

(12) SOLICITUD INTERNACIONAL PUBLICADA EN VIRTUD DEL TRATADO DE COOPERACIÓN EN MATERIA DE PATENTES (PCT)

(19) Organización Mundial de la
Propiedad Intelectual
Oficina internacional



(43) Fecha de publicación internacional
23 de julio de 2020 (23.07.2020)

WIPO | PCT

(10) Número de publicación internacional
WO 2020/148475 A2

(51) Clasificación internacional de patentes:
C12Q 1/68 (2018.01)

(21) Número de la solicitud internacional:
PCT/ES2020/070022

(22) Fecha de presentación internacional:
16 de enero de 2020 (16.01.2020)

(25) Idioma de presentación: español

(26) Idioma de publicación: español

(30) Datos relativos a la prioridad:
P201930027 17 de enero de 2019 (17.01.2019) ES

(71) Solicitantes: **SERVICIO ANDALUZ DE SALUD** [ES/ES]; Avda. de la Constitución, 18, 41071 Sevilla (ES). **UNIVERSIDAD DE SEVILLA** [ES/ES]; Pabellón de Brasil, P. de las Delicias; s/n Sevilla, 41013 Sevilla (ES).

(72) Inventores: **MONTESEIRÍN MATEO, Francisco Javier**; Servicio Andaluz de Salud, Avda. de la Constitución, 18, 41071 Sevilla (ES). **CORNEJO GARCÍA, José Antonio**; Servicio Andaluz de Salud, Avda. de la Constitución, 18, 41071 Sevilla (ES). **BLANCA GÓMEZ, Miguel**; Servicio Andaluz de Salud, Avda. de la Constitución, 18, 41071 Sevilla (ES). **VEGA RIOJA, Antonio**; Servicio Andaluz de Salud, Avda. de la Constitución, 18, 41071 Sevilla (ES). **BELLIDO LINARES, Virginia**; Servicio Andaluz de Salud, Avda. de la Constitución, 18, 41071 Sevilla (ES). **CHACÓN FERNÁNDEZ, Pedro**; Universidad de Sevilla, Pabellón de Brasil, P. de las Delicias; s/n, 41013 Sevilla (ES).

(74) Mandatario: **GUSTAVO FUSTER**; Paseo de la Castellana 140, Planta 3 Edificio Lima, 28046 Madrid (ES).

(81) Estados designados (a menos que se indique otra cosa, para toda clase de protección nacional admisible): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Estados designados (a menos que se indique otra cosa, para toda clase de protección regional admisible): ARIPO

(BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), euroasiática (AM, AZ, BY, KG, KZ, RU, TJ, TM), europea (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Publicada:

— sin informe de búsqueda internacional, será publicada nuevamente cuando se reciba dicho informe (Regla 48.2(g))

(54) Title: METHODS FOR THE DIAGNOSIS OF ATOPIC PATIENTS SENSITIVE TO THE ALLERGENIC COMPONENTS OF POLLEN FROM OLEA EUROPAEA (OLIVE)

(54) Título: MÉTODOS PARA EL DIAGNÓSTICO DE ENFERMOS ATÓPICOS SENSIBLES A COMPONENTES ALERGÉNICOS DEL POLEN DE OLEA EUROPAEA (OLIVO)

(57) Abstract: Biomarkers and method for the diagnosis, stratification, follow-up and prognosis of olive pollen allergy disorder, kit or device, and uses.

(57) Resumen: Biomarcadores y método para el diagnóstico, estratificación, seguimiento y pronóstico de la evolución de la enfermedad alérgica a polen del olivo, kit o dispositivo y usos.



WO 2020/148475 A2

Métodos para el diagnóstico de enfermos atópicos sensibles a componentes alergénicos del polen de *Olea europaea* (olivo).

La presente invención se refiere al campo de la medicina de precisión, y más concretamente, al campo del diagnóstico molecular. La presente invención describe el uso de biomarcadores para el diagnóstico y pronóstico de individuos afectados o con riesgo de desarrollar susceptibilidad a alérgenos del polen de olivo, un kit o dispositivo y sus usos.

ESTADO DE LA TÉCNICA ANTERIOR

Las enfermedades alérgicas (EAs) han aumentado en las últimas décadas en todo el mundo, constituyendo una gran carga en cuanto a pérdida de calidad de vida se refiere, y suponiendo un gran coste económico para la familia y administraciones. En los años ochenta se observó una prevalencia de la enfermedad que alcanzaba el 25-30% de la población en países de Europa occidental. En diez años se calcula que aproximadamente el 50% de la población general, con un aumento del 2% anual, puede verse afectada por un problema alérgico. La organización Mundial de la Alergia (WAO) estima que 334 millones de personas en el mundo padecen asma alérgico y según la organización mundial de la salud (WHO) 250,000 personas mueren prematuramente cada año a causa de la misma.

La factura económica que originan las EAs crece a la par que su frecuencia. Los cálculos de 2013 estiman que en la Unión Europea los costes directos originados por el asma ascendieron a 20.000 millones de euros. A éstos hay que añadir los indirectos, estimados en unos 14.000 millones de euros y los generados por la pérdida de horas de asistencia a clases, horas de trabajo, molestias para los pacientes y familiares, etc., estimados en 38.000 millones de euros, lo que suma un total de 72.000 millones de euros. En España se estima que el coste anual del asma en 2009 fue de 1.480 millones de euros considerando la prevalencia basada en el diagnóstico sintomático asociado a la hiperreactividad bronquial, y de 3.022 millones de euros si se considera únicamente el diagnóstico sintomático del asma.

El olivo (*Olea europaea*) es uno de los cultivos de mayor importancia en el área Mediterránea, donde ha sido cultivado durante milenios. El cultivo extensivo de esta planta ha conducido a la aparición de un germoplasma extremadamente amplio y variado, con al menos 262 cultivares identificados solamente en España. Hasta muy recientemente, la caracterización de los cultivares de olivo se ha realizado primariamente de acuerdo a parámetros morfológicos y agronómicos. Sin embargo, está emergiendo la utilización de técnicas moleculares con la finalidad de hacer esta caracterización varietal mucho más fiable, rápida y sencilla.

La enfermedad alérgica a polen del olivo, en adelante EAPO, es una de las EAs más prevalente, dado que el polen de olivo es de los más importantes agentes causales de fenómenos de alergia estacional en humanos, más concretamente de las alergias de tipo I, en aquellos países en los que esta planta está ampliamente distribuida. El alérgeno mayoritario de este polen, denominado Ole e1, es una proteína de 20 kDa

En nuestro país, el 80% de todos los asmáticos lo son por causa alérgica, y la sensibilización a polen de Olivo, es la segunda causa de alergia respiratoria en España tras la sensibilización a las gramíneas. En Andalucía, es la causa más frecuente de alergia respiratoria siendo Jaén, Córdoba y Sevilla las provincias más afectadas, puesto que gran parte de su extensión está dedicada a su cultivo y el 75-90% por ciento de los alérgicos están sensibilizados al olivo.

Un alérgeno es una sustancia no propia del organismo capaz de inducir una respuesta inmunológica mediada por IgE. Podemos distinguir entre alérgenos mayoritarios, aquellos frente a los cuales presentan IgE específica más del 50% de los enfermos alérgicos, y los alérgenos minoritarios, que son reconocidos por menos del 50% de los pacientes alérgicos.

En los últimos cien años el diagnóstico y el tratamiento de las enfermedades alérgicas se han basado en el empleo de extractos alérgicos provenientes de fuentes naturales. Aunque se han conseguido grandes avances en la calidad y estandarización de los extractos alérgicos, éstos siguen siendo aún una fuente heterogénea de macromoléculas y su composición y cantidad de proteínas pueden variar notablemente entre los diferentes lotes. Así, algunos alérgenos mayoritarios pueden estar escasamente representados y algunos minoritarios pueden encontrarse en exceso, con las repercusiones diagnósticas y terapéuticas que ello conlleva. Por otra parte, los extractos naturales permiten únicamente identificar la fuente alérgica, pero no las moléculas concretas a las que un paciente se encuentra sensibilizado.

Actualmente los extractos naturales están siendo sustituidos por extractos recombinantes, en los que cada componente puede ser producido, tras la clonación de los mismos. Sin embargo no todos los componentes de cada extracto alérgico se han podido clonar y lo más importante es que no siempre los alérgenos recombinantes mimetizan el modo de acción de los naturales.

En los últimos 18 años se han descrito y caracterizado 13 alérgenos del polen de olivo, denominados de Ole e1a Ole e13. La respuesta IgE frente al polen de olivo es extremadamente compleja, debido a las diferentes frecuencias de sensibilización (o prevalencias de IgE) entre los alérgenos caracterizados, a la cosensibilización en un mismo individuo a los diferentes alérgenos y al impacto que ciertos factores ambientales tienen en

la prevalencia de IgE. A modo de ejemplo, indicaremos que, algunos pacientes como por ejemplo los sensibilizados al Ole e2, presentan los cuadros más severos de asma. Por otra parte, los pacientes sensibilizados a Ole e7 y Ole e9 son menos tolerantes a las dosis habituales de inmunoterapia con extractos de polen de olivo.

- 5 Debido a ello, resulta por tanto imprescindible el conocimiento de nuevos avances en el diagnóstico y tratamiento del asma en individuos con EAPO, lo que puede tener grandes repercusiones en su correcto diagnóstico, así como en la prevención y promoción de la salud de los ciudadanos.

10 Dada la complejidad y heterogeneidad del asma y las EAs, se hace necesario estudiar las mismas desde una aproximación global a distintos niveles; a nivel molecular, nivel celular y nivel fisiológico, acoplados a un análisis bioinformático y computacional que identifique nuevos genes funcionales, elementos reguladores y metabolitos, en el contexto de redes moleculares que definen procesos biológicos asociados con sistemas fisiológicos.

15 Esta aproximación global, también conocida como biología de sistemas, permite modelar la multitud de conexiones e interdependencias entre la predisposición genética, perturbaciones ambientales e intermediarios moleculares y reguladores, que en última instancia conducen a la gran diversidad de fenotipos y diferentes respuestas a los tratamientos entre los individuos alérgicos.

20 Desde la finalización del proyecto genoma humano, la tecnología de microarrays ha hecho posible el registro del genotipo en millones de variantes genéticas de forma simultánea. Estas variaciones a lo largo del genoma, como polimorfismos de un único nucleótido (SNPs), ha posibilitado el establecimiento de asociaciones estadísticas entre genotipo y fenotipo (genome-wide association studies, GWAS), que identifican loci genéticos y factores de riesgo para el asma y las EAs. La mayoría de los GWAS han sido realizados en asma, 25 donde se ha establecido una asociación con el locus 17q21. Este locus comprende 4 genes (ORMDL3, GSDMB, ZBP2, y IKZF3), relacionados con la homeostasis del calcio en el retículo y con el plegamiento proteico. Otras asociaciones han relacionado el asma con genes que participan en el reclutamiento de células inflamatorias (TSLP, IL33, IL1RL), activación y diferenciación de células T (genes HLA, IL2RB), proteasas extracelulares 30 (ADAM33, DPP10) o proteínas desconocidas (PHF11) . Pero algunos de estos estudios desembocan en resultados contradictorios; por ejemplo ADAM33 se encuentra asociado al asma y a la hiperreactividad bronquial en algunas poblaciones pero no en otras . Estas discrepancias pueden deberse a que hasta ahora se ha tratado de identificar un "perfil genético universal" común a todos los pacientes alérgicos, para lo que se ha estado 35 estudiando una población demasiado heterogénea de individuos. Se necesitan, por tanto,

estudios de procesos clínicos y poblacionales mucho más concretos.

Existe, por tanto, la necesidad de disponer de biomarcadores que permitan el diagnóstico de EAPO y de susceptibilidad para desarrollar esta enfermedad en la población general, así como su clasificación o estratificación en los diferentes estadios o niveles determinados, el pronóstico de la evolución de la enfermedad y la evaluación de su respuesta a un determinado tratamiento, mediante un método sencillo, eficaz y no invasivo.

DESCRIPCIÓN DE LAS FIGURAS

10 **Fig. 1.** Correlación entre las variantes significativas de interacciones SNPs siguiendo un modelo codominante.

Fig. 2. Correlación entre las variantes significativas de interacciones SNPs siguiendo un modelo dominante.

15 **Fig. 3.** Correlación entre las variantes significativas de interacciones SNPs siguiendo un modelo recesivo.

Fig. 4. Representación Manhattan. Seleccione los marcadores mas significativos en cromosoma humano.

20 EXPLICACIÓN DE LA INVENCION

Los autores de la presente invención han desarrollado una herramienta para el **diagnóstico temprano, estratificación seguimiento, pronóstico, prevención y/o tratamiento** de individuos afectados o con riesgo a desarrollar enfermedades alérgicas relacionadas con el polen de olivo, de ahora en adelante EAPO.

La presente invención es el resultado de la búsqueda de nuevos polimorfismos de un único nucleótido asociados con EAPO y evaluados mediante GWAS, lo cual ha permitido la identificación de biomarcadores y dianas efectivas para esta enfermedad que permiten: 1) **discriminar entre individuos** sanos, pacientes diagnosticados de EAPO y personas con alto riesgo de desarrollar EAPO, 2) **estratificar clínicamente los individuos** en base a parámetros clínicos y biomarcadores específicos en muestras biológicas complejas, 3) **facilitar el seguimiento clínico de estos individuos**, permitiendo pronosticar clínicamente

a individuos diagnosticados de EAPO y contribuyendo al desarrollo de una medicina personalizada; 4) prevenir la aparición de EAPO en individuos previamente identificados como susceptibles o en riesgo de padecer la enfermedad y 5) tratar a los individuos diagnosticados con EAPO.

- 5 Como parte de la divulgación, los inventores proponen la **evaluación estandarizada de un panel de biomarcadores**, en este caso relacionados con EAPO, en muestras biológicas complejas de individuos, p.ej. muestras sanguíneas. Esta información permitiría, por ejemplo, el desarrollo de un chip o array para la estratificación de individuos en diversos grupos de riesgo y la definición de sus necesidades específicas de seguimiento clínico y
- 10 manejo terapéutico.

Por tanto, **un primer aspecto** de la invención se refiere al uso de, al menos, uno de los genes *SCP2*, *CTNND2* y/o *ZNF252P* o cualquiera de sus combinaciones, para diagnosticar, clasificar, predecir o pronosticar si un individuo ha desarrollado o es susceptible de desarrollar EAPO.

- 15 A no ser que se indique lo contrario, cuando se utiliza en el presente documento el término "*SCP2*", (también llamado en literatura *NLTP*, *SCPX*, *SCP-2*, *SCP-X*, *NSL-TP* y *SCP-CHI*) se refiere al gen Sterol Carrier Protein 2 [*Homo sapiens*], gen que codifica dos proteínas: la proteína portadora de esteroles X (SCPx) y la proteína transportadora de esteroles 2 (*SCP2*), como resultado del inicio de la transcripción de 2 promotores regulados
- 20 independientemente. La transcripción iniciada desde el promotor proximal codifica la proteína SCPx más larga, y la transcripción iniciada desde el promotor distal codifica la proteína *SCP2* más corta, con las 2 proteínas compartiendo un extremo C-terminal común. La evidencia sugiere que la proteína SCPx es una tiolasa asociada con el peroxisoma que está involucrada en la oxidación de los ácidos grasos de cadena ramificada, mientras que la
- 25 proteína *SCP2* se cree que es una proteína de transferencia lipídica intracelular. Este gen está altamente expresado en órganos involucrados en el metabolismo de los lípidos. El splicing alternativo de este gen produce múltiples variantes de transcripción, algunas codifican diferentes isoformas.

- En el contexto de la presente invención, "*SCP2*" se define también por una secuencia de
- 30 nucleótidos o polinucleótido, que constituye la secuencia codificante de la proteína "*SCP2*" y "*SCPx*", y que comprendería diversas variantes procedentes de:

a) moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica de la SEQ ID NO: 1,

- b) moléculas de ácido nucleico cuya cadena complementaria hibrida con la
- 35 secuencia polinucleotídica de a),

c) moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,

d) moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica con una identidad de al menos un 80%, un 90%, un 95%, un 98% o un 99% con la SEQ ID NO: 1. en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad y las características estructurales de la proteína SCP2. Entre otras posibles secuencias nucleotídicas que codifican SCP2 se encuentra, pero sin limitarse, la SEQ ID NO:2.

Gene ID: 6342

Localización del gen: 1p32.3

SEQ ID NO: 1

```
GCGCCCCACCCCGCCCCCTCCTCCTGCCCCTACGCACGCGCACTCTCACGCGCCTGT
GTTGCCGCCCCGCGGCCCTGGCTTCGGGCTTCAGGGAGCTCTGGTGCAGTCTCCGCCT
GTCAGTGCCGGCAGTCGTCCGCGGCGCCCCGCCCCGGTCCCGCACTGGTGCAGCCATG
15 TCCTCTTCCCCGTGGGAGCCTGCGACCCTGCGCCGGGTGTTTCGTGGTGGGGGTTGGC
ATGACCAAGTTTGTGAAGCCTGGAGCTGAGAATTCAAGAGACTACCCTGACTTGGCAGA
AGAAGCAGGCAAGAAGGCTTTAGCTGATGCACAGATCCCTTATTCAGCAGTGGACCAG
GCATGTGTTGGCTATGTTTTTGGTGA CTCTACCTGTGGGCAGAGGGCTATCTATCACAG
TTTGGGAATGACTGGAATTCCTATAATCAATGTCAACAATAACTGTGCTACTGGTTCTAC
20 TGCTTTGTTTATGGCCCGCCAGCTGATTCAGGGTGGTGTGGCAGAATGTGTCTTGGCTC
TTGGGTTTGAGAAGATGAGTAAGGGAAGCCTTGGAATAAAATTTTCAGATAGAACCATTG
CCACTGATAAGCATGTTGACCTCCTGATCAATAAGTATGGATTGTCTGCTCACCCAGTTG
CTCCTCAGATGTTTGGGTATGCTGGAAAAGAACATATGGAAAAATATGGAACAAAATTG
AACACTTTGCAAAAATTGGATGGAAAAATCATAAACATTCAGTTAATAACCCGTATTCCCA
25 GTTCCAAGATGAATACAGTTTAGATGAAGTGATGGCATCTAAAGAAGTTTTTGTATTTTTG
ACTATCTTACAATGTTGTCCCACTTCAGATGGTGCTGCAGCAGCAATTTTGGCCAGTGAA
GCATTTGTACAGAAGTATGGCCTGCAATCCAAAGCTGTGGAAATTTTGGCACAAGAAAT
GATGACTGATTTGCCAAGCTCGTTTGAAGAAAAAAGCATTATTAAATGGTTGGCTTTGA
TATGAGTAAAGAAGCTGCAAGAAAATGCTATGAGAAATCTGGCCTGACACCAAATGATAT
30 TGACGTAATAGAACTTCACGATTGCTTTTCTACCAACGAACTCCTTACTTATGAAGCACT
GGGACTCTGTCCAGAAGGACAAGGTGCAACGCTGGTTGATAGAGGAGATAATACATAT
GGAGGAAAGTGGGTCATAAATCCTAGTGGTGGACTGATTTCAAAGGGACACCCACTAG
GCGCTACAGGTCTTGCTCAGTGTGCAGAACTCTGCTGGCAGCTGAGAGGGGAAGCCG
GAAAGAGGCAAGTTCCTGGTGCAAAGGTGGCTCTGCAGCATAATTTAGGCATTGGAGG
35 AGCTGTGGTTGTAACACTCTACAAGATGGGTTTTCCGGAAGCCGCCAGTTCTTTTAGAA
```

CTCATCAAATTGAAGCTGTTCCAACCAGCTCTGCAAGTGATGGATTTAAGGCAAATCTTG
TTTTTAAGGAGATTGAGAAGAACTTGAAGAGGAAGGGGAACAGTTTGTGAAGAAAATC
GGTGGTATTTTTGCCTTCAAGGTGAAAGATGGCCCTGGGGGTAAAGAGGCCACCTGGG
TGGTGGATGTGAAGAATGGCAAAGGATCAGTGCTTCCTAACTCAGATAAGAAGGCTGAC
5 TGCACAATCACAATGGCTGACTCAGACTTCCTGGCTTTAATGACTGGTAAAATGAATCCT
CAGTCGGCCTTCTTTCAAGGCAAATTGAAAATCACTGGCAACATGGGTCTCGCTATGAA
GTTACAAAATCTTCAGCTTCAGCCAGGCAACGCTAAGCTCTGAAGAACTCCCTTTGGCT
ACTTTTGAAAATCAAGATGAGATATATAGATATATATCCATACATTTTATTGTCAGAATTTA
GACTGAAACTACACATTGGCAAATAGCGTGGGATAGATTTGTTTCTTAATGGGTGTGAC
10 CAATCCTGTTTTCTCTATGCTCTGGGTGAATAGAGCCTGATGGTATACTACTGCTTTGCG
GAATTGCATACAACTGTGCATTACAAAGTTAATATGGTAATTATGGTCTGGGGTAAAATT
GAGTTTCAGAATAAAATTAGGAACAGTAAATCCAAAGAACTATGTAAACAAAAAAGCTT
TTGTTTTGCTTACAAAGTATATTTAAGGATTATTCTGCTGAAGATTCAGTTTAAGAGTTTT
CCTTGGGAGAACTAAGTAAGAAACACAATGCCAACAGCTGGCCAGTAATTAGTGTGTG
15 CACTTCATGTCATTAATCAATTTCTCAATAGTTCTTAAAATTAGTGAGATTAAAAATCTAAA
AATTTTGCATTTTATGCTATCAGAAACAGTATTTTCTTCCCAAATCAAATAAAAGAAATA
TGATCAGAGCTTGAACACAGGCTTATTTTTAAAATAAAAATATTTTAAACATGGGTTTCCT
TATTGAAAAATCAGTGTATTAGTCATAAAACACCATCATTAAGAATAATTGAACAATAAAG
TTTGCTTTCAGATGCAGTTTTCAAATTATAATCTCATTTCAATTTATAACGTTCTCAGTCCT
20 TTGTTATAATTTTCTTTTTTCTATGTAAGTTTAATTATCTGCATTTATCTTTTTTCTAGTTTT
TCTAATACTAATGTTATTTCTTAAATTCAGTGAGATATAGGATAAAATAATGCTTTGAGA
AGAATGTTTAATAGAAAATTAAATAACTTTTTCTGGCCTCTCTT

SEQ ID NO: 2

MS SSPWEPATLRRVFVVGVM TKFVKPGAENSRDYPDLAEEAGKKALADAQIPYSAVDQA
25 CVGYVFGDSTCGQRAIYHSLGMTGIPIINVNNNCATGSTALFMARQLIQGGVAECVLALGFE
KMSKGS LGIKFS DRTIPTDKHVDLLINKYGLSAHPVAPQMFGYAGKEHMEKYGTKIEHFAKIG
WKNHKHSVNNPYSQFQDEYSLDEV MASKEVDFLTILQCCPTSDGAAAAILASEAFVQKYG
LQSKAVEILAQEMMTDLPSSFEEKSIIKMVGFDMSKEAARKCYEKSGLTPNDIDVIELHDCFS
TNELLTYEALGLCPEGQGATLVDRGDNTYGGKWVINPSGGLISKGHPLGATGLAQCAELCW
30 QLRGEAGKRQVPGAKVALQHN LGIGGAVVVTLYKMGFPEAASSFRTHQIEAVPTSSASDGF
KANLVFKEIEKKLEEEGEQFVKKIGGIFAFKVKDGP GGKEATWVVDVKNGKGSVLPNSDKKA
DCTITMADSDFLALMTGKMNPQSAFFQGKLGITGNMGLAMKLQNLQLQPGNAKL

Asimismo, a no ser que se indique lo contrario, cuando se utiliza en el presente documento
35 el término “CTNND2”, (también llamado en literatura GT24, NPRAP, CTNND-2 y GT-24) se

refiere al gen Catenin Delta 2 [*homo sapiens*], gen que codifica una proteína asociada a la unión adhesiva de la superfamilia armadillo / beta-catenina y está implicada en el desarrollo del cerebro y los ojos y la formación de cáncer. La proteína codificada por este gen promueve la ruptura de la unión adherentes a base de E-cadherina para favorecer la propagación de las células tras la estimulación por el factor de crecimiento de hepatocitos. Este gen reside en una región del brazo corto del cromosoma 5. El splicing alternativo resulta en múltiples variantes de transcripción que codifican diferentes isoformas.

En el contexto de la presente invención, "CTNND2" se define también por una secuencia de nucleótidos o polinucleótido, que constituye la secuencia codificante de la proteína "CTNND2", y que comprendería diversas variantes procedentes de:

a) moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica de la SEQ ID NO: 3,

b) moléculas de ácido nucleico cuya cadena complementaria hibrida con la secuencia polinucleotídica de a),

c) moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,

d) moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica con una identidad de al menos un 80%, un 90%, un 95%, un 98% o un 99% con la SEQ ID NO: 1. en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad y las características estructurales de la proteína CTNND2. Entre otras posibles secuencias nucleotídicas que codifican CTNND2 se encuentra, pero sin limitarse, la SEQ ID NO:4.

Gene ID: 1501

Localización del gen: 5p15.2

SEQ ID NO: 3

GCGTCGGGAGCCGCCTCTCGCCCGCGGCGCTCGCCCCTGCCGCCCCGCCAGCATCCCT
TGTCCCGCGGCCGCGCTCAGACAACAAAAGCGGAAGATGCTGCAGTTGGGCAAGGTCA
GGACCTTGCCCTGAAGCCGGGCGGCGCCGCGCACGCCTCTTCCCGGACTGAGGAGCT
GTCGCCGCGGAGGGTGCATGTTTGCGAGGAAGCCGCCGGGCGCCGCGCCTTTGGG
AGCTATGCCTGTTCCAGACCAGCCTTCATCAGCCTCAGAGAAGACGAGTTCCCTGAGCC
CCGGCTTAAACACCTCCAACGGGGATGGCTCTGAAACAGAAACCACCTCTGCCATCCTC
GCCTCAGTCAAAGAACAGGAATTACAGTTTGAAAGGCTGACCCGAGAGCTGGAGGCTG

AACGGCAGATCGTAGCCAGCCAGCTGGAGCGATGCAAGCTCGGATCCGAGACTGGCA
GCATGAGCAGCATGAGTTCAGCAGAAGAGCAGTTTCAGTGGCAGTCACAAGATGGTCA
AAAAGATATCGAAGATGAGCTTACAACAGGTCTCGAGCTGGTGGACTCCTGTATTAGGT
CACTACAGGAATCAGGAATACTTGACCCACAGGATTATTCTACAGGTGAAAGGCCCAGC
5 CTGCTCTCCCAGAGTGCACTTCAGCTCAATTCCAAACCTGAAGGGTCTTTCCAGTATCC
GGCCAGCTACCATAGCAACCAGACCCTGGCCCTGGGGGAAACCACCCCTTCACAGCTC
CCGGCCCCGAGGCACACAAGCCCCGAGCTACGGGCCAGAGCTTCAGCCAGCTGGTTCCC
GAGCCTCATACAGCAGCCAGCATGGGCACCTGGGCCCAGAGTTGCGGGCCCTGCAGT
CCCCAGAACACCACATAGATCCCATCTATGAAGACCGCGTCTATCAGAAGCCCCCTATG
10 AGGAGTCTCAGCCAGAGCCAGGGGGACCCTCTGCCGCCAGCACACACCGGGCACCTAC
CGCACGAGCACAGCCCCATCTTCCCCTGGTGTGCGACTCCGTCCCCTTGACGCGCACAG
GCAGCCAGCACGGCCCCACAGAATGCCGCCGCGGCCACCTTCAGAGGGGCCAGCTATG
CCGCCGGCCCAGCCTCCAATTACGCGGACCCCTACCGACAGCTGCAGTATTGTCCCTC
TGTTGAGTCTCCATACAGCAAATCCGGCCCTGCTCTCCCGCCTGAAGGCACCTTGGCC
15 AGGTCCCCGTCCATTGATAGCATTGAGAAAGATCCCAGAGAATTTGGATGGAGAGACCC
GGAAGTCCCGGAAGTGATTGAGATGTTGCAGCACCAAGTTTCCCTCGGTCCAGTCTAACC
CGGCAGCCTACTTGCAACACCTCTGTTTTGGAGACAACAAAATTAAAGCCGAGATAAGG
AGACAAGGAGGCATCCAGCTCCTGGTGGACCTGTTGGATCATCGGATGACCGAAGTCC
ACCGTAGTGCCTGTGGAGCTCTGAGAAACCTGGTGTATGGGAAGGCCAACGATGATAA
20 CAAAATTGCCCTGAAAACTGTGGTGGCATCCCAGCACTGGTGAGGTTACTCCGCAAGA
CGACTGACCTGGAGATCCGGGAGCTGGTCACAGGAGTCCTTTGGAACCTCTCCTCATG
CGATGCACTCAAAATGCCAATCATCCAGGATGCCCTAGCAGTACTGACCAACGCGGTGA
TTATCCCCCACTCAGGCTGGGAAAATTCGCCTCTTCAGGATGATCGGAAAATACAGCTG
CATTGATCACAGGTGCTGCGTAACGCCACCGGGTGCCTAAGGAATGTTAGTTCGGCCG
25 GAGAGGAGGCCCGCAGAAGGATGAGAGAGTGTGATGGGCTTACGGATGCCTTGCTGTA
CGTGATCCAGTCTGCGCTGGGGAGCAGTGAGATCGATAGCAAGACCGTTGAAAAGTGT
GTGTGCATTTTAAGGAACCTCTCGTACCGGCTGGCGGCAGAAACGTCTCAGGGACAGC
ACATGGGCACGGACGAGCTGGACGGGCTACTCTGTGGCGAGGCCAATGGCAAGGATG
CTGAGAGCTCTGGGTGCTGGGGCAAGAAGAAGAAAAAGAAATCCCAAGATCAGTG
30 GGATGGAGTAGGACCTCTTCAGACTGTGCTGAACCACCAAAGGGATCCAGATGCTG
TGGCACCCATCAATAGTCAAACCCTACCTCACACTGCTCTCTGAGTGCTCAAATCCAGA
CACGCTGGAAGGGGGCGGCAGGCGCCCTGCAGAACTTGGCTGCAGGGAGCTGGAAGTG
GTCAGTATATATCCGAGCCGCTGTCCGAAAAGAGAAAGGCCTGCCCATCCTCGTGGAG
CTGCTCCGAATAGACAATGACCGTGTGGTGTGCGCGGTGGCCACTGCGCTGCGGAACA
35 TGGCCTTGGACGTCAGAAATAAGGAGCTCATCGGCAAATACGCCATGCGAGACCTAGT
CCACAGGCTTCCAGGAGGGAACAACAGCAACAACACTGCAAGCAAGGCCATGTGGAT
GACACAGTGACAGCTGTCTGCTGCACACTGCACGAAGTGATTACCAAGAACATGGAGAA

CGCCAAGGCCTTACGGGATGCCGGTGGCATCGAGAAGTTGGTCGGCATCTCCAAAAGC
AAAGGAGATAAACACTCTCCAAAAGTGGTCAAGGCTGCATCTCAGGTCCTCAACAGCAT
GTGGCAGTACCGAGATCTGAGGAGTCTCTACAAAAGGATGGATGGTCACAATACCACT
TTGTAGCCTCGTCTTCAACCATCGAGAGGGGACCGGCAAAGGCCCTACTCCTCCTCCCG
5 CACGCCCTCCATCTCCCCTGTGCGCGTGTCTCCCAACAACCGCTCAGCAAGTGCCCCA
GCTTCACCTCGGGAAATGATCAGCCTCAAAGAAAGGAAAACAGACTACGAGTGCACCG
GCAGCAACGCCACCTACCACGGAGCTAAAGGCGAACACACTTCCAGGAAAGATGCCAT
GACAGCTCAAAACACTGGAATTTCAACTTTGTATAGGAATTCTTATGGTGCGCCCGCTGA
AGACATCAAACACAACCAGGTTTCAGCACAGCCAGTCCCACAGGAGCCCAGCAGAAAA
10 GATTACGAGACCTACCAGCCATTTCAGAATTCACAAGAAATTACGATGAGTCCTTCTTC
GAGGACCAGGTCCACCATCGCCCTCCCGCCAGCGAGTACACCATGCACCTGGGTCTCA
AGTCCACCGGCAACTACGTTGACTTCTACTCAGCTGCCCGTCCCTACAGTGAAGTGAAC
TATGAAACGAGCCACTACCCGGCCTCCCCGACTCCTGGGTGTGAGGAGCAGGGCACA
GGCGCTCCGGGAACAGTGCATGTGCATGCATACCACAAGACATTTCTTTCTGTTTTGTT
15 TTTTCTCCTGCAAATTTAGTTTGTAAAGCCTGTTCCATAGGAAGGCTGTGATAACCAGT
AAGGAAATATTAAGAGCTATTTTAGAAAGCTAAATGAATCGCAAGTTAACTTGGAATCA
GTAGAAAGCTAAAGTGATCCTAAATATGACAGTGGGCAGCACCTTTCTAGCGTGAGCTG
TAGAGTAACGAGAAGTGCTTTATACTGAACGTGGTTGATGGGAGGAGAGACGAGGCATT
CGGGCCCGGTGGGGCGTAAGGGTTATCGTTAAGCACAAGACACAGAATAGTTTACACAC
20 TGTGTGGGGGACGGCTTCTCACGCTTTGTTTACTCTCTTCATCCGTTGTGACTCTAGGC
TTCAGGTTGCATTGGGGTTCCTCTGTACAGCAAGATGTTTCTTGCCTTTTGTTAATGCAT
TGTTGTAAAGTATTTGATGTACATTACAGATTAAGAAGAAAAGCGCGTTGTGTATATTAC
ACCAATGCCGCCGTGTTTCCTCATCTATGGTTCTAAATATTGCTTCAATTTCAAACTTTTG
AAAGATGTATGGATTTCCAGTTTTTCTTTACTTTCTCCAGTATGTTTTAACAAAAA
25 AAAAAAGCAGGAAAAAAGGAATATTTAGCAGTATTGTTGTTCTGATATGTGAATTTGTTT
GTGACAACTAAACAAGGCATTACAGCAGTTTCTGACAATTAACATACATCATTCCACACTC
CTTGTCAACAAAGTGCTTTTTCACTGCCTAAAATTTTAGATGTAGATATTTGAAATAGATT
TTTTCATTTATACCAGTTTTCTTTATGATGATACAGTGTTAAAGAAAATAAATTACAATTG
ATCTGTCATCCATATTTGCTAAGAATGTCTATTTCCAAGAACCTATATCTTACAAAATACA
30 CATTCTTGCTTGTGTGTCTGTGTGTGGGTGTATGTATGTCCACGTACTAGAAAAATTA
CCTGTGTGAGATTTTATTTTGACAGAAGCTATTTCAATTTGCTGTTTCATCTTGTACAACTT
TGTTTTTGTTTTTTAGTATAAATGTACATCAGTTTGTTCCTTTTGTAAGTGTATCTATTCTT
CTGACCATCTAGTGACTCAGGATATTAGGCCAGTTGAAGTGGAGTTATAGGATTTTCACA
CATTACGCATACTTGGATATCAACCCTCTCTACAGGTCCGTTTCCTTACACATTTTATC
35 CCCTCAGAACACGATAAACCATGGCCAATTCAGTTTCACTTTGGGGCCAATTACTTGAA
GTATCCACAGAAAATGTAACTGCTCGCCTTATCCCCCAAACTTTTTGTTCATGTGGCA
GGGAGCAGTACACAACCTATGGAATAGATTATTAAACTGAACTTGAAGCTAACGTTAGAA

TAGTGGTTAACAGAGCTAAGAAATAAGGCTAACCATCGAAAAGAACATGGGCCGCGGGT
AAACTAGCTTTGCTCTTTAGATGCAAGTGGCTGAGGCCAATAGGTGCCTAGTAAATGTTA
ACTGTTCTTGAAAAAATAAAAACATCAGTAACAAAA

SEQ ID NO:4

5 MRSLSQSQGDPLPPAHTGTYRTSTAPSSPGVDSVPLQRTGSQHGPQNAAAATFQRASYAA
GPASNYADPYRQLQYCPSVESPYSKSGPALPEGLTARSPSIDSIQKDPREFGWRDPELPE
VIQMLQHQPFSVQSNAAYLQHLCFGDNKIKAEIRRQGGIQLLDLLDHRMTEVHRSACGAL
RNLVYGKANDDNKIALKNCGGIPALVRLLRKTDDLEIRELVTGVLWNLSLSCDALKMPIIQDAL
AVLTNAVIIHSGWENSPLQDDRKIQLHSSQVLRNATGCLRVSSAGEEARRRMRECDGLT
10 DALLYVIQSALGSSEIDSKTVENCVCILRNLSYRLAAETSQGGQHMGTDELDGLLCGEANGKD
AESSGCWGKKKKKKKSQDQWDGVGPLPDCAEPPKGIQMLWHPISIVKPYLTLLSECSNPDT
LEGAAGALQNLAAGSWKWSVYIRAAVRKEKGLPILVELLRIDNDRVCAVATALRNMALDVR
NKELIGKYAMRDLVHRLPGGNNSNNTASKAMSDDTVTAVCCTLHEVITKNMENAKALRDAG
GIEKLVGISKSKGDKHSPKVKAASQVLNSMWQYRDLRSLYKKDGWSQYHFVASSSTIERD
15 RQRPYSSSRTPSISPVRVSPNNRSASAPSPREMISLKERKTDYECTGSNATYHGAKGEHT
SRKDAMTAQNTGISTLYRNSYGAPAEDIKHNVSAQPVQPQEPSRKDYETYQPFQNSTRNY
DESFEDQVHHRPPASEYTMHLGLKSTGNYVDFYSAARPYSELNYETSHYPASPD SWV

Igualmente, a no ser que se indique lo contrario, cuando se utiliza en el presente documento el término “*ZNF252P*”, (también llamado en literatura *ZNF252*, *ZNF-252*, *ZNF-252P* y
20 *ZNF252-P*) se refiere al gen Zinc Finger Protein 252 Pseudogene [*homo sapiens*], gen que pertenece a la familia de proteínas Zinc Finger o dedo de zinc (ZnF), una familia amplia y diversa de proteínas que comprenden un dominio “dedo zinc”, referido a las estructuras proteicas secundarias (hélice α y hoja β) que se mantienen juntas por un ion Zn, estos dominios que típicamente sirven para interactuar con ADN, ARN, proteínas o moléculas
25 pequeñas.

En el contexto de la presente invención, “*ZNF252P*” se define también por una secuencia de nucleótidos o polinucleótido, que constituye la secuencia codificante de la proteína “*ZNF252P*”, y que comprendería diversas variantes procedentes de:

- 30 a) moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica de la SEQ ID NO: 5,
- b) moléculas de ácido nucleico cuya cadena complementaria hibrida con la secuencia polinucleotídica de a),
- c) moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,

d) moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica con una identidad de al menos un 80%, un 90%, un 95%, un 98% o un 99% con la SEQ ID NO: 5. en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad y las características estructurales de la proteína ZNF252P. Entre otras posibles secuencias nucleotídicas que codifican ZNF252P se encuentra, pero sin limitarse, la SEQ ID NO:6.

Gene ID: 286101

Localización del gen: 8q24.3

SEQ ID NO: 5

```
10  AGGCATTTCCGGCCCAGGGGGCTTGGCTCCCCGCGGCACGGGATTTAGCGTTCGCGC
    TCCTTCCCTTCCCGTGGTTCGAGCCGAGTCCTGACCTGAGGGCTGCATCAAGATCTTGTC
    ATTCCACATCGTGGTTTCTTTGAGGATGTGGCTGTACCCCTCTCCAGGAGGAGTGGG
    ACTGTCTGATCCCTGCTCAGAGGGGGCCTCTACAAGGATGTGATGATGGGGACCTATGG
    GAACCTACTCTCATTAGTAGGTGAATGGTTAAGCAAAGTGTGGTACATCCATACCATGG
15  GATACGACTCAACAATCAAAGGAACTGCCCAGACTTCACCACGATGCAATATATGCAT
    GACTTCAAGCTTCCAAACCTGATGTCATCTCCAGGCTGGAGCGGGGGGACGAACCATA
    GACCCCTCACATCCTGAGAACTCAGGGGAGCTGGAGCTGGAGGCACAAGAGAGAAGG
    CTGTGATTCTAGGATTGAAAAGAGGAAATGATTCCAAAGCAGGACATTTCTGAAGAATT
    GGAATCCCAGAGAGCAAAATCAGAAGATCATGTAAGGAATATTTTAAAGGAACTGAAG
20  AGATGAGTAAAGTGAAGGAAAGTTAGAGAATTGCTGGAGAAAATATGCAGTAGAAGGA
    GTTAAGAACTCATTCTCCAGAAAGAGCAATTTAGAGAAATTACCATGAGGTATGTGAAA
    ACCCTCTCTAGAGAGAATGGCCAGAAGTTCAATGCTGTTGGAGAAAAGTGCATTACAGA
    CTCAAATCCTGCCAAACATCTTAGAGGGTCTAGAGAGGAGAGTCTCCATCCAAGTGTGT
    CAAGTGTAGAAAAGTTCGAACAACATGAGGACCTCATTAACCTCCAGAGTTTCCAATTAG
25  GAGAAAGAGCCTATCAGACAGATGTGTTAGTGAAAGTGCCCAGACAGAGCTCAGTTCTT
    AGTGAGAATCAGAGGATGAATAATCCAGAGAGATGGTTTGAGAGTACTGGGTGTGGAAA
    AACTTACAATCAGAACAGAGCTTTTAACCAGCACCAGAGATTTTCATAGTGGAGAGAAGA
    CCTATGAGCACAATGAATGTGGAAAAGCTTTTCAGTTGGCCCTCTATCCTCAGTAAACATC
    AGAGAATCCACACTGGTAAGAACTCTACACATGTGAGGATTGTGGCAAATCTTTTCAGT
30  GTTCACTCATACTTTATTTCAGCATTGTAAAATTCACACTAGAGAAAAACCTATGAGTGTA
    TTAAATGTGGGAAAGCTTTTAGTACACATTCATCTTATGTTCAACATCTAAAAATTCATAC
    AGGAGAGAAACATCATGAGTGTAATCAGTGTGGGAAAGCCTTTAGTCATAGCTCTAATC
    TAATTCATCATCAGAGAATTCATAGTGGAGAGAAACCTTACAAGTGCAAAGAATGTGGGA
    AAGCCTTCAACAGACAATCAAACCTTATTCAGCATCAGAGAATTCATTCTGGAGAGAAAC
35  CTTATGACTGTAAGGAGTGTGGCAAAGCCTTCAGTACACAATTATTTCTCATTTCAGCATC
```

AGAGAATTCATACAGGAGAAAAGCCCTATGAATGTAATGAATGTGCAAAATCCTTTAGCC
TGAACCGAACTCTTACTGTCCATCAGAGAATTCACACTGGAGAGAAACCTTATAGGTGTA
ATGAATGTGGGAAATCCTTTAGTCAATGCTCACAAGTTATTCAACATAAGAGAATTCACA
CTGGAGAGAAACCTTATATCTGCAATGAGTGTGGAAAATCATTTGGTGCTCGTCTATCCC
5 TTATCCAGCATCAGAGAATTCACACTGGAGAGAAACCTTATGGTTGTAGTGTGTGTGGG
AAAACCTTTAGTCAAAAGGGACATCTTATTAGCATCAGTGAATTCACACAGGAGAGAAA
CCCTATGAATGTAGTGAGTGTGGAAAAGCTTTCAGCCAGAGTTTTAATCTTATTCACCAT
CAAAGAACACACAATGGTGAGAAGTCCTATGAATGTAATGAATGTGATAAAGCCTTCAGT
TTGCTTTCTTCCCTTGTTCAACATCAGAGAATACATAATGGAGACAAACCCTATGAGTGT
10 CACAAATGTGGGAAGGCCTTTAGCCAGGGGTACACCTTATTAGCATCAGAGGAGTCA
CATTGGTGAGAAACCCTATGAGTGTAAATGAGTGTGGGAAAACCTTTGGGCAGATATCCA
CCCTAATTAAGCATGAGAGAACACACAATGGAGAGAAGCCCTATGAGTGCAGTGACTGT
GGGAAGGCCTTCAGCCAGAGTGACACCTTATCCACCATCAAAGAATTCACACTGGAGA
GAATCCCTATGAGTGCAGTGAATGTGGGAAGGCCTTCAATGTTTGTTCTCTCTCATTCA
15 GCATCACAGAATTCATACTGGTGAGAAACCTTATGAATGTAGTGACTGTGGCAAGGCGT
TCAGTCAGCATTACAATTTATCCAACATCAGAGAATTCACACTGGAGAGAAACCCTACA
TGTGCAATGAGTGTGAGAAATCCTTCAGTGCATGCTTATCCCTTATCCAACACAAGAGAA
TTCACACTGGAGAGAAACCCTATGTATGTGCCAAATGTGGAAAATCCTTCTGACAAAGCT
CTCACCTTATTCAACATCAGAGAATTCACAGTGGGGAGCAACCTCATACGTGTAATCGAT
20 GTGAAAAAACCTTCAGTTAGAGAATAACTCTTAGTAGTCATGAGAAAATCCATACAATTC
ACATTAGAGAGCAAGTCTATGAATGTAGTAAGTGTGGGGAACTCTAGCGCACAGTCATC
TTTCATTCAACATTGTACAGTTCACAGTGGAGAGTAACTGATTAGTCTATAAATTCTTCAC
AGTTATACATTATATGAGAGTCATACTCGTGAGAAATGAGTCACTTTATGGTTTTCTTTT
GGAGATTAGCTCTTAAGGGATTTCACTCTCCAGTGAGCTGCTAACTAAAGGCAGGCT
25 GCTGAAGTATATAGCAGCAATCCCACAATAGCACACCTTGCTTCAGATTCAGAGGATGA
GCCCCCTTCTTGAGCCCTGTAGATATGTGAACTTAATTATTTAGATGCATTTTCATGGCA
CCAGACCCTAAGGTACAGTTAGGTATTTCACTGGAGACCACTTACTTTGCTTTGCTTTGC
TTTTTTTTTTTTTTTTTTTTTGTATGATTTATTTCAACTTCAAAGAAAGTAGTTGAGATCCTTG
CACTTCCAAGTGTGATGGAACCAGGCTAGGGTTTTCTTTGTTTGTTGGGCCTCTTCAAA
30 ACGTTTTTATTCTGTTATAAGGAATATGATACCTTAAGCTTCAATTCACAGGGAAGAGAAA
CATGATTATAAGCAAAAGTTTGTCAAGGAATAAGGTGAGTCTTTGTGTAGATTAAATAG
TTGAGACCTTGCTAAAAAGTCAGATAACACTGTAATCCACCTCCAGGTTTCCAAGAGTTA
AGACGCTGCCAGCAAGGAGCTTGCAGTCTTGTAGCTGGCCCAATATATCCAATAATTC
AAAATGGGAAATTGTACATTTTACTTAAACAGGATTTGGTAGATGGGTCTTCTAGATATG
35 TTGTCATGAAGAAAAATCAGGAAAATGTCCAAGGGTGGGAGCCTGTGGCTTATGGCTTC
ACTTACCCATCACAGATAGCTATGTTACACTGCAAGAACAGAGTTCATGGCTTTTCTGA
GGTGGCTGGAAACCTTAATCTCCTGAAGTTAAGTCAAACTTAAATGTCAATTCACCCAG

CTCTCTTGCCTGCAGGCTGATCCAGGGACTTAGAGCTTGTTACCCAGTTACAACCTTCTT
CTAGAGAAGAACTTTTTAAGAAGGAAGCAAAAAGAAATTCAGTTTCCCCCAAGGATCTCTT
CTTTGAGGTTCCATCAAGAGATGACAAAAGGAGGCTGGGCGCGGTGTCTCACACCTGT
AATCCCAGCACTTTGGGAGGCTGAAGTGGGCAGATCACGAGGTCAAGAGATCAAGACC
5 ATCCTGGCCAACATGGTGAAACCACATTTCTACAAAATGCAAAAATTAGCTGGGCATGG
TGGTGCACGCCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAC
CCAGGAGGCGGAGGTAGCAGTGAGCCAAGATTGCACCACTGCATTCCAGCATGGCCAC
AAAGCAAGACTCCAAAAAATATATATATATATGACAAAAGGGTTAGGAAGCTTTTCCTCT
GAGGCCACGATTACACAACCAAATCAGAAAATAATATTTTTTTTGTTTTAAACCTGTTTA
10 CATAAGATGGAGGTGTTTACTTAACTAAGAACCTCGGTATTTACTCAAATCCGGTTTTGG
GTTTTATGGCTAAACATTTCTCCAGATTAACCTTATCAGTTTGGAATATGTTTTTATTTTC
CATTAAATATTTTCATCCAAATCATACTGAATTTTATTTATCATGGCTTTTAATGTTCTCTAT
TCAAGTCAAATTTCTGTTTCCTAATTACTTGGCTTGGTGATGGCATTAAATATTGGTTGGG
AGCTAGATGGTCTCTGCTCTATTTGATGATTGGCATAGAGCAATCATACACTTGGTACTT
15 CCTAGCATTTTTTTTTGAGACAGTATCATGCTGTATCATGTTGCACAGGGTGGAGTGCAG
TGGTGCAGTCATAGCTCACTGTAACCTTGAACCTCGGGCTCAAGTGATTCCCCTGCCT
CAACTTCATGAGTAGCAGGGACTACAGCTGTGTACCACCACACCGGCCTCTCCTGGCTT
CTTAACCACTTACATTAATAATTGAGAGGAGAAAGGCATTTTCAGTTTCTTTAGTTAATAAA
AAGAAGCCATTTCTGGAGGAGTTTTATGCCTGTACCAGCAGAGGTTCCAGCCTTCCAGGA
20 ATCTCATCATGATCCATACTGCTGACACAGGCCTTTGTACCTGAAGCATTCTTAAATA
AGGAGACTGACATTAAACAGGACAATTGTGAACTCCACTTTGTAAGCATCATACATATCT
TACAACCTATTCTGAAGACTCCTTTATTCTTGCTCTTCCCAGAGAGCTACTGTGTTTGT
CATGCTCACTGCACTCACAAGAGAAAAAGAGTAAGTGGCATATTTGCAGGTGTTGATGG
TTACAAGAGCATTTTTTACTTTTTCTCCATGTTTGTTCTGCAGCATTTCCATTGATGCATG
25 CTATTGTAGACTTCTTAGGTTGACCTTGGCCTTGACCTCCACCACTAGCTCATGACCCAA
AACTGAAGATGTGCATCAGAAAGGGACAGATTTGGTCTTTCTGCTGTAAAGATATTCTG
CCTTAGTGCCTGCTAATTTGTTGTGCATAGTCCTGTGCTGCCAATCTTCCCCACCAATTG
TTAAGGTTAATGTGTGTTTTATATGTGTATTACAAACATTAAATGTAATCAGTGGATTG
CAAAAAAAAAAAAAAAAAAAAAAAAAA

30 SEQ ID NO:6

MIPKQDISEELESQRAKSEDHVRNIFKETEEMSKTEGKLENCWRKYAVEGVKNSFSQKSNF
REITMRYVKTLRENGQKFNAVGENCITDSNPAKHLRGSREESLHPSVSSVENLQQHEDLIN
LQSFQLGERAYQTDVLVKVPRQSSVLSENQRMNNPERWFESTGCGKTYNQNRANQHQ
FHSGEKTYEHNECGKAFSWPSILSKHQRIHTGKKLYTCEDCGKSFSVHSYFIQHCKIHTREK
35 PYECIKCGKAFSTHSSYVQHLKIHTGEKHHECNQCGKAFSHSSNLIHHQRIHSGEKPYKCKE
CGKAFNRQSNLIHQRIHSGEKPYDCKEKGKAFSTQLFLIQHQRIHTGEKPYECNECAKSFS

LNRTLTVHQRIHTGEKPYRCNECGKSFSQCSQVIQHKRIHTGEKPYICNECGKSFGARLSLI
 QHQRIHTGEKPYGCREKPYECSECGKAFSQSFNLIHHQRTHNGEKSYECNECDKAFLSS
 LVQHQRIHNGDKPYECHKCGKAFSQGSHLIQHQRSHIGEKPYESNECGKTFGQISTLIKHER
 THNGEKPYECSDCGKAFSQSAHLIHHQRIHTGENPYECSECGKAFNVCSSLIQHHRIHTGE
 5 KPYECSDCGKAFSQHSQFIQHQRRIHTGEKPYMCNECEKSFSACLSLIQHKRIHTGEKPYVCA
 KCGKSF

Más preferiblemente se refiere al uso de los polimorfismos rs10437066[G/A] y
 rs7544995[A/G] del gen *SCP2*, los polimorfismos rs6867416[C/T], rs6554641[T/C] del gen
CTNND2, y los polimorfismos rs7817276[G/A], >rs2272715[T/C], rs6599548[C/T] y
 10 rs3866997[A/C] del gen *ZNF252P*.

En esta memoria se definen los siguientes polimorfismos

>rs6867416[C/T]

ACTTGGGCATTTGAATTGACGAAAATTGGA[C/T]ATAAAATTCTCTTCTGCCACATATTAG
 15 CTA

>rs6554641[T/C]

GGGCTTTTCCTCAATTGCAAATGAACTAAT[T/C]TGAATAGGCACAGAGATGTGTAAAGG
 CAAC
 20

ZNF252P

>rs7817276[G/A]

CTACAAGACTGCAAGCTCCTTGCTGGCAGC[G/A]TCTTAACCTTGGAAACCTGGAGGTG
 GATT
 25

>rs2272715[T/C]

CTATATTGGCTCATCTTGGTGTGCCCCAAC[T/C]CTGACCTCTACTGTACCTGGTCCAGG
 GACA

30 >rs6599548[C/T]

CAGGAATGAGAGACTTAGAAGAGAACGTCT[C/T]TCTAATTGGTATTTCCCAGCAGCCAA

AACA

>rs3866997[A/C]

TTCTCTGATGTTGAACAAGGGAAGAAAGCA[A/C]ACTGAAGGCTTTATCACATTCATTACA
5 TTC

SCP2

>rs10437066[G/A]

TATTTTCATTATGGATTTTAGATCTGAGACC_TATAAGGCATGGTTACAGCCATAATCACAA

10

>rs7544995[A/G]

CTTCACATAAAATGTAGAAGCCTTGCAACC_TATAGGTCCATTCACTGACTCCCCTTTGC
C

15 El primero es el alelo de referencia y el segundo alelo es el alternativo, es el alelo de susceptibilidad.

En el sentido utilizado en esta descripción, el término "variante" se refiere a una secuencia sustancialmente homóloga a la secuencia de los genes que contienen las variantes descritas. En general, una variante incluye adiciones, deleciones o sustituciones de
20 nucleótidos. Más preferiblemente, las variantes descritas hacen referencia al cambio o sustitución de un nucleótido de la secuencia en el genoma humano de referencia en las posiciones descritas.

El término "polimorfismo" se refiere a una variación en la secuencia de nucleótidos de un
25 ácido nucleico donde cada secuencia posible está presente en una proporción igual o superior a un 1% en una población; en un caso particular, cuando dicha variación es la secuencia nucleotídica ocurre en un sólo nucleótido (A, C, T o G) se denomina SNP.

Los términos "variante alélica" o "alelo" se usan indistintamente en la presente descripción y
30 se refieren a un polimorfismo que ocurre en un mismo locus en una misma población.

El término "variación génica" o "variante génica", tal y como se utiliza en la presente

descripción, incluye a las mutaciones, polimorfismos y variantes alélicas. Una variación o variante génica se encuentra entre individuos dentro de las poblaciones y entre las poblaciones dentro de las especies.

- 5 Existen más de 100 métodos distintos de genotipado de SNPs, siendo cualquiera aplicable en la presente invención. Por ejemplo, sin limitarnos, puede detectarse empleando desde enzimas de restricción o por secuenciación simple mediante PCR, sondas ASO e hibridación en Micromatrices de ADN o esferas. Así, por ejemplo, como métodos de amplificación de SNPs se encuentran la PCR, el método Invader, la amplificación por ADN de círculo
- 10 rodante, etc. Como métodos de discriminación alélica, pueden emplearse, pero sin limitarnos, hibridación específica de alelo, ligamiento de oligonucleótidos específicos de alelo, minisequenciación o extensión de sondas (SBE, Single Base Extension), rotura invasiva específica de alelo (Invader), etc. Según el formato de detección, podemos citar, también sin limitarnos, el uso de sondas Taqman, de molecular beacons, scorpion, etc.
- 15 Respecto a reacciones en un soporte sólido, tenemos, por ejemplo, biochips y microarrays de ADN, u otros métodos que emplean partículas, como por ejemplo, Luminex 100, BeadArray, etc. También existen métodos de detección masiva, como la pirosecuenciación, o la miniaturización.
- 20 Un **segundo aspecto** de la invención se refiere a un método de obtención de datos útiles para el diagnóstico, estratificación y pronóstico de la evolución de la EAPO en un individuo, de ahora en adelante primer método de la invención, que comprende:
- a) determinar los polimorfismos de un solo nucleótido rs10437066[G/A] y rs7544995[A/G] del gen *SCP2*, los polimorfismos rs6867416[C/T], rs6554641[T/C] del
- 25 gen *CTNND2*, y los polimorfismos rs7817276[G/A], >rs2272715[T/C], rs6599548[C/T] y rs3866997[A/C] del gen *ZNF252P* en una muestra biológica aislada, o cualquiera de sus combinaciones.

Un **tercer aspecto** de la invención se refiere a un método para el diagnóstico, estratificación y pronóstico de la evolución de EAPO en un individuo, de ahora en adelante

30 segundo método de la invención, que comprende que comprende el paso (a) del primer método de la invención, y adicionalmente:

- b) incluir a los individuos que presentan la variante A del rs10437066 y/o la variante G del rs7544995 del gen *SCP2*, la variante T del rs6867416 y/o la variante C del

rs6554641[T/C] del gen *CTNND2*, y/o la A del rs7817276 la variante C del rs2272715[T/C] y/o la variante T del rs6599548 y/o la variante rs3866997 del gen *ZNF252P*, o cualquiera de sus combinaciones, en el grupo de individuos con mayor probabilidad de desarrollar EAPO..

- 5 El término "muestra biológica" o "muestra biológica aislada", tal como aquí se utiliza, se refiere, pero no se limita, a células, tejidos y/o fluidos biológicos de un organismo, obtenidos mediante cualquier método conocido por un experto en la materia, que sirva para llevar a cabo cualquiera de los métodos proporcionados por la presente invención. Preferiblemente, la muestra biológica aislada es sangre periférica
- 10 El término "individuo", tal y como se utiliza en la descripción, se refiere a animales, preferiblemente mamíferos, y más preferiblemente, humanos. El término "individuo" en esta memoria, es sinónimo de "paciente", y no pretende ser limitativo en ningún aspecto, pudiendo ser éste de cualquier edad, sexo y condición física. Los términos "sujeto humano", "sujeto" y "paciente" se usan por tanto de manera indistinta en esta memoria descriptiva.
- 15 Una "secuencia de ácido nucleico" o "polinucleótido" puede comprender las cinco bases que aparecen biológicamente (adenina, guanina, timina, citosina y uracilo) y/o bases distintas de las cinco que aparecen biológicamente. Estas bases pueden servir para distintos propósitos, por ejemplo, para estabilizar o desestabilizar la hibridación; para estimular o inhibir la degradación de la sonda; o como puntos de unión para restos detectables o restos de
- 20 apantallamiento. Por ejemplo, un polinucleótido de la invención puede contener uno o más restos de base modificados, no estándar, derivatizados, incluyendo, pero sin limitarse a 4-acetilcitidina, 5-(carboxihidroximetil)uridina, 2'-O-metilcitidina, 5-carboximetilaminometil-2-tiouridina, 5-carboximetilaminometiluridina, dihidrouridina, 2'-O-metilpseudouridina, D-galactosilqueosina, 2'-O-metilguanosina, inosina, N6-isopenteniladenosina, 1-
- 25 metiladenosina, 1-metilpseudouridina, 1-metilguanosina, 1-metilinosina, 2,2-dimetilguanosina, 2-metiladenosina, 2-metilguanosina, 3-metilcitidina, 5-metilcitidina, N6-metiladenosina, 7-metilguanosina, 5-metilaminometiluridina, 5-metoxiaminometil-2-tiouridina, beta, D-manosilqueosina, 5-metoxycarbonilmetil-2-tiouridina, 5-metoxycarbonilmetiluridina, 5-metoxiuridina, 2-metiltio-N6-isopenteniladenosina, N-((9-beta-D-ribofuranosil-2-metiltiopurin-
- 30 6-ilo)carbamoil)treonina, N-((9-beta-D-ribofuranosil purin-6-ilo)N-metilcarbamoil)treonina, 5-metoxycarbonilmetoxiuridina, 5-carboximetoxiuridina, Wybutoxosina, Pseudouridina, Queosina, 2-tiocitidina, 5-metil-2-tiouridina, 2-tiouridina, 4-tiouridina, 5-metiluridina, N-((9-

beta-D-ribofuranosilpurin-6-ilo)-carbamoil)treonina, 2'-O-metil-5-metiluridina, 2'-O-metiluridina, Wybutosina, 3-(3-amino-3-carboxipropil)uridina.

En otra realización preferida, el primer método de la invención puede comprender un paso adicional, previo al paso (a), que comprenda un pretatamiento de la muestra biológica compleja.

A efectos de la presente invención el término "pretratamiento", referido a una muestra biológica compleja, incluye cualquier cambio o transformación que pueda sufrir la muestra desde que sea extraída del individuo o llegue al laboratorio hasta que tiene una forma medible con objeto de garantizar su utilidad y adecuación con respecto al análisis, al analito y a la muestra, en uno o mas etapas. Los ejemplos de esta manipulación incluye el cambio de estado físico, cambio de propiedades químicas (salinidad, pH, potencial redox,...) cambio de concentración (por dilución o concentración), procesos de eliminación o enmascaramiento del efecto de interferentes presentes en la matriz, cambio de la matriz, adición de componentes conservantes, técnicas de extracción del analito (p.ej. extracción de ADN, ARN o proteínas) o modificaciones en la dimensión o coloración de tejidos o muestras sólidas.

En una realización preferida, tanto el primer como el segundo método de la invención descritos anteriormente pueden ser total o parcialmente automatizados y/o computerizados, por ejemplo, por medio de un equipo robótico de dispensación de líquidos en el paso (a), un equipo robótico sensor para la detección de la cantidad en el paso (b), o la comparación computerizada en el paso (c) y/o (e).

Un "método diagnóstico" (o "método de diagnóstico") se refiere a un procedimiento donde se examina o analiza a un paciente (o a una muestra obtenida del mismo) para obtener un conocimiento diferencial de una enfermedad a través de los signos y síntomas que la caracterizan. Esta discriminación tal y como es entendida por un experto en la materia no pretende ser correcta en un 100% de las muestras analizadas. Sin embargo, requiere que una cantidad estadísticamente significativa de las muestras analizadas sean clasificadas correctamente. La cantidad que es estadísticamente significativa puede ser establecida por un experto en la materia mediante el uso de diferentes herramientas estadísticas, por ejemplo, pero sin limitarse, mediante la determinación de intervalos de confianza, determinación del valor p, test de Student o funciones discriminantes de Fisher. Preferiblemente, los intervalos de confianza son al menos del 90 %, al menos del 95 %, al menos del 97 %, al menos del 98 % o al menos del 99 %. Preferiblemente, el valor de p es menor de 0.1, de 0.05, de 0.01, de 0.005 o de 0.0001. Preferiblemente, la presente

invención permite detectar correctamente la enfermedad de forma diferencial en al menos el 60 %, en al menos el 70 %, en al menos el 80 %, o en al menos el 90 % de los sujetos de un determinado grupo o población analizada.

5 Un "*método pronóstico*" (o "*método de pronóstico*") se refiere a un procedimiento donde, a partir de análisis clínicos, se identifica la enfermedad o dolencia, como la EAPO, que sufre la persona y, a partir de las experiencias con otros pacientes aquejados por la misma enfermedad o dolencia, EAPO, se predice cómo será el estado de salud del individuo en el futuro; es decir, la evolución de la enfermedad o dolencia en el tiempo. El término
10 "*pronóstico*" en la presente invención también se refiere a la capacidad de detectar pacientes o sujetos asintomáticos que presentan una alta probabilidad de padecer una enfermedad o dolencia, EAPO, aun cuando en el momento del diagnóstico no presenten dichas patologías o manifestaciones evidentes de las mismas. Como entenderán los expertos en la materia, tal valoración, aunque se prefiere que sea, normalmente puede no
15 ser correcta para el 100% de los sujetos que se va a pronosticar. El término, sin embargo, requiere que una parte estadísticamente significativa de los sujetos se pueda identificar como que padecen la enfermedad o que tienen predisposición a la misma. Si una parte es estadísticamente significativa se puede determinar sin más por el experto en la materia usando varias herramientas de evaluación estadística bien conocidas, por ejemplo,
20 determinación de intervalos de confianza, determinación de valores p, prueba *t* de *Student*, prueba de *Mann-Whitney*, etc. Los intervalos de confianza preferidos son al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 90%, al menos el 95%. Los valores de p son, preferiblemente pero sin limitarse, 0,2, 0,1, 0,05.

"Uno o más" tal como se usa también en la presente memoria descriptiva incluye uno y la
25 especificación individualizada de cualquier número que sea más de uno, tal como dos, tres, cuatro, cinco, seis, etc. "Más de uno" o "algunos" tal como se usa en la presente memoria descriptiva incluye la especificación individualizada de cualquier número que sea más de uno, tal como dos, tres, cuatro, cinco, seis, etc.

30 USOS MÉDICOS DE LA INVENCION

Un **cuarto aspecto** de la invención se refiere al uso de una composición, de ahora en adelante composición de la invención, que comprende al menos un nucleótido con la

secuencia correcta del gen *SCP2*, *CTNND2* y/o *ZNF252P*, para la prevención, mejora y/o tratamiento de la EAPO.

El individuo a tratar puede identificarse como un individuo con mayor susceptibilidad a padecer EAPO mediante cualquiera de los métodos de la invención anteriormente descritos.

- 5 Por tanto, más preferiblemente se refiere al uso de una composición que comprende al menos un nucleótido con la secuencia correcta del gen *SCP2*, *CTNND2* y/o *ZNF252P*, para la prevención, mejora y/o tratamiento de la EAPO en un individuo que ha sido identificado por el primer o segundo método de la invención.

- 10 Preferiblemente, la composición de la invención es una composición farmacéutica que comprende, además, un vector aceptable para su implementación en terapia génica de sustitución.

- El término "vector", se refiere a un sistema que ayuda en el proceso de transferencia de un gen exógeno a la célula, facilitando la entrada y biodisponibilidad intracelular del mismo, de tal modo que este pueda funcionar correctamente. En esta definición se incluyen vectores
15 tanto virales como no virales.

Por secuencia correcta se entiende en esta invención aquella que no presenta el polimorfismo identificado como de mayor riesgo para un individuo de padecer EAPO y que se han descrito anteriormente.

- El éxito de cualquier protocolo de terapia génica exige la existencia de un gen susceptible
20 de manipulación terapéutica, un vector que facilite la entrega del material genético en las células apropiadas y un método que permita el acceso de ese vector al tejido correspondiente. El material genético utilizado puede ser ADN, ácido ribonucleico (ARN), oligonucleótidos sintéticos o ribozimas (ARN catalíticos). Los métodos disponibles para la introducción del ADN en las células deseadas son variados e incluyen vectores virales
25 (adenovirus, retrovirus deficientes en replicación, virus herpes, etc) y vectores no virales (liposomas, conjugados ADN-proteína). La administración del ADN a las células somáticas se puede hacer *ex vivo* (el vector con ADN se introduce en las células de interés en cultivo y se administran al paciente) o *in vivo* (donde el ADN asociado al vector se administra directamente al individuo, empleando métodos que permitan su llegada al tejido o célula
30 blanco). Actualmente, la mayoría de los protocolos en desarrollo utilizan vectores virales para introducir ADN en células o tejidos y se espera que una mejoría significativa en la

eficiencia y la selectividad de los vectores sea clave en el futuro desarrollo y difusión de la terapia génica.

KIT O DISPOSITIVO DE DIAGNÓSTICO Y USOS

- 5 Un **quinto aspecto** de la presente invención se refiere a un kit o dispositivo, de ahora en adelante kit o dispositivo de la invención, que comprende los elementos y/o reactivos necesarios para cuantificar los polimorfismos de un solo nucleótido rs10437066[G/A] y rs7544995[A/G] del gen *SCP2*, los polimorfismos rs6867416[C/T], rs6554641[T/C] del gen *CTNND2*, y los polimorfismos rs7817276[G/A], >rs2272715[T/C], rs6599548[C/T] y
10 rs3866997[A/C] del gen *ZNF252P*, de forma individual o cualquiera de sus combinaciones, a partir de una muestra biológica aislada de un individuo.

En una realización preferida, el kit o dispositivo de la invención proporciona datos suficientes para llevar a cabo el diagnóstico, estratificación, seguimiento y/o pronóstico de la evolución de EAPO en un individuo, de forma que estos datos permitan realizar el paso (b), (c) y/o (e)
15 de cualquiera de los métodos de la invención detallados anteriormente.

En una realización adicional, el kit o dispositivo de la invención puede comprender instrucciones para realizar automáticamente (c) y/o (e) de cualquiera de los métodos de la invención detallados anteriormente.

En otra realización, el kit o dispositivo de la invención puede comprender las herramientas
20 necesarias para hacer una manipulación, acondicionamiento, almacenamiento o pretratamiento de la muestra biológica compleja procedente del individuo de estudio, que permita un análisis instrumental que de como resultado los datos del paso (a) de cualquiera de los métodos de la invención detallados anteriormente.

AUTOMATIZACIÓN DEL MÉTODO DE LA INVENCION IMPLEMENTÁNDOLO EN UN 25 PROGRAMA DE ORDENADOR

Otro **aspecto** de la invención se refiere a un programa de ordenador que comprende instrucciones de programa para hacer que un ordenador lleve a la práctica el procedimiento de acuerdo con cualquiera de los métodos de la invención.

En particular, la invención abarca programas de ordenador dispuestos sobre o dentro de una
30 portadora. La portadora puede ser cualquier entidad o dispositivo capaz de soportar el programa. Cuando el programa va incorporado en una señal que puede ser transportada directamente por un cable u otro dispositivo o medio, la portadora puede estar constituida

por dicho cable u otro dispositivo o medio. Como variante, la portadora podría ser un circuito integrado en el que va incluido el programa, estando el circuito integrado adaptado para ejecutar, o para ser utilizado en la ejecución de, los procesos correspondientes.

Por ejemplo, los programas podrían estar incorporados en un medio de almacenamiento, como una memoria ROM, una memoria CD ROM o una memoria ROM de semiconductor, una memoria USB, o un soporte de grabación magnética, por ejemplo, un disco flexible o un disco duro. Alternativamente, los programas podrían estar soportados en una señal portadora transmisible. Por ejemplo, podría tratarse de una señal eléctrica u óptica que podría transportarse a través de cable eléctrico u óptico, por radio o por cualesquiera otros medios.

La invención se extiende también a programas de ordenador adaptados para que cualquier medio de procesamiento pueda llevar a la práctica los métodos de la invención. Tales programas pueden tener la forma de código fuente, código objeto, una fuente intermedia de código y código objeto, por ejemplo, como en forma parcialmente compilada, o en cualquier otra forma adecuada para uso en la puesta en práctica de los procesos según la invención. Los programas de ordenador también abarcan aplicaciones en la nube basadas en dicho procedimiento.

Otro aspecto de la invención se refiere a un medio de almacenamiento legible por un ordenador que comprende instrucciones de programa capaces de hacer que un ordenador lleve a cabo los pasos de cualquiera de los métodos de la invención.

Otro aspecto de la invención se refiere a una señal transmisible que comprende instrucciones de programa capaces de hacer que un ordenador lleve a cabo los pasos de cualquiera de los métodos de la invención.

A lo largo de la descripción y las reivindicaciones la palabra "comprende" y sus variantes no pretenden excluir otras características técnicas, aditivos, componentes o pasos. Para los expertos en la materia, otros objetos, ventajas y características de la invención se desprenderán en parte de la descripción y en parte de la práctica de la invención. Los siguientes ejemplos y dibujos se proporcionan a modo de ilustración, y no se pretende que sean limitativos de la presente invención.

A lo largo de la descripción y las reivindicaciones la palabra "comprende" y sus variantes no pretenden excluir otras características técnicas, aditivos, componentes o pasos. Además, la palabra "comprende" incluye el caso "consiste en". Para los expertos en la materia, otros objetos, ventajas y características de la invención se desprenderán en parte de la

descripción y en parte de la práctica de la invención. Los siguientes ejemplos y dibujos se proporcionan a modo de ilustración, y no se pretende que sean limitativos de la presente invención. Los signos numéricos relativos a los dibujos y colocados entre paréntesis en una reivindicación, son solamente para intentar aumentar la comprensión de la reivindicación, y no deben ser interpretados como limitantes del alcance de la protección de la reivindicación. Además, la presente invención cubre todas las posibles combinaciones de realizaciones particulares y preferidas aquí indicadas.

EXPOSICION DETALLADA DE MODOS DE REALIZACIÓN / EJEMPLOS

10 **Ejemplo 1. Estudio para la identificación de componentes clave para desarrollar métodos clínicos en la enfermedad alérgica al polen de olivo (EAPO).**

En el presente estudio se analizó la posible asociación entre 306670 SNPs usando la comparación "CONTROL" "CASE", con un número total de muestras de 670, siendo 135 controles y 535 casos. En dicho estudio se utilizó la plataforma estadística R , junto a librerías altamente especializadas en estudios de asociación genotípico, "genetics" y "SNPassoc".

- Número de sujetos estudiados: Número total de sujetos estudiados de 670, siendo 135 controles y 535 casos.

- Criterios de inclusión en el estudio:

- 20 A. Historia clínica compatible.
- B. Tests cutáneos e IgE específica positivos al polen de olivo.
- C. IgE específica positiva por lo menos a uno de los siguientes componentes del polen de olivo: Ole e 1, Ole e 2, Ole e 7 y Ole e 9.
- D. Edad entre 18-60 años.
- 25 E. Sin tratamiento previo con inmunoterapia alérgica específica.
- F. Se excluirán del estudio aquellos sujetos que no cumplan la totalidad de los requisitos anteriores.

- Número de muestras empleadas: Número total de muestras de 670, siendo 135 controles y 535 casos.

30 - Tipo de muestra: DNA obtenido de 15 ml de sangre periférica.

- Procedimiento de obtención y preparación de muestras: Una vez que el individuo otorgue

su consentimiento informado, se obtendrán 15 ml de sangre periférica mediante venopunción periférica. Esta intervención la realizará personal cualificado del Hospital Universitario Virgen Macarena. Los viales para esta extracción están pre-codificados. Es decir, incluyen un código de barras coincidente con el código de la hoja de recogida de datos de cada individuo. Mediante este sistema de pre-codificación el personal del laboratorio nunca tiene acceso a los datos personales de los individuos incluidos en el estudio. Por tanto, se garantiza que las muestras obtenidas están anonimizadas antes de su envío al laboratorio.

10 Pre-procesado y control de calidad:

Para el estudio de pre-procesado y control de calidad, tanto de las muestras como de las variantes a analizar, fue basado por el artículo usando el paquete estadístico snpsStast, SNPRelate y plyr.

15 Tal como indican en se utilizó como umbrales óptimos para la selección de variantes aquellas con una frecuencia alélica mínimo de 0.01 y máxima de 0.95, estos umbrales permite eliminar la mayor fuente de errores en el genotipado y baja variabilidad . Respecto a las muestras, se realizó diferentes evaluaciones para medir el grade de contaminación y correlación poblacional . Después de este paso quedaron 251.952 variantes y 668 muestras.

20 Posteriormente se analizó el equilibrio de Hardy-Weinberg (HWE) para cada variante. Este equilibrio establece que la composición genética de una población permanece en equilibrio mientras o actúe la selección natural ni ningún otro factor y no se produzca ninguna mutación, este principio es básico y sobre él se basa es resto del estudio estadístico de asociación, por lo tanto, es considerado como un paso importante para el filtrado de variantes. Las variantes que no cumplen los requerimientos mínimos para el estudio de asociación. En el presente estudio se analizó HWE considerando todas las muestras y mediante estratificación por sexo y fenotipo. Se tomó como referencia aquellas variantes que cumplieran con el HWE en todas las inferencias, utilizando el valor 0.1 como valor umbral de significancia estadística. La información sobre el equilibrio de HWE puede ser visualizado en el excel de resultados,

Estudio genético descriptivo:

Tanto el estudio de significancia estadística como el descriptivo está contenido en el Excel de resultados, dentro de la carpeta resultados. Allí está contenida la descripción sobre los

alelos descritos, la frecuencia del alelo mayoritario y el porcentaje de datos perdidos o que no se han podido evaluar.

Análisis de asociación y Odds ratio:

- 5 Partiendo de las variantes que cumplían con los requisitos de HWE descritos en 2.1, se realizó un estudio de Asociación Fenotipo-Genotipo, tanto a nivel global como estratificando por sexo, aplicando en cada caso diferentes modelos genéticos. Estos fueron:
- Codominante
 - Dominante
 - 10 • Recesivo
 - Log-Aditive

Los datos de significancia para cada modelo pueden ser visualizados en el Excel de resultados, dentro de la carpeta de resultados.

- 15 Posteriormente se realizó el estudio de los Odds ratio para cada variante, este valor es un indicador del aumento o disminución del riesgo de padecer el fenotipo a estudio asociado a las variantes estudiadas. Dicho análisis se aplicó a todas las variantes iniciales que cumplían los requisitos de HWE.

- 20 Estudio de Epistasia:

Para el estudio de epistasia se tomaron las variantes significativas estadísticamente con un umbral mínimo de 0.005 de pValue. Se estudió la posible epistasia utilizando los modelos, codominante, dominante y recesivo. Los resultados para el modelo codominante pueden ser visualizados en la Fig. 1, el modelo dominante en la Fig.2 y en la Fig. 3 el modelo recesivo.

- 25 Para considerar una correlación significativa se realizó el 95 percentil para cada modelo. Las variantes con epistasis significativas en cada modelo están descritas en el Excel de resultados.

Estudio Funcional de las variantes estadísticamente significativas:

- 30 El presente estudio se realizó el estudio funcional de variantes significativas mediante la información proporcionada por la base de datos Ensembl , dSNPs , Interpro , ClinVar y

uniprot . La información fue complementada por los algoritmos SIFT , PolyPhen . Para los estudios de frecuencia poblacional se utilizó las bases de datos 1000 Genomas y ExAC . Para estudios de predicción de eventos de splicing y motivos regulatorios se utilizó los algoritmos propuesto por . Para la conservación evolutiva de la posición de cada variante se utilizó el algoritmo propuesto por .

Selección de variantes más significativas:

Para la selección de variantes significativas se utilizó un umbral de pValue de 0.00001. Este umbral ha sido observado en otros trabajos similares. Se puede observar el mahattan plot en la Fig. 4.

Influencia de diferentes poblaciones:

El dataset inicial está contenido por dos poblaciones diferentes, de Sevilla y Málaga. Para poder corroborar que no existe desplazamiento y que las variantes significativas lo son en ambas poblaciones se añadió el pValue estratificando por población y se realizó un estudio de PCA. Para todas las dimensiones de PCA se realizó el test no paramétrico wilcoxon entre las dos poblaciones y los pValue fueron normalizados por fdr. El resultado fue negativo, por lo que la influencia de la población no parece significativa en los marcadores denominados significativos.

Variantes más significativas en modelo dominante:

A. *SCP2*

B. *CTNND2*

C. *ZNF252P*

REIVINDICACIONES

- 1.- El uso de, al menos, uno de los genes *SCP2*, *CTNND2* y/o *ZNF252P* o cualquiera de sus combinaciones, para diagnosticar, clasificar, predecir o pronosticar si un individuo ha desarrollado o es susceptible de desarrollar la enfermedad alérgica a polen del olivo (EAPO).
- 2.- El uso de los polimorfismos rs10437066[G/A] y rs7544995[A/G] del gen *SCP2*, los polimorfismos rs6867416[C/T], rs6554641[T/C] del gen *CTNND2*, y los polimorfismos rs7817276[G/A], rs2272715[T/C], rs6599548[C/T] y rs3866997[A/C] del gen *ZNF252P*.
- 3.- Un método de obtención de datos útiles para el diagnóstico, estratificación, seguimiento y pronóstico de la evolución de la EAPO en un individuo, de ahora en adelante primer método de la invención, que comprende:
- a) determinar los polimorfismos de un solo nucleótido rs10437066[G/A] y rs7544995[A/G] del gen *SCP2*, los polimorfismos rs6867416[C/T], rs6554641[T/C] del gen *CTNND2*, y los polimorfismos rs7817276[G/A], rs2272715[T/C], rs6599548[C/T] y rs3866997[A/C] del gen *ZNF252P* en una muestra biológica aislada, o cualquiera de sus combinaciones.
- 4.- Un método para el diagnóstico, estratificación, seguimiento y pronóstico de la evolución de EAPO en un individuo, que comprende el paso (a) del primer método de la invención, y adicionalmente:
- b) incluir a los individuos que presentan la variante A del rs10437066 y/o la variante G del rs7544995 del gen *SCP2*, la variante T del rs6867416 y/o la variante C del rs6554641[T/C] del gen *CTNND2*, y/o la A del rs7817276 la variante C del rs2272715[T/C] y/o la variante T del rs6599548 y/o la variante rs3866997 del gen *ZNF252P*, o cualquiera de sus combinaciones, en el grupo de individuos con mayor probabilidad de desarrollar EAPO.
- 5.- El método según cualquiera de las reivindicaciones 3-4, donde la muestra biológica es sangre.
- 6.- El método según cualquiera de las reivindicaciones 3-5, donde la determinación de los

polimorfismos se realiza mediante PCR.

7.- Uso de una composición, de ahora en adelante composición de la invención, que comprende al menos un nucleótido con la secuencia correcta del gen *SCP2*, *CTNND2* y/o
5 *ZNF252P*, para la prevención, mejora y/o tratamiento de la EAPO

8.- Uso de una composición según la reivindicación anterior, donde el individuo se ha identificado como susceptible de padecer EAPO mediante un método según las reivindicaciones 3-6.

10

9.- Un kit o dispositivo que comprende una o varias secuencias nucleotídicas necesarias para detectar los polimorfismos rs10437066[G/A] y rs7544995[A/G] del gen *SCP2*, los polimorfismos rs6867416[C/T], rs6554641[T/C] del gen *CTNND2*, y los polimorfismos rs7817276[G/A], rs2272715[T/C], rs6599548[C/T] y rs3866997[A/C] del gen *ZNF252P*.

15

10.- El uso de un kit o dispositivo según la reivindicación 7 para el diagnóstico, estratificación, seguimiento y pronóstico de la evolución de EAPO en un individuo

FIGURA 1

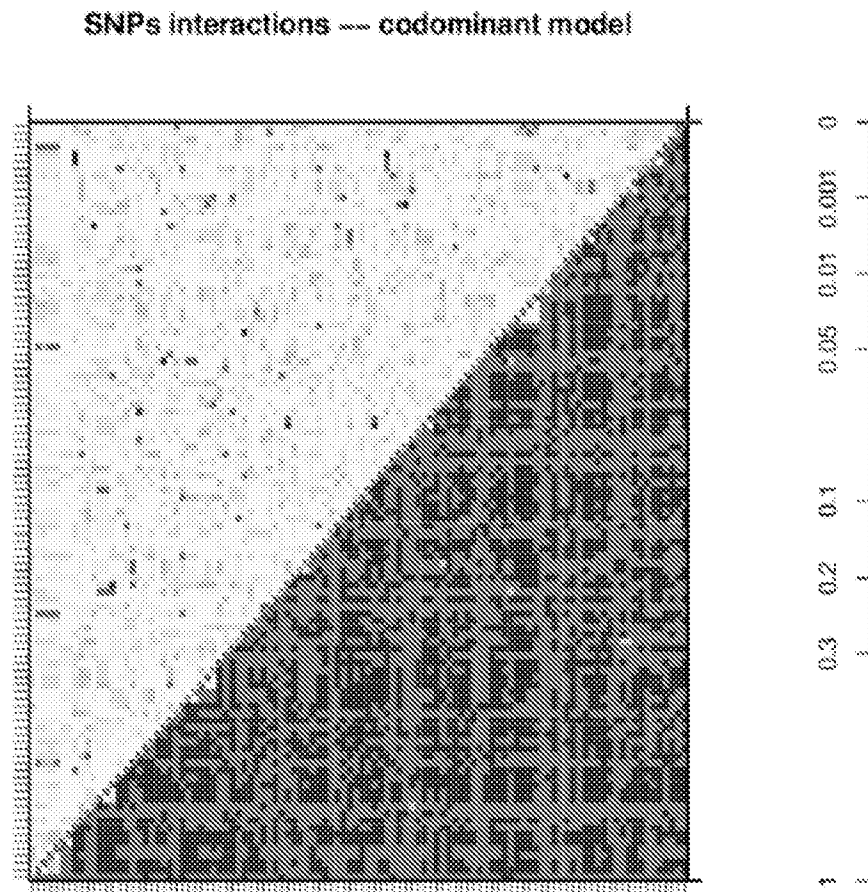


Figura1: Correlación entre las variantes significativas siguiendo un modelo codominante.

FIGURA 2

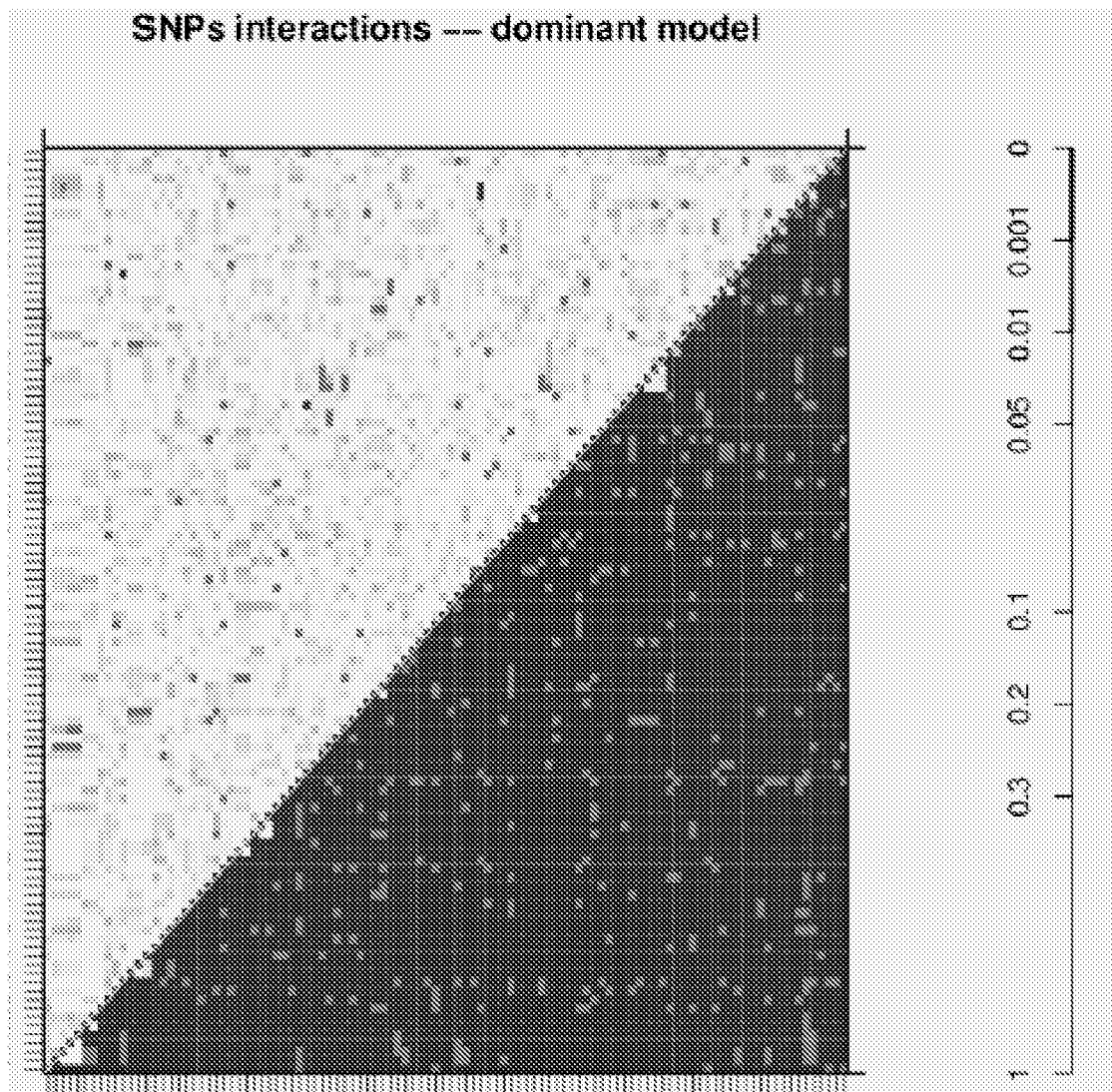


Figura2: Correlación entre las variantes significativas siguiendo un modelo dominante.

FIGURA 3

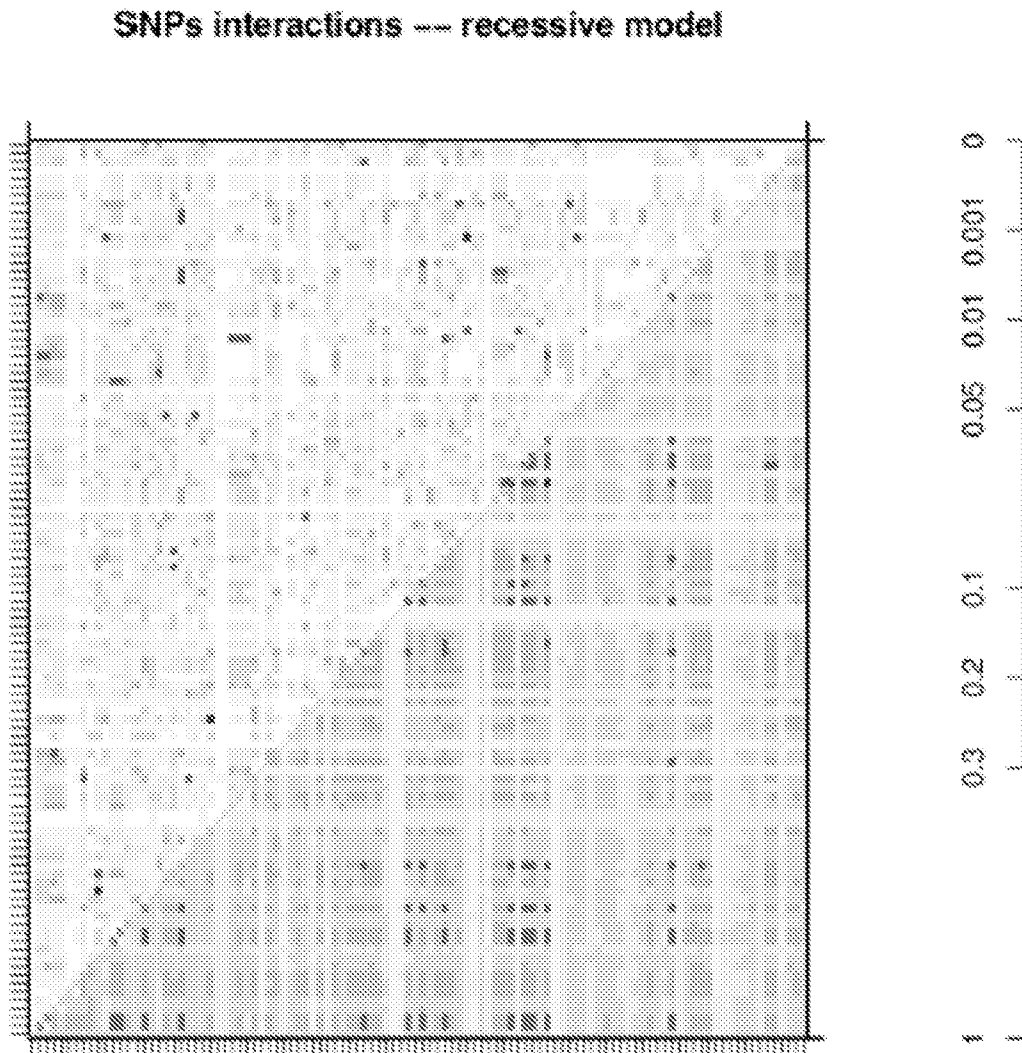


Figura3: Correlación entre las variantes significativas siguiendo un modelo recesivo.

FIGURA 4

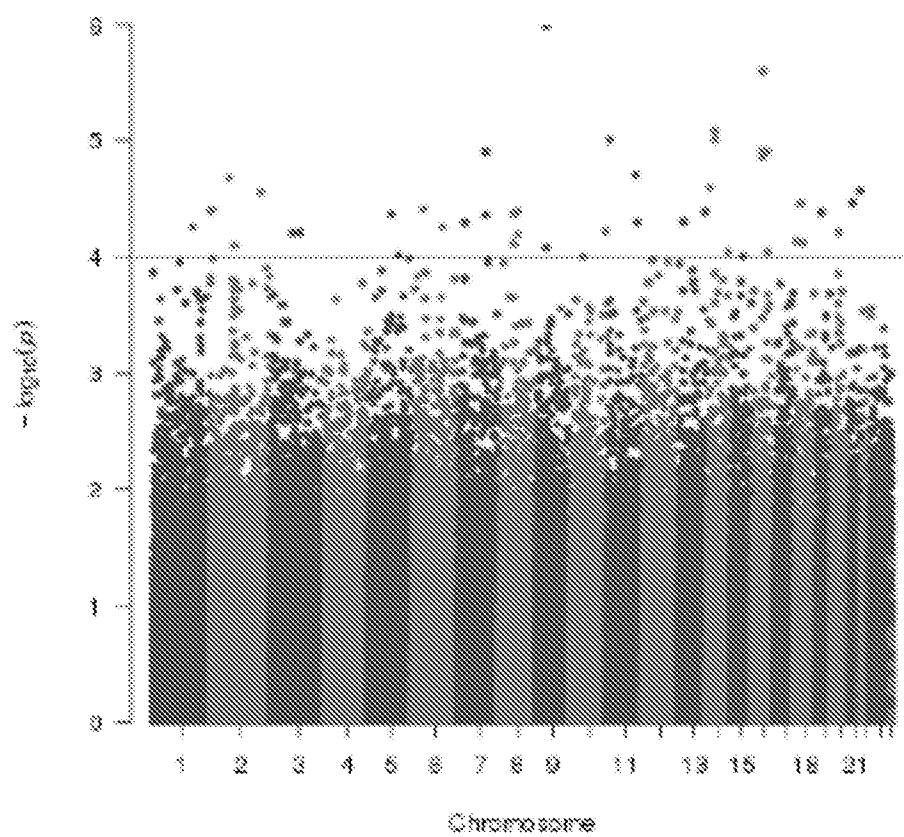


Figura4. Manhattan plot. Selección de los marcadores más significativos