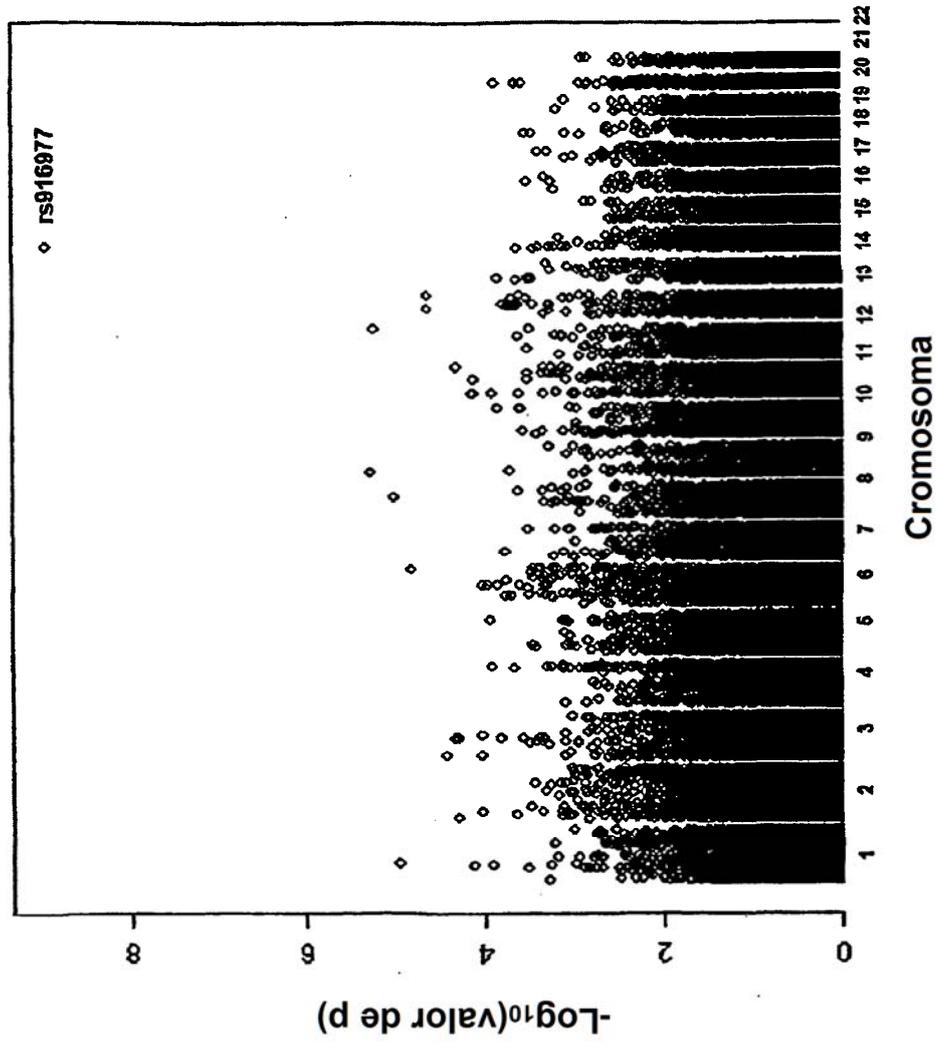


Fig. 1A

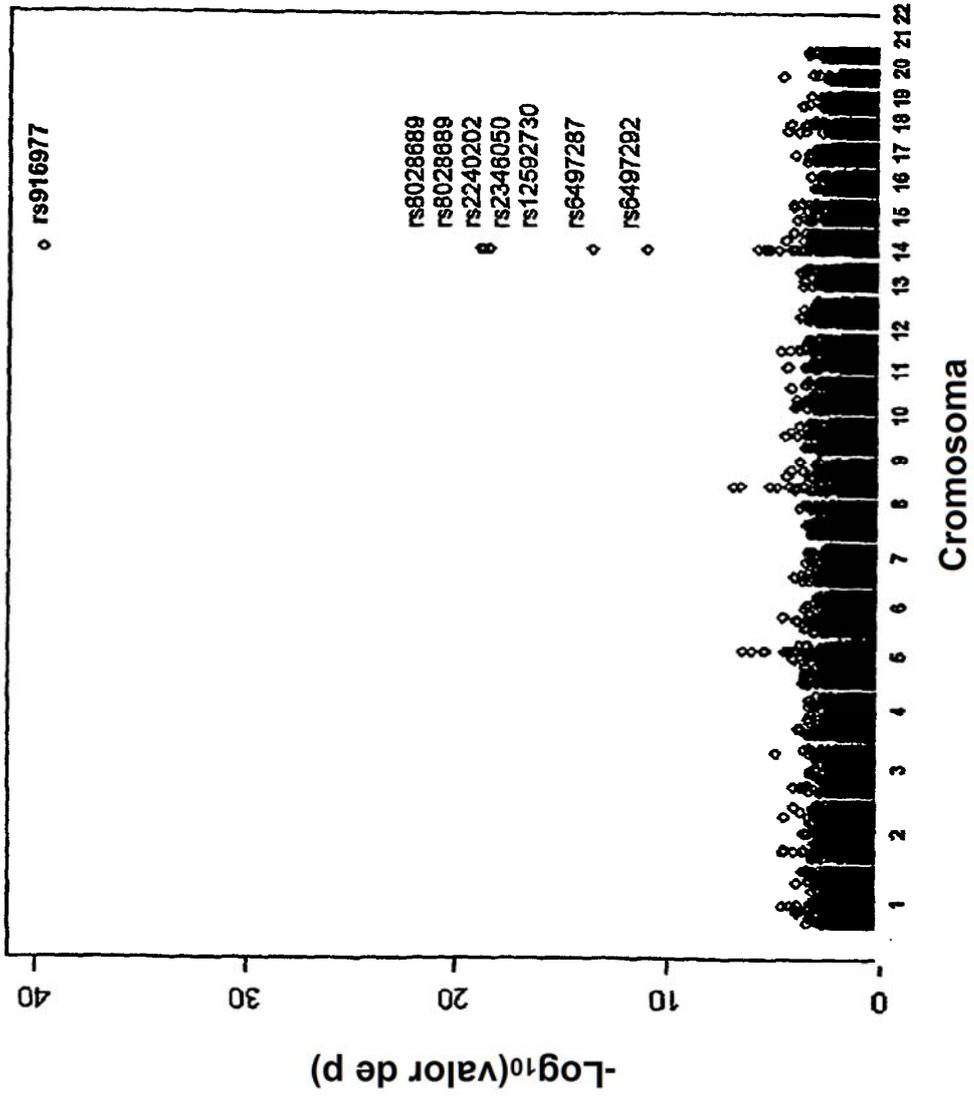


Fig. 1B

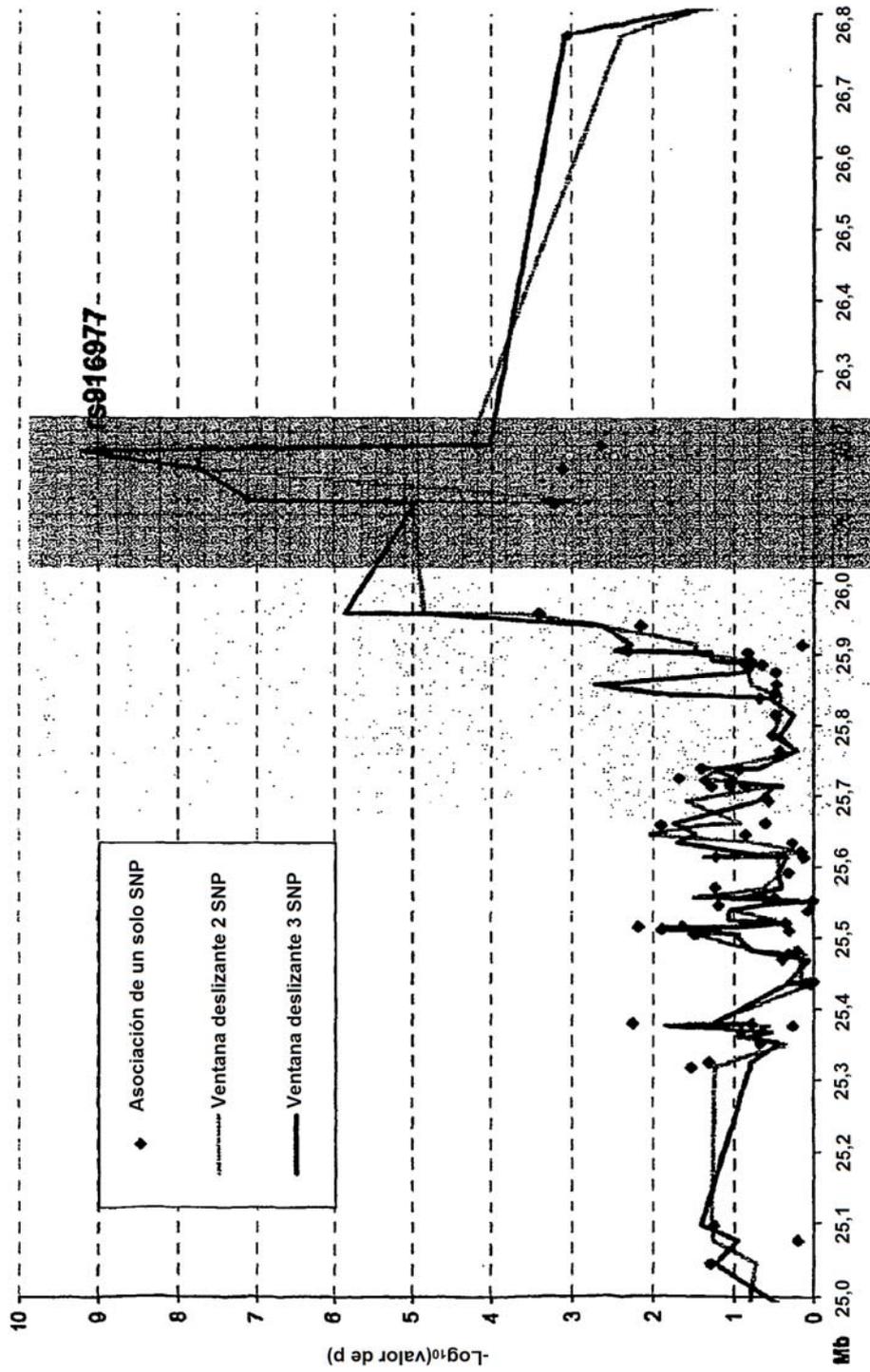


Fig. 2A

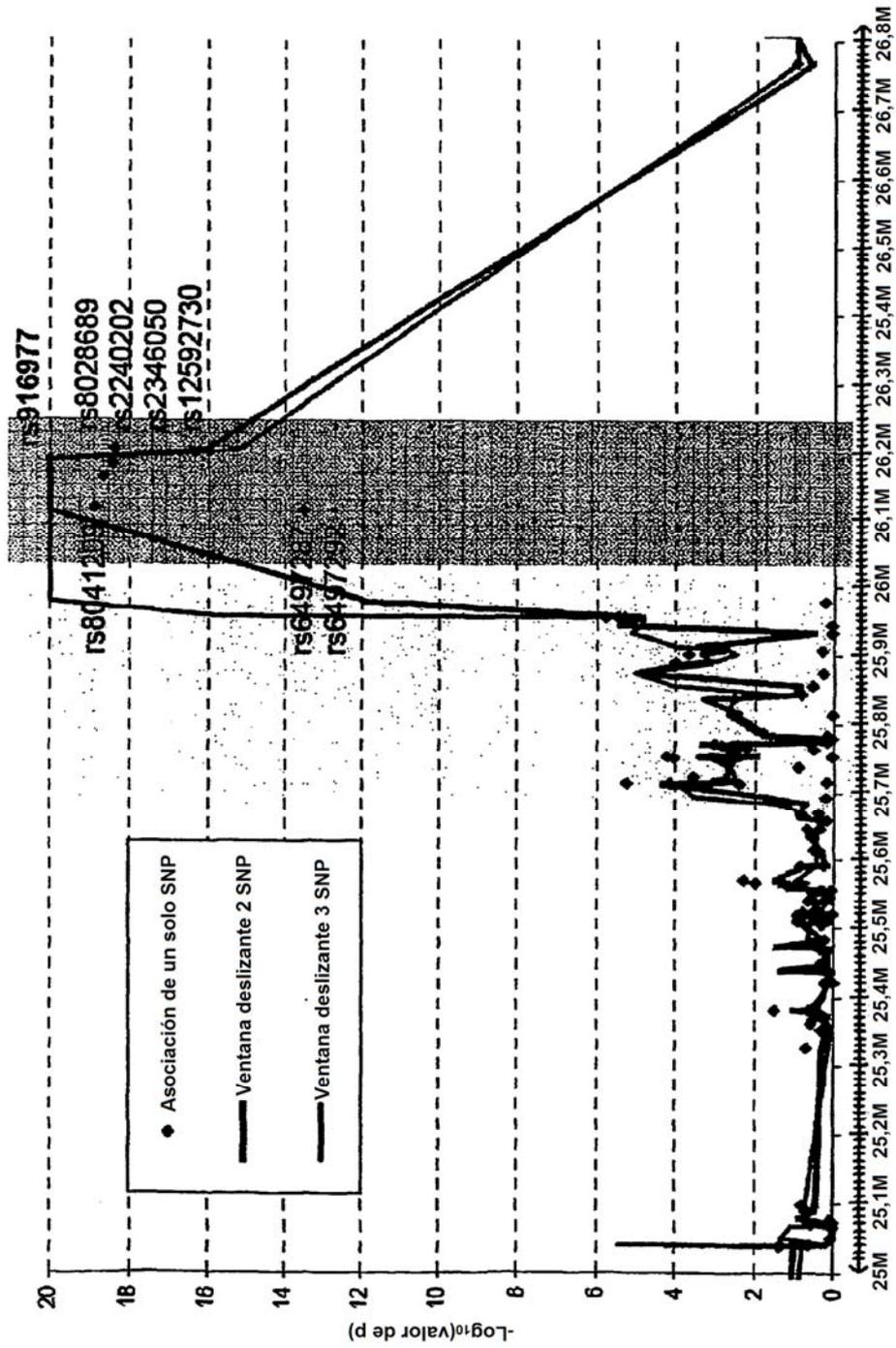


Fig. 2B

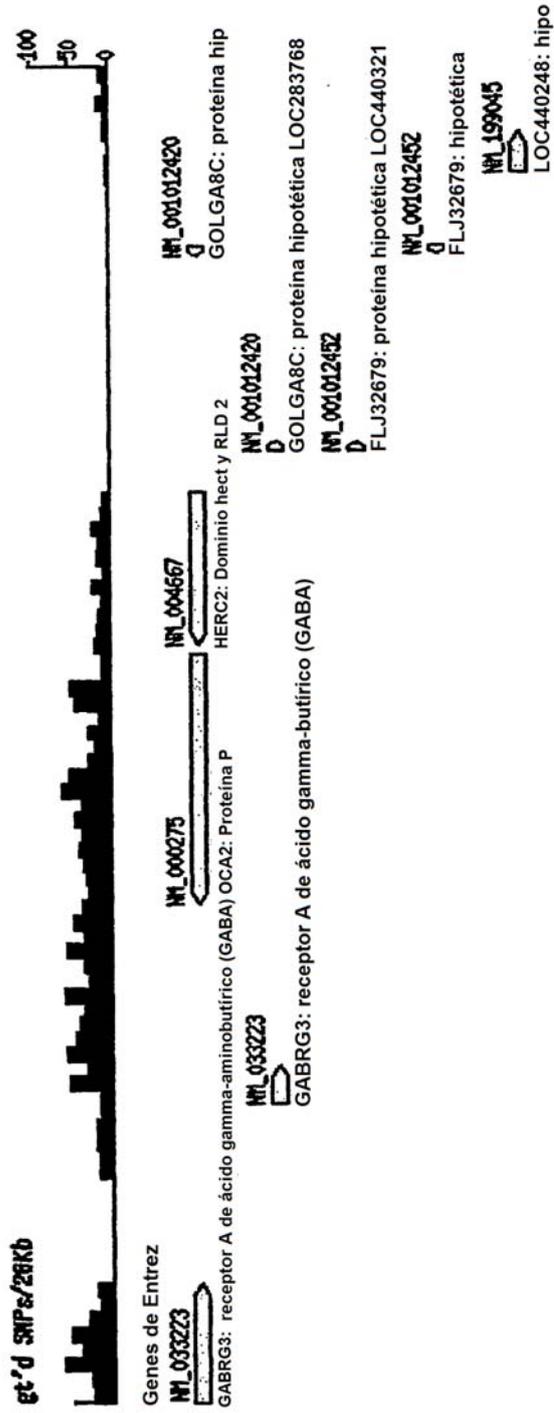


Figura 2B, continuación

Fig. 3

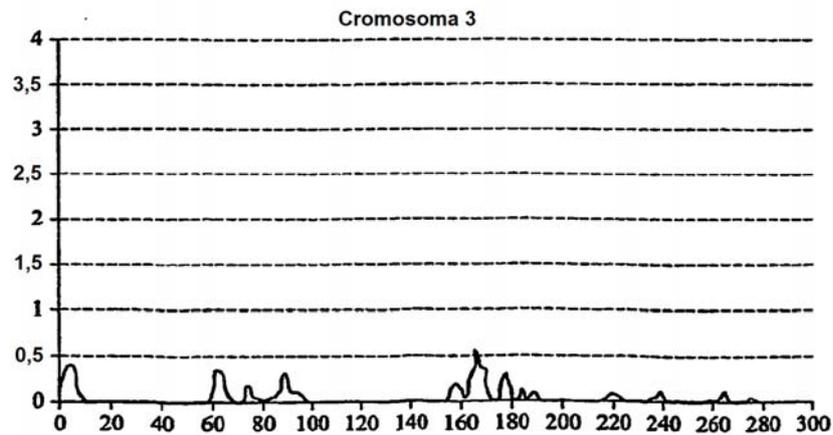
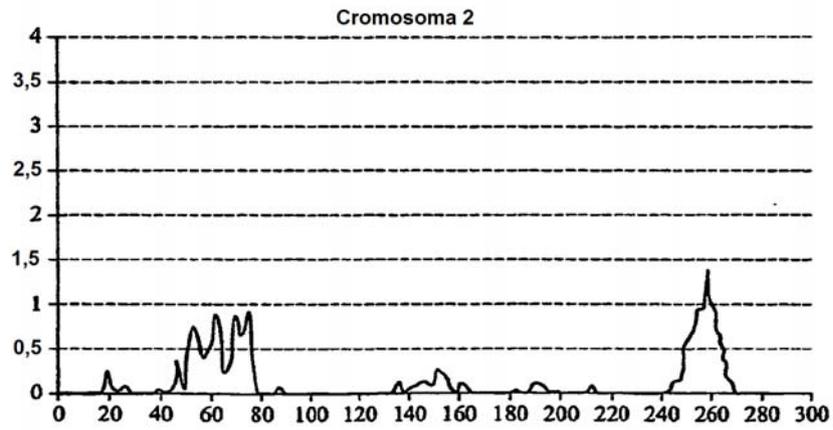
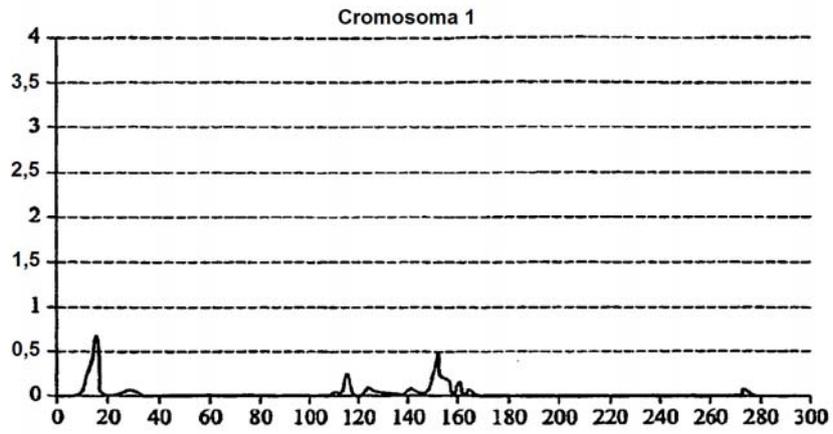


Fig. 3 continuación

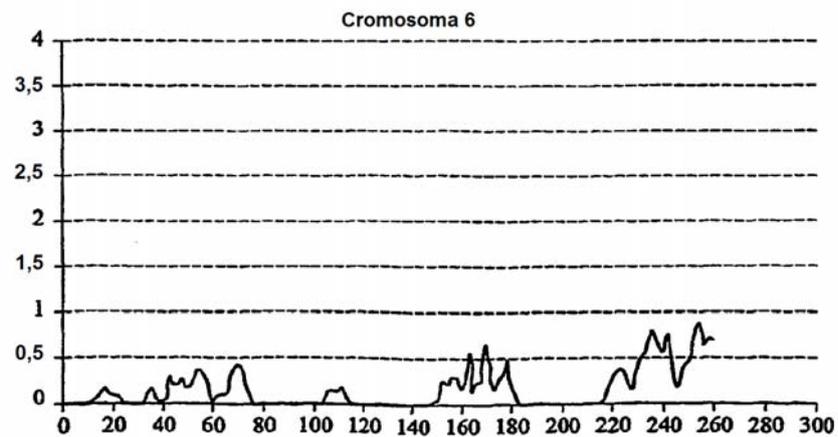
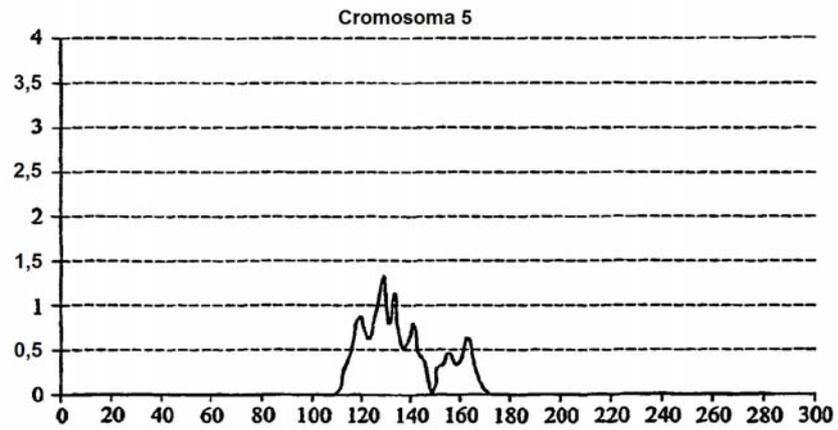
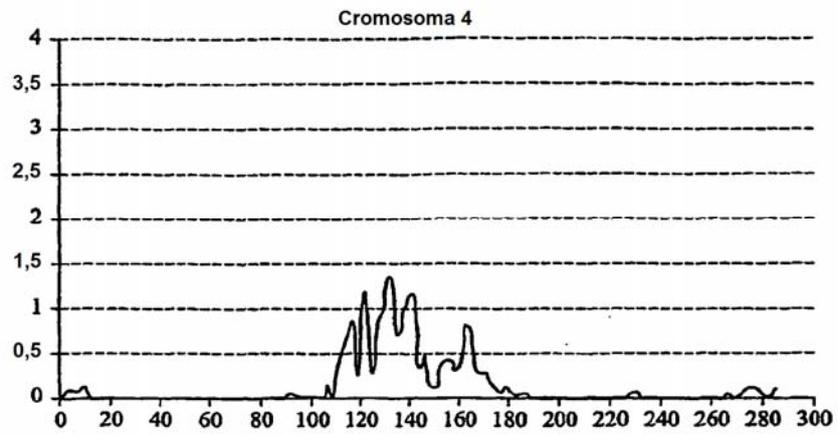


Fig. 3 continuación

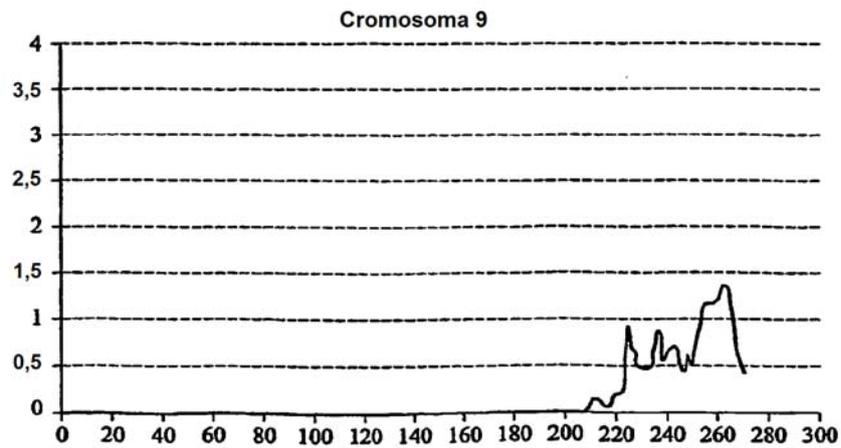
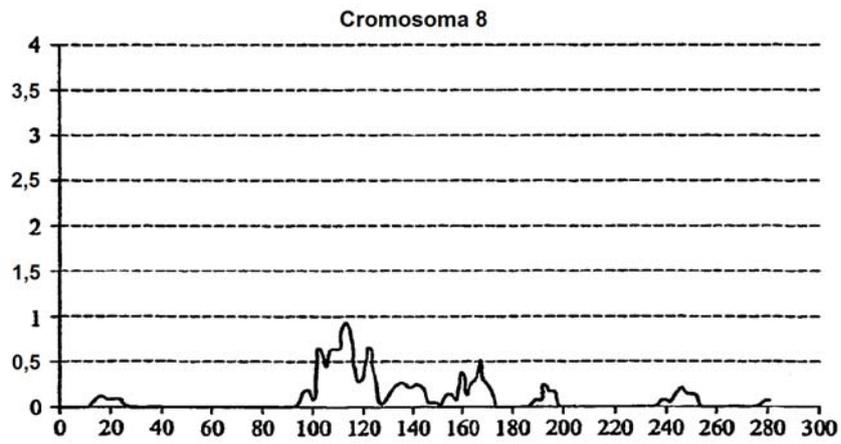
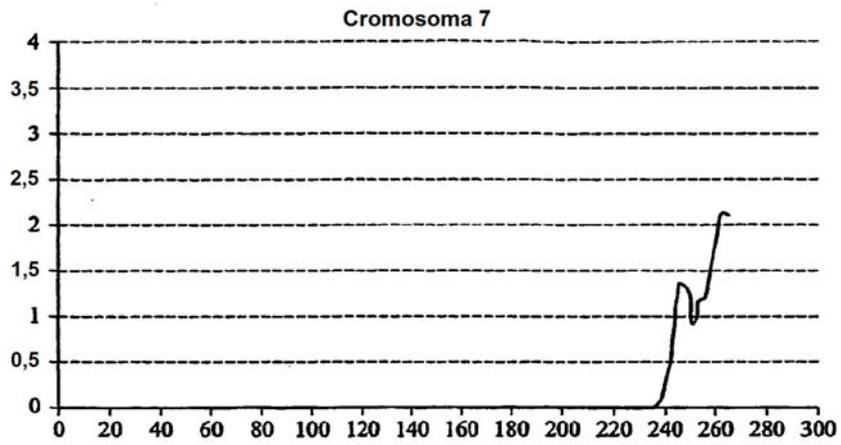


Fig. 3 continuación

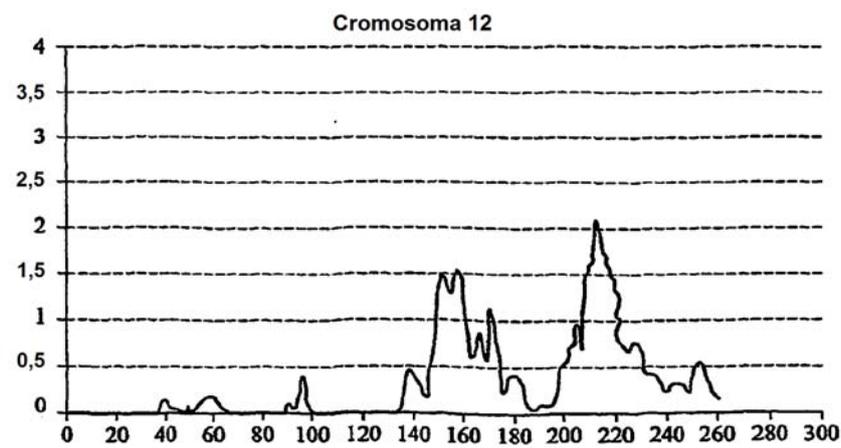
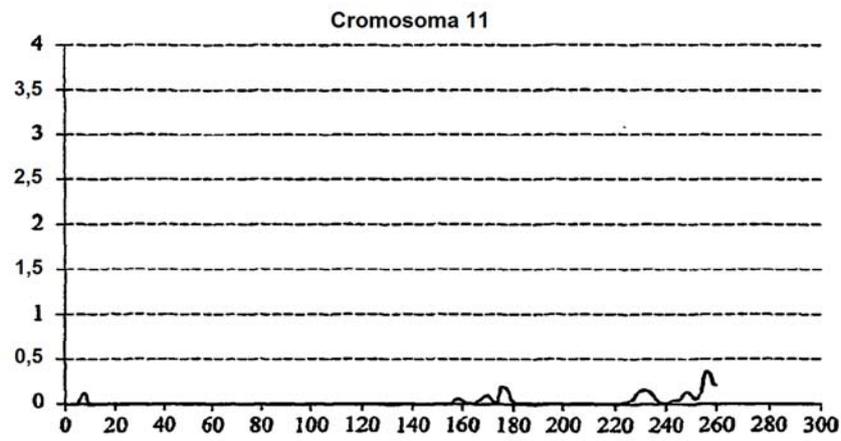
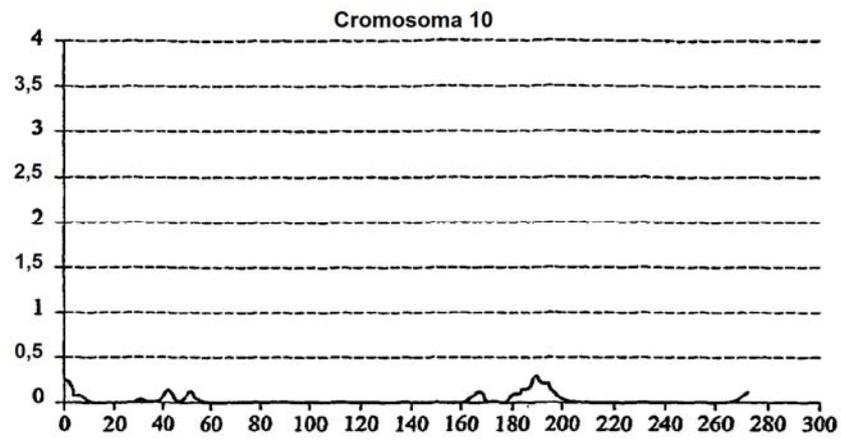


Fig. 3 continuación

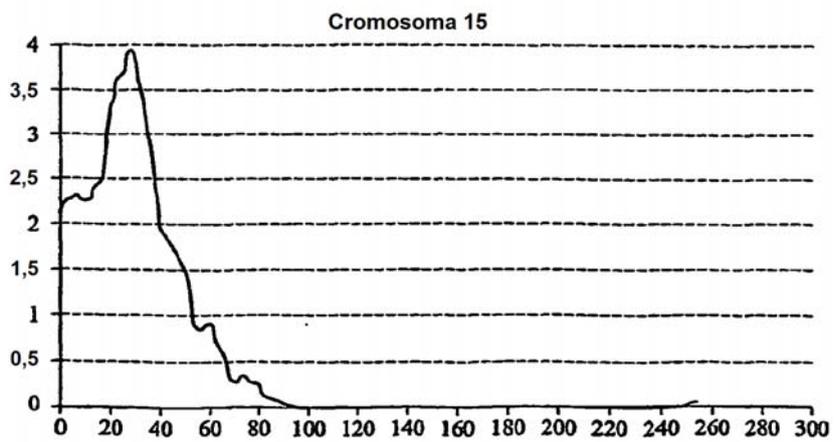
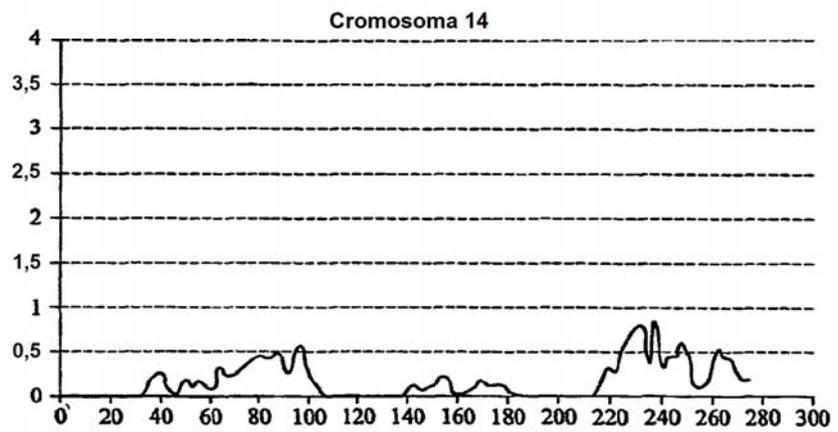
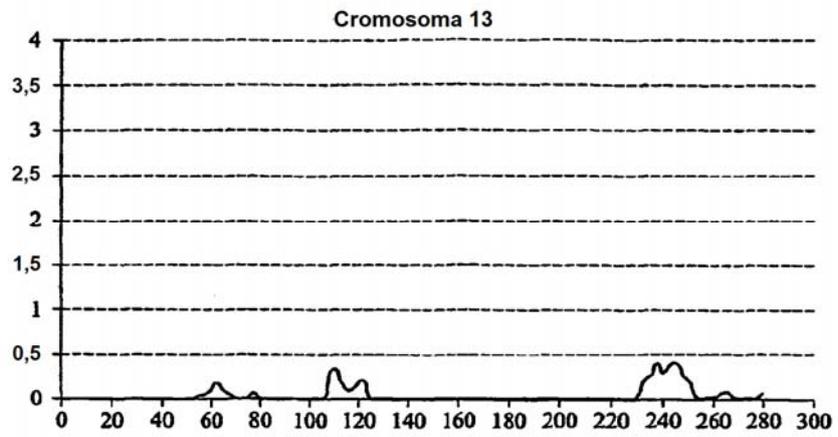


Fig. 3 continuación

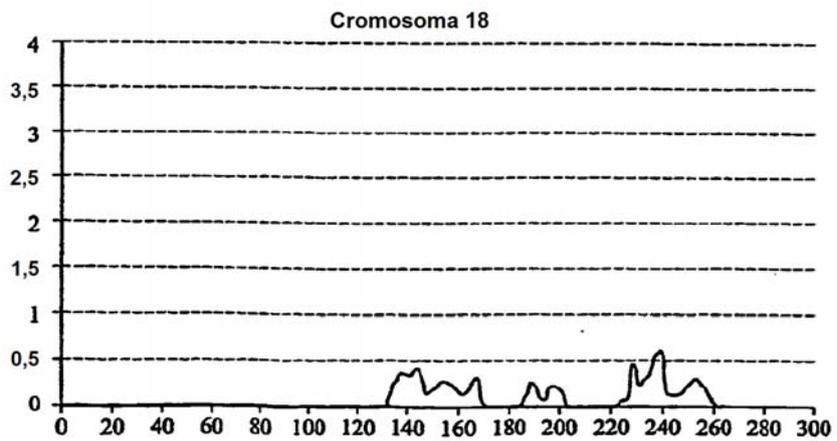
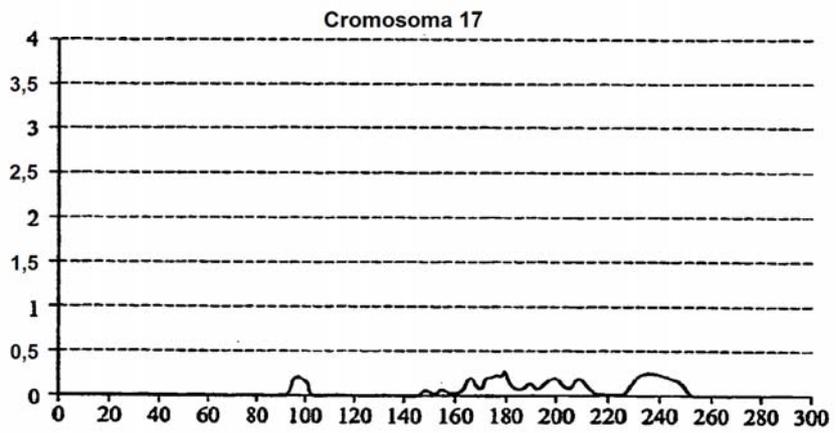
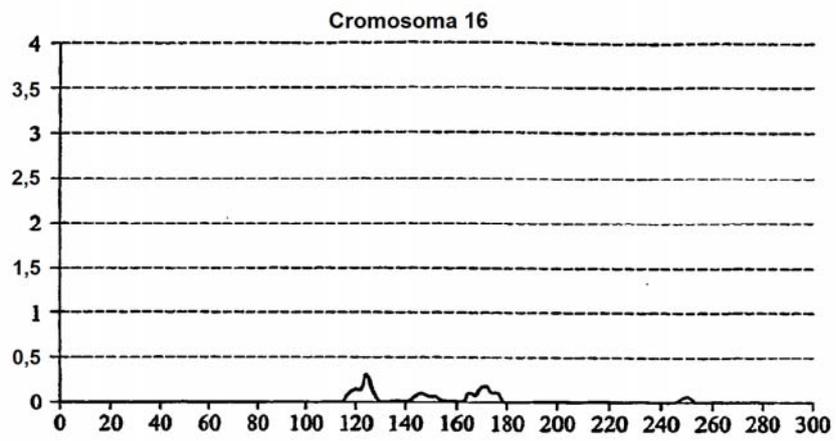
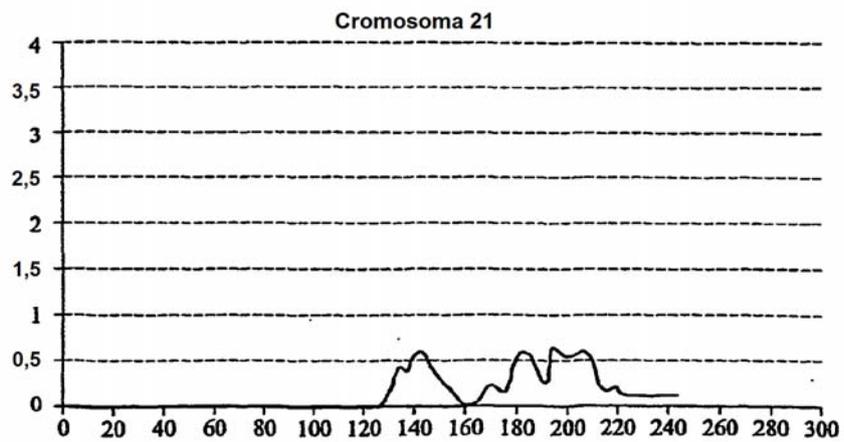
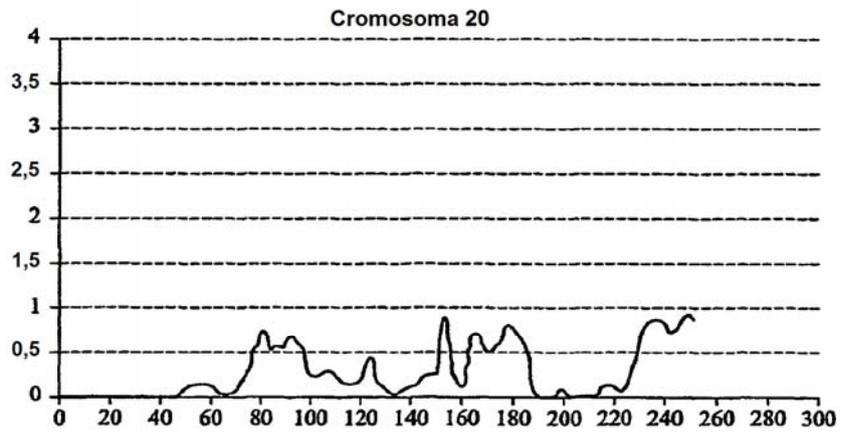
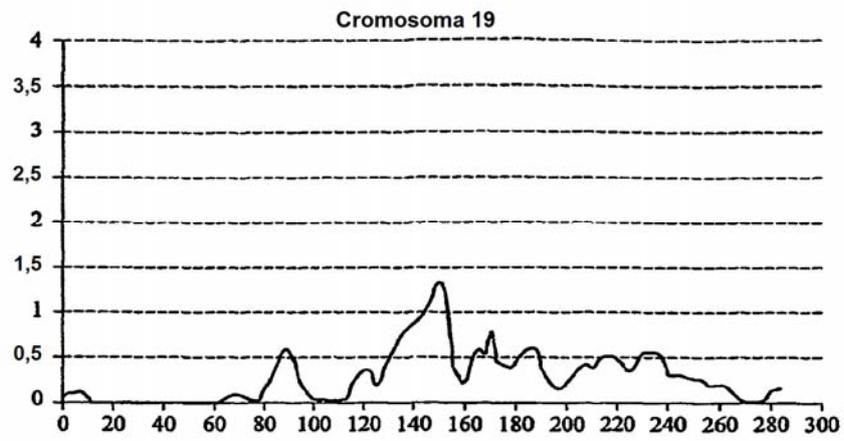
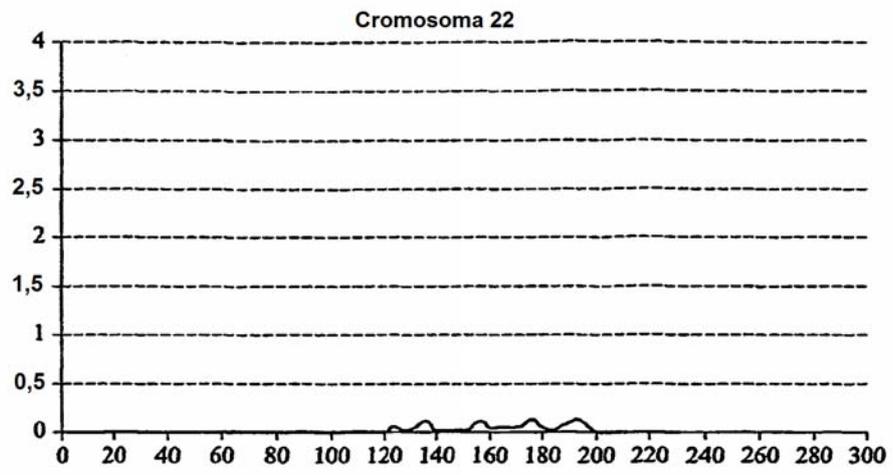


Fig. 3 continuación



**Fig. 3 continuación**



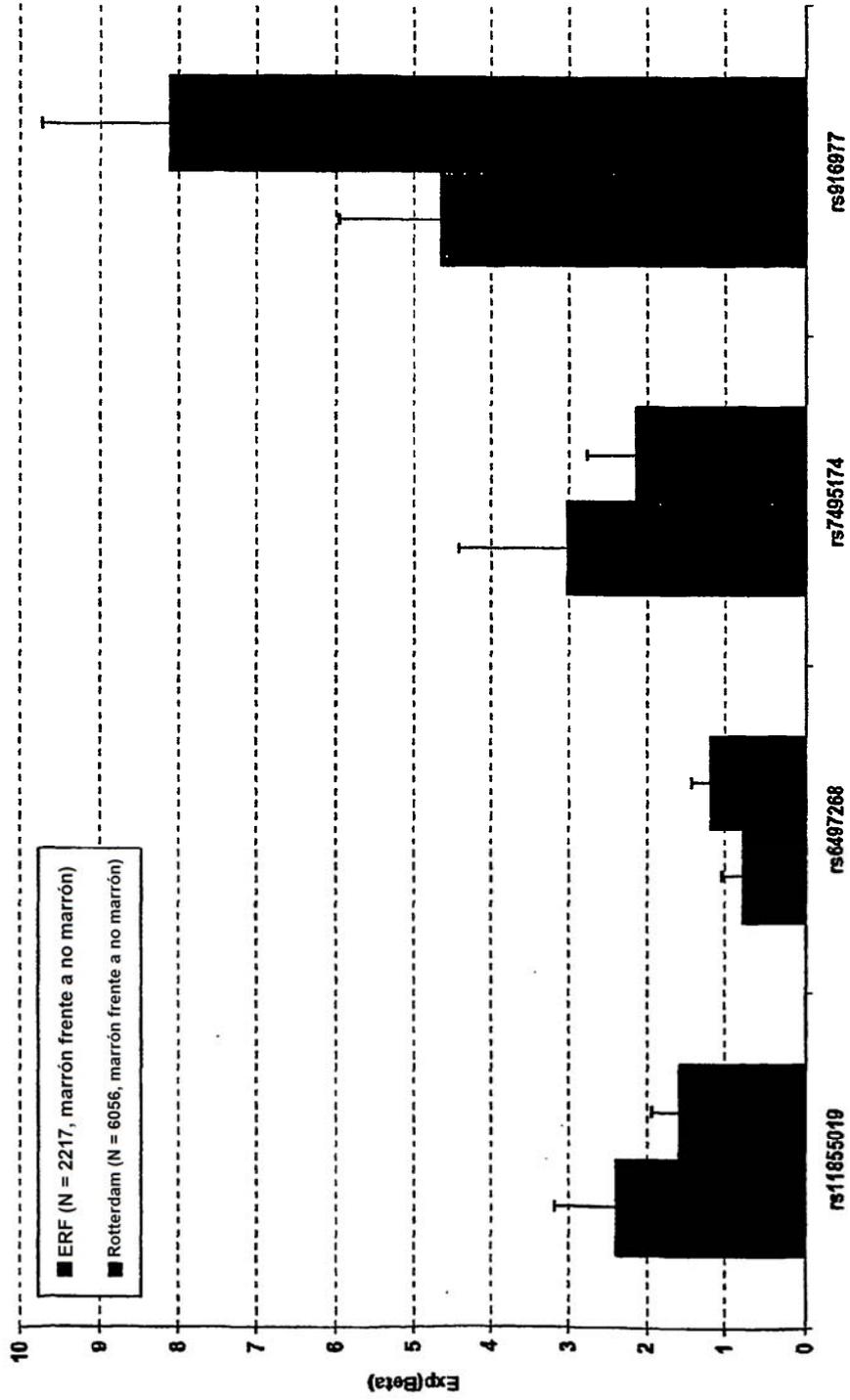


Fig. 4

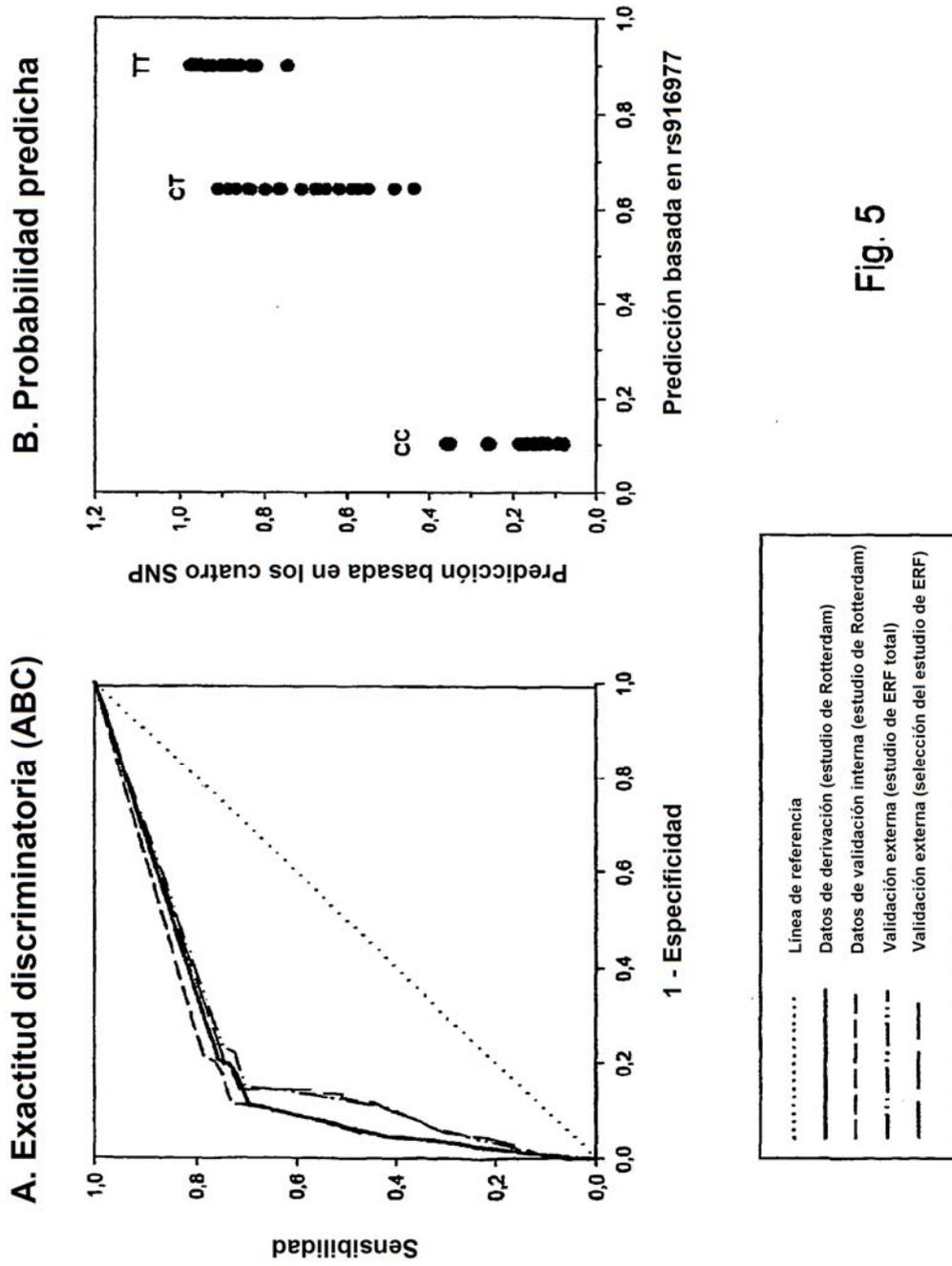


Fig. 5

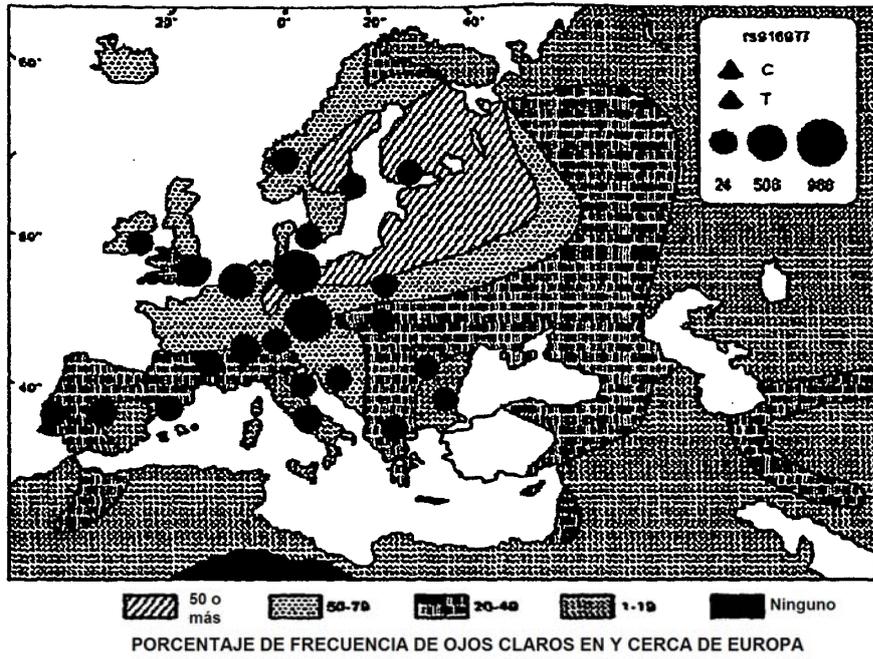


Fig. 6A

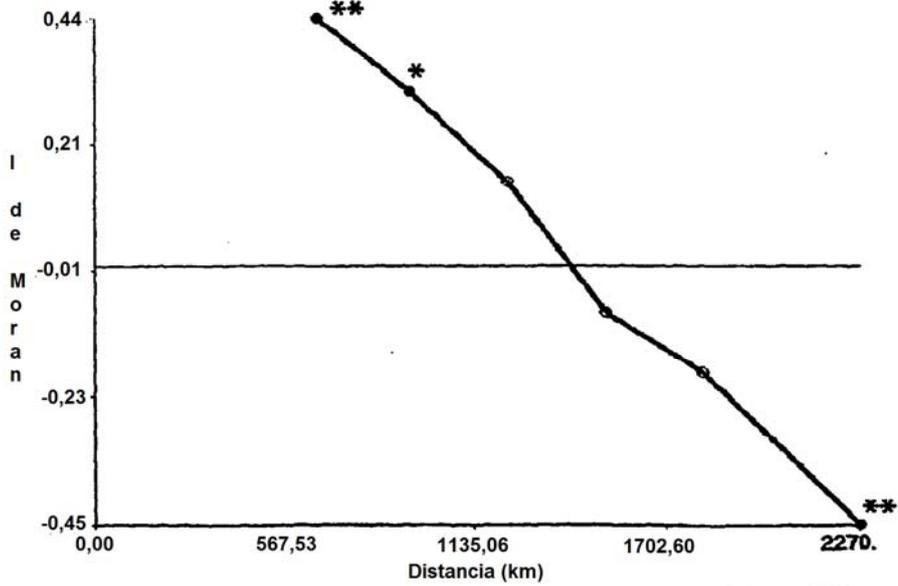


Fig. 6B