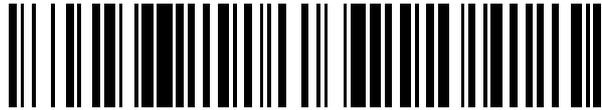


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 463 766**

51 Int. Cl.:

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **11.08.2009 E 09807152 (5)**

97 Fecha y número de publicación de la concesión europea: **23.04.2014 EP 2324126**

54 Título: **Método de identificación de factores de riesgo de la enfermedad de Alzheimer**

30 Prioridad:

12.08.2008 US 88203 P
12.06.2009 US 186673 P
10.07.2009 US 224647 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
29.05.2014

73 Titular/es:

ZINFANDEL PHARMACEUTICALS, INC. (100.0%)
One Science Drive
Durham, NC 27708, US

72 Inventor/es:

ROSES, ALLEN D.

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 463 766 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

Método de identificación de factores de riesgo de la enfermedad de Alzheimer

Campo de la invención

5 La presente invención se refiere al campo de la genómica, la genética, la farmacogenética y la bioinformática, incluyendo el análisis del genoma y el estudio de la variación de la secuencia de ADN. La invención también se refiere a estudios de asociación entre variaciones en las secuencias de ADN y la previsión de la susceptibilidad de un individuo a una enfermedad, trastorno, o condición concretos y/o la respuesta a un fármaco o tratamiento concretos.

Antecedentes de la invención

10 La búsqueda de marcadores genéticos asociados a enfermedades complejas está en curso. Los estudios de exploración del genoma completo con matrices de SNP siguen destacando la región ApoE como la zona más importante para la investigación en el estudio de la enfermedad de Alzheimer (Coon et al., J. Clin Psychiatry 68: 613-8 (2007); Li et al., Arch. Neurol 65: 45-53 (2007)).

15 La isoforma 4 de ApoE ha sido previamente asociada fuertemente a un mayor riesgo de desarrollar la enfermedad de Alzheimer de inicio tardío. (Pericak-Vance et al., Am. J. Hum. Genet. 48, 1034-1050 (1991); Martin et al, 2000, Patente de los Estados Unidos Núm. 6.027.896 de Roses, et al., Patente de los Estados Unidos Núm. 5.716.828 de Roses et al.). La relación es dependiente de la dosis (Yoshizawa et al, 1994; Schellenberg, 1995). Es decir, es más probable que un portador de dos alelos ApoE 4 desarrolle la enfermedad de Alzheimer de inicio tardío (EAT) que un portador de un solo alelo ApoE 4, y a una edad más temprana (Corder et al., Science 261, 921-3 (1993)).

20 Sin embargo, los alelos E4 solo representan aproximadamente 50% de la enfermedad de Alzheimer hereditaria. Una explicación es que ApoE 4 sirve únicamente como un marcador sustituto de algún desequilibrio de ligamiento próximo. Alternativamente, considerando el reciente descubrimiento de un papel mecánico para ApoE 4 en la toxicidad mitocondrial, los efectos negativos de la ApoE 4 se pueden anular o exacerbar por otro producto génico codificado en las inmediaciones (Chang et al., 2005).

25 Como el estado de la ApoE también está asociado con el riesgo de enfermedad arterial coronaria y probablemente también una serie de otras enfermedades y trastornos, las implicaciones del estudio de la región ApoE no se limitan a la enfermedad de Alzheimer, sino que son potencialmente de largo alcance (Mahley et al., Proc. Natl. Acad. Sci. USA 103: 5644-51 (2006)). En sentido más amplio, el examen de secuencias variantes para los procesos o rutas que rodean genes en desequilibrio de ligamiento con otras regiones genéticas que se sabe que están implicadas en procesos de enfermedades complejas proporcionará información valiosa para descifrar los mecanismos de esas enfermedades.

Compendio de la invención

35 En la presente memoria se describe un método para identificar una variante genética que está asociada con el desarrollo de una afección de interés (p. ej., inicio más temprano o más tardío de una enfermedad de interés), que comprende: (a) determinar a partir de muestras biológicas que contienen ADN la secuencia de nucleótidos portada por una pluralidad de sujetos humanos individuales en un locus genético de interés, en donde los sujetos incluyen tanto (i) sujetos afectados por la afección de interés como (ii) sujetos no afectados por la afección de interés; (b) identificar las variantes genéticas en dicho locus genético a partir de secuencias de nucleótidos observadas en dicha pluralidad de sujetos (p. ej., utilizando un análisis de alineación de secuencias múltiples); (c) mapear dichas variantes genéticas mediante la construcción de un árbol filogenético de dichas secuencias de nucleótidos de dichos sujetos, comprendiendo dicho árbol ramas que identifican los cambios de las variantes entre dichos sujetos (p. ej., cambios de variantes en el mismo cistrón); (d) examinar las variantes genéticas representadas como ramas en dicho árbol y determinar la proporción de sujetos afectados y no afectados para identificar aquellos cambios que llevan a una razón de cambio de los sujetos afectados con respecto a los no afectados (preferiblemente en donde el punto de partida es la variante genética que representa el mayor número de sujetos); y a continuación (e) identificar una variante genética o grupo de variantes (un haplotipo) en el que la razón de sujetos afectados con respecto a no afectados es sustancialmente diferente de una o más variantes adyacentes de dicho árbol (p. ej., al menos 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, o 90% diferente) para identificar de ese modo una variante genética asociada con el desarrollo de dicha afección de interés.

50 En algunas realizaciones, todos los sujetos portan un mismo polimorfismo conocido que se asocia con la afección de interés. En algunas realizaciones, la afección de interés es una enfermedad neurodegenerativa. En algunas realizaciones, la enfermedad de interés es una enfermedad en la que la ApoE y/o TOMM40 están implicados en la patogénesis de la enfermedad.

55 En algunas realizaciones, la afección de interés se asocia con un aumento o disminución de la disfunción mitocondrial. También se describe que la afección de interés es la esquizofrenia. También se describe que la afección de interés es una enfermedad de la arteria coronaria. También se describe que la afección de interés es la

diabetes mellitus, tipo II. También se describe que la afección de interés es la enfermedad de Parkinson. En algunas realizaciones, la afección de interés es la enfermedad de Alzheimer.

En algunas realizaciones, el factor de riesgo de polimorfismo conocido es el alelo E de la apolipoproteína (p. ej., ApoE 2, ApoE 3 o ApoE 4).

- 5 En algunas realizaciones, el locus genético de interés está en desequilibrio de ligamiento con el polimorfismo conocido. En algunas realizaciones, el locus genético de interés está en el mismo cromosoma y a menos de 10, 20, 30, 40, o 50 kilobases del polimorfismo conocido.

En algunas realizaciones, el locus genético es TOMM40.

- 10 También se describe un método para determinar un mayor riesgo para el desarrollo de una afección de interés, que comprende: (a) determinar en una muestra biológica que contiene ADN una variante genética identificada por el método de cualquiera de los párrafos anteriores portada por un sujeto individual; y a continuación (b) determinar si el sujeto tiene mayor riesgo de desarrollar la afección de interés cuando la variante genética está presente.

- 15 Además se proporciona un método para determinar el riesgo de desarrollar enfermedad de Alzheimer en un sujeto (p. ej., un sujeto que porta al menos un alelo Apo E3), que comprende: (a) detectar en una muestra biológica que contiene ADN tomada del sujeto la presencia o ausencia de una variante genética del gen TOMM40 asociada con un aumento o disminución del riesgo de enfermedad de Alzheimer, en donde dicha variante es un polimorfismo por delección/inserción en el intrón 6 o el intrón 9 del gen TOMM40; y (b) determinar si el sujeto tiene mayor o menos riesgo de enfermedad de Alzheimer cuando la variante genética está presente o ausente.

- 20 En algunas realizaciones, se determina si el sujeto es un sujeto Apo E2/E2, E2/E3, E2/E4, E3/E3, E3/E4, o E4/E4. En algunas realizaciones, se determina si el sujeto es un sujeto Apo E3/E3 o E3/E4.

- 25 También se proporciona una cantidad eficaz de tratamiento de un agente activo anti-enfermedad de Alzheimer seleccionado del grupo que consiste en inhibidores de acetilcolinesterasa, antagonistas del receptor de NMDA, agonistas o moduladores de receptores activados por proliferadores de peroxisomas (PPAR), anticuerpos, proteínas de fusión, moléculas de ARN terapéuticas, y combinaciones de los mismos, para su uso en un método de tratamiento de la enfermedad de Alzheimer caracterizado porque el método comprende la etapa de determinar si un sujeto tiene riesgo o no de desarrollar la enfermedad de Alzheimer, en donde el riesgo se determina de acuerdo con los métodos para determinar el riesgo descritos en la presente memoria.

- 30 En algunas realizaciones, la etapa de administración se lleva a cabo en el sujeto a una edad más temprana cuando se determina que el sujeto tiene mayor riesgo por la presencia o ausencia de la variante genética en comparación con un sujeto en el que la variante genética no está presente o ausente (p. ej., para un sujeto ApoE 4/4, comenzando a los 45, 46, 47, 48, 49, 50, 51, 52, o 53 años, y de forma continua a lo largo de cada año a partir de entonces, en lugar de comenzar a los 55 o más años; para un sujeto ApoE 4/3, a la edad de 50, 51, 52, 53, 54, 55, 56, 57, o 58 años, y de forma continua a lo largo de cada año a partir de entonces, en lugar de comenzar a los 60 o más años; para un sujeto ApoE 3/3, a la edad de 55, 56, 57, 58, 59, 60, 61, 62, o 63 años, y de forma continua a lo largo de cada año a partir de entonces, en lugar de empezar a los 65 años de edad o más; y para un sujeto ApoE 2/3, a los 60, 61, 62, 63, 64, 65, 66, 67, o 68 años, y de forma continua a lo largo de cada año a partir de entonces, en lugar de comenzar a los 70 años o más).

- 40 En algunas realizaciones, el agente activo se selecciona del grupo que consiste en inhibidores de acetilcolinesterasa, antagonistas del receptor de NMDA, agonistas o moduladores de PPAR (p. ej., fármacos de las clases de tiazolidinodiona o glitazar), anticuerpos, proteínas de fusión, moléculas de ARN terapéuticas, y combinaciones de los mismos. En algunas realizaciones, el agente activo es la rosiglitazona o una sal farmacéuticamente aceptable de la misma.

En algunas realizaciones, la variante genética de TOMM40 es una variante enumerada en la Tabla 1 mostrada a continuación.

- 45 También se describe un método para tratar a un sujeto (p. ej., un sujeto que tiene al menos un alelo ApoE 3) de la enfermedad de Alzheimer administrando un agente activo anti-enfermedad de Alzheimer al sujeto en una cantidad eficaz como tratamiento; comprendiendo la mejora: administrar el agente activo al sujeto a una edad más temprana cuando el sujeto porta una variante genética del gen TOMM40 asociada con un aumento del riesgo de enfermedad de Alzheimer en comparación con un sujeto correspondiente que no porta la variante genética (p. ej., para un sujeto ApoE 4/4, comenzando a los 45, 46, 47, 48, 49, 50, 51, 52, o 53 años, y de forma continua a lo largo de cada año a partir de entonces, en lugar de comenzar a los 55 o más años; para un sujeto ApoE 4/3, a la edad de 50, 51, 52, 53, 54, 55, 56, 57, o 58 años, y de forma continua a lo largo de cada año a partir de entonces, en lugar de comenzar a los 60 o más años; para un sujeto ApoE 3/3, a la edad de 55, 56, 57, 58, 59, 60, 61, 62, o 63 años, y de forma continua a lo largo de cada año a partir de entonces, en lugar de empezar a los 65 años de edad o más; y para un sujeto ApoE 2/3, a los 60, 61, 62, 63, 64, 65, 66, 67, o 68 años, y de forma continua a lo largo de cada año a partir de entonces, en lugar de comenzar a los 70 años o más).

En algunas realizaciones, el sujeto es un sujeto Apo E3/E3 o E3/E4.

5 En algunas realizaciones, el agente activo se selecciona del grupo que consiste en inhibidores de acetilcolinesterasa, antagonistas del receptor de NMDA, agonistas o moduladores de PPAR (p. ej., fármacos de las clases de tiazolidinodiona o glitazar), anticuerpos, proteínas de fusión, moléculas de ARN terapéuticas, y combinaciones de los mismos. En algunas realizaciones, el agente activo es la rosiglitazona o una sal farmacéuticamente aceptable de la misma.

En algunas realizaciones, la variante genética del gen TOMM40 es un polimorfismo de delección/inserción (DIP). En algunas realizaciones, el DIP es un polimorfismo de inserción. En algunas realizaciones, el DIP es un polimorfismo de delección/inserción de poli-T (p. ej., entre 5 y 100, o 10 y 80, o 20 y 50 pb de poli-T).

10 En algunas realizaciones, la variante genética de TOMM40 es una variante enumerada en la Tabla 1 como se expone a continuación. En algunas realizaciones, el DIP es rs10524523, rs10602329 o DIP3. En algunas realizaciones, el DIP es rs10524523. Adicionalmente se describe un método de tratamiento de una afección de interés, en donde la afección de interés está asociada con ApoE y/o TOMM40, de un paciente que lo necesite, incluyendo el método las etapas de: (a) determinar la presencia o ausencia de una variante genética identificada por el método del párrafo 1-12 llevado a cabo por un sujeto individual para generar un perfil genético del paciente; y a 15 continuación, si el perfil es indicativo de que el paciente es sensible a un agente activo, (b) administrar el agente activo al sujeto a una cantidad eficaz como tratamiento para tratar la afección de interés.

20 En algunas realizaciones, el agente activo se selecciona del grupo que consiste en inhibidores de acetilcolinesterasa, antagonistas del receptor de NMDA, agonistas o moduladores de PPAR (p. ej., fármacos de las clases de tiazolidinodiona o glitazar), anticuerpos, proteínas de fusión, moléculas de ARN terapéuticas, y combinaciones de los mismos. En algunas realizaciones, el agente activo es la rosiglitazona o una sal farmacéuticamente aceptable de la misma.

25 En algunas realizaciones, la variante genética del gen TOMM40 es un polimorfismo de delección/inserción (DIP). En algunas realizaciones, el DIP es un polimorfismo de inserción. En algunas realizaciones, el DIP es un polimorfismo de delección/inserción de poli-T (p. ej., una inserción entre 5 y 100, o 10 y 80, o 20 y 50 pb de poli-T).

En algunas realizaciones, la variante genética de TOMM40 es una variante enumerada en la Tabla 1 expuesta a continuación. En algunas realizaciones, el DIP es rs10524523, rs10602329 o DIP3. En algunas realizaciones, el DIP es rs10524523.

Asimismo se describe un método de tratamiento para la enfermedad de Alzheimer en un sujeto, incluyendo:

30 (a) detectar en una muestra biológica que contiene ADN tomada del sujeto la presencia o ausencia de una variante genética del gen TOMM40 asociada con la sensibilidad a un agente activo; y, si la variante genética está presente, (b) administrar el agente activo al sujeto a una cantidad eficaz como tratamiento para tratar la enfermedad de Alzheimer.

35 En algunas realizaciones, el sujeto porta al menos un alelo ApoE 3. En algunas realizaciones, el sujeto es un sujeto Apo E3/E3 o E3/E4.

40 En algunas realizaciones, el agente activo se selecciona del grupo que consiste en inhibidores de acetilcolinesterasa, antagonistas del receptor de NMDA, agonistas o moduladores de PPAR (p. ej., fármacos de las clases de tiazolidinodiona o glitazar), anticuerpos, proteínas de fusión, moléculas de ARN terapéuticas, y combinaciones de los mismos. En algunas realizaciones, el agente activo es la rosiglitazona o una sal farmacéuticamente aceptable de la misma.

En algunas realizaciones, la variante genética del gen TOMM40 es un polimorfismo de delección/inserción (DIP). En algunas realizaciones, el DIP es un polimorfismo de inserción. En algunas realizaciones, el DIP es un polimorfismo de delección/inserción de poli-T (p. ej., entre 5 y 100, o 10 y 80, o 20 y 50 pb de poli-T).

45 En algunas realizaciones, la variante genética del gen TOMM40 es una variante enumerada en la Tabla 1 expuesta a continuación. En algunas realizaciones, el DIP es rs10524523, rs10602329 o DIP3. En algunas realizaciones, el DIP es rs10524523.

50 En algunas realizaciones se proporciona el uso de un agente activo anti-enfermedad de Alzheimer para la preparación de un medicamento para llevar a cabo un método de tratamiento para la enfermedad de Alzheimer de acuerdo con los párrafos expuestos anteriormente. Asimismo se proporciona el uso de un agente activo anti-enfermedad de Alzheimer para llevar a cabo un método de tratamiento para la enfermedad de Alzheimer.

Se proporciona un método para determinar un pronóstico para un paciente en riesgo de desarrollar la enfermedad de Alzheimer, que incluye obtener un perfil del paciente, en donde la obtención del perfil de paciente incluye: detectar la presencia o ausencia de al menos un alelo ApoE en una muestra biológica del paciente, y detectar la presencia o ausencia de al menos un polimorfismo de delección/inserción (DIP) de TOMM40 situado en el intrón 6 o

el intrón 9 del gen TOMM40, y a continuación, convertir el perfil de paciente en el pronóstico, en donde la presencia del alelo ApoE y la presencia del al menos un polimorfismo DIP de TOMM40 identifica al paciente como un paciente que tiene riesgo de desarrollar la enfermedad de Alzheimer.

5 En algunas realizaciones, el DIP es un polimorfismo de inserción. En algunas realizaciones, el DIP es un polimorfismo de delección/inserción de poli-T (p. ej., entre 5 y 100, o 10 y 80, o 20 y 50 poli-T).

En algunas realizaciones, el DIP es rs10524523, rs10602329 o DIP3. En algunas realizaciones, el DIP es rs10524523.

En algunas realizaciones, el método incluye adicionalmente detectar si el sujeto es un Apo sujeto E2/E2, E2/E3, E2/E4, E3/E3, E3/E4, E4/E4. En algunas realizaciones, el sujeto es un sujeto Apo E3/E3 o E3/E4.

10 También se describe un método para estratificar un sujeto en un grupo de una prueba clínica de una terapia para el tratamiento de la enfermedad de Alzheimer, incluyendo el método: detectar la presencia o ausencia de al menos un alelo ApoE en una muestra biológica del paciente, y detectar la presencia o ausencia de al menos un polimorfismo de delección/inserción (DIP) de TOMM40 situado en el intrón 6 o el intrón 9 del gen TOMM40, en donde el sujeto se estratifica en el subgrupo para la prueba clínica de la terapia basada en la presencia o ausencia de al menos un
15 alelo ApoE y/o DIP de TOMM40.

En algunas realizaciones, el DIP es un polimorfismo de inserción. En algunas realizaciones, el DIP es un polimorfismo de inserción de poli-T (p. ej., inserción entre 5 y 100, o 10 y 80, o 20 y 50 poli-T).

En algunas realizaciones, el DIP es rs10524523, rs10602329 o DIP3. En algunas realizaciones, el DIP es rs10524523.

20 En algunas realizaciones, el método incluye adicionalmente detectar si el sujeto es un sujeto Apo E2/E2, E2/E3, E2/E4, E3/E3, E3/E4, E4/E4. En algunas realizaciones, el sujeto es un sujeto Apo E3/E3 o E3/E4.

Adicionalmente se describe un método para identificar un paciente en una prueba clínica de un tratamiento para la enfermedad de Alzheimer que incluyendo: (a) identificar un paciente y (b) determinar un pronóstico para el paciente diagnosticado con la enfermedad de Alzheimer que comprende obtener un perfil del paciente, en donde el perfil del
25 paciente comprende (i) detectar la presencia o ausencia de al menos un alelo ApoE en una muestra biológica del paciente, (ii) detectar la presencia o ausencia de al menos un polimorfismo de delección/inserción (DIP) de TOMM40 situado en el intrón 6 o el intrón 9 del gen TOMM40, y (iii) convertir el perfil del paciente en el pronóstico, incluyendo el pronóstico una predicción de si el paciente es un candidato para la prueba clínica para el tratamiento de la enfermedad de Alzheimer.

30 En algunas realizaciones, el DIP es un polimorfismo de inserción. En algunas realizaciones, el DIP es un polimorfismo de delección/inserción de poli-T (p. ej., entre 5 y 100, o 10 y 80, o 20 y 50 poli-T).

En algunas realizaciones, el DIP es rs10524523, rs10602329 o DIP3. En algunas realizaciones, el DIP es rs10524523.

35 En algunas realizaciones, el método incluye adicionalmente detectar si el sujeto es un sujeto Apo E2/E2, E2/E3, E2/E4, E3/E3, E3/E4, o E4/E4 sujeto. En algunas realizaciones, el sujeto es un sujeto Apo E3/E3 o E3/E4.

Asimismo se describe un kit para determinar si un sujeto tiene un aumento del riesgo de desarrollar la enfermedad de Alzheimer de inicio tardío, que incluye: (A) al menos un reactivo que detecta específicamente ApoE 3, ApoE 4, o ApoE 2, en donde el reactivo se selecciona del grupo que consiste en anticuerpos que se unen selectivamente a ApoE 3, ApoE 4, o ApoE 2, y sondas de oligonucleótidos que se unen selectivamente al ADN que codifica los
40 mismos; (B) al menos un reactivo que detecta específicamente la presencia o ausencia de al menos un polimorfismo de delección/inserción (DIP) de TOMM40 situado en el intrón 6 o el intrón 9 del gen TOMM40; y (C) instrucciones para determinar que el sujeto tiene un aumento del riesgo de desarrollar la enfermedad de Alzheimer de inicio tardío mediante: (i) la detección de la presencia o ausencia de una isoforma de ApoE en el sujeto con el al menos un reactivo; (ii) la detección de la presencia o ausencia de al menos un polimorfismo de delección/inserción (DIP) de
45 TOMM40 situado en el intrón 6 o el intrón 9 del gen TOMM40; y (iii) la observación de si el sujeto tiene o no un aumento del riesgo de desarrollar la enfermedad de Alzheimer de inicio tardío observando si la presencia de la isoforma de ApoE 3 y el DIP de TOMM40 es detectada a no con el al menos un reactivo, en donde la presencia de la isoforma de ApoE y el DIP de TOMM40 indica que el sujeto tiene un aumento del riesgo de desarrollar la enfermedad de Alzheimer de inicio tardío.

50 En algunas realizaciones, el al menos un reactivo y las instrucciones están empaquetados en un único recipiente.

En algunas realizaciones, el DIP es un polimorfismo de inserción. En algunas realizaciones, el DIP es un polimorfismo de delección/inserción de poliT (p. ej., entre 5 y 100, o 10 y 80, o 20 y 50 pb de poli-T).

En algunas realizaciones, el DIP es rs10524523, rs10602329 o DIP3. En algunas realizaciones, el DIP es rs10524523.

En algunas realizaciones, la etapa de determinación incluye adicionalmente detectar si el sujeto es un sujeto Apo E2/E2, E2/E3, E2/E4, E3/E3, E3/E4, o E4/E4. En algunas realizaciones, el sujeto es un sujeto Apo E3/E3 o E3/E4.

5 Se describe un kit para determinar si un sujeto es sensible al tratamiento para una afección de interés, en donde la afección de interés está asociada con ApoE y/o TOMM40, con un agente activo, incluyendo el kit: (A) al menos un reactivo que detecta específicamente ApoE 3, ApoE 4, o ApoE 2, en donde el reactivo se selecciona del grupo que consiste en anticuerpos que se unen selectivamente a ApoE 3, ApoE 4, o ApoE 2, y sondas de oligonucleótidos que se unen selectivamente al ADN que codifica los mismos; (B) al menos un reactivo que detecta específicamente la presencia o ausencia de al menos un polimorfismo de delección/inserción (DIP) de TOMM40 situado en el intrón 6 o el intrón 9 del gen TOMM40; y (C) instrucciones para determinar que el sujeto es sensible al tratamiento para la afección de interés con el agente activo de interés mediante: (i) la detección de la presencia o ausencia de una isoforma de ApoE en el sujeto con el al menos un reactivo; (ii) la detección de la presencia o ausencia de al menos un polimorfismo de delección/inserción (DIP) de TOMM40 situado en el intrón 6 o el intrón 9 del gen TOMM40; y (iii) la determinación de si el sujeto es sensible o no al tratamiento observando si la presencia de la isoforma de ApoE 3 y el DIP de TOMM40 es detectada a no con el al menos un reactivo, en donde la presencia de ApoE y el DIP de TOMM40 indica que el sujeto es sensible al tratamiento con el agente activo.

En algunas realizaciones, el al menos un reactivo y las instrucciones están empaquetados en un único recipiente.

En algunas realizaciones, el DIP es un polimorfismo de inserción. En algunas realizaciones, el DIP es un polimorfismo de delección/inserción de poliT (p. ej., entre 5 y 100, o 10 y 80, o 20 y 50 pb de poli-T).

20 En algunas realizaciones, el DIP es rs10524523, rs10602329 o DIP3. En algunas realizaciones, el DIP es rs10524523.

En algunas realizaciones, la etapa de determinación incluye adicionalmente detectar si el sujeto es un sujeto Apo E2/E2, E2/E3, E2/E4, E3/E3, E3/E4, o E4/E4. En algunas realizaciones, el sujeto es un sujeto Apo E3/E3 o E3/E4.

25 Se entenderá que todas las realizaciones anteriores se pueden combinar de cualquier manera y/o en cualquier combinación. Los anteriores y otros objetos y aspectos de la presente invención se explican con mayor detalle en los dibujos proporcionados con la presente y en la memoria descriptiva expuesta a continuación.

Breve descripción de los dibujos

La Figura 1 muestra un diagrama de flujo general para identificar una variante genética en una región predeterminada de la secuencia genómica en un locus genético de interés, que puede estar asociada con una afección de interés, de acuerdo con algunas realizaciones.

30 La Figura 2 muestra un gráfico de la edad media de inicio de la enfermedad de Alzheimer como una función de la herencia de los cinco genotipos ApoE comunes, y que representa ApoE 4 como un factor de riesgo para la enfermedad de Alzheimer (1993).

35 La Figura 3 muestra las regiones A, B, y C en el Cromosoma 19, que son loci genéticos ilustrativos de interés. El gen TOMM40 se encuentra en estrecha proximidad con el gen ApoE y codifica una proteína de 40 kD dirigida a la membrana mitocondrial externa. TOMM40 interactúa con ApoE directamente en la regulación de la importación de proteínas mitocondriales, y una hipótesis actual es que la presencia de una o varias variantes de TOMM40 concretas exacerba el aumento de riesgo de enfermedad de Alzheimer asociado con la presencia dependiente de la dosis del alelo ApoE 3.

40 La Figura 4 muestra, de manera esquemática, un mapa evolutivo de la Región B como se indica en la Figura 3 creado de acuerdo con algunas realizaciones detalladas en la presente memoria. Cada óvalo horizontal representa una variante de secuencia observada (p. ej., un cambio de nucleótido), y el tamaño del óvalo representa su frecuencia. Cada variante se denomina un nodo del árbol, y la conexión de un nodo con otro nodo por una rama representa una observación de las dos variantes en cis portadas por un sujeto individual.

45 La Figura 5 es un diagrama esquemático del árbol filogenético basado en la Región B construido para TOMM40, que muestra los porcentajes de fenotipos de ApoE en dos agrupamientos principales, o clados, de las variantes de TOMM40 en esta región.

50 La Figura 6 es una visión esquemática del locus TOMM40-ApoE que incluye un gráfico LD que muestra bloques de haplotipos y regiones sujetas a secuenciación primaria en estudios exploratorios (R1) (23 Kb) y confirmatorios (R2) (10 Kb) (NCBI Build 36.3). El gráfico LD se muestra para los datos HapMap (panel de análisis CEU), definición de bloques de haplotipos mediante "solid spine", r^2 valores con esquema de color D'/LOD representados por diferentes características de la línea.

La Figura 7 muestra representaciones de los árboles filogenéticos con separación de las variantes. 7A: Variantes SNP, clado A vs B, E6 - E10 representan exones TOMM40 y las líneas verticales indican las ubicaciones aproximadas de los SNP. La separación de las dos ramas principales tiene un fuerte apoyo Bootstrap (973/1000).

7B: polimorfismos de longitud rs10524523. Se proporcionan estadísticas descriptivas para cada grupo de polimorfismos de longitud. Varios haplotipos largos que formaban grupos externos individuales, en el árbol o muy pequeños clados, se encuentran en el grupo identificado como "Resto".

5 La Figura 8 presenta histogramas de la longitud del polimorfismo de longitud rs10524523 estratificado por genotipos de ApoE 3/3 (8A), 3/4 (8B), y 4/4 (8C). N = 210 haplotipos (cohorte AS).

La Figura 9 muestra la asociación entre la edad de inicio de la EA y la longitud del polimorfismo rs10524523 para pacientes de EA con inicio entre los 60 y 86 años. Los diagramas de cajas indican el rango del 95% (líneas verticales), la mediana (línea horizontal en la caja) y el rango intercuartílico (caja).

Descripción detallada

10 La presente invención se explica con mayor detalle a continuación. No se pretende que esta descripción sea un catálogo detallado de todas las diferentes maneras en las cuales se puede implementar la invención, o de todas de las características que se pueden añadir a la presente invención. Por ejemplo, las características ilustradas con respecto a una realización se pueden incorporar a otras realizaciones, y las características ilustradas con respecto a una realización concreta, se pueden eliminar de esa realización. Además, serán evidentes numerosas variaciones y adiciones a las diversas realizaciones sugeridas en la presente memoria para los expertos en la técnica a la luz de la presente descripción que no se apartan de la presente invención. Por lo tanto, se pretende que la siguiente memoria descriptiva ilustre algunas realizaciones concretas de la invención, y especifique de manera no exhaustiva todas las permutaciones, combinaciones y variaciones de las mismas.

20 Según se utiliza en la descripción de la invención y en las reivindicaciones adjuntas, se pretende que las formas en singular "un", "una", "el" y "la" incluyan también las formas en plural, a menos que el contexto indique claramente lo contrario. También, según se utiliza en la presente memoria, "y/o" se refiere a e incluye cualquiera y todas las posibles combinaciones de uno o más de los elementos enumerados, así como la carencia de combinaciones cuando se interpreta en la alternativa ("o").

25 La presente invención está dirigida a métodos para revelar la variación genética en regiones de particular interés para enfermedades y trastornos complejos. También se refiere al descubrimiento de los marcadores genéticos más informativos sobre la base de asociaciones con información del fenotipo. En una realización, la invención se puede utilizar para localizar marcadores genéticos asociados con la susceptibilidad a una determinada enfermedad, trastorno, o afección. En otra realización, los datos respecto a las respuesta del sujeto a un tratamiento o fármaco candidato se pueden incluir en un análisis filogenético de la ubicación de marcadores genéticos asociados con una respuesta beneficiosa al tratamiento o fármaco (es decir, farmacogenética). Los métodos se pueden aplicar a cualquier conjunto de datos de variación genética de un locus concreto. Véase la Figura 1 para un diagrama de flujo del enfoque para encontrar factores de riesgo genético de acuerdo con la presente invención.

30 En un aspecto, el análisis de la variación genética se basa en datos de secuencias variantes. En un segundo aspecto, la estructura se descubrió utilizando datos de genotipos diploides, evitando así la necesidad de inferir experimentalmente o computacionalmente los haplotipos componentes (véase, p. ej., la patente de los Estados Unidos Núm. 6.027.896 de Roses et al.). En otro aspecto, el presente método se puede aplicar a la variación alélica no caracterizada que resulta de la interrogación de un ácido nucleico diana con un procedimiento experimental que proporciona un registro de la variación de la secuencia presente, pero no proporciona realmente la secuencia completa. La estructura subyacente de la variación genética también es útil para la deducción de los haplotipos constitutivos a partir de los datos de genotipos diploides.

35 Se prefiere y contempla que los métodos descritos en la presente memoria se utilicen junto con otra información de diagnóstico clínico conocida o descrita en la técnica que se utilizan en la evaluación de los sujetos con enfermedades o trastornos (p. ej., los que se cree que están implicados en la disfunción mitocondrial (p. ej., enfermedad de Alzheimer u otras enfermedades neurodegenerativas)) o para la evaluación de sujetos que se sospecha que se encuentran en riesgo de desarrollar dicha enfermedad. Tal invención también es aplicable al descubrimiento de factores de riesgo genéticos para otras enfermedades, trastornos o afecciones complejas.

1. Definiciones. Las siguientes definiciones se utilizan en la presente memoria:

40 "Afección de interés" se refiere a una afección, enfermedad o trastorno específicos designados para el estudio filogenético y/o el diagnóstico o pronóstico posterior. "Afección" según se utiliza en la presente memoria incluye, pero no se limita a, afecciones asociadas con ApoE y/o TOMM40 y/o disfunción mitocondrial, p. ej., enfermedades neurodegenerativas, enfermedades metabólicas, trastornos psiquiátricos, y cáncer.

45 Los ejemplos de las afecciones en las que han sido implicados ApoE y/o TOMM40 incluyen, pero no se limitan a, enfermedades cardiovasculares; enfermedades metabólicas; enfermedades neurodegenerativas; trauma o enfermedad neurológicas; enfermedad autoinmunitaria (p. ej., esclerosis múltiple (Pinholt M, et al. Apo E in multiple sclerosis and optic neuritis: the apo E-epsilon4 allele is associated with progression of multiple sclerosis. *Mult Scler.* 11:511-5 (2005); Masterman, T. y Hillert, J. The telltale scan: APOE ε4 in multiple sclerosis. *Lancet Neurol.* 3: 331 (2004), lupus eritematoso generalizado neuropsiquiátrico (Pullmann Jr. R, et al. Apolipoprotein E polymorphism in

patients with neuropsychiatric SLE. *Clin. Rheumatol.* **23**: 97-101 (2004), etc.); infección viral (p. ej., enfermedad viral asociada con infección por hepatitis C (Wozniak MA, et al. Apolipoprotein E-ε4 protects against severe liver disease caused by hepatitis C virus. *Hepatology.* **36**: 456-463 (2004)), enfermedad por VIH (Burt TD, et al. Apolipoprotein (apo) E4 enhances HIV-1 cell entry in vitro, and the APOE epsilon4/epsilon4 genotype accelerates HIV disease progression. *Proc Natl Acad Sci USA.* **105**:8718-23 (2008)), etc.); fractura/osteoporosis de cadera (Pluijm SM, et al. Effects of gender and age on the association of apolipoprotein E epsilon4 with bone mineral density, bone turnover and the risk of fractures in older people. *Osteoporos Int.* **13**: 701-9 (2002)); enfermedades mitocondriales (Chang S, et al. Lipid- and receptor-binding regions of apolipoprotein E4 fragments act in concert to cause mitochondrial dysfunction and neurotoxicity. *Proc Natl Acad Sci USA.* **102**:18694-9 (2005)); envejecimiento (Schachter F, et al. Genetic associations with human longevity at the APOE and ACE loci. *Nat Genet.* **6**:29-32 (1994); Rea IM, et al., Apolipoprotein E alleles in nonagenarian subjects in the Belfast Elderly Longitudinal Free-living Ageing Study (BELFAST). *Mech. Ageing and Develop.* **122**: 1367-1372 (2001)); inflamación (Li L, et al., Infection induces a positive acute phase apolipoprotein E response from a negative acute phase gene: role of hepatic LDL receptors. *J Lipid Res.* **49**:1782-93 (2008)); y disfunción de la memoria (Caselli RJ, et al. Longitudinal modeling of age-related memory decline and the APOE epsilon4 effect. *N Engl J Med.* **361**:255-63 (2009)).

"Enfermedad cardiovascular", según se utiliza en la presente memoria se refiere a una enfermedad que implica al corazón y/o los vasos sanguíneos, incluyendo, pero no limitada a, enfermedad de la arteria coronaria (Song Y, et al Meta-analysis: apolipoprotein E genotypes and risk for coronary heart disease. *Ann Intern Med.* **141**:137-47 (2004); Bennet AM, et al., Association of apolipoprotein E genotypes with lipid levels and coronary risk. *JAMA* **298**:1300-11 (2007)), aterosclerosis (Norata GD, et al. Effects of PCSK9 variants on common carotid artery intima media thickness and relation to ApoE alleles. *Atherosclerosis* (2009) 27 de Jun. [publicación electrónica antes de impresión], doi:10.1016/j.atherosclerosis.2009.06.023; Paternoster L, et al. Association Between Apolipoprotein E Genotype and Carotid Intima-Media Thickness May Suggest a Specific Effect on Large Artery Atherothrombotic Stroke. *Stroke* **39**:48-54 (2008)), enfermedad cardíaca isquémica (Schmitz F, et al., Robust association of the APOE 4 allele with premature myocardial infarction especially in patients without hypercholesterolaemia: the Aachen study. *Eur. J. Clin. Investigation* **37**: 106-108 (2007)), enfermedad vascular tal como accidente cerebrovascular isquémico (Peck G, et al. The genetics of primary haemorrhagic stroke, subarachnoid haemorrhage and ruptured intracranial aneurysms in adults. *PLoS One.* **3**:e3691 (2008); Paternoster L, et al. Association Between Apolipoprotein E Genotype and Carotid Intima-Media Thickness May Suggest a Specific Effect on Large Artery Atherothrombotic Stroke. *Stroke* **39**:48-54 (2008)), demencia vascular (Bang OY, et al. Important link between dementia subtype and apolipoprotein E: a meta-analysis. *Yonsei Med J.* **44**:401-13 (2003); Baum L, et al. Apolipoprotein E epsilon4 allele is associated with vascular dementia. *Dement Geriatr Cogn Disord.* **22**:301-5 (2006)), etc.

"Enfermedad neurodegenerativa" según se utiliza en la presente memoria, se refiere a la enfermedad de Alzheimer (Corder EH, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**:921-3 (1993).; Corder EH, et al. There is a pathologic relationship between ApoE-epsilon 4 and Alzheimer's disease. *Arch Neurol.* **52**:650-1 (1995)), enfermedad de Parkinson (Huang X, et al. Apolipoprotein E and dementia in Parkinson disease: a meta-analysis. *Arch Neurol.* **63**:189-93 (2006); Huang X et al. APOE-[epsilon]2 allele associated with higher prevalence of sporadic Parkinson disease. *Neurology* **62**:2198-202 (2004); Martinez, M. et al. Apolipoprotein E4 is probably responsible for the chromosome 19 linkage peak for Parkinson's disease. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **136B**, 172-174 (2005)), enfermedad de Huntington, y una pluralidad de enfermedades y trastornos menos comunes que hacen que las neuronas disminuyan, . ej., degeneración macular relacionada con la edad (Thakkestian A, et al. Association between apolipoprotein E polymorphisms and age-related macular degeneration: A HuGE review and meta-analysis. *Am J Epidemiol.* **164**:813-22 (2006); Bojanowski CM, et al. An apolipoprotein E variant may protect against age-related macular degeneration through cytokine regulation. *Environ Mol Mutagen.* **47**:594-602 (2006)).

"Trauma o enfermedad neurológicos" incluye, pero no se limita a, los resultados después de lesión en la cabeza (Zhou W, et al. Meta-analysis of APOE4 allele and outcome after traumatic brain injury. *J Neurotrauma.* **25**:279-90 (2008); Lo TY, et al. Modulating effect of apolipoprotein E polymorphisms on secondary brain insult and outcome after childhood brain trauma. *Childs Nerv Syst.* **25**:47-54 (2009)), migraña (Gupta R, et al. Polymorphism in apolipoprotein E among migraineurs and tension-type headache subjects. *J Headache Pain.* **10**:115-20 (2009)), edema vasogénico (James ML, et al. Apolipoprotein E modifies neurological outcome by affecting cerebral edema but not hematoma size after intracerebral hemorrhage in humans. *J Stroke Cerebrovasc Dis.* **18**:144-9 (2009); James ML, et al. Pharmacogenomic effects of apolipoprotein e on intracerebral hemorrhage. *Stroke* **40**:632-9 (2009)), etc.

"Enfermedad metabólica" según se utiliza en la presente memoria incluye, pero no se limita a, dislipidemia (Willer CJ, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet.* **40**:161-9 (2008); Bennet AM, et al., Association of apolipoprotein E genotypes with lipid levels and coronary risk. *JAMA* **298**:1300-11 (2007)), enfermedad renal en fase terminal (Oda H, et al. Apolipoprotein E polymorphism and renal disease. *Kidney Int Suppl.* **71**:S25-7 (1999); Hubacek JA, et al. Apolipoprotein E Polymorphism in Hemodialyzed Patients and Healthy Controls. *Biochem Genet.* (2009) 30 de Jun. [publicación electrónica antes de impresión] DOI 10.1007/s10528-009-9266-y.), enfermedad renal crónica (Yoshida T, et al. Association of a polymorphism of the apolipoprotein E gene with chronic kidney disease in Japanese individuals with metabolic syndrome. *Genomics* **93**:221-6 (2009); Leiva E, et al. Relationship between Apolipoprotein E polymorphism and nephropathy in type-2 diabetic patients. *Diabetes Res Clin Pract.* **78**:196-201 (2007)), enfermedad de la vesícula

- biliar (Boland LL, et al. Apolipoprotein E genotype and gallbladder disease risk in a large population-based cohort. *Ann Epidemiol.* **16**:763-9 (2006); Andreotti G, et al. Polymorphisms of genes in the lipid metabolism pathway and risk of biliary tract cancers and stones: a population-based case-control study in Shanghai, China. *Cancer Epidemiol Biomarkers Prev.* **17**:525-34 (2008)), diabetes mellitus (tipo II) (Elosua R, et al. Obesity Modulates the Association among APOE Genotype, Insulin, and Glucose in Men. *Obes Res.* **11**:1502-1508 (2003); Moreno JA, et al. The Apolipoprotein E Gene Promoter (-219G/T) Polymorphism Determines Insulin Sensitivity in Response to Dietary Fat in Healthy Young Adults. *J. Nutr.* **135**:2535-2540 (2005)), síndrome metabólico, colestiasis (Abu Abeid S, et al. Apolipoprotein-E genotype and the risk of developing cholelithiasis following bariatric surgery: a clue to prevention of routine prophylactic cholecystectomy. *Obes Surg.* **12**:354-7 (2002)), etc.
- 5
- 10 "Trastorno Psiquiátrico" según se utiliza en la presente memoria, se refiere a la esquizofrenia (Kampman O, et al. Apolipoprotein E polymorphism is associated with age of onset in schizophrenia. *J Hum Genet.* **49**:355-9 (2004); Dean B. et al., Plasma apolipoprotein E is decreased in schizophrenia spectrum and bipolar disorder. *Psychiatry Res.* **158**:75-78 (2008)), trastorno obsesivo compulsivo (TOC), conducta adictiva (la adicción al tabaco, la adicción al alcohol, etc.), trastorno bipolar (Dean B. et al., Plasma apolipoprotein E is decreased in schizophrenia spectrum and bipolar disorder. *Psychiatry Res.* **158**:75-78 (2008)), y otras enfermedades, trastornos o afecciones de naturaleza psiquiátrica.
- 15
- "Desarrollo de una afección" según se utiliza en la presente memoria, se refiere tanto a un diagnóstico inicial de una enfermedad, trastorno u otra afección médica, como a la exacerbación de una enfermedad, trastorno o afección médica existente para los cuales el sujeto ya ha sido diagnosticado.
- 20
- 25 "Diagnóstico" o "pronóstico", según se utiliza en la presente memoria, se refieren a la utilización de la información (p. ej., información genética o datos de otros ensayos moleculares sobre muestras biológicas, signos y síntomas, descubrimientos en exámenes físicos, resultados de rendimiento cognitivo, etc.) para preveer los resultados más probables, los marcos temporales, y/o la respuesta a un tratamiento concreto para una enfermedad, trastorno o afección dados, basándose en comparaciones con una pluralidad de individuos que comparten secuencias de nucleótidos, síntomas, signos, antecedentes familiares comunes, u otros datos relevantes para la consideración del estado de salud de un paciente.
- 30
- "Muestra biológica" según se utiliza en la presente memoria, se refiere a un material que se sospecha que contiene un ácido nucleico de interés. Las muestras biológicas que contienen ADN incluyen pelo, piel, frotis bucal, y fluidos biológicos tales como sangre, suero, plasma, esputo, fluido linfático, semen, moco vaginal, heces, orina, fluido espinal, y similares. El aislamiento de ADN a partir de dichas muestras es bien conocido por los expertos en la técnica.
- 35
- Un "sujeto", de acuerdo con algunas realizaciones es un individuo cuyo genotipo o genotipos o haplotipo o haplotipos están por determinar y registrar junto con la afección del individuo (es decir, la enfermedad o el estado del trastorno) y/o la respuesta a un fármaco o tratamiento candidatos. Las secuencias de nucleótidos de una pluralidad de sujetos se utilizan para construir un árbol filogenético, y a continuación se pueden comparar las secuencias de nucleótidos análoga de un sujeto individual con las del árbol filogenético para fines de diagnóstico o pronóstico.
- 40
- "Gen" según se utiliza en la presente memoria significa un segmento de ADN que contiene toda la información para la biosíntesis regulada de un producto de ARN, incluyendo promotores, exones, intrones, y otras regiones no traducidas que controlan la expresión.
- 45
- "Locus genético" o "locus" según se utiliza en la presente memoria significan una ubicación en un cromosoma o una molécula ADN, que a menudo corresponde a un gen o un rasgo físico o fenotípico o a un nucleótido o tramo de nucleótidos concretos. Loci es la forma en plural de locus.
- 50
- "Amplificación", tal como se aplica a los ácidos nucleicos en la presente memoria se refiere a cualquier método que da como resultado la formación de una o más copias de un ácido nucleico, donde preferiblemente la amplificación es exponencial. Uno de tales métodos para la amplificación enzimática de secuencias específicas de ADN es conocido como reacción en cadena de la polimerasa (PCR), descrito por Saiki et al., 1986, *Science* 230:1350-1354. Los cebadores utilizados en la PCR normalmente varían en longitud de aproximadamente 10 a 50 o más nucleótidos, y se seleccionan típicamente para que sean al menos aproximadamente 15 nucleótidos para garantizar la especificidad suficiente. El fragmento de doble cadena que se produce se conoce como "amplicón", y puede variar en longitud de tan pocos como aproximadamente 30 nucleótidos, a 20.000 o más.
- 55
- Un "marcador" o "marcador genético" según se utiliza en la presente memoria es una variación conocida de una secuencia de ADN en un locus concreto. La variación puede estar presente en un individuo debido a la mutación o la herencia. Un marcador genético puede ser una secuencia corta de ADN, tal como una secuencia en torno a un solo cambio de un solo par de bases (polimorfismo de un solo nucleótido, SNP), o una larga, como los minisatélites. Los marcadores se pueden utilizar para estudiar la relación entre una enfermedad heredada y su causa genética (p. ej., una mutación concreta de un gen que da como resultado una forma defectuosa o de otro modo indeseable de la proteína).
- Un "factor de riesgo genético" según se utiliza en la presente memoria significa un marcador genético que se asocia

con un aumento de susceptibilidad a una afección, enfermedad, o trastorno. También puede referirse a un marcador genético que se asocia con una respuesta concreta a un fármaco o tratamiento seleccionado de interés.

5 "Asociado a" según se utiliza en la presente memoria significa la aparición conjunta de dos o más de características más a menudo de lo que cabría esperar solo por azar. Un ejemplo de asociación implica un rasgo sobre la superficie de los glóbulos blancos denominado HLA (HLA significa antígeno leucocitario humano). Un tipo de HLA concreto, el tipo de HLA B-27, se asocia con un mayor riesgo de numerosas enfermedades incluyendo espondilitis anquilosante. Es 87 veces más probable que aparezca espondilitis anquilosante en personas con HLA B-27 que en la población general.

10 Un sujeto" que tiene mayor riesgo de desarrollar una enfermedad", debido a un factor de riesgo genético es aquel que está predispuesto a la afección, tiene susceptibilidad genética para la afección, y/o tiene más probabilidades de desarrollar la afección que los sujetos en los que el factor de riesgo genético está ausente. Por ejemplo, un sujeto que "tiene mayor riesgo de desarrollar la enfermedad de Alzheimer" debido a la presencia de uno o dos alelos ApoE 4 es más probable que desarrolle la enfermedad de Alzheimer que un sujeto que no porta un alelo ApoE 4.

15 "Polimorfismo" según se utiliza en la presente memoria se refiere a la existencia de dos o más secuencias de nucleótidos diferentes en un locus concreto en el ADN del genoma. Los polimorfismos pueden servir como marcadores genéticos y también puede ser denominado como variantes genéticas. Los polimorfismos incluyen sustituciones, inserciones, deleciones de nucleótidos y microsatélites, y pueden, pero no es necesario, dar como resultado diferencias detectables en la expresión génica o la función de la proteína. Un sitio polimórfico es una
20 posición de nucleótido dentro de un locus en el que la secuencia de nucleótidos varía con respecto a una secuencia de referencia en al menos un individuo en una población. Un "polimorfismo de deleción/inserción" o "DIP" según se utiliza en la presente memoria es una inserción de uno o más nucleótidos en una versión de una secuencia con respecto a otra. Si se sabe cuál de los alelos representa el alelo minoritario, el término "deleción" se utiliza cuando el alelo minoritario es una deleción de un nucleótido, y el término "inserción" se utiliza cuando el alelo minoritario es una adición de un nucleótido. El término "polimorfismo de deleción/inserción" también se utiliza cuando existen
25 múltiples formas o longitudes y el alelo minoritario no es evidente. Por ejemplo, para los polimorfismos de poli-T descritos en la presente memoria, se observan múltiples longitudes de polimorfismos.

30 "Datos de polimorfismo" según se utiliza en la presente memoria significa la información relacionada con uno o más de los siguientes para un gen específico: la ubicación de los sitios polimórficos; la variación de la secuencia en esos sitios; la frecuencia de polimorfismos en una o más poblaciones; los diferentes genotipos y/o haplotipos determinados para el gen; la frecuencia de uno o más de estos genotipos y/o haplotipos en una o más poblaciones; y cualquiera de las asociaciones conocidas entre un rasgo y un genotipo o haplotipo para el gen.

35 "Haplotipo" según se utiliza en la presente memoria se refiere a una variante genética o combinación de variantes portadas en al menos uno de los cromosomas en un individuo. Un haplotipo a menudo incluye múltiples loci polimórficos contiguos. Todas las partes de un haplotipo según se utiliza en la presente memoria se producen en la misma copia de un cromosoma o una molécula de ADN haploide. A falta de pruebas de lo contrario, se presume que un haplotipo representa una combinación de múltiples loci que es probable que sean transmitidos juntos durante la meiosis. Cada ser humano porta un par de haplotipos para cualquier locus genético dado, que consiste en secuencias heredadas en los cromosomas homólogos de los dos padres. Estos haplotipos pueden ser idénticos o
40 pueden representar dos variantes genéticas diferentes para el locus dado. La haplotipificación es un procedimiento para determinar uno o más haplotipos en un individuo. La haplotipificación puede incluir el uso de las genealogías familiares, técnicas moleculares y/o inferencia estadística.

45 Una "variante", "varianza," o "variante genética" según se utiliza en la presente memoria, se refiere a una isoforma específica de un haplotipo en una población, difiriendo la forma específica de otras formas del mismo haplotipo en la secuencia de al menos uno, y con frecuencia más de uno, sitios de variante o nucleótidos dentro de la secuencia del gen. Las secuencias de estos sitios variantes que difieren entre los diferentes alelos de un gen se denominan "variantes de la secuencia génica", "alelos", "variaciones" o "variantes". El término "forma alternativa" se refiere a un alelo que se puede distinguir de otros alelos por tener al menos uno, y con frecuencia más de uno, sitios de variantes dentro de la secuencia del gen. Otros términos conocidos en la técnica por ser equivalentes a "variaciones" o "variantes" incluyen mutaciones y polimorfismos de un único nucleótido (SNP). La referencia a la presencia de una
50 variación o variaciones significa variaciones concretas, es decir, nucleótidos concretos en sitios polimórficos concretos, en lugar de solo la presencia de cualquier variación en el gen.

55 "Isoforma" según se utiliza en la presente memoria significa una forma concreta de un gen, ARNm, ADNc o la proteína codificada por los mismos, que se distingue de otras formas por su secuencia y/o estructura concretas. Por ejemplo, la isoforma ApoE 4 de la apolipoproteína E en oposición a las isoformas apoE2 o ApoE 3.

"Cistrón" según se utiliza en la presente memoria significa una sección de ADN que se encuentra en un solo cromosoma que contiene el código genético para un único polipéptido y funciona como una unidad hereditaria. Un cistrón incluye exones, intrones, y elementos reguladores relacionados con una sola unidad funcional (es decir, un gen). El término deriva del ensayo cis-trans clásico para determinar si los elementos genéticos eran capaces de

interactuar funcionalmente con independencia de si estaban localizados en la misma molécula de ADN (complementación "trans") o solo cuando estaban localizados en la misma molécula de ADN (elementos que actúan en "cis").

5 El término "genotipo" en el contexto de esta invención se refiere a la forma alélica concreta de un gen, que puede ser definida por el nucleótido o los nucleótidos concretos presentes en una secuencia de ácido nucleico en uno o varios sitios concretos. El genotipo también puede indicar el par de alelos presentes en uno o más loci polimórficos. Para los organismos diploides, tales como los seres humanos, dos haplotipos constituyen un genotipo.

10 La genotipificación es cualquier proceso para la determinación de un genotipo de un individuo, p. ej., mediante amplificación de ácido nucleico, unión de anticuerpos, u otro análisis químico. El genotipo resultante puede ser "unphased", lo que significa que no se sabe si las secuencias encontradas derivan de un cromosoma parental o del otro.

15 "Desequilibrio de ligamiento" según se utiliza en la presente memoria significa la asociación no aleatoria de los alelos en dos o más loci. El desequilibrio de ligamiento describe una situación en la que algunas combinaciones de alelos o marcadores genéticos se producen más o menos frecuentemente en una población de lo que cabría esperar de una formación aleatoria de haplotipos a partir de alelos basándose en sus frecuencias. Las asociaciones no aleatorias entre polimorfismos en diferentes loci se miden por el grado de desequilibrio de ligamiento.

20 "Alineamiento de secuencias múltiples" o "MSA" según se utiliza en la presente memoria significa el alineamiento de tres o más secuencias de nucleótidos a partir de ADN genómico derivado de una pluralidad de individuos para determinar la homología y heterología entre las secuencias. En general, se supone que el conjunto de entrada de secuencias problema tiene una relación evolutiva por la que comparten un linaje y descienden de un antepasado común. Los algoritmos informáticos son los más utilizados para realizar el análisis de las secuencias alineadas.

25 Algunas realizaciones de la presente invención se describen con referencia a diagramas de bloques que ilustran los métodos (p. ej., Figura 1), que pueden incluir etapas ejecutadas por un ordenador y/o productos de programas de ordenador. Se entenderá que cada bloque de los diagramas de bloques y/o las ilustraciones operativas, y la combinación de bloques de los diagramas de bloques y/o las ilustraciones operativas, se pueden implementar por medio de un equipo analógico y/o digital, y/o las instrucciones de programa informático. Estas instrucciones del programa informático se pueden proporcionar a un procesador de un ordenador de propósito general, ordenador de propósito especial, ASIC, y/u otro aparato de procesamiento de datos programable, de tal manera que las instrucciones, que se ejecutan a través del procesador del ordenador y/u otro dispositivo de procesamiento de datos programable, crean los medios para implementar las funciones/actos especificados en los diagramas de bloques y/o ilustraciones operativas. En consecuencia, se apreciará que los diagramas de bloques y las ilustraciones operativas se apoyan en dispositivos, métodos y productos de programas informáticos.

30 También se puede incluir otro soporte lógico, tal como un sistema operativo. Se apreciará adicionalmente que la funcionalidad del módulo de alineamiento de secuencias múltiples, el módulo de mapeo y/u otros módulos descritos en la presente memoria se pueden incorporar, al menos en parte, utilizando componentes discretos del equipo, uno o más circuitos integrados de aplicación específica (ASIC) y/o uno o más procesadores digitales y/u ordenadores para fines especiales.

35 "Mapeo" según se utiliza en la presente memoria significa la creación de un árbol filogenético mediante la asignación de un nodo a cada nueva variante de la secuencia de nucleótidos observada, conectando ese nodo a otro nodo que representa una secuencia conocida portada por el mismo individuo en el mismo cromosoma o cistrón, y contando el número de cada tipo de sujetos representado en cada nodo. Véase la Figura 4 para un ejemplo de un árbol filogenético desarrollado de esta manera.

40 "Filogenético" significa relacionado con el estudio de las conexiones evolutivas entre diferentes grupos de organismos o individuos dentro de una especie. Antes de que la información genética estuviera fácilmente disponible, la filogenia se basaba principalmente en la observación fenotípica. El "Mapeo filogenético", según se utiliza en la presente memoria, significa la utilización de datos de secuencia de ADN para conectar variantes de secuencia relacionadas portadas por una pluralidad de individuos con el fin de determinar las conexiones evolutivas y la cronología de la divergencia. Un "árbol filogenético" es el resultado del mapeo de las conexiones entre las variantes.

45 "Nodo", según se utiliza en la presente memoria significa un punto de datos de polimorfismo en un árbol filogenético que representa una secuencia variante real portada por al menos un sujeto. Un nodo está conectado por medio de una rama a otro nodo que representa una secuencia variante portada por el mismo individuo en el mismo cromosoma y en el mismo cistrón pero en un locus genético diferente dentro del cistrón. La presencia de un nodo indica que al menos un sujeto portaba tanto la secuencia indicada por el nodo, como la secuencia representada por el nodo vecino al que está conectado por una rama.

50 "Rama" según se utiliza en la presente memoria significa una conexión entre dos nodos que representan dos secuencias variantes distintas o haplotipos, en donde las dos variantes se encuentran en el mismo cromosoma y en el mismo cistrón de un sujeto individual. "Punto de ramificación" significa cualquier nodo desde el que se extienden

más de dos ramas, pero se utiliza especialmente en la presente memoria para referirse a un nodo raíz a partir del cual se extienden tres o más nodos. Un "nodo raíz" representa la secuencia genética de un ancestro evolutivo común a partir del cual la divergencia genética ha generado la variedad de variantes de secuencia próximas representadas por los nodos conectados.

5 "Iterativamente" según se utiliza en la presente memoria, se refiere al cálculo repetitivo de valores para cada carácter en una serie. Por ejemplo, cada nodo en un árbol filogenético se analiza para calcular la razón del número de sujetos afectados con una afección de interés (tal como la enfermedad de Alzheimer) con respecto a los sujetos control no afectados; esta razón se compara con los nodos conectados para localizar correlaciones con aumento o disminución del riesgo de desarrollar una enfermedad, trastorno, o afección de interés. Un cambio sustancial en esta razón entre un nodo y el siguiente indica la presencia de una variante que aumenta o disminuye el riesgo de un inicio más temprano de la enfermedad. "Examinar iterativamente las variantes genéticas" significa comenzar el análisis con nodos que representan las secuencias compartidas por el mayor número de sujetos individuales y analizar sucesivamente los nodos conectados por ramas que se extienden desde ese nodo, seguido del segundo nivel de nodos, y así sucesivamente. El análisis se desplaza a continuación en general de las raíces del árbol hacia las ramas exteriores y nodos del árbol.

"Tratamiento" según se utiliza en la presente memoria incluye cualquier fármaco, procedimiento, cambio de estilo de vida, u otro ajuste introducido en un intento de efectuar un cambio en un aspecto concreto de la salud de un sujeto (es decir, dirigido a una enfermedad, trastorno, o afección concretos).

20 "Fármaco" según se utiliza en la presente memoria se refiere a una entidad química o producto biológico, o combinación de entidades químicas o productos biológicos, administrados a una persona para tratar o prevenir o controlar una enfermedad o afección. El término "fármaco" según se utiliza en la presente memoria es sinónimo de los términos "medicina", "medicamento", "intervención terapéutica" o "producto farmacéutico". Lo más preferentemente, el fármaco es aprobado por una agencia gubernamental para el tratamiento de al menos una enfermedad o afección específicas.

25 "Enfermedad", "trastorno", y "afección" son comúnmente reconocidos en la técnica y designan la presencia de signos y/o síntomas en un individuo o paciente que son generalmente reconocidos como anómalos y/o no deseables. Las enfermedades o afecciones pueden ser diagnosticadas y clasificadas basándose en cambios patológicos. La enfermedad o la afección se pueden seleccionar entre los tipos de enfermedades enumerados en los textos convencionales tales como Harrison Principios of Internal Medicine, 1997, o Robbins Patologic Basis of Diseases, 1998.

30 "Disfunción mitocondrial" según se utiliza en la presente memoria significa cualquier anomalía perjudicial de las mitocondrias dentro de una célula o células. Algunas enfermedades, trastornos o afecciones actualmente conocidos en la técnica por estar asociados con la disfunción mitocondrial incluyen la enfermedad de Alzheimer, enfermedad de Parkinson, y otras enfermedades neurodegenerativas, la lesión por isquemia-reperusión en el accidente cerebrovascular y ataque al corazón, la epilepsia, la diabetes, y el envejecimiento. Muchas otras enfermedades, trastornos y afecciones se han asociado con la disfunción mitocondrial en la técnica. De hecho, la mitocondria es crítica para el funcionamiento adecuado de la mayoría de tipos de células, y la disminución mitocondrial a menudo conduce a la muerte celular. Esta disfunción mitocondrial causa lesión daño y muerte celular por comprometer la producción de ATP, interrumpir la homeostasis del calcio y aumentar el estrés oxidativo. Además, la lesión mitocondrial puede conducir a la muerte celular por apoptosis, causando la liberación de citocromo C y otros factores pro-apoptóticos en el citoplasma (para una revisión, véanse Wallace, 1999; Schapira, 2006). En cuanto a un ejemplo específico que se encuentra en la presente memoria, se especula que las isoformas ApoE 3 y ApoE 4 ocasionan disfunción mitocondrial a través de interacciones con TOMM40. Algunas variantes de TOMM40 pueden actuar sinérgicamente con la isoforma ApoE 3 para acelerar la disminución mitocondrial. Se cree que este mecanismo mitocondrial contribuye a muchas enfermedades, trastornos y afecciones genéticos complejos.

35 "Sujetos" según se utiliza en la presente memoria son preferiblemente, pero no se limitan a, los sujetos humanos. Los sujetos pueden ser hombres o mujeres y pueden ser de cualquier raza o grupo étnico, incluyendo, pero no limitado a, caucásico, afroamericano, africano, asiático, hispano, indio, etc. Los sujetos pueden tener cualquier edad, incluyendo recién nacidos, neonatos, bebés, niños, adolescentes, adultos y ancianos. Los sujetos también pueden incluir sujetos animales, en particular sujetos mamíferos, tales como cánidos, félidos, bóvidos, cápridos, équidos, óvidos, súidos, roedores (p. ej., ratas y ratones), lagomorfos, primates (incluyendo primates no humanos), etc., escrutados con fines de desarrollo de fármacos de medicina veterinaria o farmacéuticos.

40 "Tratar", "tratando" o "tratamiento" según se utiliza en la presente memoria se refiere a cualquier tipo de medida que confiere un beneficio a un paciente aquejado de una enfermedad, incluyendo la mejora en el estado del paciente (p. ej., en uno o más síntomas), el retraso en la aparición o el progreso de la enfermedad, etc.

45 La "Enfermedad de Alzheimer de inicio tardío" o "EAT" según se utiliza en la presente memoria es conocida en la técnica, y es la clasificación utilizada si la enfermedad de Alzheimer tiene un inicio o se diagnostica después de la edad de 65. Es la forma más común de la enfermedad de Alzheimer.

2. Métodos para la identificación de variantes genéticas

Si bien las listas de las asociaciones derivadas de los barridos de todo el genoma son útiles, son generalmente inadecuadas para explicar la complejidad de la enfermedad. Las familias, las rutas y las interacciones de los genes pueden proporcionar especificidades. El mapeo de variaciones de alta resolución puede revelar respuestas a interacciones genéticas complejas. Esto es particularmente aplicable cuando un factor de riesgo genético conocido que no explica completamente por sí mismo una asociación con la enfermedad, trastorno, o afección de interés puede presentar un excelente locus genético candidato para investigaciones más detalladas. Por otra parte, la farmacogenética, si bien es útil para el desarrollo de fármacos, puede también ampliar la relevancia biológica. El análisis de los datos de secuencia de un gran número de individuos para descubrir las variaciones en la secuencia génica entre los individuos en una población dará como resultado la detección de una mayor fracción de todas las variantes en la población.

La información de la secuencia inicial que se va a analizar por medio del método de la presente invención deriva del ADN genómico de una pluralidad de sujetos. El organismo puede ser cualquier organismo para el que están disponibles múltiples secuencias, pero es preferiblemente de ser humano. Al identificar nuevas variaciones a menudo es útil escrutar diferentes grupos de población basándose en la raza, el grupo étnico, el género, y/o el origen geográfico, debido a que las variaciones concretas pueden diferir en la frecuencia entre tales grupos. Lo más preferiblemente, para las enfermedades o trastornos que se cree que multigénicos (enfermedades/trastornos genéticamente complejos), los fenotipos representados por la población de sujetos son de los extremos de un espectro. Las muestras biológicas que contienen ADN pueden ser sangre, semen, frotis bucal, etc. El aislamiento de ADN a partir de tales muestras es bien conocido en la técnica.

En algunas realizaciones, la invención se refiere al análisis de datos de la secuencia de nucleótidos de una pluralidad de sujetos que tienen al menos un factor de riesgo conocido para una enfermedad, trastorno, o afección dados (genético o de otro tipo). Las secuencias de nucleótidos se analizan para generar datos de haplotipos, y los haplotipos o variantes genéticas se mapean en un árbol filogenético para demostrar la evolución de las secuencias representadas. Al comparar este árbol con los datos del fenotipo acerca de la pluralidad de sujetos, es posible un pronóstico o diagnóstico para un sujeto individual que porta haplotipos observados en el árbol filogenético. En otras realizaciones, la invención se refiere a los campos de la farmacogenética y la farmacogenómica y al uso de la información del haplotipo genético para pronosticar la susceptibilidad de un individuo a una enfermedad y/o su respuesta a un fármaco o fármacos concretos, de manera que se puedan desarrollar fármacos adaptados a las diferencias genéticas de los grupos de población y/o administrar a individuos con el perfil genético apropiado. La información de la secuencia de nucleótidos deriva de ADN genómico. Los datos de la secuencia genómica utilizados se pueden obtener a partir de animales clínicos o no humanos o a partir de células cultivadas o estudios de tejidos aislados. El organismo puede ser cualquier organismo para el que cual se encuentran disponibles múltiples secuencias, pero es preferiblemente humano. En la identificación de las nuevas variaciones a menudo es útil escrutar diferentes grupos de población basándose en la raza, grupo étnico, género, y/u origen geográfico, debido a que las variaciones concretas pueden diferir en la frecuencia entre tales grupos. Lo más preferiblemente, para las enfermedades o trastornos que se cree que son multigénicos (enfermedades/trastornos genéticamente complejos), los fenotipos representados por la población sujeto están en extremos opuestos.

Las muestras biológicas que contienen ADN pueden ser sangre, semen, frotis bucal, etc. El aislamiento de ADN a partir de tales muestras es bien conocido en la técnica. Los métodos para la determinación de la secuencia de ADN en un locus genético concreto de interés son también conocidos en la técnica. La secuenciación automática está ahora ampliamente disponible y requiere solo una muestra de ADN aislado y al menos un cebador que esté diseñado específicamente para reconocer una secuencia altamente conservada dentro de o en estrecha proximidad con el locus genético de interés.

De acuerdo con algunas realizaciones, una región genética definida o locus de interés (p. ej., que se define por un conjunto de cebadores de PCR directos e inversos) se secuencia cuidadosamente a partir de una cohorte de personas que incluyen pacientes que están bien caracterizados para un trastorno concreto.

Se determina una secuencia consenso, y todas las variantes de secuencia observadas para un locus genético dado se recopilan en una lista. Los loci que tienen el mayor número de variantes observadas representan la divergencia evolutiva de un ancestro común. Como tales, estos loci están conectados en cis a loci que tienen solamente una o muy pocas variantes observadas. Durante las fases iniciales de la investigación al menos, se prefiere que las poblaciones se analicen en grupos de sujetos que comparten un fenotipo general común que representa una ascendencia similar. De otro modo, el análisis de estos datos a través de la construcción de la construcción de árboles filogenéticos requerirá un número de sujetos prohibitivamente grande.

3. Alineamientos de secuencias múltiples.

La determinación de la presencia de una variación, o una pluralidad de variaciones concretas en un gen o región génica de una población se puede realizar de diferentes maneras, todas las cuales implican la localización de un locus genético concreto, mediante direccionamiento de secuencias dentro de la región de interés que se sabe que está altamente conservada. A partir del locus altamente conservado, se obtienen fácilmente las secuencias

contiguas a través de una de los muchos mecanismos bien conocidos en la técnica.

La primera etapa en el análisis de secuencias de ADN paralelas de una pluralidad de sujetos es el alineamiento de secuencias múltiples ("MSA"). El MSA se utiliza típicamente para presentar el alineamiento de secuencias de muestras homólogas con diferencias polimórficas en los genes o regiones génicas para mostrar las zonas conservadas y las secuencias variantes. Los MSA de la información de la secuencia obtenida en el locus de interés se pueden construir utilizando una o más técnicas conocidas diversas y el soporte lógico a disposición del público, y están a disposición del público a partir de muchas fuentes, incluyendo Internet. Los métodos para el análisis de alineamientos de secuencias múltiples conocidos en la técnica incluyen, p. ej., los descritos en la Patente de los Estados Unidos 6.128.587 de Sjolander; la Patente de los Estados Unidos 6.291.182 de Schork et al.; y la Patente de los Estados Unidos 6.401.043 de Stanton et al.

4. Árboles y análisis filogenéticos.

Se conocen en la técnica varios métodos para la construcción de "árboles filogenéticos" (véase, p. ej., Sanderson, 2008). Sun et al. utilizaron análisis de "bloques de haplotipos" para estudiar las asociaciones entre las variantes de los receptores tipo Toll (TLR) y el cáncer de próstata (2005) y Bardel et al. (2005) usaron un enfoque de análisis cladístico para investigar la asociación entre variantes del gen CARD15 y la enfermedad de Crohn. Sin embargo, ninguno utilizó loci genéticos previamente asociados con la enfermedad para investigar los ligamientos.

Los árboles filogenéticos de acuerdo con algunas realizaciones se pueden construir con una topología en la que las variantes de secuencia del haplotipo observadas en sujetos humanos individuales estudiados forman nodos (representando cada secuencia observada en los datos) en un árbol. Los nodos pueden estar unidos a otros nodos, y el antepasado común se encuentra en el sitio de ramificación, raíz común o nodo raíz del árbol. Un árbol filogenético refleja la relación evolutiva entre los loci genéticos para los que se analizan los datos (véanse Sanderson, 2008; Tzeng, 2005; Seltman, 2003). La Figura 4 muestra un árbol filogenético detallado construido para la Región B del locus genético mostrado en la Figura 3.

El punto de partida para la estimación de árbol filogenético es generalmente un MSA (véase más arriba). Se encuentran disponibles múltiples aplicaciones de soporte lógico para la construcción de árboles filogenéticos basados en datos de la secuencia. Véanse, p. ej., la Patente de los Estados Unidos 7.127.466 y la Patente de los Estados Unidos 6.532.467 de Brocklebank, et al. La premisa básica es que un locus genético que presenta muchas variantes está representado por estas variantes conectadas en cis. Los polimorfismos crean puntos de ramificación (nodos) en el árbol que definen grupos de secuencias o haplotipos relacionados.

El árbol filogenético se utiliza para la información de forma iterativa mediante el examen de las proporciones de sujetos afectados con una afección con respecto a sujetos de control no afectados; los cálculos comienzan con los nodos observados en el mayor número de sujetos y se desplazan hacia la periferia del árbol hacia los nodos observados en menos sujetos. El objetivo es encontrar un punto de ramificación, rama, o nodo donde haya un cambio sustancial en la proporción de sujetos afectados con la afección de interés con respecto a los sujetos de control no afectados. Tal punto de ramificación representa la divergencia evolutiva de los sujetos de mayor riesgo de los sujetos de menor riesgo o viceversa.

El análisis estadístico del árbol filogenético generado se puede llevar a cabo de acuerdo con los métodos conocidos en la técnica. Un método reconocido en la técnica es el cálculo de los niveles de confianza bootstrap (véase Efron et al., Proc. Natl. Acad. Sci. USA 93, 13429-13434 (1996)).

5. Evaluación del paciente.

Una vez que se ha generado un árbol filogenético para un locus genético concreto, se puede evaluar un sujeto individual mediante la comparación de su secuencia de ADN con las secuencias que comprenden el árbol filogenético. La presencia de haplotipos o variantes de secuencia que corresponden con las regiones del árbol que representan los sujetos con mayor incidencia de la afección de interés (es decir, mayores proporciones de sujetos afectados con la enfermedad o trastorno con respecto a los sujetos de control no afectados) significaría que el sujeto individual también tiene un mayor riesgo. Por el contrario, las razones sustancialmente más bajas corresponden a la reducción del riesgo de desarrollar la afección de interés.

Los árboles filogenéticos también se pueden analizar basándose en la sensibilidad de la afección de interés al tratamiento con un agente activo o un método de tratamiento de interés de acuerdo con algunas realizaciones.

6. APOE y TOMM40.

Los fenotipos y genotipos de ApoE son bien conocidos en la técnica. El sistema de nomenclatura establecido así como los fenotipos y genotipos para ApoE son descritos, por ejemplo, por Zannis et al., 1982.

Los fenotipos y genotipos de TOMM40 (la subunidad del canal de la membrana mitocondrial externa, 40 kDa) también son conocidos. TOMM40 funciona como una subunidad formadora de canal de la translocasa encontrada en la membrana mitocondrial que es esencial para la importación de proteínas a las mitocondrias.

Los de exploración de la asociación de todo el genoma de estudios de pacientes con enfermedad de Alzheimer han identificado de manera inequívoca la región de desequilibrio de ligamiento que contiene el gen de la apolipoproteína E (ApoE). La variante ApoE 4 ha sido ampliamente replicada como un gen de susceptibilidad confirmada desde las primeras publicaciones en 1993 (véase, p. ej., Corder et al.). Sin embargo, los datos de exploración de la asociación de todo el genoma dieron como resultado una "coincidencia" notable observada en los estudios de biología celular que implicaba la co-localización ApoE y TOMM40 en la membrana mitocondrial externa. Este otro gen, TOMM40, se encontró por primera vez durante los estudios de modelado del desequilibrio de ligamiento alrededor ApoE en 1998. Los polimorfismos se localizaron adyacentes a ApoE en una pequeña región de desequilibrio de ligamiento.

ApoE se co-localiza en la membrana mitocondrial externa, lo que sugiere interacciones específicas de la isoforma que conducen a un papel potencial para la apoptosis mitocondrial inducida por ApoE como una etapa temprana en la expresión de la enfermedad de Alzheimer. Los datos biológicos han demostrado que la proporción de mitocondrias móviles en cultivo de células neuronales, así como la velocidad a la que se mueven y la distancia que atraviesan, son factores que afectan al aumento de la apoptosis mitocondrial. Los datos filogenéticos sugieren un efecto genético independiente sobre el desarrollo de la enfermedad de Alzheimer para TOMM40.

ApoE se une específicamente a la mitocondria en cultivos neuronales humanos (Chang, 2005), y la secuenciación de esta región de desequilibrio de ligamiento en cientos de pacientes de enfermedad de Alzheimer y los controles adaptados, combinada con el mapeo de la evolución de la variante genética de TOMM40, define una región de especial interés para la interacciones ApoE-TOMM40, como se muestra en la Figura 3. Estos datos evolutivos apoyan adicionalmente la asociación genética entre ApoE y TOMM40, y sugieren que la disfunción mitocondrial podría ser responsable de la muerte neuronal que se produce lentamente a lo largo de muchos años. La edad de inicio de la distribución de la enfermedad de Alzheimer (véase, p. ej., la patente de los Estados Unidos 6.027.896 de Roses et al.) podría reflejar la herencia de variantes estrechamente vinculadas de dos proteínas que interactúan bioquímicamente que conducen a la expresión clínica de la enfermedad.

Como se detalla en la presente memoria, la interacción entre múltiples haplotipos de variantes de TOMM40 y alelos de ApoE contribuye a la patogénesis de la enfermedad de Alzheimer; en particular, los haplotipos de TOMM40 en el ligamiento al alelo E 3 de ApoE contribuyen a la patogénesis de la enfermedad. Varias de las variantes del gen de TOMM40 evolucionaron solamente conectadas en cis a ApoE 3. (De un modo similar, las variantes específicas de TOMM40 pueden haber evolucionado segregando en cis a ApoE 4 o ApoE 2.) Por lo tanto, cualquier efecto genético añadido de las variantes de TOMM40 segrega independientemente de ApoE 4 pero los dos productos de la proteína variante pueden interactuar funcionalmente, en trans, para producir un fenotipo observable o rasgo dado. Por lo tanto, cualquier efecto genético agregado de las variantes de TOMM40 segrega independientemente de ApoE 4. Esta "coincidencia" de genes que interactúan adyacentes puede representar los datos estadísticos de asociación extraordinariamente importante que se encuentran en todos los estudios de exploración de asociación del genoma completo de la enfermedad de Alzheimer. Es interesante observar que las plataformas de exploración de asociación del genoma completo disponibles en el mercado iniciales no contenían ningún polimorfismo de ApoE, pero fueron identificadas con SNP de TOMM40 y ApoC1 - pero la región es casi siempre referida como "región ApoE".

Estos datos, que combinan la genética de la enfermedad y los supuestos mecanismos moleculares de la patogénesis, también se pueden observar en el contexto de la farmacogenética. Debido al fuerte efecto genético de la herencia del alelo ApoE 4, ApoE 4 ha sido referido como un gen de susceptibilidad complejo durante más de una década. Repeticiones consistentes de la edad de las distribuciones de inicio como una función del genotipo ApoE confirman que el papel de la herencia de ApoE 3 no es totalmente benigno, sino que es un factor de riesgo menor observado a una tasa de inicio de la enfermedad más lenta. Existen variantes genéticas de TOMM40 que se encuentran solo en las cadenas de ADN que contienen ApoE 3 en las regiones de desequilibrio de ligamiento (Roses et al., datos no publicados), y de este modo no en equilibrio de Hardy-Weinberg como se requería para los SNP en paneles de asociación de genoma completo. Los cambios evolutivos en las secuencias de TOMM40 que están conectadas en cis únicamente a ApoE 3 actúan aumentando el riesgo de la enfermedad de Alzheimer asociada con ApoE 3, mientras que otras variantes de TOMM40 conectadas en cis a ApoE 3 disminuyen el riesgo asociado con ApoE 3. Un ensayo genético independiente consistiría en determinar si esos polimorfismos de TOMM40 asociados con menos enfermedad de Alzheimer segregan a una edad más tardía en la edad de los gráficos de distribución del comienzo para los genotipos que contienen ApoE 3 [ApoE 3/3 o ApoE 4/3].

La detección de la presencia o ausencia de haplotipos ApoE 2, 3 o 4, y/o de TOMM40 o del ADN que codifica los mismos (incluyendo, en algunas realizaciones, el número de alelos para cada uno) en un sujeto puede llevarse a cabo ya sea directamente o indirectamente mediante cualquier método adecuado. Una variedad de técnicas son conocidas por los expertos en la técnica. Todas implican generalmente la etapa de recolección de una muestra de material biológico que contiene ADN o la proteína del sujeto, y a continuación, la detección de si el sujeto posee o no el haplotipo de interés. Por ejemplo, la etapa de detección con respecto a la ApoE se puede llevar a cabo recogiendo una muestra de ApoE del sujeto (p. ej., de fluido cerebroespinal, o cualquier otro fluido o tejido que contenga ApoE), y a continuación determinando la presencia o ausencia de una isoforma ApoE 2, 3, o 4 en la muestra de ApoE (p. ej., mediante isoelectroenfoque o inmunoanálisis).

La determinación de la presencia o ausencia de ADN que codifica una isoforma de ApoE y/o TOMM40 se puede llevar a cabo mediante secuenciación directa de la región del ADN genómico de interés, con una sonda

oligonucleotídica marcada con un grupo detectable adecuado, y/o por medio de una reacción de amplificación tal como una reacción en cadena de la polimerasa o reacción en cadena de la ligasa (cuyo producto de la reacción de amplificación se puede detectar a continuación con una sonda oligonucleotídica marcada u otras diversas técnicas). Adicionalmente, la etapa de detección puede incluir la etapa de detección de si el sujeto es heterocigoto u homocigoto para el gen que codifica un haplotipo de ApoE y/o TOMM40. Se conocen numerosos formatos de análisis de sondas oligonucleotídicas que se puede emplear para llevar a cabo la presente invención. Véanse, p. ej., la Patente de los Estados Unidos. Núm. 4.302.204 de Wahl et al.; la Patente de los Estados Unidos. Núm. 4.358.535 de Falkow et al.; la Patente de los Estados Unidos. Núm. 4.563.419 de Ranki et al.; y la Patente de los Estados Unidos. Núm. 4.994.373 de Stavrianopoulos et al. En algunas realizaciones, la detección puede incluir la amplificación múltiple del ADN (p. ej., PCR fluorescente específica de alelo). En algunas realizaciones, la detección puede incluir la hibridación con una micromatriz (chip, cuentas, etc.). En algunas realizaciones, la detección puede incluir la secuenciación de las porciones apropiadas del gen que contienen los haplotipos que se pretende detectar. En algunas realizaciones, se pueden usar para la detección haplotipos que cambian la susceptibilidad a la digestión por una o más enzimas endonucleasas de restricción. Por ejemplo, se puede utilizar un polimorfismo de longitud del fragmento de restricción (RFLP), que se refiere al patrón de digestión cuando se aplican diversas enzimas de restricción al ADN. En algunas realizaciones, la presencia de uno o más haplotipos se puede determinar mediante la amplificación específica del alelo. En algunas realizaciones, la presencia de haplotipos se puede determinar mediante extensión del cebador. En algunas realizaciones, la presencia de haplotipos se puede determinar mediante ligación de oligonucleótidos. En algunas realizaciones, la presencia de haplotipos se puede determinar mediante hibridación con una sonda marcada de forma detectable. Véanse, p. ej., la Publicación de la Solicitud de Patente de los Estados Unidos Núm. 2008/0153088 de Sun et al.; Kobler et al., Identification of an 11T allele in the polypyrimidine tract of intron 8 of the CFTR gene, *Genetics in Medicine* 8(2): 125-8 (2006); Costa et al., Multiplex Allele-Specific Fluorescent PCR for Haplotyping the IVS8 (TG)m(t)N Locus in the CFTR gene, *Clin. Chem.*, 54:1564-1567 (2008); Johnson et al., A Comparative Study of Five Technologically Diverse CFTR Testing Platforms, *J. Mol. Diagnostics*, 9(3) (2007); Pratt et al., Development of Genomic Reference Materials for Cystic Fibrosis Genetic Testing, *J. Mol. Diagnostics*, 11:186-193 (2009).

La amplificación de una secuencia de ácido nucleico seleccionada, o diana, se puede llevar a cabo por medio de cualquier método adecuado sobre ADN aislado de muestras biológicas. Véanse, en general D. Kwoh y T. Kwoh, 1990. Los ejemplos de las técnicas de amplificación adecuadas incluyen, pero no se limitan a, reacción en cadena de polimerasa, reacción en cadena de la ligasa, amplificación por desplazamiento de cadena (véanse, en general Walker et al, 1992a; Walker et al., 1992b), amplificación basada en la transcripción (véase Kwoh et al., 1989), la replicación autosostenida de secuencia (o "3SR") (véase Guatelli et al., 1990), el sistema de la replicasa Q β (véase Lizardi et al., 1988), la amplificación basada en la secuencia de ácido nucleico (o "NASBA") (véase Lewis, 1992), la reacción de reparación en cadena (o "RCR") (véase Lewis, más arriba), y la amplificación de ADN bumerang (o "BDA") (véase Lewis, más arriba). Actualmente se prefiere la reacción en cadena de la polimerasa.

Las técnicas de amplificación de ADN, tales como las anteriores pueden implicar el uso de una sonda, un par de sondas, o dos pares de sondas que se unen específicamente al ADN que codifica ApoE 4, pero no se unen al ADN que codifica ApoE 2 o ApoE 3 en las mismas condiciones de hibridación, y que sirven como cebador o cebadores para la amplificación del ADN de ApoE 4 o una porción del mismo en la reacción de amplificación. Asimismo, se puede utilizar una sonda, un par de sondas, o dos pares de sondas que se unen específicamente al ADN que codifica ApoE 2, pero no se unen a ADN que codifica ApoE 3 o ApoE 4 en las mismas condiciones de hibridación, y que sirven como cebador o cebadores para la amplificación del ADN de ApoE 2 o una porción del mismo en la reacción de amplificación; y se puede utilizar una sonda, un par de sondas, o dos pares de sondas que se unen específicamente al ADN que codifica ApoE 3, pero no se unen a ADN que codifica ApoE 2 o ApoE 4 en las mismas condiciones de hibridación, y que sirven como cebador o cebadores para la amplificación del ADN de ApoE 3 o una porción del mismo en la reacción de amplificación.

Del mismo modo, se puede utilizar una sonda, un par de sondas, o dos pares de sondas que se unen específicamente al ADN que codifica un haplotipo de TOMM40 de interés, pero no se unen a otros haplotipos de TOMM40 en las mismas condiciones de hibridación, y que sirven como cebador o cebadores para la amplificación del ADN TOMM40 o una porción del mismo en la reacción de amplificación. En general, una sonda oligonucleotídica que se utiliza para detectar el ADN que codifica los haplotipos de ApoE y/o TOMM40 es una sonda oligonucleotídica que se une al ADN que codifica el haplotipo de interés, pero no se une al ADN que codifica otros haplotipos en las mismas condiciones de hibridación. La sonda oligonucleotídica se marca con un grupo detectable adecuado, tal como el expuesto más abajo en conexión con los anticuerpos.

La reacción en cadena de la polimerasa (PCR) se puede llevar a cabo de acuerdo con técnicas conocidas. Véanse, p. ej., las Patente de los Estados Unidos. Núms. 4.683.195; 4.683.202; 4.800.159; y 4.965.188. En general, la PCR implica, tratar primero una muestra de ácido nucleico (p. ej., en presencia de una ADN polimerasa térmicamente estable) con un cebador oligonucleotídico para cada hebra de la secuencia específica que se va a detectar en las condiciones de hibridación, de modo que se sintetice un producto de extensión de cada cebador que es complementario a cada hebra de ácido nucleico, con los cebadores suficientemente complementarios a cada hebra de la secuencia específica para hibridarse con ella de modo que el producto de extensión sintetizado a partir de cada cebador, cuando se separa de su complemento, pueda servir como un molde para la síntesis del producto de extensión del otro cebador, y a continuación tratar la muestra en condiciones de desnaturalización para separar los

5 productos de extensión del cebador de sus moldes si la secuencia o secuencias que se van a detectar están presentes. Estas etapas se repiten cíclicamente hasta que se obtiene el grado deseado de amplificación. La detección de la secuencia amplificada puede llevarse a cabo mediante la adición al producto de reacción de una sonda oligonucleotídica capaz de hibridarse con el producto de reacción (p. ej., una sonda oligonucleotídica de la presente invención), llevando la sonda una marca detectable, y a continuación la detección de la marca de acuerdo con técnicas conocidas, o mediante visualización directa sobre un gel.

10 Cuando las condiciones de PCR permiten la amplificación de todos los tipos alélicos de ApoE, los tipos se pueden distinguir mediante hibridación con una sonda específica del alelo, mediante digestión con endonucleasas de restricción, mediante electroforesis en gel de gradiente desnaturante, u otras técnicas. Un protocolo de PCR para determinar el genotipo ApoE es descrito por Wenham et al. (1991). Los ejemplos de los cebadores eficaces para la amplificación e identificación de las isoformas de ApoE se describen allí. Se pueden emplear cebadores específicos para la región polimórfica de ApoE (ya sea ApoE 4, E3 o E2). En Wenham, por ejemplo, se emplean cebadores de PCR que amplifican una región de 227 pb de ADN que abarca los sitios polimórficos de ApoE (codones 112 y 158, que contienen los nucleótidos 3745 y 3883). Los fragmentos amplificados se someten a continuación a la endonucleasa de restricción CfoI que proporciona diferentes fragmentos de restricción de los seis posibles genotipos de ApoE que pueden ser reconocibles en un gel de electroforesis. Véanse también, Hixon et al. (1990); Houlston et al. (1989); Wenham et al. (1991); y Konrula et al. (1990) para métodos adicionales. Además de la enfermedad de Alzheimer, existen otras diversas enfermedades y trastornos genéticamente complejos para los que los métodos de la presente invención proporcionan ventajas sobre los análisis existentes. Por ejemplo, los datos de múltiples estudios genéticos de diabetes mellitus tipo 2 apoyan la opinión de que será necesaria una gran serie de casos clínicos/de control para proporcionar significación estadística para loci definidos por los estudios de asociación de todo el genoma.

7. Agentes activos, composiciones y tratamiento.

25 Como se señaló anteriormente, los árboles filogenéticos creados utilizando los métodos detallados en la presente memoria también pueden ser analizados basándose en la capacidad de respuesta de la afección de interés al tratamiento con un agente activo o al método de tratamiento de interés de acuerdo con algunas realizaciones, y las decisiones de tratamiento para un sujeto o paciente se pueden basar en las variantes genéticas específicas identificadas.

30 Agentes activos. Los agentes activos incluyen los conocidos para el tratamiento de una afección de interés, e incluyen agentes activos anti-enfermedad de Alzheimer, incluyendo, pero no limitados a, inhibidores de la acetilcolinesterasa, antagonistas del receptor de NMDA, y agonistas o moduladores del receptor activado por el proliferador de peroxisoma (PPAR), incluyendo pero no limitados a los fármacos de las clases de tiazolidinodiona o glitazar. El agente activo también podría ser un producto biofarmacéutico, por ejemplo un anticuerpo (p. ej., anticuerpos monoclonales, policlonales, derivados o modificados tales como Anticuerpos de Dominio™, bapineuzumab, etc.), proteínas de fusión o moléculas de ARN terapéuticas. El agente activo también podría ser una combinación de cualquiera de estos productos.

40 Los ejemplos de los inhibidores de la acetilcolinesterasa incluyen, pero no se limitan a, donepezilo (disponible comercialmente como Aricept), galantamina (disponible comercialmente como RAZADYNE), y rivastigmina (disponible comercialmente como Exelon) y las sales farmacéuticamente aceptables de los mismos. Los ejemplos adicionales incluyen, pero no se limitan a, los descritos en las Patentes de los Estados Unidos Núms. 6.303.633; 5.965.569; 5.595.883; 5.574.046; y 5.171.750.

45 Los ejemplos de los antagonistas de los receptores de NMDA incluyen, pero no se limitan a, memantina (disponible comercialmente como AKATINOL, AXURA, EBIXIA/ABIXIA, MEMOX y NAMENDA) y las sales farmacéuticamente aceptables de los mismos. Los ejemplos adicionales incluyen, pero no se limitan a, los descritos en las Patentes de los Estados Unidos Núms. 6.956.055; 6.828.462; 6.642.267; 6.432.985; y 5.990.126.

Los ejemplos de las tiazolidinodionas incluyen, pero no se limitan a, rosiglitazona (disponible comercialmente como AVANDIA) y las sales farmacéuticamente aceptables de la misma. Los ejemplos adicionales incluyen, pero no se limitan a:

50 5-(4-[2-(N-metil-N-(2-benzotiazolil)amino)etoxi]bencil)-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-(2-benzotiazolil)amino)etoxi]benciliden)-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-(2-benzoxazolil)amino)etoxi]bencil)-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-(2-benzoxazolil)amino)etoxi]benciliden)-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-(2-pirimidinil)amino)etoxi]bencil)-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-(2-pirimidinil)amino)etoxi]benciliden)-2,4-tiazolidinodiona; 5-(4-(2-(N-metil-N-[2-(4,5-dimetiltiazolil)]amino)etoxi]bencil)-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-[2-(4,5-dimetiltiazolil)]amino)etoxi]benciliden)-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-(2-tiazolil)amino)etoxi]bencil)-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-(2-tiazolil)amino)etoxi]benciliden)-2,4-tiazolidinodiona; 5-(4-(2-(N-metil-N-(2-(4-feniltiazolil)]amino)etoxi]bencil)-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-(2-(4-feniltiazolil)]amino)etoxi]benciliden)-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-[2-(4-fenil-5-metiltiazolil)]amino)etoxi]bencil)-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-[2-(4-fenil-5-metiltiazolil)]amino)etoxi]benciliden)-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-[2-(4-metiltiazolil)]amino)etoxi]bencil)-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-[2-(4-metiltiazolil)]amino)etoxi]benciliden)-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-[2-(4-metiltiazolil)]amino)etoxi]bencil)-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-[2-(4-metiltiazolil)]amino)etoxi]benciliden)-2,4-tiazolidinodiona; 5-(4-(2-(N-metil-N-

[2-(5-feniloxazolil)]amino)etoxi]bencil]-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-[2-(5-feniloxazolil)]amino)etoxi]benciliden)-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-[2-(4,5-dimetiloxazolil)]amino)etoxi]bencil)-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-[2-(4,5-dimetiloxazolil)]amino)etoxi]benciliden)-2,4-tiazolidinodiona; 5-[4-(2-(2-pirimidinilamino)etoxi) bencil]-2,4-tiazolidinodiona; 5-[4-(2-(2-pirimidinilamino)etoxi) benciliden]-2,4-tiazolidinodiona; 5-(4-[2-(N-acetil-N-(2-pirimidinil)amino)etoxi]bencil)-2,4-tiazolidinodiona; 5-(4-(2-(N-(2-benzotiazolil)-N-bencilamino)etoxi) benciliden)-2,4-tiazolidinodiona; 5-(4-(2-(N-(2-benzotiazolil)-N-bencilamino)etoxi) bencil)-2,4-tiazolidinodiona; 5-(4-[3-(N-metil-N-(2-benzoxazolil)amino)propoxi]bencil)-2,4-tiazolidinodiona; 5-(4-[3-(N-metil-N-(2-benzoxazolil)amino)propoxi]benciliden)-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-(2-piridil)amino)etoxi]bencil)-2,4-tiazolidinodiona; 5-(4-[2-(N-metil-N-(2-piridil)amino)etoxi]benciliden)-2,4-tiazolidinodiona; 5-(4-[4-(N-metil-N-(2-benzoxazolil)amino)butoxi]benciliden)-2,4-tiazolidinodiona; 5-(4-[4-(N-metil-N-(2-benzoxazolil)amino)butoxi]bencil)-2,4-tiazolidinodiona; 5-(4-[2-(N-(2-benzoxazolil)amino)etoxi]benciliden)-2,4-tiazolidinodiona; 5-(4-[2-(N-(2-benzoxazolil)amino)etoxi]bencil)-2,4-tiazolidinodiona; 5-(4-[2-(N-isopropil-N-(2-benzoxazolil)amino)etoxi]bencil)-2,4-tiazolidinodiona, y las sales farmacéuticamente aceptables de los mismos. Véase, por ejemplo, la Patente de los Estados Unidos Núm. 5.002.953.

Los agentes activos descritos en la presente memoria se pueden preparar, como se señaló anteriormente, en forma de sus sales farmacéuticamente aceptables. Las sales farmacéuticamente aceptables son sales que conservan la actividad biológica deseada del compuesto parental y no confieren efectos toxicológicos no deseados. Los ejemplos de tales sales son (a) sales de adición de ácido formadas con ácidos inorgánicos, por ejemplo ácido clorhídrico, ácido bromhídrico, ácido sulfúrico, ácido fosfórico, ácido nítrico y similares; y sales formadas con ácidos orgánicos tales como, por ejemplo, ácido acético, ácido oxálico, ácido tartárico, ácido succínico, ácido maleico, ácido fumárico, ácido glucónico, ácido cítrico, ácido málico, ácido ascórbico, ácido benzoico, ácido tánico, ácido palmítico, ácido algínico, ácido poliglutámico, ácido naftalenosulfónico, ácido metanosulfónico, ácido p-toluenosulfónico, ácido naftalenodisulfónico, ácido poligalacturónico, y similares; (b) sales formadas a partir de aniones elementales tales como cloro, bromo, y yodo, y (c) sales derivadas de bases, tales como sales de amonio, sales de metales alcalinos tales como las de sodio y potasio, sales de metales alcalinotérreos tales como las de calcio y magnesio, y sales con bases orgánicas tales como diciclohexilamina y N-metil-D-glucamina.

Los agentes activos se pueden administrar como profármacos. "Profármacos" según se utiliza en la presente memoria se refiere a aquellos profármacos de los compuestos de la presente invención que son, dentro del alcance del criterio médico lógico, adecuadas para uso en contacto con los tejidos de seres humanos y animales inferiores sin toxicidad, irritación, respuesta alérgica indebidas y similares, conmensurados con una razón riesgo/beneficio razonable, y eficaces para su uso pretendido, así como las formas zwitteriónicas, cuando sea posible, de los compuestos de la invención. El término "profármaco" se refiere a compuestos que se transforman rápidamente in vivo para producir el compuesto parental de las fórmulas anteriores, por ejemplo, por hidrólisis en la sangre. Una discusión detallada es proporcionada por T. Higuchi y V. Stella, en *Prodrugs as Novel Delivery Systems*, Vol. 14 de la ACS Symposium Series, y por Edward B. Roche, ed., en *Bioreversible Carriers in Drug Design*, American Pharmaceutical Association and Pergamon Press, 1987. Véase también la Patente de los Estados Unidos Núm. 6.680.299. Los ejemplos incluyen un profármaco que es metabolizado in vivo por un sujeto a un fármaco activo que tiene la actividad de los compuestos activos descritos en la presente memoria, en donde el profármaco es un éster de un grupo alcohol o ácido carboxílico, si tal grupo está presente en el compuesto; un acetal o cetal de un grupo alcohol, si tal grupo está presente en el compuesto; una base de N-Mannich o una imina de un grupo amina, si tal grupo tal está presente en el compuesto; o una base de Schiff, oxima, acetal, éster de enol, oxazolidina, o tiazolidina de un grupo carbonilo, si tal grupo tal está presente en el compuesto, tal como se describe en la Patente de los Estados Unidos Núm. 6.680.324 y la Patente de los Estados Unidos Núm. 6.680.322.

Composiciones. Los agentes activos descritos anteriormente pueden formularse para la administración en un vehículo farmacéutico de acuerdo con técnicas conocidas. Véase, p. ej., Remington, *The Science and Practice of Pharmacy* (9ª Ed. 1995). En la fabricación de una formulación farmacéutica de acuerdo con la invención, el compuesto activo (incluyendo las sales fisiológicamente aceptables del mismo) se mezcla típicamente, entre otros, con un vehículo aceptable. El vehículo debe ser, por supuesto, aceptable en el sentido de ser compatible con cualquier otro ingrediente de la formulación y no debe ser perjudicial para el paciente. El vehículo puede ser un sólido o un líquido, o ambos, y preferiblemente se formula con el compuesto en forma de una formulación de dosis unitaria, por ejemplo, un comprimido, que puede contener de 0,01 o 0,5% a 95% o 99% en peso del compuesto activo. Uno o más compuestos activos pueden incorporarse a las formulaciones de la invención, que se puede preparar mediante cualquiera de las técnicas bien conocidas de farmacia que comprenden mezclar los componentes, incluyendo opcionalmente uno o más ingredientes accesorios.

Las formulaciones de la invención incluyen las adecuadas para la administración oral, rectal, tópica, bucal (p. ej., sublingual), vaginal, parenteral (p. ej., subcutánea, intramuscular, intradérmica, o intravenosa), tópica (es decir, superficies tanto de la piel como mucosas, incluyendo superficies de las vías respiratorias) y transdérmica, aunque la ruta más adecuada en cualquier caso dado dependerá de la naturaleza y gravedad de la afección que se esté tratando y de la naturaleza del compuesto activo concreto que se está utilizando.

Las formulaciones adecuadas para la administración oral pueden presentarse en unidades discretas, tales como cápsulas, sellos, grageas o comprimidos, conteniendo cada uno una cantidad predeterminada del compuesto activo; en forma de un polvo o de gránulos; en forma de una solución o una suspensión en un líquido acuoso o no acuoso;

o en forma de una emulsión de aceite-en-agua o agua-en-aceite. Tales formulaciones se pueden preparar mediante cualquier método de farmacia adecuado que incluye la etapa de asociar el compuesto activo y un vehículo adecuado (que puede contener uno o más ingredientes accesorios como se ha indicado anteriormente). En general, las formulaciones de la invención se preparan mezclando uniforme e íntimamente del compuesto activo con un vehículo líquido o sólido finamente dividido, o ambos, y a continuación, si es necesario, conformando la mezcla resultante. Por ejemplo, se puede preparar un comprimido por compresión o moldeo de un polvo o gránulos que contienen el compuesto activo, opcionalmente con uno o más ingredientes accesorios. Los comprimidos se pueden preparar comprimiendo, en una máquina adecuada, el compuesto en una forma de flujo libre, tal como un polvo o gránulos opcionalmente mezclados con un aglutinante, lubricante, diluyente inerte, y/o uno o varios agentes tensioactivos/dispersantes. Los comprimidos moldeados se pueden elaborar moldeando, en una máquina adecuada, el compuesto en polvo humedecido con un aglutinante líquido inerte.

Las formulaciones adecuadas para la administración bucal (sublingual) incluyen grageas que comprenden el compuesto activo en una base aromatizada, normalmente sacarosa y goma arábiga o tragacanto; y pastillas que comprenden el compuesto en una base inerte tal como gelatina y glicerina o sacarosa y goma arábiga.

Las formulaciones de la presente invención adecuadas para la administración parenteral comprenden soluciones inyectables acuosas y no acuosas estériles del compuesto o los compuestos activos, cuyas preparaciones son preferiblemente isotónicas con la sangre del receptor previsto. Estas preparaciones pueden contener antioxidantes, tampones, bacteriostáticos y solutos que hacen la formulación isotónica con la sangre del receptor previsto. Las suspensiones estériles acuosas y no acuosas pueden incluir agentes suspensores y agentes espesantes. Las formulaciones se pueden presentar en recipientes de una unidad/dosis o de múltiples dosis, por ejemplo ampollas y viales sellados, y se pueden almacenar en estado secado por congelación (liofilizado) que requiere solo la adición del vehículo líquido estéril, por ejemplo, solución salina o agua para inyectables inmediatamente antes de su uso. Las soluciones y suspensiones inyectables extemporáneas se pueden preparar a partir de polvos, gránulos y comprimidos estériles del tipo descrito previamente. Por ejemplo, en un aspecto de la presente invención, se proporciona una composición inyectable, estable, estéril que comprende uno o varios agentes activos, o una sal de los mismos, en una forma de dosificación unitaria en un recipiente sellado. El compuesto o sal se proporcionan en forma de un producto liofilizado que es susceptible de ser reconstituido con un vehículo adecuado farmacéuticamente aceptable para formar una composición líquida adecuada para su inyección en un sujeto. La forma de dosificación unitaria comprende típicamente de aproximadamente 10 mg a aproximadamente 10 gramos del compuesto o sal. Cuando el compuesto o sal son sustancialmente insolubles en agua, se puede emplear una cantidad suficiente de agente emulsionante que sea fisiológicamente aceptable en una cantidad suficiente para emulsionar el compuesto o la sal en un vehículo acuoso. Uno de tales agentes emulsionantes útiles es la fosfatidilcolina.

Las formulaciones adecuadas para aplicación tópica a la piel adoptan preferiblemente la forma de un pomada, crema, loción, pasta, gel, pulverización, aerosol o aceite. Los portadores que se pueden utilizar incluyen vaselina, lanolina, polietilenglicoles, alcoholes, potenciadores transdérmicos, y combinaciones de dos o más de los mismos.

Las formulaciones adecuadas para la administración transdérmica se pueden presentar en forma de parches discretos adaptados para permanecer en contacto íntimo con la epidermis del receptor durante un período prolongado de tiempo. Las formulaciones adecuadas para la administración transdérmica pueden suministrarse también mediante iontoforesis (véase, por ejemplo, *Pharmaceutical Research* 3 (6): 318 (1986)) y típicamente adoptan la forma de una solución acuosa opcionalmente tamponada del compuesto activo. Las formulaciones adecuadas comprenden tampón citrato o bis/Tris (pH 6) o etanol/agua y contienen ingrediente activo de 0,1 a 0,2 M.

Además del compuesto o los compuestos activos, las composiciones farmacéuticas pueden contener otros aditivos, tales como aditivos de ajuste del pH. En particular, los agentes de ajuste del pH útiles incluyen ácidos, tales como ácido clorhídrico, bases o tampones, tales como lactato de sodio, acetato de sodio, fosfato de sodio, citrato de sodio, borato de sodio, o gluconato de sodio. Adicionalmente, las composiciones pueden contener conservantes microbianos. Los conservantes microbianos útiles incluyen metilparabeno, propilparabeno, y alcohol bencílico. El conservante microbiano se emplea típicamente cuando la formulación se coloca en un vial diseñado para uso multidosis. Por supuesto, como se indica, las composiciones farmacéuticas de la presente invención se pueden liofilizar utilizando mecanismos bien conocidos en la técnica.

Dosificación. La dosificación terapéuticamente eficaz de cualquier agente activo específico, cuyo uso se encuentra en el alcance de la presente invención, variará algo de compuesto a compuesto, y de paciente a paciente, y dependerá de la afección del paciente y la ruta de liberación. Para la administración oral, se puede utilizar una dosificación diaria total de 1, 2 o 3 mg, hasta 30, 40 o 50 mg, proporcionada como una dosis diaria única o dividida en dos o tres dosis diarias.

Tratamiento. Las variantes genéticas descritas en la presente memoria o descubiertas utilizando los métodos ilustrados en la presente memoria se pueden utilizar para determinar el curso del tratamiento de un paciente afectado con una afección (p. ej., una afección asociada con ApoE y/o TOMM40), determinado, p. ej., cuál agente activo y/o curso de tratamiento administrar basándose en presencia o ausencia de la variante o variantes genéticas. La presencia o ausencia de las variantes genéticas pueden indicar la eficacia de un agente activo y/o el curso del

tratamiento para el paciente, predecir la edad de aparición de una afección, indicar los regímenes de dosificación preferidos, etc. Se puede generar un perfil genético para un paciente, y consultar el perfil para determinar si el paciente se encuentra entre un grupo de pacientes que es probable que sea sensible a un agente activo concreto.

5 Las instrucciones de uso se pueden empaquetar o asociar de otro modo con un agente activo indicando las recomendaciones para el tratamiento, el tiempo de tratamiento, los regímenes de dosificación, etc., basándose en la presencia o ausencia de las variantes genéticas.

8. Métodos para determinar la predicción del riesgo de enfermedad o pronóstico.

10 Para determinar una predicción del riesgo de enfermedad para un individuo no sintomático o un pronóstico (la posibilidad de la afección o el curso de la enfermedad como se había anticipado a partir del curso habitual de una enfermedad o peculiaridades del caso) de acuerdo con algunas realizaciones de la presente invención, los datos de diagnóstico, incluyendo el diagnóstico o historia clínica y los datos genéticos del paciente, tales como el genotipo del paciente (p. ej., genotipo ApoE y/o TOMM40), pueden ser procesados para proporcionar opciones terapéuticas y las predicciones de los resultados. El procesamiento puede incluir la obtención de un "perfil del paciente", tal como la recogida de la historia clínica del paciente, incluyendo la edad y el género, la genotipificación de los loci de interés

15 (p. ej., utilizando los cebadores diseñados adecuadamente y utilizando una etapa de amplificación mediante RT-PCR o PCR y/o fenotipificación, p. ej., utilizando un método mediado por anticuerpo o un ensayo enzimático), y análisis estadísticos u otros análisis que convierten estos datos en bruto en un pronóstico. El pronóstico puede incluir una predicción de la edad de aparición de la enfermedad de un paciente, la respuesta a la terapia con el fármaco, el tiempo para el tratamiento, la eficacia del tratamiento, etc. En algunas realizaciones, el pronóstico puede

20 incluir el uso de un programa de soporte lógico informático para analizar los datos del paciente y ejecutar controles cruzados estadísticos en bases de datos relacionales con el fin de convertir los datos o el perfil del paciente en un pronóstico.

Un "perfil del paciente" incluye datos y/o materiales pertenecientes al paciente para el cual se está realizando el análisis predictivo y/o pronóstico. Los datos pueden incluir información sobre el diagnóstico, la edad, el sexo, y/o el genotipo del paciente. El perfil del paciente también puede incluir sustancias del paciente tales como sangre, muestras de proteína de suero, líquido cefalorraquídeo, o ARN o ADN purificado.

25

9. Estratificación del genotipo en pruebas clínicas.

La detección del genotipo ilustrada en la presente memoria o determinada con los métodos de la presente memoria se puede utilizar en la realización de una prueba clínica de la misma manera que se utiliza otra información del genotipo para llevar a cabo una prueba clínica, como se describe, p. ej., en la Patente de los Estados Unidos Núms. 6.573.049, 6.368.797 y 6.291.175.

30

En algunas realizaciones, tales métodos estratifican ventajosamente o permiten el perfeccionamiento de la población de pacientes (p. ej., mediante la división de la población en uno o más subgrupos) de modo que se pueden detectar con mayor exactitud las ventajas de los regímenes de tratamiento concretos, particularmente con respecto a subpoblaciones concretas de pacientes con genotipos concretos. En algunas realizaciones, tales métodos comprenden la administración de un agente activo o una terapia de ensayo a una pluralidad de sujetos (administrándose típicamente una terapia de control o placebo a una pluralidad de sujetos separados pero de características similares) y detectar la presencia o ausencia de un genotipo (p. ej., ApoE y/o TOMM40) como se ha descrito anteriormente en la pluralidad de sujetos. El genotipo se puede detectar antes de, después de, o simultáneamente a la etapa de administración de la terapia de ensayo. La influencia de uno o más alelos detectados en la terapia de ensayo se puede determinar con cualquier parámetro adecuado o resultado o consecuencia potenciales del tratamiento, incluyendo, pero no limitados a, la eficacia de dicha terapia, la falta de efectos secundarios de la terapia, etc.

35

40

Se puede configurar una prueba clínica para someter a ensayo la eficacia de los compuestos de ensayo para el tratamiento de cualquier número de enfermedades para las se ha determinado que está asociado un genotipo concreto, para los sujetos que han sido diagnosticados con la enfermedad o están en riesgo de desarrollar la enfermedad. Si los sujetos son genotipificados después de la finalización de la prueba clínica, los análisis aún pueden estar dirigidos a determinar una relación entre un tratamiento para una enfermedad y el alelo para el que se va a evaluar la eficacia. Alternativamente, si un sujeto sintomático o asintomático todavía no ha sido diagnosticado con la enfermedad, pero se ha determinado que tiene riesgo de desarrollar la enfermedad, se puede llevar a cabo una prueba clínica similar a la prueba clínica descrita anteriormente.

45

50

Los mecanismos biológicos subyacentes también pueden ser considerados cuando se diseñan los grupos de tratamiento. Por ejemplo, el fragmento ApoE 4 (1-272) se une a las mitocondrias, disminuye la dinámica celular mitocondrial y disminuye la sinaptogénesis más que ApoE 3 (1-272). La rosiglitazona, un fármaco candidato para el tratamiento de la enfermedad de Alzheimer, aumenta la mitogénesis y aumenta la sinaptogénesis - en oposición a los efectos de la unión del fragmento ApoE - para ApoE 3 más que con ApoE 4. Por lo tanto, el fármaco o el tratamiento candidatos (p. ej., rosiglitazona) se pueden seleccionar basándose en un mecanismo de acción subyacente, ya que se refiere a los marcadores genéticos utilizados para las estratificaciones (p. ej., variantes ApoE

55

2, E 3, E 4 y/o TOMM40).

La evaluación de la eficacia de un fármaco seleccionado para la prueba puede incluir el control del sujeto durante un período de tiempo y el análisis del retraso del inicio de la enfermedad y la intensidad de la enfermedad en el momento de inicio, así como la medición del inicio de los síntomas con los que está asociada la enfermedad. Un fármaco que, en una prueba clínica, elimina o retrasa el inicio de la enfermedad, o reduce los síntomas de la enfermedad puede ser un fármaco beneficioso para su uso en pacientes diagnosticados con la enfermedad o en riesgo de desarrollar la enfermedad. Los compuestos de ensayo que se pueden utilizar en tales pruebas incluyen los agentes descritos anteriormente, incluyendo los previamente aprobados para uso clínico y nuevos compuestos aún no aprobados para su uso, o aprobados para el tratamiento de una enfermedad concreta. De este modo, en algunas realizaciones, la prueba clínica puede incluir la optimización de la administración del fármaco, incluyendo la dosificación, el momento de la administración, las toxicidades o efectos secundarios, la ruta de administración, y la eficacia del tratamiento.

10. Kits útiles para la detección de variantes de genotipo en loci de interés.

Los kits para determinar si un sujeto tiene mayor riesgo de desarrollar una enfermedad, desarrollar una enfermedad a una edad de inicio más temprana, y/o un candidato para un tratamiento concreto, donde la enfermedad está asociada con ApoE y/o TOMM40 (p. ej., enfermedad de Alzheimer de inicio tardío), se describen en la presente memoria. Los kits incluyen al menos un reactivo específico para detectar la presencia o ausencia de una variante ApoE y/o TOMM40 como se describe en la presente memoria, y pueden incluir instrucciones para ayudar a determinar si el sujeto tiene mayor riesgo de desarrollar la enfermedad. El kit puede incluir opcionalmente un ácido nucleico para la detección de un gen de ApoE (p. ej., ApoE 2, ApoE 3 y/o ApoE 4) o instrucciones para métodos de isoelectroenfoque para detectar el genotipo de ApoE; y/o un ácido nucleico para la detección de una variante de TOMM40 como se describe en la presente memoria. El kit puede incluir opcionalmente uno o más anticuerpos que se unen a ApoE 2, ApoE 3, ApoE 4, o a isoformas de TOMM40. El kit de ensayo se puede empaquetar de cualquier manera adecuada, típicamente con todos los elementos en un único contenedor junto con una hoja de instrucciones impresas para llevar a cabo el ensayo.

El kit puede contener opcionalmente tampones, enzimas y reactivos para la amplificación de los ácidos nucleicos genómicos a través de la amplificación dirigida por cebadores. El kit también puede incluir uno o más dispositivos para detectar la presencia o ausencia de los haplotipos concretos en el ácido nucleico amplificado. Tales dispositivos pueden incluir una o más sondas que hibridan a un ácido nucleico del haplotipo, que puede estar unido a un bio-chip o dispositivo de micromatriz, tal como cualquiera de los descritos en la Patente de los Estados Unidos. Núm. 6.355.429. El bio-chip o dispositivo de micromatriz tiene opcionalmente al menos una sonda de captura unida a una superficie que puede hibridar con una secuencia del haplotipo. El bio-chip o micromatriz contiene múltiples sondas, y lo más preferiblemente contiene al menos una sonda para una secuencia del haplotipo que, si estuviera presente, se amplificaría mediante un conjunto de cebadores limitrofes. Por ejemplo, si se utilizan cinco pares de cebadores limitrofes para la amplificación, el dispositivo contendría al menos una sonda de haplotipo para cada producto amplificado, o al menos cinco sondas. El kit también incluye preferiblemente las instrucciones para utilizar los componentes del kit.

La presente invención se explica con mayor detalle en los siguientes Ejemplos no limitantes.

Ejemplo 1: Construcción de árboles filogenéticos

Todos los estudios de exploración del genoma completo conocidos demuestran valores de p extremadamente significativos en las inmediateces del locus de la apolipoproteína CI [ApoCI]. (Mahley et al, Proc. Natl. Acad. Sci. USA 103: 5644-51 (2006), Coon et al., J. Clin. Psychiatry 68: 613-8 (2007); Li et al, Arch. Neurol. 65: 45-53 (2007)). De igual importancia es que cada serie identificara un gen candidato significativo límite "predilecto" fuera de la zona de desequilibrio de ligamiento de ApoE, pero estos genes candidato predilectos fueron diferentes en cada estudio. TOMM40 está cerca de ApoCI y en desequilibrio de ligamiento con ApoE. Se cree que las interacciones entre ApoE 3 o ApoE 4 y diferentes isoformas de TOMM40 están asociadas con el aumento o la disminución del riesgo de desarrollar la enfermedad de Alzheimer en un intervalo de edad más temprana. La edad de las curvas de inicio para los genotipos Apo 4/4, 3/4, 3/3, 2/4, y 2/3 se muestra en la Figura 2, indicando un intervalo de riesgo para el desarrollo precoz de la enfermedad, dependiendo del perfil de ApoE. Sin embargo ApoE sola no parece explicar todos los datos en estas curvas de edad de inicio.

Se han propuesto varios métodos para el perfilado polimórfico del riesgo de enfermedad de Alzheimer asociado con los diferentes alelos de ApoE (véanse, p. ej., la Solicitud de los Estados Unidos de Cox et al, Núm. 20060228728; la Solicitud de Estados Unidos de Li y Grupe, Núm. 20080051318). En la presente memoria se muestra un enfoque filogenético al rompecabezas de ApoE 4.

Muestras biológicas, aislamiento de ADN, amplificación de loci de interés. Un total de 340 sujetos incluyeron 135 casos de enfermedad de Alzheimer y 99 controles de la misma edad en el grupo A, así como 57 casos y 49 controles en el Grupo B. Todos los sujetos portaban los genotipos de ApoE previamente asociados con un mayor riesgo de inicio más temprano de la enfermedad (es decir, 3/3, 3/4, o 4/4). Las muestras biológicas que contenían el

ADN se recogieron de todos los sujetos. El ADN genómico se aisló a continuación, de acuerdo con los métodos convencionales para la secuenciación de los loci genéticos en el Cromosoma 19.

La Figura 3 muestra las regiones genéticas en el cromosoma 19 elegidas como diana para el estudio utilizando datos de exploración de todo el genoma a partir de varios informes. La región se engloba dentro de la secuencia de referencia de GenBank AF050154. Se utilizó un soporte lógico para generar alineamientos de secuencias múltiples para los loci de las variantes (p. ej., ClustalW2, European Bioinformatic Institute). Posteriormente, se analizaron los alineamientos de secuencias múltiples utilizando el soporte lógico para el desarrollo de los árboles filogenéticos (p. ej., MEGA versión 2.1, Center for Evolutionary Functional Genomics, TREEVOLVE, Departamento de Zoología de la Universidad de Oxford, o un soporte lógico para la construcción basada en la parsimonia tal como PAUP, Sinauer Associates). Los análisis estadísticos se pueden realizar, p. ej., con Genetic Data Analysis (GDA: Software for the Analysis of Discrete Genetic Data, The Bioinformatics Research Center of North Carolina State University). Los resultados del análisis de la Región B se muestran en el árbol filogenético de la Figura 4.

Cada dato en la Figura 4 representa una variante de la secuencia observada. Estas variantes pueden ser sustituciones, inserciones, supresiones de nucleótidos o microsatélites y pueden o no dar como resultado diferencias detectables en la expresión génica o la función de las proteínas. Cada nodo representa una variante (o un número de variantes) que se produce en más de un cromosoma. Los nodos adyacentes definen los límites de las secuencias que están en cis, y por lo tanto es más probable que sean heredadas como una unidad, en la región de interés en el cromosoma de un sujeto. Los nodos que preceden al mayor número de nodos siguientes representan variantes evolutivamente ancestrales a partir de las cuales se ha producido la divergencia genética a lo largo del tiempo.

La presencia de haplotipos o variantes de secuencia correspondientes a estas regiones del árbol que representa sujetos con una incidencia sustancialmente mayor de enfermedad de Alzheimer (es decir, mayores proporciones de sujetos afectados con la enfermedad con respecto a sujetos de control no afectados) significaría que el sujeto individual también tiene un riesgo mayor. Por el contrario, las proporciones sustancialmente más bajas corresponden a la reducción del riesgo de desarrollar la enfermedad de Alzheimer.

TOMM40 interactúa con ApoE directamente en la regulación de la importación de la proteína mitocondrial, y una hipótesis actual es que la expresión de una o varias variantes concretas de TOMM40 exacerba el riesgo relativamente moderado de enfermedad de Alzheimer asociada con la presencia dependiente de la dosis del alelo ApoE 3. Tal variante de TOMM40 se descubrió dentro de la Región B utilizando los métodos de la presente invención.

El ensayo de nuevos fármacos en sujetos humanos acarrea un riesgo inmenso (véase Kenter y Cohen, *Lancet*, 368: 1387-1391 (2006)). El uso de árboles filogenéticos para anticipar la respuesta individual a un fármaco o tratamiento de interés tiene el potencial de aliviar ese riesgo de manera significativa. Los estudios preliminares indicaron que la rosiglitazona (Avandia) puede tener una eficacia específica del perfil genético en el tratamiento de la enfermedad de Alzheimer (véanse Risner et al, *The Pharmacogenomics Journal* 6, 246-254 (2006); Brodbeck et al., *Proc. Nat. Acad. Sci.* 105, 1343-6 (2008)). Los datos de las clínicas en fase II indican que los pacientes con enfermedad de Alzheimer sin un alelo ApoE 4 respondían mejor a la rosiglitazona que los pacientes que portaban 1 o 2 alelos ApoE 4 (datos no mostrados). Esto apoya la hipótesis de que las variantes identificadas con los métodos ilustrados en la presente memoria se pueden utilizar para preveer la respuesta individual al tratamiento basado en el genotipo.

Ejemplo 2: Identificación de variantes de TOMM40 de interés

Se alinearon 174 secuencias (2 de cada uno de 87 sujetos) utilizando el programa CLUSTAL X (versión 2.0.10, Larkin et al, *Clustal W y Clustal X versión 2.0 Bioinformatics*, 23:2947-2948 (2007)). Se utilizó el alineamiento de secuencias múltiples para construir un árbol filogenético utilizando un algoritmo de vecino más próximo (Saitou y Nei, *The neighbor joining method: A new method for reconstructing phylogenetic trees Mol. Evol. Biol.*, 4:406-425 (1987)) tal como se aplica en el sitio web del Instituto Europeo de Bioinformática (EBI).

El árbol filogenético resultante tiene una estructura de dos grandes grupos (A, B) en la primera divergencia. Las frecuencias del genotipo ApoE para estos grupos se tabulan y se muestran en la Figura 5. Está claro que el grupo B contiene haplotipos de sujetos de los genotipos de ApoE principalmente ϵ_3/ϵ_3 y ϵ_3/ϵ_4 y casi ninguno ϵ_4/ϵ_4 . El grupo A contiene casi todos los haplotipos de sujetos con el genotipo ϵ_4/ϵ_4 .

La lista de los polimorfismos generados por la plataforma de descubrimiento de SNP (polimórfica) se utilizó para identificar las variantes específicas en el gen TOMM40 que separaba los datos en los dos grupos. Se utilizó en ensayo de la razón de verosimilitud para identificar variantes significativas con un valor de p de menos de 0,005.

La lista de variantes se resume en la Tabla 1. En la tabla, el término "delección" se utiliza cuando el alelo menor es una delección de un nucleótido, y el término "inserción" se utiliza cuando el alelo menor es una adición de un nucleótido. El término "polimorfismo de delección/inserción" se utiliza cuando hay más de dos formas posibles y el alelo menor no es evidente. Por ejemplo, para los polimorfismos de poli-T, se observan múltiples polimorfismos de longitud. La segunda columna de la tabla proporciona información sobre las identidades de los alelos específicos asociados con la variante que dividen las secuencias en los dos grupos. Por ejemplo, T> A indica que el alelo T segrega secuencias en el grupo "A" en el árbol filogenético. Cuando se indican dos alelos, p. ej., G> B; A> A, cada

alelo segrega únicamente los datos de secuencia en los dos grupos, mientras que cuando se enumera un solo alelo éste está asociado con la separación predominante de los datos, y el alelo restante no solo separa los datos en un grupo homogéneo, sino en una mezcla de ambos grupos.

Tabla 1. Variantes TOMM40 asociadas con grupos en el árbol filogenético que se distribuyen por el genotipo de ApoE.

5

Variante	Alelo>grupo árbol	Localización Genómica (NCBI Build 36.3)	Función	Clasificación UCSC
50.092.565	T>A	50.092.565	Intrón 6	sencillo
50.092.587	T>A	50.092.587	Intrón 6	sencillo
rs8106922	G>B; A>A	50.093.506	Intrón 6	sencillo
rs34896370, rs55821237, rs56290633	T12_C_T15, T12_C_T16, T13_C_T14, T13_C_T15, T13_C_T16>A; T14_C_T14, T14_C_T15>B	50.093.609	Intrón 6	complejo
rs34878901	T>B; C>A	50.094.317	Intrón 6	sencillo
rs35568738	C>B	50.094.558	Intrón 6	sencillo
rs10602329	T16,17,18>A T14,15>B	50.094.716	Intrón 6	inserción/delección
50,094,733	->A	50.094.733	Intrón 7	inserción
rs10524523	T12,14,15,16,17>B T21,22,26,27,28,29, 30,31,32,33,34 35,36>A	50.094.889	Intrón 6	inserción/delección
rs1160985	T>B; C>A	50.095.252	Intrón 6	sencillo
50.095.506	T>A	50.095.506	Intrón 6	sencillo
rs760136	A>A; G>B	50.095.698	Intrón 6	sencillo
rs1160984	T>B	50.095.764	Intrón 6	sencillo
rs741780	C>B; T>A	50.096.271	Intrón 8	sencillo
rs405697	A>A	50.096.531	Intrón 9	sencillo
50.096.647 (DIP3)	->A	50.096.647	Intrón 9	delección
50.096.697	C>A	50.096.697	Intrón 9	sencillo
rs1038025	C>B; T>A	50.096.812	Intrón 9	sencillo
rs1038026	G>B; A>A	50.096.902	Intrón 9	sencillo
rs1305062	C>B; G>A	50.097.361	Intrón 9	sencillo
rs34215622	G>B; ->A	50.098.378	Exón 10	inserción
rs10119	A>A	50.098.513	Exón 10	sencillo
rs7259620	G>A; A>B	50.099.628	desconocida	sencillo

Ejemplo 3: Dos formas distintas de ApoE 3: las ligadas a los haplotipos de TOMM40 que aumentan el riesgo y disminuyen la edad de inicio, y las que disminuyen el riesgo

10

La asociación de los genotipos de la apolipoproteína E (ApoE), en particular ApoE ε4 (APOE 4), con el riesgo y la edad de aparición de la enfermedad de Alzheimer (EA) sigue siendo la asociación genética más confirmada para cualquier enfermedad compleja. Las estimaciones de la heredabilidad de ApoE 4 para la EA de inicio tardío oscilan de 58% a 79%, y el riesgo atribuible a la población debido al alelo ApoE 4 se encuentra entre 20% y 70%. Estas estimaciones sugieren que otras variantes genéticas y/o interacciones entre las variantes incurrir en un riesgo de

enfermedad adicional y modifican las distribuciones de la edad de inicio.

Los resultados de asociación de exploración del genoma completo para la EA han reproducido de manera coherente la extraordinaria asociación de la región LD que contiene ApoE. TOMM40, la proteína translocasa de la membrana mitocondrial externa, tiene un elevado LD con ApoE, y codifica el canal de membrana a través del cual atraviesan los péptidos y proteínas citoplásmicos con el fin de sintetizar nuevas mitocondrias. Los objetivos de los autores de la presente invención fueron la identificación de haplotipos adicionales dentro de la región LD que aumentan las estimaciones de heredabilidad.

Métodos: Los autores de la presente invención examinaron la región LD que contenía tanto ApoE como TOMM40 utilizando una secuenciación primaria profunda (10X) en pacientes con EA y controles. Los autores de la presente invención realizaron análisis filogenéticos de la región LD que abarcaba TOMM40 y ApoE en 66 pacientes y 66 controles emparejados por edad con respecto al riesgo y la distribución de la edad de inicio.

Conclusión: Los autores de la presente invención encontraron que familias heredadas únicas y distintas de diferentes variantes de TOMM40 se encuentran en el mismo intervalo genómico que ApoE 3, pero no en el intervalo genómico que contiene ApoE 4, y pueden aumentar o disminuir la distribución de la edad de riesgo de la EA.

Por lo tanto, la herencia genética de estas variantes de TOMM40 es independiente de la herencia de ApoE 4, proporcionando eficazmente una diferenciación de dos formas distintas de ApoE 3: aquellas ligadas a los haplotipos de TOMM40 que aumentan el riesgo y disminuyen la edad de inicio, y aquellas que disminuyen el riesgo. Estos datos aumentan la exactitud de la edad genética de riesgo de inicio, dependiente de la edad, de los genotipos de ApoE y TOMM40 y brindan la oportunidad de definir un alto riesgo de EA en los próximos 5-7 años, frente a un menor riesgo de EA.

Ejemplo 4: Análisis de tres variantes DIP de TOMM40 identificadas.

Tres de las variantes de TOMM40 identificadas en esta solicitud son polimorfismos de delección/inserción (DIP) situados en el intrón 6 o intrón 9. Estos DIP se identifican como rs10524523 y rs10602329 en la base de datos del Centro Nacional para la Información Biotecnológica dbSNP, y un polimorfismo no descrito anteriormente, denominado DIP3. Estos polimorfismos se encuentran en chr19:50.094.889, chr19:50.094.731, y chr19:50.096.647, respectivamente, de acuerdo con NCBI Build 36. Esta invención describe la identificación de estos DIP utilizando el análisis filogenético del gen TOMM40, específicamente de un fragmento de 10 Kb del gen, y que los DIP están asociados con diferentes grupos evolutivos determinados por análisis filogenético. La presente invención describe adicionalmente la utilidad de estos DIP para (1) determinar el riesgo de que una persona sana desarrolle enfermedad de Alzheimer en el futuro, y (2) pronosticar la edad de inicio de la EA dentro de un marco temporal de de aproximadamente 8 años.

Los tres polimorfismos DIP caracterizados en la presente memoria corresponden a diferentes longitudes de repeticiones de poli-T en DIP en el gen TOMM40. La asociación de las variantes de poli-T en DIP con el riesgo de enfermedad tiene precedencia. Por ejemplo, una variante de poli-T en el intrón 8 del gen regulador de la conductancia transmembrana de la fibrosis quística (CFTR) está asociada con la omisión del exón 9 y el desarrollo de la fibrosis quística (Groman et al, Am J Hum Genet 74 (1): 176 -9 (2004)). En la presente memoria se describen: (1) el uso del nuevo método - análisis de asociación filogenética (descrito anteriormente) - para identificar los DIP que son predictivos del riesgo de enfermedad y/o diferencias en la edad de inicio de la enfermedad, (2) la identidad de tres DIP específicos asociados con diferencias en la edad de inicio de la EA y el riesgo de EA, (3) el uso de estos SNP individualmente, juntos, o con otras variantes de secuencia en TOMM40 o ApoE para diagnosticar la enfermedad o predecir o determinar características de la enfermedad tales como la edad de aparición de la enfermedad, el pronóstico de la enfermedad, los sub-tipos de la enfermedad, la gravedad de la enfermedad, y también para analizar o determinar la respuesta a los fármacos.

El análisis filogenético revela la distribución de los DIP rs10524523 y rs10602329 en dos clados diferentes. Este análisis revela que las longitudes de poli-T más cortas en éstos loci se mapean para los clados identificados filogenéticamente en el grupo B, el grupo que también comprende un mayor porcentaje de sujetos de genotipo ApoE E3/E3, efectivamente pocos (0%) sujetos ApoE E4/E4 y proporciones menores de casos/control (es decir, riesgo de enfermedad EA) (Figura 5). La asociación entre la longitud del DIP y el grupo filogenético es estadísticamente significativa ($p < 0,0001$) por la prueba de razón de verosimilitud o la prueba Chi-cuadrado de Pearson.

Debido a la arquitectura genómica, el alto desequilibrio de ligamiento y las relaciones evolutivas indicadas en el análisis filogenético, entre los dos genes, y la supuesta interacción física entre los dos productos génicos, es probable que la influencia del genotipo TOMM40 que amplíe a otras enfermedades que están influenciadas por el genotipo ApoE. Estas enfermedades incluyen, pero no se limitan a, la enfermedad de Parkinson, esclerosis múltiple, enfermedad cardiovascular, dislipidemia, recuperación de una lesión cerebral traumática, recuperación de eventos isquémicos cerebrales, respuesta a los anestésicos, y respuesta a los fármacos utilizados para tratar la EA y las enfermedades enumeradas aquí.

Estos polimorfismos podrían utilizarse también en los esfuerzos de descubrimiento de fármacos para el escrutinio de compuestos útiles para el tratamiento de enfermedades influenciadas por las variaciones en TOMM40 o variantes de

la proteína o el gen de ApoE. Además, las variantes pueden influir o determinar las terapias basadas en productos biofarmacéuticos específicamente dirigidos como los ilustrados por los anticuerpos monoclonales y las moléculas de ARNip.

5 Los polimorfismos DIP en TOMM40 que se describen en la presente memoria pueden ser identificados en una muestra de un ADN de un individuo utilizando muchas metodologías moleculares diferentes de análisis de nucleótidos, incluyendo, pero no limitadas a, secuenciación de ADN con los cebadores indicados en la Tabla 4 enumerada a continuación.

Ejemplo 5: Las extensiones de poli-T más largas en rs10524523 se correlacionan significativamente con la edad más temprana de inicio de la EAT

10 El análisis filogenético se ha utilizado para identificar las relaciones genómicas entre las variantes genéticas de baja frecuencia y para agrupar los haplotipos relacionados evolutivamente (Hahn et al. Population genetic and phylogenetic evidence for positive selection on regulatory mutations at the factor VII locus in humans. *Genetics* **167**, 867-77 (2004)). Esta metodología se utilizó para explorar el bloque LD ApoE-TOMM40 para determinar la existencia de nuevos factores determinantes de riesgo para la EAT. En un estudio exploratorio, se amplificaron y se secuenciaron 23 Kb de ADN que contenían los genes de TOMM40 y ApoE, y se determinaron los haplotipos resueltos por fases, para 72 casos de EAT y 60 controles emparejados por edad (Li et al. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch Neurol* **65**, 45-53 (2008)). Fue posible construir un árbol filogenético distinto para 10 Kb, que codificaban los exones 2-10, de esta región. Se distinguieron dos clados (A y B) con un fuerte apoyo bootstrap (98%, 1.000 repeticiones). Hubo una diferencia significativa en la distribución de los genotipos de ApoE entre los dos clados de haplotipos de TOMM40 sobre este árbol filogenético, lo que sugiere que esta región podría ser significativa funcionalmente. Ambos clados contenían sujetos con el genotipo $\epsilon 3/\epsilon 3$, pero el 98% de todos los haplotipos del clado B aparecían en cis con el alelo ApoE $\epsilon 3$ ($P = 1,2 \times 10^{-18}$, test exacto de Fisher, de dos colas).

25 La estructura filogenética de esta región de 10 Kb de TOMM40, la herencia específica de ApoE $\epsilon 3$ de haplotipos concretos, y la identificación de los polimorfismos específicos de un clado se confirmaron posteriormente en dos cohortes de casos de EAT/control independientes, incluyendo una cohorte con estado de AD confirmado en autopsia y datos de la edad de inicio de la enfermedad. La asociación entre los dos clados y el riesgo de enfermedad y la edad de inicio de la enfermedad, cuando se disponía de los datos, también se exploró para estas dos cohortes. La primera cohorte (AS) comprendía casos de EA ($n = 74$) y controles ($n = 31$) determinados en el Centro de Investigación de la Enfermedad de Alzheimer de Arizona (ADRC por sus siglas inglesas). La segunda cohorte (DS) se recopiló en el Bryan ADRC de Duke y comprendía sujetos ApoE $\epsilon 3/\epsilon 4$ solamente (40 casos confirmados por autopsia con una edad conocida de inicio de la enfermedad y 33 controles) (Tabla 2). Aunque la secuenciación del ADN fue un éxito para un subgrupo de la cohorte DS que tenía un inicio de la enfermedad entre 50-68 años de edad, los análisis de asociación se limitaron a un subgrupo de pacientes que desarrollaron AD después de la edad de 60.

35 Tabla 2. Composiciones de cohortes. El número de casos y controles, la edad media y el porcentaje que eran mujeres se muestran para cada serie. La edad media se proporciona como la edad al diagnóstico de EA para los casos y la edad del examen para los controles. La desviación típica de la media se proporciona entre paréntesis.

Serie	n		Edad Media (DT)				% Mujeres	
	Casos	Controles	Casos		Controles		Casos	Controles
AS	74	31	81,7	(8,01)	77	(8,93)	56,3	46,7
DS	40	33	69,3	(8,3)	71,9	(7,5)	70	66,7

40 Se desarrolló un árbol filogenético de estructura similar al generado en el estudio exploratorio con un fuerte apoyo bootstrap (97%, 1.000 repeticiones) para la cohorte de AS. Los sujetos ApoE $\epsilon 4/\epsilon 4$ aparecieron solamente en el clado A (98% de separación entre grupos, $P = 2,0 \times 10^{-4}$ en la prueba exacta de Fisher de dos colas), mientras que los genotipos de ApoE restantes se distribuyeron entre clados A y B (Figura 6). Es decir, ApoE $\epsilon 4$ siempre estuvo en LD con las variantes del clado A mientras que ApoE $\epsilon 3$ apareció en los haplotipos tanto del clado A como del clado B. El examen de la distribución de los pocos sujetos ApoE $\epsilon 2/\epsilon 4$ en el árbol filogenético sugiere que los haplotipos ApoE $\epsilon 2$ -TOMM40 comparten una historia evolutiva similar con los haplotipos ApoE $\epsilon 3$ -TOMM40 (datos no mostrados). Para verificar la estructura filogenética utilizando un método separado, y para garantizar que la recombinación dentro del intervalo genético no confundía la estructura del árbol filogenético desarrollado para la cohorte AS, también se construyeron redes de haplotipos utilizando parsimonia estadística (TCS versión 1.21 (Clement et. al. TCS: a computer program to estimate gene genealogies. *Mol Ecol* **9**, 1657-9 (2000))). Las principales agrupaciones de haplotipos de sujetos derivadas de los dos métodos (parsimonia máxima y TCS) fueron congruentes.

El clado A se asoció con mayor frecuencia con los casos de EA que el clado B (OR = 1,44, IC 95% = 0,76-2,70). Los heterocigotos para ApoE $\epsilon 3/\epsilon 4$ ($n = 36$) se analizaron para estimar el riesgo de enfermedad asociado con los

haplotipos del clado A la vez que se controlaba el efecto de ApoE $\epsilon 4$. Hubo una tendencia a una mayor incidencia de EAT para el subgrupo que era homocigoto para TOMM40 en el clado A con respecto al subgrupo que era heterocigoto para el clado A y el clado B (OR = 1,36, IC 95% = 0,40-4,61) y por lo tanto se postuló que al menos algunas de las variantes de TOMM40 que definen el clado A confieren un riesgo EAT independientes de ApoE $\epsilon 4$.

5 El análisis de los datos de secuencia de la cohorte AS identificó 39 sitios polimórficos en la región TOMM40 de 10 Kb, de los cuales había 30 sitios informativos de parsimonia (al menos dos nucleótidos diferentes, cada uno representado en al menos dos secuencias). De los 30 sitios informativos de parsimonia, 18 tuvieron una frecuencia de alelo menor (MAF) > 0,10 y seis SNP se encontraban fuera de los límites del gen TOMM40. Diez SNP se produjeron exclusivamente en el contexto de ApoE $\epsilon 3$ ($P = 6,07 \times 10^{-50}$, prueba exacta de Fisher, de dos colas, $n =$
10 210) y nunca se observaron en sujetos homocigotos para ApoE $\epsilon 4/\epsilon 4$ ($n = 16$). La mayoría de las variantes de TOMM40 específicas de $\epsilon 3$ se localizaron en regiones intrónicas.

La Figura 7 ilustra los 10 SNP y los 6 polimorfismos de inserción/delección que distinguen clados A y B de TOMM40 (a $P < 0,001$) para los sujetos ApoE $\epsilon 3/\epsilon 3$ de la cohorte de AS. Estos polimorfismos se sometieron a ensayo individualmente y como haplotipos para la asociación con el riesgo de EAT (Tabla 3). Las probabilidades relativas para el riesgo de enfermedad para cada alelo del clado B, en todos los casos el alelo menor, sugieren que los alelos del clado B son protectores frente al de EA en la cohorte AS, sin embargo, en cada caso, la asociación perdió por poco la significación. Para tener en cuenta el efecto de la ApoE $\epsilon 4$ en las probabilidades relativas presentadas en la Tabla 3, se construyó un grupo equilibrado de 48 casos de EA y 48 controles de EA mediante la selección de secuencias al azar de los sujetos ApoE $\epsilon 3/\epsilon 4$ de las cohortes AS y DS agrupadas. De nuevo los SNP individuales no se asociaron significativamente con la EAT en este grupo de datos equilibrados. Sin embargo, los alelos menores de cuatro de los SNP (rs8106922, rs1160985, rs760136, rs741780) que distinguen el clado B de TOMM40 se analizaron previamente en tres estudios de asociación del genoma de casos de EAT/control y se encontró que eran protectores del riesgo de enfermedad (OR < 1 en cada caso), lo cual es consistente con la tendencia observada en el estudio de los autores de la presente invención (Abraham et al. A genome-wide association study for late-onset Alzheimer's disease using ADN pooling. *BMC Med Genomics* 1, 44 (2008); Carrasquillo et al. Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease. *Nat Genet* 41, 192-198 (2009); Takei et al. Genetic association study on in and around the ApoE in late-onset Alzheimer disease in Japanese. *Genomics* 93, 441-448 (2009)).

Núm. de Expediente del Agente 9719-2wo

Tabla 3. Estadística descriptiva y resultados de la asociación alélica y genotípica para los SNP individuales.

Todos los genotipos ApoE																			
ID SNP	Posición	Allele	Alelo clado B	MAF (todo)	MAF (casos)	MAF (control)	EAT (M)	EAT (m)	Control (M)	Control (m)	EAT (MM)	EAT (M/m)	EAT (mm)	Control (MM)	Control (M/m)	Control (mm)	OR	IC inferior 95%	IC superior 95%
rs1038025	50096812	T/c	c	0,31	0,28	0,37	106	41	39	23	40	27	7	11	17	3	0,66	0,35	1,23
rs1038026	50096902	A/g	g	0,31	0,28	0,37	106	41	39	23	40	27	7	11	17	3	0,66	0,35	1,23
rs1160985	50095252	C/t	t	0,30	0,28	0,37	107	41	39	23	40	27	7	11	17	3	0,65	0,34	1,19
rs1305062	50097361	G/c	c	0,28	0,26	0,31	106	38	43	19	43	24	7	13	17	1	0,81	0,42	1,56
rs34215622	50098378	-g	g	0,28	0,26	0,34	110	38	40	21	42	26	6	12	17	2	0,66	0,35	1,25
rs34878901	50094317	C/t	t	0,26	0,25	0,28	105	35	44	17	45	23	6	15	15	1	0,86	0,44	1,70
rs7259620	50099628	G/a	a	0,30	0,27	0,37	108	40	39	23	41	26	7	11	17	3	0,63	0,33	1,18
rs741780	50096271	T/c	c	0,30	0,28	0,37	107	41	39	23	40	27	7	11	17	3	0,65	0,34	1,19
rs760136	50095698	A/g	g	0,30	0,28	0,37	107	41	39	23	40	27	7	11	17	3	0,65	0,34	1,19
rs8106922	50093506	A/g	g	0,28	0,26	0,31	109	39	43	19	42	25	7	13	17	1	0,81	0,42	1,55
APOE ε3/ε4																			
ID SNP	Posición	Allele	alelo clado B	MAF (todos)	MAF (casos)	MAF (control)	EAT(M)	EAT(m)	Control (M)	Control (m)	EAT (MM)	EAT (M/m)	EAT (mm)	Control (MM)	Control (M/m)	Control (mm)	OR	IC inferior 95%	IC superior 95%
rs1038025	50096812	T/c	c	0,28	0,25	0,38	68	28	63	33	22	24	2	17	29	2	0,79	0,43	1,45
rs1305062	50097361	G/c	c	0,27	0,24	0,38	69	25	64	32	25	21	2	17	30	1	0,72	0,39	1,35
rs34215622	50098378	-g	g	0,28	0,25	0,38	70	26	64	32	24	22	2	17	30	1	0,74	0,40	1,38
rs34878901	50094317	C/t	t	0,24	0,20	0,38	69	25	61	31	25	21	2	18	29	1	0,71	0,38	1,34
rs8106922	50093506	A/g	g	0,28	0,25	0,38	70	25	64	32	25	21	2	17	30	1	0,71	0,38	1,33

Otro polimorfismo que distingue los dos clados y, por lo tanto, dos grupos de haplotipos ApoE ϵ 3, fue una variante poli-T (rs 10524523) situada en el intrón 6 de TOMM40. En los cromosomas de ApoE ϵ 4, la variante era relativamente larga, con una distribución unimodal, estrecha de longitudes (21-30 residuos de T, media = 26,78, DT = 2,60, $n = 32$), mientras que en los cromosomas ApoE ϵ 3, se puso de manifiesto una distribución bimodal de longitudes con picos a 15,17 (DT = 0,85, $n = 36$) y 33,15 (DT = 2,09, $n = 55$) residuos de T (Figura 8). Las longitudes más largas de poli-T ($T \geq 27$) segregaban casi exclusivamente en el clado A, el clado de mayor riesgo, en la cohorte AS ($P = 7,6 \times 10^{-46}$, $n = 210$, prueba exacta de Fisher de dos colas). El razón de casos/control para la categoría que contenía las dos longitudes más cortas, más comunes (15 o 16 residuos de T) fue 1,46 (IC 95% = 1,25-1,75), y la razón de casos/control para la categoría de mayor longitud (28, 29, 33 y 34 residuos de t) fue de 2,02 (CI 95% = 1,13-2,87) Estos datos mostraron una tendencia a una asociación entre la mayor longitud de poli-T de rs10524523 y la EA (OR = 1,38, IC 95% = 0,80-2,39).

Aunque solo hubo tendencias hacia la asociación de haplotipos de TOMM40 o polimorfismos individuales con la EAT para la cohorte AS, hubo una asociación significativa entre la categoría de longitud de poli-T de rs10524523 y la edad de inicio de la EAT. Esto se sometió a ensayo utilizando la cohorte DS de sujetos ApoE ϵ 3/ ϵ 4 confirmados mediante autopsia para lo cuales había datos de inicio de la enfermedad. Los alelos poli-T más largos (≥ 27 residuos de T) se asociaron significativamente con el inicio de la enfermedad a una edad mucho más temprana (70,5 años \pm 1,2 frente a 77,6 años \pm 2,1, $P = 0,02$, $n = 34$) (Fig. 5).

Este polimorfismo, por lo tanto, repercutió significativamente en la edad de inicio de la enfermedad para los individuos que portaban un alelo ApoE ϵ 3. Otros tres polimorfismos de longitud poli-T situados en el intrón 6 (rs34896370, rs56290633 y rs10602329) también distinguían los clados A y B, pero estos polimorfismos no se asociaron con la edad de inicio de la enfermedad. Del mismo modo, no hubo relación entre los haplotipos de los SNP que distinguían el clado y la edad de la EAT, o para el único SNP, rs8106922, que se había sido asociado de manera significativa con el riesgo de EA en tres estudios de asociación del genoma completo (Abraham et al. A genome-wide association study for late-onset Alzheimer's disease using ADN pooling. *BMC Med Genomics* **1**, 44 (2008); Carrasquillo et al. Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease. *Nat Genet* **41**, 192-198 (2009); Takei et al. Genetic association study on in and around the ApoE in late-onset Alzheimer disease in Japanese. *Genomics* **93**, 441-448 (2009)) (datos no mostrados).

Los autores de la presente invención llegaron a la conclusión de que las extensiones de poli-T más largas en rs 10524523 se correlacionaban significativamente con una edad más temprana de inicio de la EAT. La longitud de esta variante es relativamente homogénea, y relativamente larga, en los cromosomas ApoE ϵ 4, mientras que existen dos categorías de longitudes de poli-T ligadas a ApoE ϵ 3. Los cromosomas de ApoE ϵ 2 también parecen portar repeticiones de poli-T de longitud variable de una manera similar a los cromosomas de ϵ 3, pero se necesita una investigación adicional para verificar este hallazgo preliminar y para determinar si la repetición poli-T repercute en la edad muy tardía de inicio de la enfermedad para los portadores de ApoE ϵ 2.

Si bien es posible que haya otras variantes que influyan en la edad de inicio de la EAT para los individuos que no son homocigóticos para ApoE ϵ 4, la longitud del polimorfismo poli-T en el intrón 6 TOMM40 parece ser el predictor genético más potente en esta región de ligamiento y debe ser validado de manera prospectiva. Estos datos sugieren que la edad estratificada para el genotipo ApoE de las curvas de inicio (Corder et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921-3 (1993); Li et al. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch Neurol* **65**, 45-53 (2008)) son, en realidad, grupos de curvas en los que cada curva refleja una interacción específica de los polimorfismos ligados en ApoE y TOMM40. Por lo tanto, estos datos añaden resolución a la predicción de la edad de inicio de la EAT, en una ventana de 5-7 años, para los individuos de más de 60 años de edad. El estudio para validar la asociación de los genotipos de ApoE y los haplotipos de TOMM40 o rs10524523 con la edad de inicio de la enfermedad está siendo planeado en la actualidad. Este estudio será un estudio basado en la población, prospectivo, de 5 años, realizado en varios grupos étnicos, y se combinará con una prueba de fármacos para la prevención o retraso del inicio de la enfermedad.

Métodos

Las dos cohortes analizadas en este estudio fueron del Centro de Investigación de la Enfermedad de Alzheimer de Arizona (ADRC), Fénix, Arizona y del Bryan ADRC de Duke, Durham, Carolina del Norte. Todos los sujetos fueron de ascendencia europea. Los estudios de Arizona y Duke fueron aprobados por las juntas de revisión institucional y se obtuvo el consentimiento informado correspondiente de todos los participantes. Los datos de edad y género para los casos y los controles en cada cohorte se muestran en la Tabla 2. Para la cohorte de Duke, la edad de aparición de la enfermedad se determinó a posteriori y el diagnóstico de la enfermedad fue confirmado por autopsia.

Las muestras se cultivaron en placa sobre placas de 96 pocillos para una PCR de largo alcance y una secuenciación del ADN en el Polymorphic ADN Technologies (Alameda, CA).

La PCR de largo alcance PCR se realizó utilizando polimerasa Taq Takara LA (Takara Mirus Bio). La mezcla de reacción y las condiciones de PCR fueron las mismas que las recomendadas por el fabricante. La PCR se llevó a cabo en un volumen de 50 μ l con 2,5 U de Taq LA y 200-400 ng de ADN genómico humano. Los ciclos térmicos se

ES 2 463 766 T3

llevaron a cabo con las siguientes condiciones: 94°C, 1 min durante 1 ciclo; 94°C, 30 seg; 57°C, 30 seg; 68°C, 9 min durante 14 ciclos; 94°C, 30 seg; 57°C, 30 seg; 68°C, 9 min + 15 seg/ciclo durante 16 ciclos; 72°C, 10 min durante 1 ciclo. Los cebadores para la PCR de largo alcance se muestran en la Tabla 4.

Núm. de Expediente del Agente 9719-2wo

Tabla 4. Se enumeran los cebadores de secuenciación directos e inversos. La fila sombreada indica los cebadores directos e inversos utilizados para la PCR de largo alcance de R2 (Figura 2)

Cebadores Directos				Cebadores Inversos			
Secuencia	Coordenada UCSC del Cebador (del extremo 3' del cebador)	Posición del Cebador en el Producto de la PCR Clonado (del extremo 3' del cebador)	SEQ ID NO:	Secuencia	Coordenada UCSC del Cebador (del extremo 3' del cebador)	Posición del Cebador en el Producto de la PCR Clonado (del extremo 3' del cebador)	SEQ ID NO:
AACTCAGAGGGCCAGAGATTCTAAGT	50.092.429	25	1	AACAGCCTAATCCCAGCACATTTTAC	50.101.560	9.156	2
CAGGAAACAGCTATGAC	50.092.292	-112	3	CCCACGTGGTTTGA	50.093.034	630	4
GTGTGATGGTGATTCAAC	50.093.038	634	5	GAATAGGGGCCCTTTCA	50.093.282	878	6
CTGCAGGTATGAAAG	50.093.287	883	7	CAATCTCCTAGGGTGC	50.093.512	1108	8
GTCTCTGCAGATGTG	50.093.601	1197	9	CGGAAAGTTGCAGTAAG	50.093.706	1302	10
TACTGCAACTTCCGC	50.093.722	1318	11	AAGGTCAAAGGTTACACT	50.094.318	1914	12
TCTCTGTTGCCACG	50.094.289	1885	13	ACAAGCCTAGGTGACAT	50.094.790	2386	14
CCCAACTAATTTTGTATTGG	50.094.609	2205	15	CCTGTAATCCCAGCTAT	50.095.002	2598	16
ACATTTGTGGCCTGTAC	50.095.129	2725	17	TCATCTCTCTGTGAACCCTAA	50.095.324	2920	18
CCACATGGGCTTGTGT	50.095.603	3199	19	GGCAAAATGACGATCAGT	50.095.804	3400	20
CCCAGATGCCCAAATC	50.096.082	3678	21	GCAGCACCCAGCTAGT	50.096.218	3814	22
AACTCTGAGTGGATGTG	50.096.471	4067	23	GATGGTCTCAATCTCCCTTA	50.096.620	4216	24
CTATAGTCCCAACTACTGA	50.096.730	4326	25	TTTTTTCCAAAGCATAAAACATAGTA	50.096.863	4459	26
AGTCCCCGCTACTTA	50.097.080	4676	27	GGGGATGGACAAAAGCT	50.097.268	4864	28
ACCACAGGTGTATGCC	50.097.451	5047	29	TGAAAAGCCCCTCTAGAC	50.097.898	5494	30
GAACAGATTTCATCCGCA	50.097.864	5460	31	CACCCACGATCCAGTT	50.098.141	5737	32
TGTGGATAGCAACTGGAT	50.098.148	5744	33	CAAAGCCACACTGAAAACCTT	50.098.231	5827	34
GGGATTCTGAGTAGCA	50.098.469	6065	35	CAGAAATCCTCGCGT	50.098.526	6122	36
TGCTGCCTTAAGTCCG	50.098.937	6533	37	ACACTTGAGAAAACGGG	50.098.797	6393	38

Cebadores Directos				Cebadores Inversos			
Secuencia	Coordenada UCSC del Cebador (del extremo 3' del cebador)	Posición del Cebador en el Producto de la PCR Clonado (del extremo 3' del cebador)	SEQ ID NO:	Secuencia	Coordenada UCSC del Cebador (del extremo 3' del cebador)	Posición del Cebador en el Producto de la PCR Clonado (del extremo 3' del cebador)	SEQ ID NO:
CTGGGGTCAGCTGAT	50.099.350	6946	39	ACAAAGTCCTCTATAGCC	50.099.077	6673	40
TGAAACATCTGGGATTTATAAC	50.099.679	7275	41	TAACCTGGGGTTGGTT	50.099.429	7025	42
CTGGAACACCACAATACC	50.099.990	7586	43	AAGTTCCTTTGCTCATCAG	50.099.829	7425	44
ATCTGGGCTCACTGTA	50.100.261	7857	45	GCAAGAGGGGAGACTGT	50.100.207	7803	46
GTCAAAAGACCCTCTATGC	50.100.739	8335	47	TGTGCCTGGATGAATGTA	50.100.567	8163	48
AGGACTCCACGAGT	50.101.197	8793	49	TGAGCTCATCCCCCGT	50.100.960	8556	50
				CCGTGTTCCATTATGAG	50.101.328	8924	51
				GTAAAACGACGGCCAG	50.101.681	9277	52

Los productos de PCR se hicieron circular en un gel de agarosa al 0,8%, se visualizaron mediante tinción con cristal violeta, se compararon con patrones de tamaño, se separaron del gel mediante corte, y se extrajeron con materiales de purificación incluidos con el kit TOPO XL PCR Cloning (Invitrogen). Los productos de la PCR de largo alcance se clonaron en un vector de clonación TOPO XL PCR. Este sistema utiliza un vector de clonación TA y se recomienda para las inserciones de hasta 10 kb. Siguiendo las instrucciones del fabricante, las células electrocompetentes (del mismo kit) se transformaron por medio del vector, se cultivaron en placa en presencia de antibiótico, y se incubaron. Se seleccionaron diez clones de cada placa y se cultivaron en un formato de 96 pocillos.

Los cultivos diluidos se transfirieron a un tampón desnaturalizante que formaba parte del kit TempliPhi ADN Sequencing Template Amplification (GE Healthcare/Amersham Biosciences). Este tampón provoca la liberación de ADN plasmídico, pero no de ADN bacteriano. Los cultivos se calentaron, se enfriaron, se centrifugaron, y se transfirieron a nuevas placas que contenían la enzima TempliPhi y otros componentes. Esta mezcla se incubó a 30°C durante 18 horas para promover la amplificación de los moldes del plásmidos. Estos productos se centrifugaron a continuación y se calentaron a 65°C para destruir la enzima.

Los moldes del plásmido se utilizaron en reacciones de secuenciación de ADN utilizando el kit de secuenciación Big Dye, versión 3.1 (Applied Biosystems). Para cada reacción, se utilizó un cebador de secuenciación apropiado (Tabla 4) que se diseñó para que hibridara con localización única del molde. La secuenciación en ciclos se llevó a cabo con una temperatura de hibridación de 50°C, una temperatura de elongación de 60°C, y una temperatura de desnaturalización de 96°C, durante un total de 30 ciclos. Los productos de reacción de la secuenciación se hicieron circular en un secuenciador de ADN ABI 3730XL con un conjunto de capilares dev 50 cm utilizando el modo de ejecución convencional.

Se utilizó un programa de análisis de secuenciación patentado denominado "Agent" (desarrollado por Celera) para alinear las lecturas de secuenciación para la secuencia de referencia apropiada, y producir "cúntigos" asociados con cada clon. El sistema proporciona las puntuaciones de calidad estimadas para todas las bases para las que existe cualquier variación de cualquiera de las muestras. Se analizó el informe de secuenciación para cada muestra para determinar la presencia de los SNP que estuvieran correlacionados en un patrón de haplotipos para un subgrupo de clones y en un patrón de haplotipos diferente para los clones restantes. Se preparó un archivo de referencia para la región de interés enumerando las variaciones conocidas para esa región disponibles públicamente de NCBI dbSNP. Se creó un archivo de genotipos para la región de interés investigando cada informe de haplotipos del sujeto para todas las variaciones entre la secuencia de referencia conocida y las secuencias de haplotipos consenso.

La magnitud del error de lectura de la longitud para las variantes de poli-T (p. ej., rs10524523) se estimó mediante el examen de las longitudes observadas de los 10 clones que se habían preparado para las muestras que tenían un único haplotipo. Para una muestra típica con una longitud corta de poli-T corta de 16, la desviación típica de los 10 clones fue de 0,97. Para una muestra típica con una longitud de poli-T mayor, p. ej., 27, la desviación típica fue de 1,58.

Se realizó un análisis filogenético. Se llevó a cabo un alineamiento de secuencias múltiples de las secuencias utilizando el programa ClustalW2 (versión 2.0.10) utilizando los parámetros por defecto. El manual de ajuste de los alineamientos se completó utilizando Genedoc (versión 2.7.000). Se construyeron árboles filogenéticos utilizando reconstrucciones Bayesianas, dev máxima verosimilitud y basadas en la distancia. El soporte lógico para la construcción del árbol filogenético utilizado fue Paup* (versión 4.0b10), ClustalX2 (métodos del vecino más próximo, versión 2.0.10) y Mr. Bayes (versión 3.1.2).

Se utilizó el intercambio de ramas con bisección y reconexión del árbol en todos los métodos. El mejor modelo de ajuste de evolución de la secuencia se estimó utilizando el programa Modeltest (versión 3.7), que proporcionó estimaciones para los siguientes determinantes clave: matriz de tasas, forma de la distribución gamma y proporción de sitios invariantes. Se realizó un análisis bootstrap utilizando 1000 réplicas para determinar el soporte estadístico para la morfología específica del árbol.

Asimismo se construyeron redes de haplotipos a partir de los datos de secuencia utilizando el programa TCS (versión 1.21 (Clement et al. TCS: a computer program to estimate gene genealogies. *Mol Ecol* **9**, 1657-9 (2000))) para comparar los árboles filogenéticos con los cladogramas estimados utilizando la parsimonia estadística. Los árboles filogenéticos y las redes de haplotipos se construyeron dos veces, con los espacios tratados como datos perdidos para el primer caso y como un quinto carácter para el segundo caso. Se calculó la diversidad de nucleótidos en la región de interés utilizando ADNSP (versión 5.00.02 (Librado et al v5 ADNSP: a software for comprehensive analysis of ADN polymorphism data. *Bioinformatics* **25**, 1451-2 (2009))).

Después de la construcción de los árboles filogenéticos, la red de haplotipos, y la finalización del análisis de la diversidad de nucleótidos en la región de interés, se compararon los resultados de los diferentes métodos y se conciliaron a un árbol consenso. Se supuso que los grupos de secuencias que compartían una mutación de enfermedad reciente segregaban más de cerca en el árbol filogenético, sin embargo, los casos esporádicos debidos a fenocopias, dominancia y epistasis pueden introducir ruido en la relación fenotipo-haplotipo (Tachmazidou et al. Genetic association mapping via evolution-based clustering of haplotypes. *PLoS Genet* **3**, e111 (2007)).

5 Sin embargo, los casos esporádicos debidos a fenocopias, dominancia y epistasis pueden introducir ruido en la relación fenotipo-haplotipo. Este análisis filogenético se centró en una agregación de alto nivel de clados con el fin de minimizar estos efectos. Los clados determinados en la primera división en el árbol filogenético se utilizaron para someter a ensayo la hipótesis de que los haplotipos de sujetos TOMM40 del clado "B" estaban asociados con el inicio de la EA a una edad más tardía que los haplotipos del sujeto del clado "A", (cada sujeto contribuyó con dos haplotipos a la señal de asociación de la edad de inicio de la EA). El número de ensayos de asociación que se realizó utilizando de este enfoque tuvo órdenes de magnitud menores que en los estudios de asociación del genoma completo típicos ya que el análisis filogenético identificó categorías de haplotipos de sujetos relacionados evolutivamente. Si las pruebas de asociación confirmaban que los diferentes clados clasificaban los datos de haplotipo del sujeto por edad de inicio, se realizó adicionalmente un análisis estadístico para identificar las variantes que separaban las secuencias en cada clado. Efectivamente, este análisis evaluó la significación de cada variante como un factor que influye en la edad de inicio utilizando una serie de ensayos de un grado de libertad guiados por la estructura de árbol. Los análisis filogenéticos se llevaron a cabo utilizando polimorfismos de un solo nucleótido y de inserción/delección. Las pruebas estadísticas de asociación se ajustaron con una corrección de Bonferroni para el número de sitios polimórficos incluidos en el análisis.

10 Los informes de haplotipos del soporte lógico de análisis Polymorphic y los informes del soporte lógico ADNSP (versión 5.00.02 (Librado et al ADNSP v5: a software for comprehensive analysis of ADN polymorphism data. *Bioinformatics* **25**, 1451-2 (2009))) se utilizaron para posteriores análisis estadísticos. Los autores de la presente invención analizaron variantes SNP de TOMM40 individuales, haplotipos de TOMM40 y longitudes de repeticiones de poli-T para determinar la asociación con el riesgo de EAT para la cohorte AS y la edad de inicio de EAT para la cohorte de DS. Las diferencias en las proporciones de los alelos de TOMM40 específicos asociadas a cada alelo ApoE o genotipo ApoE se compararon utilizando una prueba exacta de Fisher (dos colas). Partiendo de 30 sitios informativos de parsimonia y $\alpha = 0,05$, una corrección de Bonferroni para la significación de una asociación alélica específica requeriría un valor de p de 0,001. Las probabilidades relativas (OR) se calcularon como (número de alelos de menor importancia en los casos/número de alelos de menor importancia en los controles)/(número de alelos principales en los casos/número de alelos principales en los controles) y se refirieron con intervalos de confianza de 95%. Las medias para los grupos de edad de comienzo de EAT definidos se compararon mediante pruebas t, de dos colas. Se realizó un ensayo F convencional en las varianzas de los grupos para determinar si la prueba t se calculaba suponiendo varianzas iguales o desiguales. El análisis estadístico se completó utilizando el soporte logico JMP (versión 8, SAS Institute, Cary, NC).

Códigos de acceso: GenBank: TOMM40, translocasa del homólogo 40 de la membrana mitocondrial externa, 10452; ApoE, apolipoproteína E, 348

Lo anterior es ilustrativo de la presente invención, y no se deben considerar limitante de la misma. La invención se define por las siguientes reivindicaciones, incluyendo en ella equivalentes de las reivindicaciones.

35

Listado de secuencias

<110> Shiraz Pharmaceuticals, Inc. Roses, Alan D.

5 <120> MÉTODO DE IDENTIFICACIÓN DE FACTORES DE RIESGO DE ENFERMEDADES

<130> 9719-2WO

<150> US 61/224,647

10 <151> 2009-07-10

<150> US 61/186,673

<151> 2009-06-12

15 <150> US 61/088,203

<151> 2008-08-12

<160> 52

20 <170> Patent In version 3.3

<210> 1

<211> 25

<212> ADN

25 <213> Artificial

<220>

<223> Cebador para la PCR de largo alcance

30 <400> 1

aactcagagg ccagagattc taagt 25

<210> 2

35 <211> 25

<212> ADN

<213> Artificial

<220>

40 <223> Cebador para la PCR de largo alcance

<400> 2

aacagcctaa tcccagcaca ttac 25

45 <210> 3

<211> 17

<212> ADN

<213> Artificial

50 <220>

<223> Cebador para la secuenciación

<400> 3

55 <210> 4

<211> 15

<212> ADN

<213> Artificial

60 <220>

<223> Cebador para la secuenciación

65 <210> 4

<211> 15

<212> ADN

<213> Artificial

<220>

<223> Cebador para la secuenciación

65 <210> 4

<211> 15

<212> ADN

<213> Artificial

<400> 4

cccactgggt gttga 15

5 <210> 5
 <211> 18
 <212> ADN
 <213> Artificial

10 <220>
 <223> Cebador para la secuenciación

<400> 5

15 **gtgtgatggt gattcaac** 18

<210> 6
 <211> 16
 <212> ADN
 <213> Artificial

20 <220>
 <223> Cebador para la secuenciación

25 <400> 6

gaataggggc cttca 16

<210> 7
 <211> 15
 <212> ADN
 <213> Artificial

30 <220>
 <223> Cebador para la secuenciación

35 <400> 7

ctgcaggtat gaaag 15

40 <210> 8
 <211> 16
 <212> ADN
 <213> Artificial

45 <220>
 <223> Cebador para la secuenciación

<400> 8

50 **caatctccta ggggtgc** 16

<210> 9
 <211> 15
 <212> ADN
 <213> Artificial

55 <220>
 <223> Cebador para la secuenciación

60 <400> 9

gtctctgcag atgtg 15

<210> 10
 <211> 16
 <212> ADN
 <213> Artificial
 5
 <220>
 <223> Cebador para la secuenciación
 <400> 10
 10 **cggaagttgc agtaag** **16**
 <210> 11
 <211> 15
 <212> ADN
 <213> Artificial
 15
 <220>
 <223> Cebador para la secuenciación
 20
 <400> 11
 <210> 12
 <211> 17
 <212> ADN
 <213> Artificial
 25
 <220>
 <223> Cebador para la secuenciación
 30
 <400> 12
 35 **aaggccaagg ttacact** **17**
 <210> 13
 <211> 15
 <212> ADN
 <213> Artificial
 40
 <220>
 <223> Cebador para la secuenciación
 45
 <400> 13
 <210> 14
 <211> 17
 <212> ADN
 <213> Artificial
 50
 <220>
 <223> Cebador para la secuenciación
 55
 <400> 14
 60 **acaagcctag gtgacat** **17**
 <210> 15
 <211> 21
 <212> ADN
 <213> Artificial

<220>
<223> Cebador para la secuenciación

5 <400> 15

cccaactaat tttgtattc g 21

<210> 16
<211> 17
<212> ADN
<213> Artificial

<220>
<223> Cebador para la secuenciación

15 <400> 16

cctgtaatcc cagctat 17

20 <210> 17
<211> 17
<212> ADN
<213> Artificial

25 <220>
<223> Cebador para la secuenciación

<400> 17

30 **acatttggg cctgtac 17**

<210> 18
<211> 20
<212> ADN
<213> Artificial

35 <220>
<223> Cebador para la secuenciación

40 <400> 18

tcattctct gtgaacctaa 20

45 <210> 19
<211> 16
<212> ADN
<213> Artificial

50 <220>
<223> Cebador para la secuenciación

<400> 19

55 **ccacatgggc ttgtgt 16**

<210> 20
<211> 18
<212> ADN
<213> Artificial

60 <220>
<223> Cebador para la secuenciación

<400> 20

ggcaaaatga cgatcagt 1E

5 <210> 21
 <211> 16
 <212> ADN
 <213> Artificial

10 <220>
 <223> Cebador para la secuenciación

<400> 21

15 **cccagatgcc caaatc** 16

<210> 22
 <211> 15
 <212> ADN
 <213> Artificial

20 <220>
 <223> Cebador para la secuenciación

25 <400> 22

gcagcaccag ctagt 15

30 <210> 23
 <211> 17
 <212> ADN
 <213> Artificial

35 <220>
 <223> Cebador para la secuenciación

<400> 23

40 **aactctgagt ggatgtg** 17

<210> 24
 <211> 19
 <212> ADN
 <213> Artificial

45 <220>
 <223> Cebador para la secuenciación

<400> 24

50 **gatggtctca atctcctta** 1E

55 <210> 25
 <211> 19
 <212> ADN
 <213> Artificial

60 <220>
 <223> Cebador para la secuenciación

<400> 25

ctatagtccc aactactga 1E

5 <210> 26
<211> 25
<212> ADN
<213> Artificial

<220>
<223> Cebador para la secuenciación

<400> 26

10 **tttttccaa gcataaaaca tagta** 25

<210> 27
<211> 15
<212> ADN
<213> Artificial

<220>
<223> Cebador para la secuenciación

20 <400> 27

agtccccgct actta 15

25 <210> 28
<211> 16
<212> ADN
<213> Artificial

30 <220>
<223> Cebador para la secuenciación

<400> 28

35 **ggggatggac aaagct** 16

<210> 29
<211> 16
<212> ADN
<213> Artificial

40 <220>
<223> Cebador para la secuenciación

45 <400> 29

accacaggtg tatgcc 16

50 <210> 30
<211> 17
<212> ADN
<213> Artificial

<220>
<223> Cebador para la secuenciación

55 <400> 30

tgaaaagccc tctagac 17

60 <210> 31
<211> 17
<212> ADN
<213> Artificial

<220>
 <223> Cebador para la secuenciación
 5 <400> 31
 gaacagattc atccgca 17
 10 <210> 32
 <211> 16
 <212> ADN
 <213> Artificial
 15 <220>
 <223> Cebador para la secuenciación
 <400> 32
 cacccacgat ccagtt 16
 20 <210> 33
 <211> 18
 <212> ADN
 <213> Artificial
 25 <220>
 <223> Cebador para la secuenciación
 <400> 33
 30 <210> 34
 <211> 19
 <212> ADN
 <213> Artificial
 35 <220>
 <223> Cebador para la secuenciación
 <400> 34
 caaagccaca ctgaaactt 19
 45 <210> 35
 <211> 16
 <212> ADN
 <213> Artificial
 50 <220>
 <223> Cebador para la secuenciación
 <400> 35
 55 <210> 36
 <211> 13
 <212> ADN
 <213> Artificial
 60 <220>
 <223> Cebador para la secuenciación

<400> 36

cagaatcctg cgt 13

5 <210> 37
<211> 16
<212> ADN
<213> Artificial

10 <220>
<223> Cebador para la secuenciación

<400> 37

15 **tgctgcctta agtccg** 16

<210> 38
<211> 16
<212> ADN
20 <213> Artificial

<220>
<223> Cebador para la secuenciación

25 <400> 38

acacttgaga aaacgg 16

<210> 39
30 <211> 15
<212> ADN
<213> Artificial

<220>
35 <223> Cebador para la secuenciación

<400> 39

ctggggtcag ctgat 15

40 <210> 40
<211> 18
<212> ADN
<213> Artificial

45 <220>
<223> Cebador para la secuenciación

<400> 40

50 **acaaagtcct ctatagcc** 18

<210> 41
<211> 22
55 <212> ADN
<213> Artificial

<220>
<223> Cebador para la secuenciación
60 <400> 41

tgaaacatct gggattata ac 22

5 <210> 42
<211> 16
<212> ADN
<213> Artificial

<220>
<223> Cebador para la secuenciación

10 <400> 42

taacctgggg ttggtt 16

15 <210> 43
<211> 17
<212> ADN
<213> Artificial

20 <220>
<223> Cebador para la secuenciación

<400> 43

25 **ctggaacca caatacc** 17

30 <210> 44
<211> 19
<212> ADN
<213> Artificial

<220>
<223> Cebador para la secuenciación

35 <400> 44

aagttccttt gctcatcag 19

40 <210> 45
<211> 16
<212> ADN
<213> Artificial

45 <220>
<223> Cebador para la secuenciación

<400> 45

atctcggctc actgta 16

50 <210> 46
<211> 16
<212> ADN
<213> Artificial

55 <220>
<223> Cebador para la secuenciación

<400> 46

60 **gcaagagggga gactgt** 16

<210> 47

<211> 18
 <212> ADN
 <213> Artificial

5 <220>
 <223> Cebador para la secuenciación

<400> 47

10 **gtcaaaagac ctctatgc** 18

<210> 48
 <211> 18
 <212> ADN
 <213> Artificial

15 <220>
 <223> Cebador para la secuenciación

20 <400> 48

tgtagcctgga tgaatgta 18

<210> 49
 <211> 14
 <212> ADN
 <213> Artificial

25 <220>
 <223> Cebador para la secuenciación

30 <400> 49

aggactccac gagt 14

35 <210> 50
 <211> 15
 <212> ADN
 <213> Artificial

40 <220>
 <223> Cebador para la secuenciación

45 <400> 50

tgagctcatc cccgt 15

<210> 51
 <211> 18
 <212> ADN
 <213> Artificial

50 <220>
 <223> Cebador para la secuenciación

55 <400> 51

ccgtgtcca ttatgag 18

60 <210> 52
 <211> 16
 <212> ADN
 <213> Artificial

ES 2 463 766 T3

<220>

<223> Cebador para la secuenciación

<400> 52

5

gtaaacgac ggccag 16

REIVINDICACIONES

1. Un método para determinar el riesgo de desarrollar la enfermedad de Alzheimer en un sujeto, que comprende:
 - (a) detectar en una muestra biológica que contiene ADN tomada de dicho sujeto la presencia o ausencia de una variante genética del gen TOMM40 asociada con un aumento o disminución del riesgo de enfermedad de Alzheimer, en donde dicha variante es un polimorfismo de delección/inserción en el intrón 6 o intrón 9 del gen TOMM40; y
 - (b) determinar si dicho sujeto tiene mayor o menor riesgo de enfermedad de Alzheimer cuando dicha variante genética está presente o ausente.
2. El método de la reivindicación 1, que comprende adicionalmente detectar si dicho sujeto es un sujeto Apo E2/E2, E2/E3, E2/E4, E3/E3, E3/E4, o E4/E4.
3. El método de la reivindicación 1, que comprende adicionalmente detectar si dicho sujeto es un sujeto de Apo E3/E3 o E3/E4.
4. Una cantidad eficaz como tratamiento de un agente activo anti-enfermedad de Alzheimer seleccionado del grupo que consiste en inhibidores de acetilcolinesterasa, antagonistas del receptor de NMDA, agonistas o moduladores del receptor activado por el proliferador de peroxisomas (PPAR), anticuerpos, proteínas de fusión, moléculas de ARN terapéuticas, y combinaciones de los mismos, para su uso en un método de tratamiento de la enfermedad de Alzheimer caracterizado porque el método comprende la etapa de determinar si el sujeto está en riesgo o no de desarrollar la enfermedad de Alzheimer, en donde el riesgo se determina de acuerdo con el método de una cualquiera de las reivindicaciones 1-3.
5. La cantidad eficaz como tratamiento de un agente activo anti-enfermedad de Alzheimer para su uso como en la reivindicación 4, en donde dicho agente es un agonista o modulador del receptor activado por el proliferador de peroxisomas.
6. La cantidad eficaz como tratamiento de un agente activo anti-enfermedad de Alzheimer para su uso como en la reivindicación 5, en donde dicho modulador o agonista es una tiazolidinodiona.
7. La cantidad eficaz como tratamiento de un agente activo anti-enfermedad de Alzheimer para su uso como en la reivindicación 5, en donde el modulador o agonista es un fármaco de tiazolidinodiona.
8. La cantidad eficaz como tratamiento de un agente activo anti-enfermedad de Alzheimer para su uso como en una cualquiera de las reivindicaciones 4-7, en donde dicho tratamiento es para retrasar el inicio de la enfermedad de Alzheimer.
9. La cantidad eficaz como tratamiento de un agente activo anti-enfermedad de Alzheimer para su uso como en la reivindicación 4 o la reivindicación 5, en donde dicha variante genética del gen TOMM40 es una variante de inserción de poli-T en la localización genómica 50094889 de NCBI Build 36.3_ (rs10524523).
10. Una cantidad eficaz como tratamiento de un agente activo anti-enfermedad de Alzheimer para su uso como en la reivindicación 6 o la reivindicación 7, en donde dicha variante genética del gen TOMM40 es una variante de inserción de poli-T en la localización genómica 50094889 de NBCI Build 36.3 (rs10524523).
11. La cantidad eficaz como tratamiento de un agente activo anti-enfermedad de Alzheimer para su uso de acuerdo con cualquiera de las reivindicaciones 4-8, en donde dicho polimorfismo de delección/inserción (DIP) es un polimorfismo de delección/inserción de poli-T.
12. La cantidad eficaz como tratamiento de un agente activo anti-enfermedad de Alzheimer para su uso de acuerdo con cualquiera de las reivindicaciones 4-8, en donde dicha variante genética del gen TOMM40 es rs10524523.
13. La cantidad eficaz como tratamiento de un agente activo anti-enfermedad de Alzheimer para su uso de acuerdo con la reivindicación 4 o la reivindicación 5, en donde dicho DIP es un polimorfismo de delección/inserción de poli-T que tiene un poli-T de entre 20 y 50 pares de bases contiguos.
14. La cantidad eficaz como tratamiento de un agente activo anti- enfermedad de Alzheimer para su uso de acuerdo con la reivindicación 6 o la reivindicación 7, en donde dicho DIP es un polimorfismo de delección/inserción de poli-T que tiene un poli-T de entre 20 y 50 pares de bases contiguos en rs10524523.
15. La cantidad eficaz como tratamiento de un agente activo anti-enfermedad de Alzheimer para su uso de acuerdo con la reivindicación 10, en donde dicho tratamiento comprende retrasar el inicio de la enfermedad de Alzheimer.
16. Un método para determinar un pronóstico del riesgo de desarrollar la enfermedad de Alzheimer en un paciente, que comprende la obtención de un perfil de un paciente, en donde dicha obtención de un perfil de un paciente comprende:

detectar la presencia o ausencia de al menos un alelo ApoE 2, ApoE 3, o ApoE 4 en una muestra biológica de dicho paciente, y

detectar la presencia o ausencia de al menos un polimorfismo de delección/inserción (DIP) de TOMM40 localizado en el intrón 6 o intrón 9 del gen TOMM40 y, a continuación,

5 convertir dicho perfil del paciente en dicho pronóstico, en donde la presencia de dicho alelo ApoE 2, ApoE 3 o ApoE 4 y la presencia de dicho al menos un polimorfismo DIP de TOMM40 identifican a dicho paciente como un paciente que tiene riesgo de desarrollar la enfermedad de Alzheimer.

17. El método de la reivindicación 16, en donde dicho DIP es un polimorfismo de delección/inserción de poli-T.

18. El método de la reivindicación 16, en donde dicho DIP es rs10524523.

10 19. El método de la reivindicación 16, en donde la variante genética del gen TOMM40 es un polimorfismo de delección/inserción rs10524523 que tiene un poli-T de entre 20 y 50 pares de bases contiguos.

20. El método de la reivindicación 16, que comprende adicionalmente:

15 detectar en una muestra biológica que contiene ADN tomada de dicho sujeto la presencia o ausencia de una variante genética del gen TOMM40 asociada con un aumento o disminución del riesgo de enfermedad de Alzheimer, en donde dicha variante es un polimorfismo de delección/inserción en el intrón 6 o intrón 9 del gen TOMM40; y

determinar si dicho sujeto tiene un riesgo mayor o menor de enfermedad de Alzheimer cuando dicha variante genética está presente o ausente.

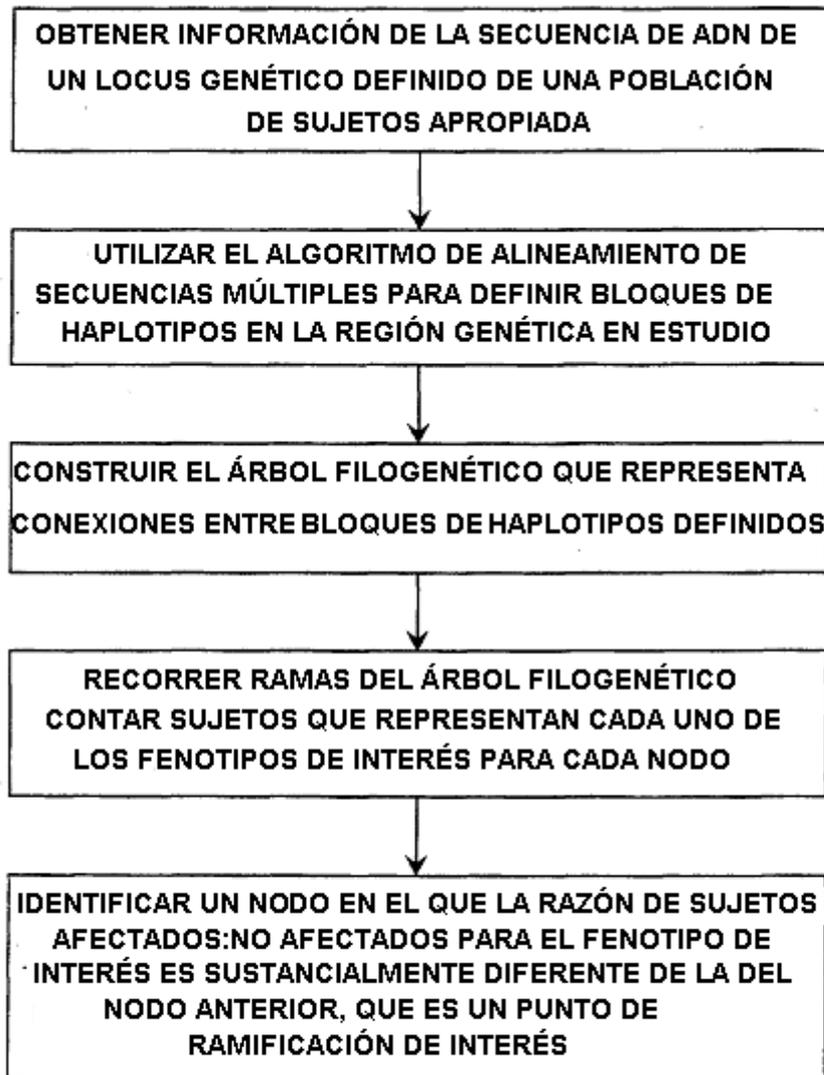


FIG. 1

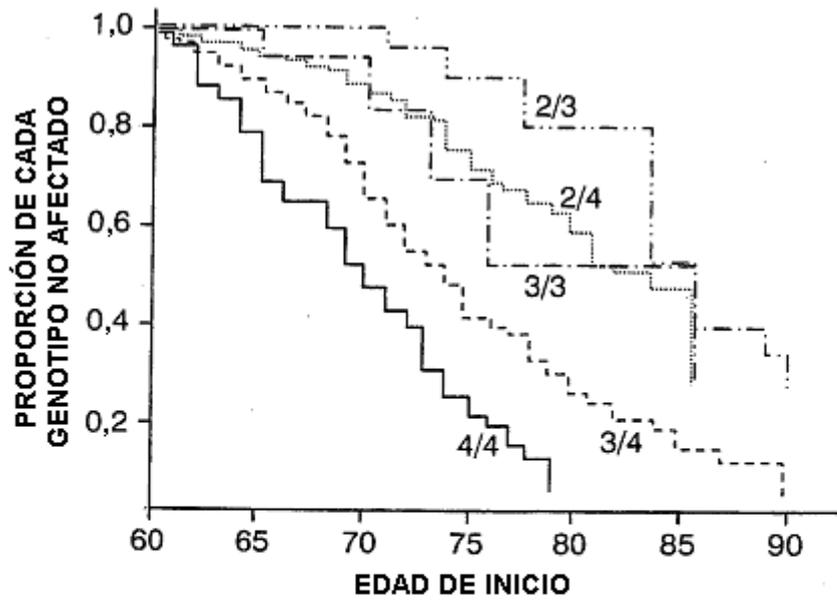


FIG. 2

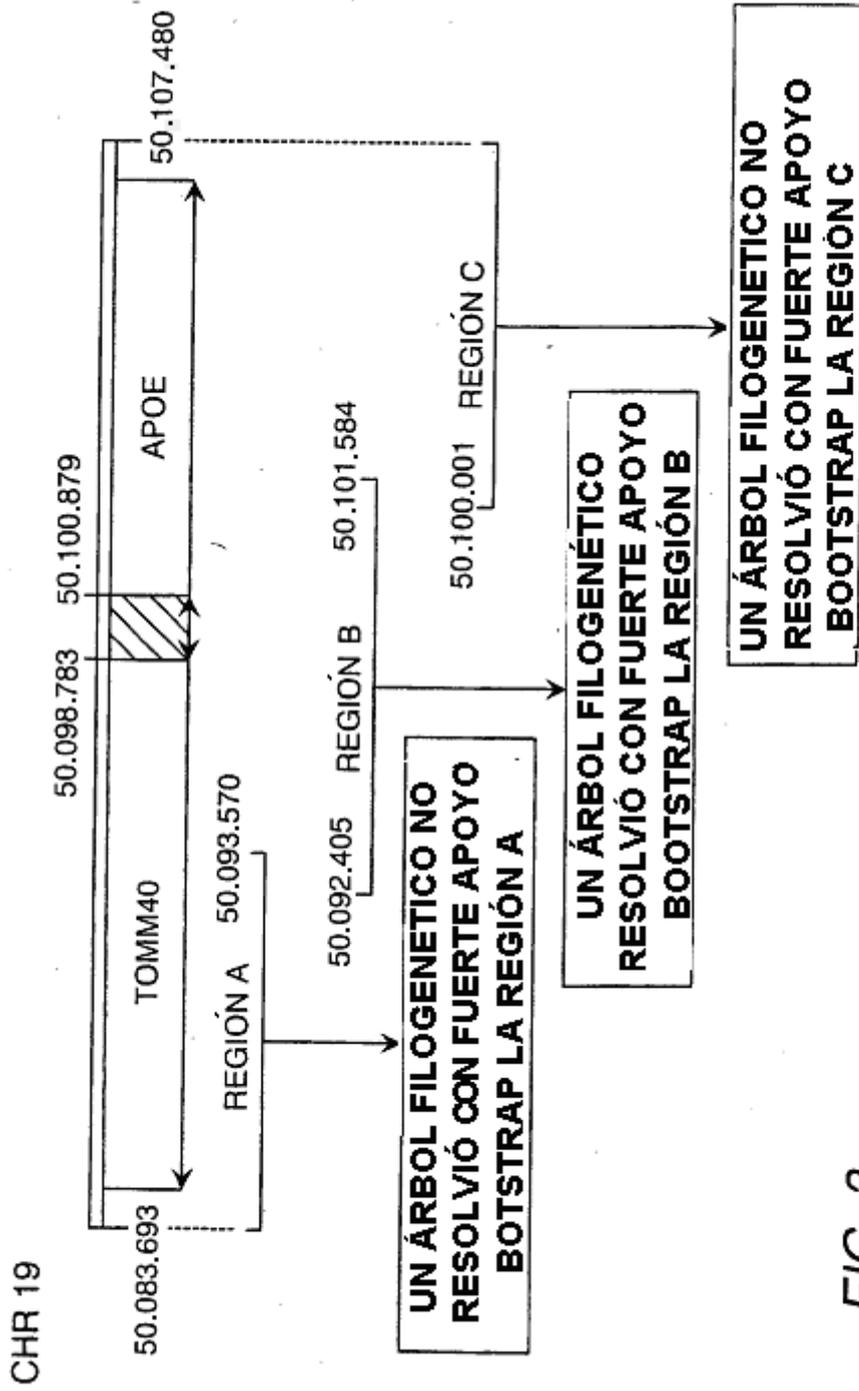


FIG. 3

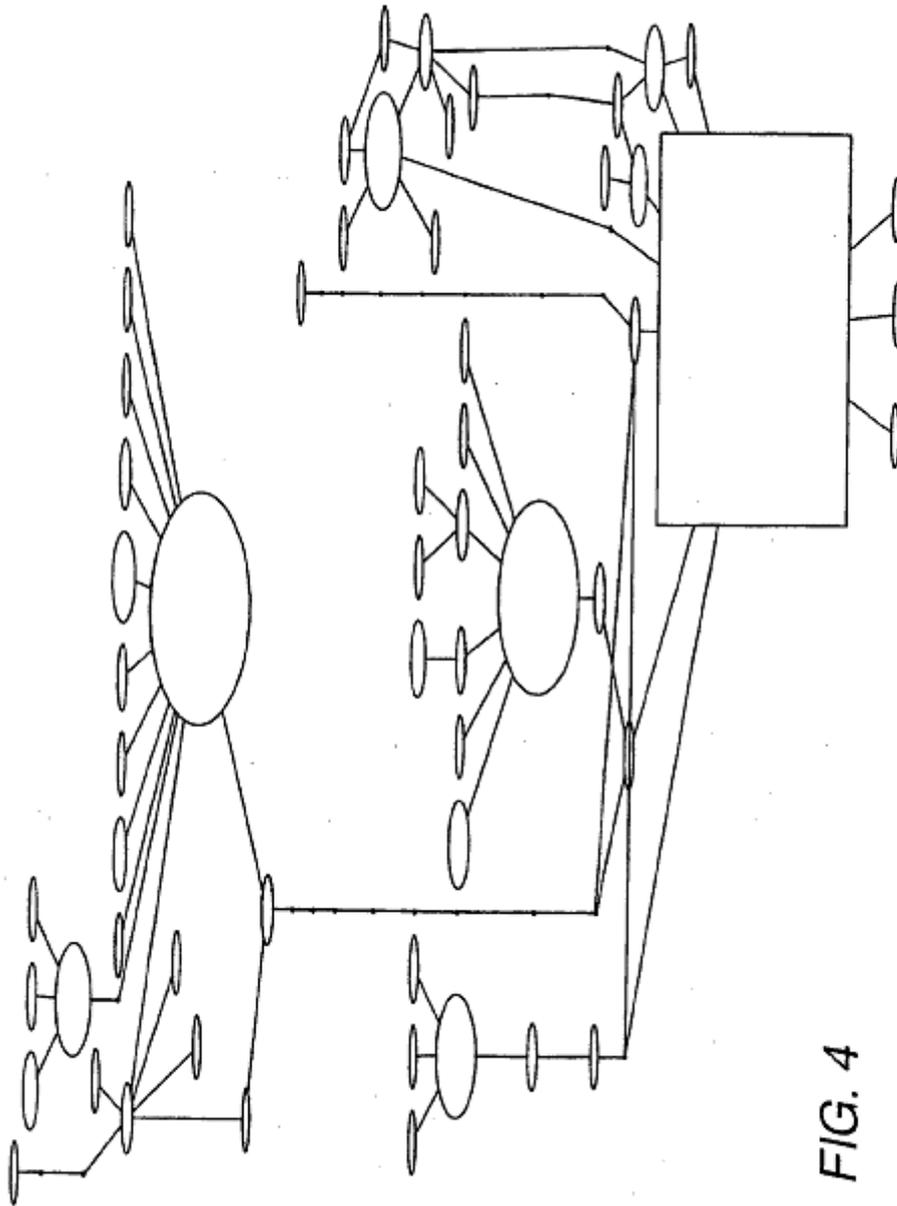


FIG. 4

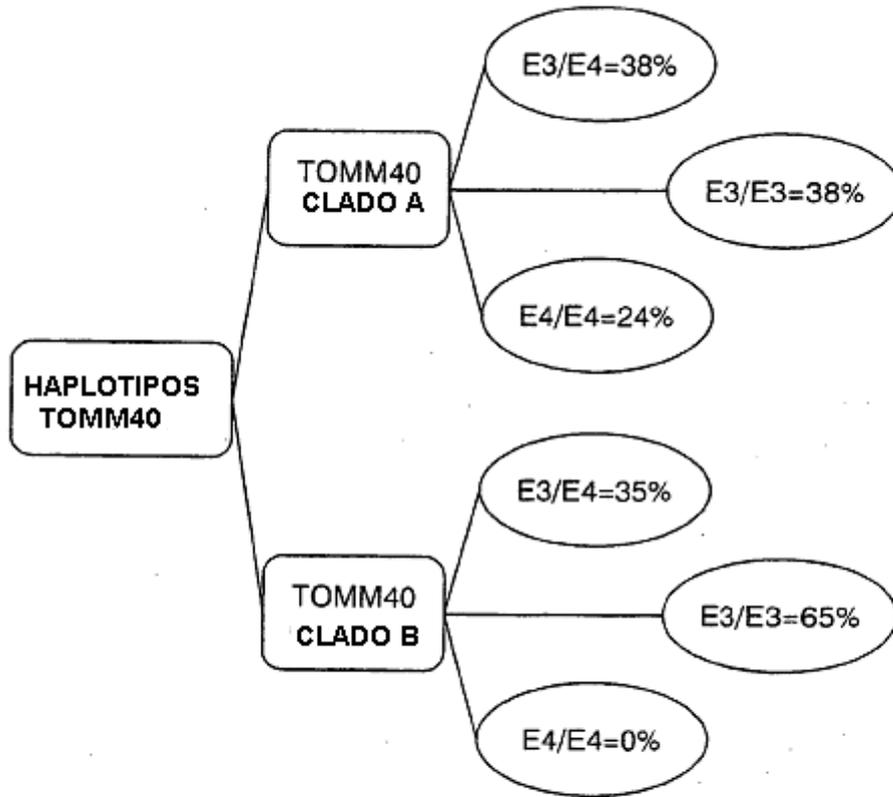


FIG. 5

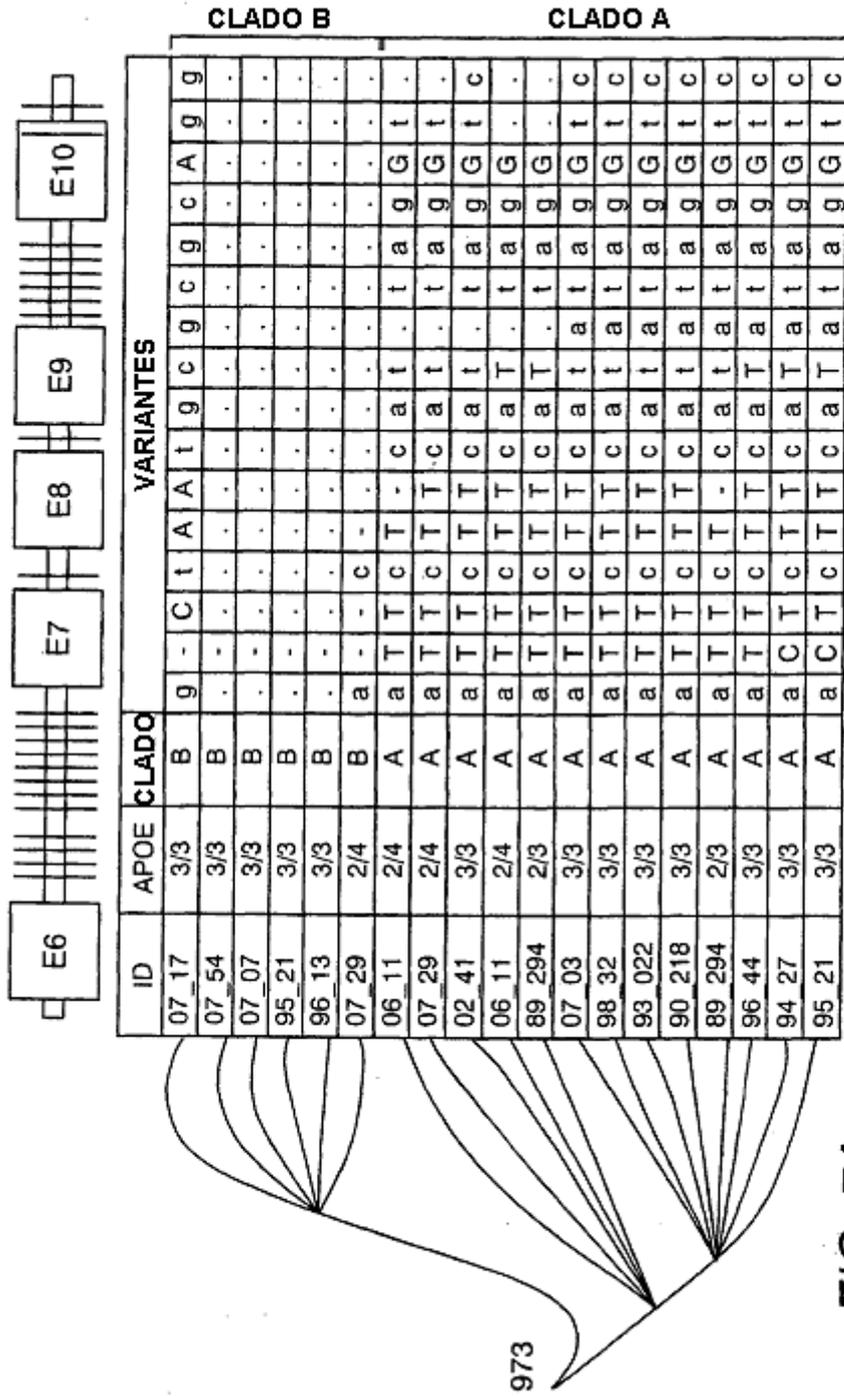


FIG. 7A

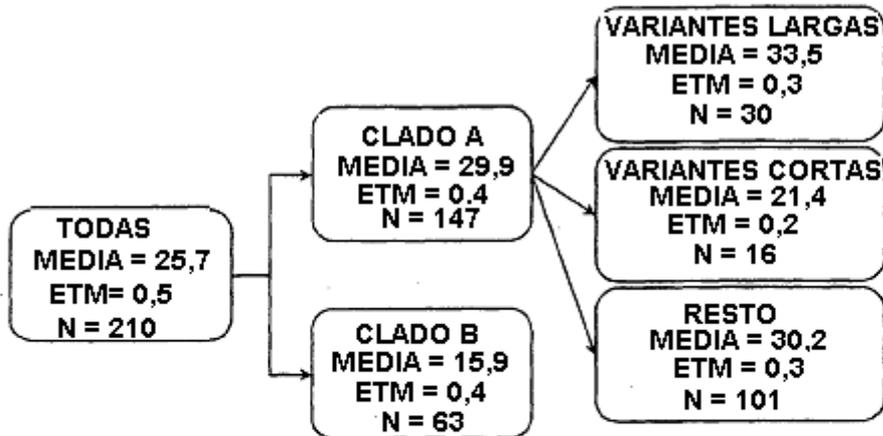


FIG. 7B

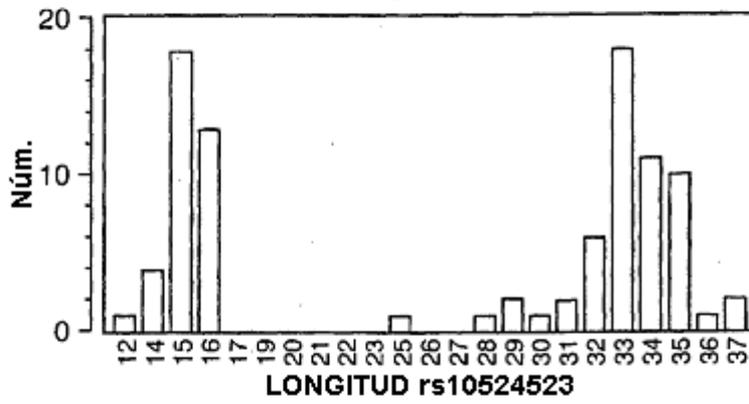


FIG. 8A

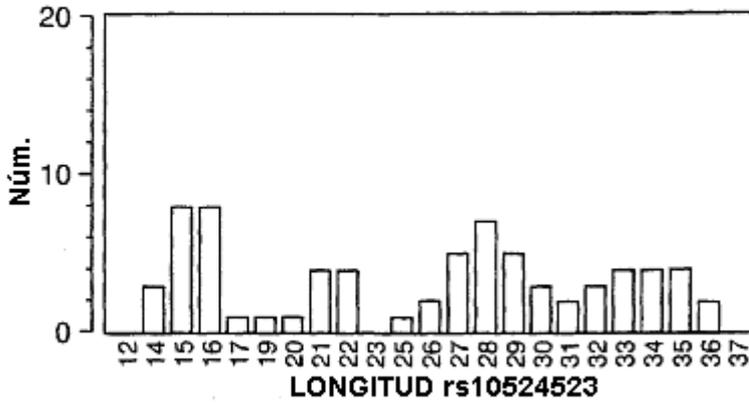


FIG. 8B

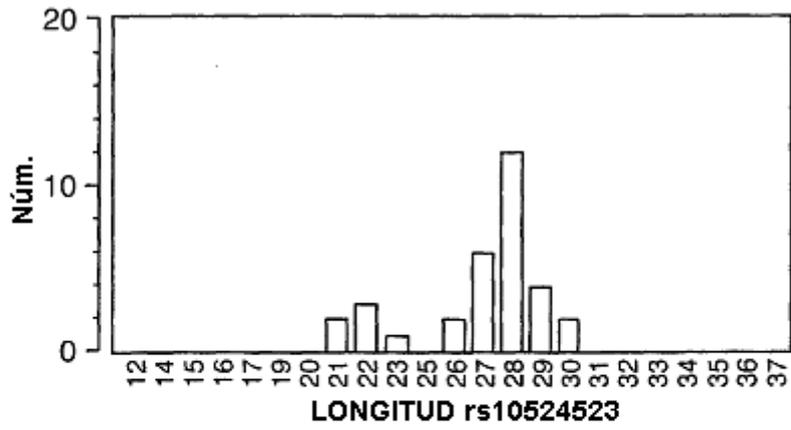


FIG. 8C

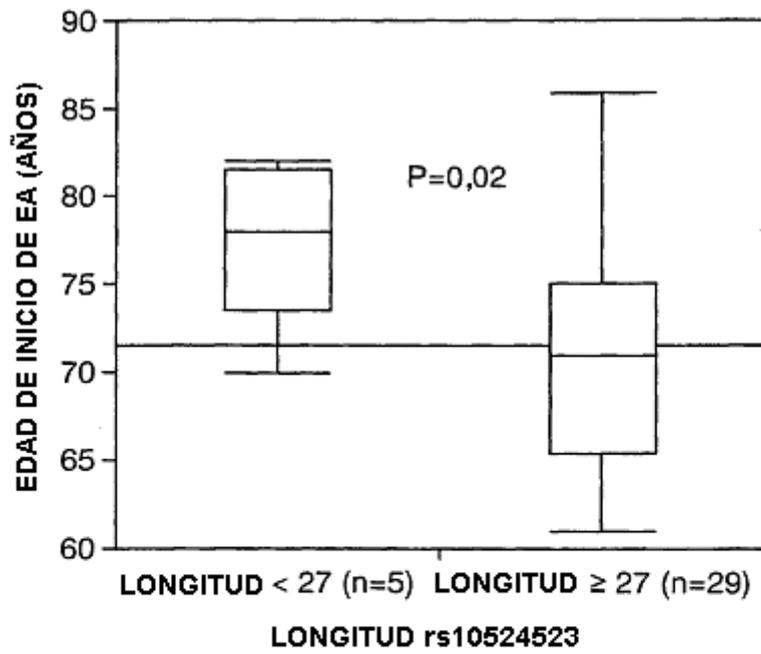


FIG. 9