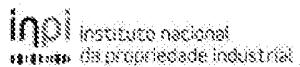

(11) Número de Publicação: **PT 2324126 E**



(51) Classificação Internacional:
C12Q 1/68 (2014.01)

(12) FASCÍCULO DE PATENTE DE INVENÇÃO

(22) Data de pedido: **2009.08.11**

(30) Prioridade(s): **2008.08.12 US 88203 P
2009.06.12 US 186673 P
2009.07.10 US 224647 P**

(43) Data de publicação do pedido: **2011.05.25**

(45) Data e BPI da concessão: **2014.04.23
105/2014**

(73) Titular(es):

**ZINFANDEL PHARMACEUTICALS, INC.
ONE SCIENCE DRIVE DURHAM, NC 27708 US**

(72) Inventor(es):

ALLEN D. ROSES US

(74) Mandatário:

**ANTÓNIO INFANTE DA CÂMARA TRIGUEIROS DE ARAGÃO
RUA DO PATROCÍNIO, Nº 94 1399-019 LISBOA PT**

(54) Epígrafe: **MÉTODO DE IDENTIFICAÇÃO DE FACTORES DE RISCO DA DOENÇA DE ALZHEIMER**

(57) Resumo:

É AQUI PROPORCIONADO UM MÉTODO PARA IDENTIFICAR UMA VARIANTE GENÉTICA QUE ESTÁ ASSOCIADA COM DESENVOLVIMENTO DE UMA CONDIÇÃO DE INTERESSE (E. G., DOENÇA DE ALZHEIMER), E VARIANTES GENÉTICAS ENTÃO IDENTIFICADAS. MÉTODOS DE TRATAMENTO COM UM AGENTE ACTIVO (E. G., COM UM AGENTE ACTIVO PARTICULAR E/OU NUMA IDADE PRECOCE) SÃO TAMBÉM PROPORCIONADOS, APÓS DETECTAR UMA VARIANTE GENÉTICA AQUI DESCrita. EM ALGUMAS FORMAS DE REALIZAÇÃO, A VARIANTE GENÉTICA É UM POLIMORFISMO DE DELECCÃO/INSERÇÃO (DIP) DO GENE TOMM40. KITS PARA DETERMINAR SE UM INDIVÍDUO ESTÁ EM RISCO AUMENTADO DE DESENVOLVER DOENÇA DE ALZHEIMER DE INÍCIO TARDIO SÃO TAMBÉM PROPORCIONADOS. KITS PARA DETERMINAR SE UM INDIVÍDUO É RESPONSIVO A TRATAMENTO PARA UMA CONDIÇÃO DE INTERESSE COM UM AGENTE ACTIVO SÃO AINDA PROPORCIONADOS.

RESUMO

"MÉTODO DE IDENTIFICAÇÃO DE FACTORES DE RISCO DA DOENÇA DE ALZHEIMER"

É aqui proporcionado um método para identificar uma variante genética que está associada com desenvolvimento de uma condição de interesse (e. g., doença de Alzheimer), e variantes genéticas então identificadas. Métodos de tratamento com um agente activo (e. g., com um agente activo particular e/ou numa idade precoce) são também proporcionados, após detectar uma variante genética aqui descrita. Em algumas formas de realização, a variante genética é um polimorfismo de deleção/inserção (DIP) do gene TOMM40. Kits para determinar se um indivíduo está em risco aumentado de desenvolver doença de Alzheimer de início tardio são também proporcionados. Kits para determinar se um indivíduo é responsivo a tratamento para uma condição de interesse com um agente activo são ainda proporcionados.

DESCRIÇÃO

"MÉTODO DE IDENTIFICAÇÃO DE FACTORES DE RISCO DA DOENÇA DE ALZHEIMER"

Campo da Invenção

A presente invenção refere-se ao campo da genómica, genética, farmacogenética e bioinformática, incluindo análise do genoma e ao estudo da variação da sequência de ADN. A invenção também se refere a estudos de associação entre variações em sequências de ADN e antecipação da susceptibilidade de um indivíduo a uma doença, distúrbio ou condição e/ou resposta particular, a um tratamento ou fármaco particular.

Antecedentes da Invenção

A procura por marcadores genéticos associados a doenças complexas está em curso. Estudos de varrimento de genoma completo com matrizes de SNP continuam a destacar a região de ApoE como a área mais importante para investigação no estudo da doença de Alzheimer (Coon et al., *J. Clin. Psychiatry* **68**: 613-8 (2007); Li et al., *Arch. Neurol.* **65**: 45-53 (2007)).

A isoforma de ApoE 4 tem sido fortemente associada com maior risco de desenvolver doença de Alzheimer de início tardio. (Pericak-Vance et al., *Am. J. Hum. Genet.* **48**, 1034-50 (1991); Martin et al., 2000, Patente US N° 6027896 de Roses, et al., Patente US N° 5716828 de Roses et al.). A relação é dependente

da dose (Yoshizawa *et al.*, 1994; Schellenberg, 1995). Ou seja, um veículo de dois alelos de ApoE 4 é mais provável desenvolver doença de Alzheimer de início tardio (LOAD) do que um veículo de apenas um alelo de ApoE 4 e numa idade mais precoce (Corder *et al.*, *Science* **261**, 921-3 (1993)).

No entanto, alelos E4 representam apenas cerca de 50% da doença de Alzheimer hereditária. Uma explicação é que ApoE 4 serve apenas como um marcador substituto para algo em desequilíbrio de ligação na proximidade. Em alternativa, considerando a recente descoberta de um papel mecanicista para ApoE 4 na toxicidade mitocondrial, os efeitos negativos de ApoE 4 podem ser anulados ou exacerbados por outro produto génico codificado na proximidade (Chang *et al.*, 2005).

Como o estado de ApoE está também associado com o risco de doença arterial coronária e também, provavelmente, com uma série de outras doenças e distúrbios, as implicações do estudo da região de ApoE não estão limitadas à doença de Alzheimer, mas são potencialmente vastas (Mahley *et al.*, *Proc. Natl. Acad. Sci. USA* **103**: 5644-51 (2006)). De um modo mais geral, o exame de sequências variantes de processos ou vias circundantes de genes em desequilíbrio de ligação com outras regiões genéticas conhecidas por estarem envolvidas em processos de doenças complexas irá proporcionar informação valiosa na explicação dos mecanismos dessas doenças.

Sumário da Invenção

É aqui divulgado um método para identificar uma variante genética que está associada com o desenvolvimento de uma

condição de interesse (e. g., antes ou depois do aparecimento de uma doença de interesse), compreendendo: (a) determinar, a partir de amostras biológicas contendo ADN, sequências nucleotíidas transportadas por uma pluralidade de indivíduos humanos únicos num *locus* genético de interesse, em que os indivíduos incluem (i) indivíduos afectados com a condição de interesse e (ii) indivíduos não afectados com a condição de interesse; (b) identificar variantes genéticas no referido *locus* genético a partir de sequências nucleotíidas observadas na referida pluralidade de indivíduos (e. g., utilizando uma análise de alinhamento de sequências múltiplas); (c) mapear as referidas variantes genéticas construindo uma árvore filogenética das referidas sequências nucleotíidas dos referidos indivíduos, a referida árvore compreendendo ramificações que identificam alterações nas variantes entre os referidos indivíduos (e. g., alterações nas variantes no mesmo cistrão); (d) examinar as variantes genéticas representadas como ramificações na referida árvore e determinar a proporção de indivíduos afectados e não afectados para identificar as alterações que levam a uma proporção alterada de indivíduos afectados e não afectados (de um modo preferido, em que o ponto de partida é a variante genética que representa o maior número de indivíduos); e em seguida (e) identificar uma variante genética ou grupo de variantes (um haplótipo), onde a proporção de indivíduos afectados e não afectados é substancialmente diferente de uma ou mais variantes adjacentes na referida árvore (e. g., pelo menos 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, ou 90% diferente) para, desse modo, identificar uma variante genética associada ao desenvolvimento da referida condição de interesse.

Em algumas formas de realização, todos os indivíduos que transportam um mesmo polimorfismo conhecido que está associado com a condição de interesse.

Em algumas formas de realização, a condição de interesse é uma doença neurodegenerativa, uma doença metabólica (e. g., dislipidemia), uma doença cardiovascular, um distúrbio psiquiátrico ou cancro. Em algumas formas de realização, a doença de interesse é uma doença na qual o ApoE e/ou TOMM40 estão implicados na patogénese da doença.

Em algumas formas de realização, a condição de interesse está associada com disfunção mitocondrial aumentada ou diminuída. Em algumas formas de realização, a condição de interesse é esquizofrenia. Em algumas formas de realização, a condição de interesse é doença arterial coronária. Em algumas formas de realização, a condição de interesse é diabetes mellitus, tipo II. Em algumas formas de realização, a condição de interesse é doença de Parkinson. Em algumas formas de realização, a condição de interesse é doença de Alzheimer.

Em algumas formas de realização, o factor de risco do polimorfismo conhecido é o alelo da Apolipoproteína E (e. g., ApoE 2, ApoE 3 ou ApoE 4).

Em algumas formas de realização, o locus genético de interesse está em desequilíbrio de ligação com o polimorfismo conhecido. Em algumas formas de realização, o locus genético de interesse está no mesmo cromossoma e a menos que 10, 20, 30, 40, ou 50 kilobases de distância do polimorfismo conhecido. Em algumas formas de realização, o locus genético é TOMM40.

É também proporcionado um método para determinar o risco aumentado de desenvolver uma condição de interesse, compreendendo: (a) determinar, a partir de uma amostra biológica contendo ADN, uma variante genética identificada pelo método de qualquer dos parágrafos precedentes conduzido por um único indivíduo; e, em seguida, (b) determinar o indivíduo que está com risco aumentado de desenvolvimento da condição de interesse quando a variante genética está presente.

É ainda proporcionado um método para determinar o risco de desenvolvimento da doença de Alzheimer num indivíduo (e. g., um indivíduo transportando, pelo menos, um alelo de Apo E3), compreendendo: (a) detectar, a partir de uma amostra biológica contendo ADN retirado do indivíduo, a presença ou ausência de uma variante genética do gene TOMM40 associado com o risco aumentado ou diminuído da doença de Alzheimer; e (b) determinar o indivíduo que está em risco aumentado ou diminuído da doença de Alzheimer quando a variante genética está presente ou ausente.

Em algumas formas de realização, é determinado se o indivíduo é um indivíduo Apo E2/E2, E2/E3, E2/E4, E3/E3, E3/E4, ou E4/E4. Em algumas formas de realização, é determinado se o indivíduo é um indivíduo Apo E3/E3 ou E3/E4.

É também proporcionado uma quantidade de tratamento eficaz de um agente activo Anti-doença de Alzheimer seleccionado do grupo consistindo de inibidores da acetilcolinesterase, antagonistas do receptor de NMDA, agonistas ou moduladores de receptor activado por proliferadores de peroxissoma (PPAR), proteínas de fusão, moléculas de ARN terapêuticas e suas combinações, para utilização num método de tratamento da doença

de Alzheimer caracterizado em que o método compreende o passo de determinar se o indivíduo está ou não em risco de desenvolver a doença de Alzheimer, em que o risco é determinado de acordo com os métodos para determinar o risco, como aqui divulgado.

Em algumas formas de realização, o passo de administração é realizado no indivíduo numa idade precoce quando o indivíduo é determinado para ter um risco aumentado pela presença ou ausência da variante genética em relação a um indivíduo no qual a variante genética não está presente ou ausente (e. g., para um indivíduo ApoE 4/4, começando aos 45, 46, 47, 48, 49, 50, 51, 52 ou 53 anos, e de forma contínua através de cada ano subsequente, em vez de começar aos 55 anos ou mais; para um indivíduo ApoE 4/3, aos 50, 51, 52, 53, 54, 55, 56, 57 ou 58 anos, e de forma contínua através de cada ano subsequente, em vez de começar aos 60 anos ou mais; para um indivíduo ApoE 3/3, aos 55, 56, 57, 58, 59, 60, 61, 62 ou 63 anos, e de forma contínua através de cada ano subsequente, em vez de começar aos 65 anos ou mais; e para um indivíduo ApoE 2/3, aos 60, 61, 62, 63, 64, 65, 66, 67 ou 68 anos, e de forma contínua através de cada ano subsequente, em vez de começar aos 70 anos ou mais).

Em algumas formas de realização, o agente activo é seleccionado do grupo consistindo de inibidores da acetilcolinesterase, antagonistas do receptor de NMDA, agonistas ou moduladores de PPAR (e. g., fármacos nas classes de tiazolidinediona ou glitazar), anticorpos, proteínas de fusão, moléculas de ARN terapêuticas e suas combinações. Em algumas formas de realização, o agente activo é rosiglitazona ou um seu sal farmaceuticamente aceitável.

Em algumas formas de realização, a variante genética do TOMM40 é uma variante listada na **Tabela 1** como definido abaixo.

É também divulgado um método de tratamento de um indivíduo (e. g., um indivíduo possuindo, pelo menos, um alelo ApoE 3) para a doença de Alzheimer pela administração de um agente activo anti-doença de Alzheimer ao indivíduo numa quantidade eficaz de tratamento; a melhoria compreendendo: administrar o agente activo ao indivíduo numa idade precoce quando o indivíduo transporta uma variante genética do gene TOMM40 associado com risco aumentado da doença de Alzheimer em comparação com um indivíduo correspondente que não transporta a variante genética (e. g., para um indivíduo ApoE 4/4, começando aos 45, 46, 47, 48, 49, 50, 51, 52 ou 53 anos, e de forma contínua através de cada ano subsequente, em vez de começar aos 55 anos ou mais; para um indivíduo ApoE 4/3, aos 50, 51, 52, 53, 54, 55, 56, 57 ou 58 anos, e de forma contínua através de cada ano subsequente, em vez de começar aos 60 anos ou mais; para um indivíduo ApoE 3/3, aos 55, 56, 57, 58, 59, 60, 61, 62 ou 63 anos, e de forma contínua através de cada ano subsequente, em vez de começar aos 65 anos ou mais; e para um indivíduo ApoE 2/3, aos 60, 61, 62, 63, 64, 65, 66, 67 ou 68 anos, e de forma contínua através de cada ano subsequente, em vez de começar aos 70 anos ou mais).

Em algumas formas de realização, o indivíduo é um indivíduo Apo E2/E2, E2/E3, E2/E4, E3/E3, E3/E4, E4/E4. Em algumas formas de realização, o indivíduo é um indivíduo Apo E3/E3 ou E3/E4.

Em algumas formas de realização, o agente activo é seleccionado do grupo consistindo de inibidores da acetilcolinesterase, antagonistas do receptor de NMDA, agonistas ou moduladores de PPAR (e. g., fármacos nas classes de

tiazolidinediona ou glitazar), anticorpos, proteínas de fusão, moléculas de ARN terapêuticas e suas combinações. Em algumas formas de realização, o agente activo é rosiglitazona ou um seu sal farmaceuticamente aceitável.

Em algumas formas de realização, a variante genética do gene TOMM40 é um polimorfismo de deleção/inserção (DIP). Em algumas formas de realização, o DIP é um polimorfismo de inserção. Em algumas formas de realização, o DIP é polimorfismo de deleção/inserção poli-T (e. g., entre 5 e 100, ou 10 e 80, ou 20 e 50 pb poli-T).

Em algumas formas de realização, a variante genética do TOMM40 é uma variante listada na **Tabela 1**, como definido abaixo. Em algumas formas de realização, o DIP é rs10524523, rs10602329 ou DIP3. Em algumas formas de realização, o DIP é rs10524523.

É ainda divulgado um método de tratamento para uma condição de interesse, em que a condição de interesse está associada com ApoE e/ou TOMM40, para um doente que dele necessite, o método incluindo os passos: (a) determinar a presença ou ausência da variante genética identificada pelo método do parágrafo 1-12 realizado por um único indivíduo para gerar um perfil genético do doente; e, em seguida, se o perfil é indicativo do doente ser responsável a um agente activo, (b) administrar o agente activo ao indivíduo numa quantidade de tratamento eficaz para tratar a condição de interesse. Em algumas formas de realização, o agente activo é seleccionado do grupo consistindo de inibidores da acetilcolinesterase, antagonistas do receptor de NMDA, agonistas ou moduladores de PPAR (e. g., fármacos nas classes de tiazolidinediona ou glitazar), anticorpos, proteínas de fusão, moléculas de ARN terapêuticas e suas combinações. Em algumas

formas de realização, o agente activo é rosiglitazona ou um seu sal farmaceuticamente aceitável. Em algumas formas de realização, a variante genética do gene TOMM40 é um polimorfismo de deleção/inserção (DIP). Em algumas formas de realização, o DIP é um polimorfismo de inserção. Em algumas formas de realização, o DIP é é polimorfismo de deleção/inserção poli-T (e. g., entre 5 e 100, ou 10 e 80, ou 20 e 50 pb inserção poli-T).

Em algumas formas de realização, a variante genética do TOMM40 é uma variante de TOMM40 listada na **Tabela 1** como definido abaixo. Em algumas formas de realização, o DIP é rs10524523, rs10602329 ou DIP3. Em algumas formas de realização, o DIP é rs10524523.

É também divulgado um método de tratamento para a doença de Alzheimer num indivíduo, incluindo:

(a) detectar, a partir de uma amostra biológica contendo ADN retirado de um indivíduo, a presença ou ausência de uma variante genética do gene TOMM40 associada com responsividade a um agente activo; e, se a variante genética está presente, (b) administrar o agente activo ao indivíduo numa quantidade de tratamento eficaz para tratar a doença de Alzheimer.

Em algumas formas de realização, o indivíduo transporta, pelo menos, um alelo de ApoE 3. Em algumas formas de realização, o indivíduo é um indivíduo Apo E3/E3 ou E3/E4.

Em algumas formas de realização, o agente activo é seleccionado do grupo consistindo de inibidores da

acetilcolinesterase, antagonistas do receptor de NMDA, agonistas ou moduladores de PPAR (e. g., fármacos nas classes de tiazolidinediona ou glitazar), anticorpos, proteínas de fusão, moléculas de ARN terapêuticas e suas combinações. Em algumas formas de realização, o agente activo é rosiglitazona ou um seu sal farmaceuticamente aceitável.

Em algumas formas de realização, a variante genética do gene TOMM40 é um polimorfismo de deleção/inserção (DIP). Em algumas formas de realização, o DIP é um polimorfismo de inserção. Em algumas formas de realização, o DIP é polimorfismo de deleção/inserção poli-T (e. g., entre 5 e 100, ou 10 e 80, ou 20 e 50 pb poli-T).

Em algumas formas de realização, a variante genética do TOMM40 é uma variante de TOMM40 listada na **Tabela 1** como definido abaixo. Em algumas formas de realização, o DIP é rs10524523, rs10602329 ou DIP3. Em algumas formas de realização, o DIP é rs10524523.

É ainda proporcionado a utilização de um agente activo anti-doença de Alzheimer para a preparação de um medicamento para realizar um método de tratamento da doença de Alzheimer, de acordo com os parágrafos acima estabelecidos. É também proporcionado a utilização de um agente activo anti-doença de Alzheimer para realizar um método de tratamento da doença de Alzheimer.

É proporcionado um método de determinação de um prognóstico para um doente em risco de desenvolver a doença de Alzheimer, incluindo obter um perfil do doente, em que o obter um perfil do doente inclui: detectar a presença ou ausência de, pelo menos,

um alelo de ApoE numa amostra biológica do doente, e detectar a presença ou ausência de, pelo menos, um polimorfismo de deleção/inserção (DIP) de TOMM40 localizado no intrão 6 ou intrão 9 do gene TOMM40 e, em seguida, converter o perfil do doente no prognóstico, em que a presença do alelo do ApoE e a presença de, pelo menos, um polimorfismo DIP de TOMM40 identifica o doente como um doente em risco de desenvolver a doença de Alzheimer.

Em algumas formas de realização, o DIP é um polimorfismo de inserção. Em algumas formas de realização, o DIP é polimorfismo de deleção/inserção poli-T (e. g., entre 5 e 100, ou 10 e 80, ou 20 e 50 poli-T).

Em algumas formas de realização, o DIP é rs10524523, rs10602329 ou DIP3. Em algumas formas de realização, o DIP é rs10524523.

Em algumas formas de realização, o método inclui ainda detectar se o indivíduo é um indivíduo Apo E2/E2, E2/E3, E2/E4, E3/E3, E3/E4, E4/E4. Em algumas formas de realização, o indivíduo é um indivíduo Apo E3/E3 ou E3/E4.

É ainda proporcionado um método para estratificar um indivíduo num sub-grupo de um ensaio clínico de uma terapia para o tratamento da doença de Alzheimer, o método incluindo: detectar a presença ou ausência de, pelo menos, um alelo de ApoE numa amostra biológica de um doente e detectar a presença ou ausência de, pelo menos, um polimorfismo de deleção/inserção (DIP) de TOMM40 localizado no intrão 6 ou intrão 9 do gene TOMM40, em que o indivíduo é estratificado no sub-grupo para o

ensaio clínico da terapia com base na presença ou ausência de, pelo menos, um alelo de ApoE e/ou DIP de TOMM40.

Em algumas formas de realização, o DIP é um polimorfismo de inserção. Em algumas formas de realização, o DIP é polimorfismo de deleção/inserção poli-T (e. g., entre 5 e 100, ou 10 e 80, ou 20 e 50 inserção poli-T).

Em algumas formas de realização, o DIP é rs10524523, rs10602329 ou DIP3. Em algumas formas de realização, o DIP é rs10524523.

Em algumas formas de realização, o método inclui ainda detectar se o indivíduo é um indivíduo Apo E2/E2, E2/E3, E2/E4, E3/E3, E3/E4, E4/E4. Em algumas formas de realização, o indivíduo é um indivíduo Apo E3/E3 ou E3/E4.

É ainda divulgado um método para identificar um doente num ensaio clínico de um tratamento para a doença de Alzheimer incluindo: a) identificar um doente diagnosticado com a doença de Alzheimer; e b) determinar um prognóstico para o doente diagnosticado com a doença de Alzheimer compreendendo obter um perfil do doente, em que o perfil do doente compreende i) detectar a presença ou ausência de, pelo menos, um alelo ApoE numa amostra biológica do doente, ii) detectar a presença ou ausência de, pelo menos, um polimorfismo de deleção/inserção (DIP) de TOMM40 localizado no intrão 6 ou intrão 9 do gene TOMM40, e iii) converter o perfil do doente no prognóstico, o prognóstico incluindo uma previsão se o doente é um candidato para o ensaio clínico para o tratamento da doença de Alzheimer.

Em algumas formas de realização, o DIP é um polimorfismo de inserção. Em algumas formas de realização, o DIP é polimorfismo de deleção/inserção poli-T (e. g., entre 5 e 100, ou 10 e 80, ou 20 e 50 poli-T).

Em algumas formas de realização, o DIP é rs10524523, rs10602329 ou DIP3. Em algumas formas de realização, o DIP é rs10524523.

Em algumas formas de realização, o método inclui ainda detectar se o indivíduo é um indivíduo Apo E2/E2, E2/E3, E2/E4, E3/E3, E3/E4, E4/E4. Em algumas formas de realização, o indivíduo é um indivíduo Apo E3/E3 ou E3/E4.

Também é divulgado um kit para determinar se um indivíduo está em risco aumentado de desenvolver doença de Alzheimer de início tardio, incluindo: (A) pelo menos um reagente que detecta, especificamente, ApoE 3, ApoE 4 ou ApoE 2, em que o reagente é seleccionado do grupo consistindo de anticorpos que ligam, selectivamente, ApoE 3, ApoE 4 ou ApoE 2, e sondas oligonucleotídicas que se ligam, selectivamente, ao ADN que codifica o mesmo; (B) pelo menos um reagente que detecta, especificamente, a presença ou ausência de, pelo menos, um polimorfismo de deleção/inserção (DIP) de TOMM40 localizado no intrão 6 ou intrão 9 do gene TOMM40; e (C) instruções para determinar se indivíduo está ou não em risco aumentado de desenvolver doença de Alzheimer de início tardio por: (i) detectar a presença ou ausência de uma isoforma de ApoE num indivíduo com, pelo menos, um reagente; (ii) detectar a presença ou ausência de, pelo menos, um polimorfismo de deleção/inserção (DIP) de TOMM40 localizado no intrão 6 ou intrão 9 do gene TOMM40; e (iii) observar se o indivíduo está ou não em risco

aumentado de desenvolver doença de Alzheimer de início tardio por observação se a presença da isoforma de ApoE e o DIP de TOMM40 é ou não detectado com, pelo menos, um reagente, em que a presença da isoforma ApoE e o DIP de TOMM40 indica se o indivíduo está em risco aumentado de desenvolver doença de Alzheimer de início tardio.

Em algumas formas de realização, o pelo menos um reagente e as instruções estão embaladas num único recipiente.

Em algumas formas de realização, o DIP é um polimorfismo de inserção. Em algumas formas de realização, o DIP é um polimorfismo de deleção/inserção poli-T (e. g., entre 5 e 100, ou 10 e 80, ou 20 e 50 poli-T).

Em algumas formas de realização, o DIP é rs10524523, rs10602329 ou DIP3. Em algumas formas de realização, o DIP é rs10524523.

Em algumas formas de realização, o método inclui ainda detectar se o indivíduo é um indivíduo Apo E2/E2, E2/E3, E2/E4, E3/E3, E3/E4, E4/E4. Em algumas formas de realização, o indivíduo é um indivíduo Apo E3/E3 ou E3/E4.

É divulgado um kit para determinar se um indivíduo é responsável ao tratamento para a condição de interesse, em que a condição de interesse está associada com ApoE e/ou TOMM40, com um agente activo, o kit incluindo: (A) pelo menos um reagente que detecta, especificamente, ApoE 3, ApoE 4 ou ApoE 2, em que o reagente é seleccionado do grupo consistindo de anticorpos que ligam, selectivamente, ApoE 3, ApoE 4 ou ApoE 2, e sondas oligonucleotídicas que se ligam, selectivamente, ao ADN que

codifica o mesmo; (B) pelo menos um reagente que detecta, especificamente, a presença ou ausência de, pelo menos, um polimorfismo de deleção/inserção (DIP) de TOMM40 localizado no intrão 6 ou intrão 9 do gene TOMM40; e (C) instruções para determinar se o indivíduo é responsivo ao tratamento para a condição de interesse com o agente activo de interesse por: (i) detectar a presença ou ausência de uma isoforma de ApoE num indivíduo com, pelo menos, um reagente; (ii) detectar a presença ou ausência de, pelo menos, um polimorfismo de deleção/inserção (DIP) de TOMM40 localizado no intrão 6 ou intrão 9 do gene TOMM40; e (iii) determinar se o indivíduo é ou não responsivo ao tratamento por observação se a presença da isoforma de ApoE e do DIP de TOMM40 é ou não detectada com, pelo menos, um reagente, em que a presença de ApoE 3 e o DIP de TOMM40 indica que o indivíduo é responsivo ao tratamento com o agente activo.

Em algumas formas de realização, o pelo menos um reagente e as instruções estão embaladas num único recipiente.

Em algumas formas de realização, o DIP é um polimorfismo de inserção. Em algumas formas de realização, o DIP é um polimorfismo de deleção/inserção poli-T (e. g., entre 5 e 100, ou 10 e 80, ou 20 e 50 pb poli-T).

Em algumas formas de realização, o DIP é rsl0524523, rsl0602329 ou DIP3. Em algumas formas de realização, o DIP é rsl0524523.

Em algumas formas de realização, o método inclui ainda detectar se o indivíduo é um indivíduo Apo E2/E2, E2/E3, E2/E4, E3/E3, E3/E4, E4/E4. Em algumas formas de realização, o indivíduo é um indivíduo Apo E3/E3 ou E3/E4.

Será entendido que todas as formas de realização anteriores podem ser combinadas de qualquer forma e/ou combinação. Os anteriores e outros objectivos e aspectos da presente invenção são explicados com maior detalhe nos desenhos juntamente proporcionados e na descrição estabelecida a seguir.

Breve Descrição dos Desenhos

A **Figura 1** mostra um fluxograma geral para identificar a variante genética numa região predeterminada de sequência genómica num locus genético de interesse, que pode ser associado com uma condição de interesse, de acordo com algumas formas de realização.

A **Figura 2** mostra um gráfico da média de idades da doença de Alzheimer de início tardio, como uma função da herança dos cinco genótipos comuns de ApoE, e representando ApoE 4 como um factor de risco para a doença de Alzheimer (1993).

A **Figura 3** mostra Regiões A, B, e C no Cromossoma 19, que são loci genéticos de interesse exemplares. O gene TOMM40 está em estreita proximidade ao gene ApoE e codifica uma proteína de 40 kD dirigida à membrana mitocondrial externa. O TOMM40 interage com ApoE directamente na regulação do importação da proteína mitocondrial e a presente hipótese é que a presença de uma(s) variante(s) particular(es) de TOMM40 exacerbam o risco aumentado para doença de Alzheimer associado com a presença dependente da dose do alelo de ApoE 3.

A **Figura 4** mostra, de forma esquemática, um mapa evolutivo da Região B como indicado na Figura 3, criado de acordo com algumas formas de realização aqui em detalhe. Cada oval horizontal representa uma variante de sequência observada (e. g., uma alteração nucleotídica) e o tamanho da oval representa a sua frequência. Cada variante é chamada um nó na árvore e a ligação de um nó a outro nó por uma ramificação representa uma observação de duas variantes em *cis* transportadas por um único indivíduo.

A **Figura 5** é um diagrama esquemático da árvore filogenética com base na Região B construída para TOMM40, mostrando as percentagens de fenótipos de ApoE em dois grandes grupos ou clados, das variantes de TOMM40 nesta região.

A **Figura 6** é uma visão esquemática do locus de *TOMM40-ApoE* incluindo uma representação gráfica LD mostrando blocos de haplótipo e regiões submetidas a sequenciamento primário nos estudos exploratórios (R1) (23 Kb) e confirmatórios (R2) (10 Kb) (NCBI Build 36.3). A representação gráfica é mostrada para dados Hapmap (painel de análise CEU), definição de bloco de haplótipo de espinha sólida, valores r^2 com esquema de cores D'/LOD representado por características de linha diferentes.

A **Figura 7** mostra representações das árvores filogenéticas com separação de variantes. **7A:** variantes de SNP, clado A vs. B, E6 - E10 representam exões de *TOMM40* e linhas verticais indicam as localizações aproximadas dos SNP. Separação das duas ramificações principais tem forte suporte “bootstrap” (973/1000). **7B:** polimorfismos de comprimento rs10524523. As estatísticas descritivas são proporcionadas para cada grupo de polimorfismos de comprimento. Diversos haplótipos longos que

formaram grupos externos únicos, na árvore ou clados muito pequenos, estão no grupo identificado como "Restante".

A **Figura 8** apresenta histogramas do comprimento do polimorfismo de comprimento rs10524523 estratificado por genótipos de ApoE 3/3 (**8A**), 3/4 (**8B**), e 4/4 (**8C**). N=210 haplótipos (coorte AS).

A **Figura 9** mostra a associação entre idade de início de AD e comprimento do polimorfismo rs10524523 para doentes com início de AD entre os 60 e 86 anos. Os diagramas de caixa indicam a gama de 95% (linhas verticais), mediana (linha horizontal na caixa) e gama interquartil (caixa).

Descrição Detalhada

A presente invenção é explicada com maior detalhe em baixo. Esta descrição não pretende ser uma lista pormenorizada de todas as diferentes formas nas quais a invenção pode ser implementada, ou todas as características que podem ser adicionadas à presente invenção. Por exemplo, características ilustradas em relação a uma forma de realização podem ser incorporadas em outras formas de realização e características ilustradas em relação a uma forma de realização particular podem ser eliminadas a partir da referida forma de realização. Além disso, numerosas variações e adições às várias formas de realização aqui sugeridas serão aparentes aos especialistas na técnica à luz da presente divulgação que não se afastem da presente invenção. Assim, a seguinte descrição destina-se a ilustrar algumas formas de realização particulares da invenção e não para especificar

exaustivamente todas as suas permutações, combinações e variações.

Como utilizado na descrição da invenção e nas reivindicações, as formas singulares "um", "uma" e "o/a" pretendem também incluir as formas de plural, a menos que o contexto indique claramente o contrário. Também, como aqui utilizado, "e/ou" refere-se a e abrange qualquer e todas as possíveis combinações de um ou mais dos artigos listados, bem como a falta de combinações quando interpretadas no alternativo ("ou").

A presente invenção é dirigida a métodos para revelar a variação genética em regiões de particular interesse para doenças e distúrbios complexos. Esta também se refere à descoberta de marcadores genéticos mais informativos com base em associações com informação fenotípica. Numa forma de realização, a invenção pode ser utilizada para localizar marcadores genéticos associados com susceptibilidade a uma doença, distúrbio ou condição particular. Noutra forma de realização, dados relativos à resposta de um indivíduo a um tratamento ou fármaco candidato podem ser incluídos numa análise filogenética para a localização de marcadores genéticos associados com uma resposta benéfica ao tratamento ou fármaco referido (*i. e., farmacogenética*). Os métodos podem ser aplicados em qualquer conjunto de dados de variação genética a partir de um *locus* particular. Ver **Figura 1** para um fluxograma da abordagem para encontrar factores de risco genéticos de acordo com a presente invenção.

Num aspecto, a análise da variação genética é baseada em dados de sequências variantes. Num segundo aspecto, a estrutura

é exposta utilizando dados genotípicos diplóides, evitando, desse modo, a necessidade de inferir experimental ou computacionalmente os haplótipos componentes (ver, e. g., Patente U.S. Nº 6027896 de Roses *et al.*). Noutro aspecto, o presente método pode ser aplicado em variação aléllica não caracterizada que resulta da interrogação de um ácido nucleico alvo com um procedimento experimental que proporciona um registo da variação da sequência presente mas não proporcionam, efectivamente, a sequência inteira. A estrutura subjacente da variação genética é também útil para a dedução dos haplótipos constituintes a partir de dados genotípicos diplóides.

É preferido e contemplado que os métodos aqui descritos sejam utilizados em conjunto com outras informações de diagnóstico clínico conhecidas ou descritas na técnica que são utilizadas na avaliação de indivíduos com doenças ou distúrbios (e. g., aqueles que se acredita estarem envolvidos na disfunção mitocondrial (e. g., doença de Alzheimer ou outras doenças neurodegenerativas)) ou para avaliação de indivíduos que se suspeita estarem em risco de desenvolver tal doença. A invenção é também aplicável para a verificação de factores de risco genéticos para outras doenças, distúrbios ou condições complexas.

1. Definições. As definições seguintes são aqui utilizadas:

“Condição de interesse” refere-se a uma condição, doença, ou distúrbio específico designado para estudo filogenético e/ou subsequente diagnóstico ou prognóstico. “Condição”, como aqui utilizado, inclui, mas não está limitada a, condições associadas com ApoE e/ou TOMM40 e/ou disfunção mitocondrial, e. g., doenças

neurodegenerativas, doenças metabólicas, distúrbios psiquiátricos e cancro.

Exemplos de condições nas quais ApoE e/ou TOMM40 têm sido implicados incluem, mas não estão limitadas a, doença cardiovascular; doença metabólica; doença neurodegenerativa; trauma neurológico ou doença; doença auto-imune (e. g., esclerose múltipla (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. & Hillert, J. The telltale scan: APOE epsilon4 in multiple sclerosis. *Lancet Neurol*. **3**: 331 (2004), lupus eritematoso sistémico neuropsiquiátrico (Pullmann Jr. R, et al., Apolipoprotein E polymorphism in patients with neuropsychiatric SLE. *Clin Rheumatol*. **23**: 97-101 (2004)), etc.); infecção viral (e. g., doença hepática associada com infecção por hepatite C (Wozniak MA, et al., Apolipoprotein E-E4 protects against severe liver disease caused by hepatitis C virus. *Hepatol*. **36**: 456-463 (2004)), doença HIV (Burt TD, et al., Apolipoprotein (apo) E4 enhances HIV-I cell entry *in vitro*, and the APOE epsilon4/epsilon4 genotype accelerates HIV disease progression. *Proc Natl Acad Sci USA*. **105**:8718-23 (2008)), etc.); fratura da anca/osteoporose (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)); doenças mitocondriais (Chang S, et al., Lipid- and receptor-binding regions of apolipoprotein E4 fragments act in concert to cause mitochondrial dysfunction e neurotoxicity. *Proc Natl Acad Sci USA*. **102**:18694-9 (2005)); envelhecimento (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 indivíduos in the Belfast Elderly Longitudinal Free-living Ageing Study (BELFAST). *Mech. Aging and Develop.* **122**: 1367-1372 (2001)); inflamação (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)); e disfunção da memória (Caselli RJ, et al., Longitudinal modeling of age-related memory decline and the APOE epsilon4 effect. *N Engl J Med.* **361** :255-63 (2009)).

"Doença cardiovascular", como aqui utilizado, refere-se a uma doença envolvendo o coração e/ou vasos sanguíneos, incluindo, mas não limitado a, doença arterial coronária (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)), aterosclerose (Norata GD, et al., Effects of PCSK9 variants on common carotid artery intima media thickness and relation to ApoE alleles. *Atherosclerosis* (2009) 27 de Junho. [publicado eletronicamente antes da impressão], doi:10.1016/j.atherosclerosis.2009.06.023; Paternoster L, et al., Association Between Apolipoprotein E Genotypes and Carotid Intima-Media Thickness May Suggest a Specific Effect on Large Artery Atherothrombotic Stroke. *Stroke* **39**:48-54 (2008)), doença cardiovascular 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)), doença vascular, tal como acidente vascular cerebral 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 Genotypes and Carotid Intima-Media Thickness May Suggest a Specific Effect on Large Artery Atherothrombotic Stroke. *Stroke* **39**:48-54 (2008)), demência 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.

"Doença neurodegenerativa", como aqui utilizado, refere-se a doença 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)), doença de Parkinson (Huang X, et al., Apolipoprotein E and dementia in Parkinson's 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's disease. *Neurology* **62**:2198-202 (2004); Martinez, M. et al., Apolipoprotein E4 is probably responsible for the cromossoma 19 linkage peak for Parkinson's disease. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **136B**, 172-174 (2005)), doença de Huntington e uma pluralidade de doenças e distúrbios menos comuns que causam a diminuição de neurônios, e. g., degeneração macular relacionada com a idade (Thakkinstian 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)).

"Doença ou trauma neurológico" inclui, mas não está limitada a, resultado após lesão na cabeça (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)), enxaqueca (Gupta R, et al., Polymorphism in apolipoprotein E among migraineurs and tension-type headache subjects. *J Headache Pain.* **10**:115-20 (2009)), 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 and on intracerebral hemorrhage. *Stroke* **40**:632-9 (2009)), etc.

"Doença metabólica", como aqui utilizado inclui, mas não está limitada a, dislipidemia (Wilier 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)), doença renal 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 Junho. [publicado electronicamente antes da impressão] DOI 10.1007/s10528-009-9266-y.), doença renal crónica (Yoshida T, et al., Association of a polymorphism of the apolipoprotein E gene with chronic

kidney disease in Japanese individuos 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)), doença da 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 Biomarcadores Prev.* **17**:525-34 (2008)), diabetes mellitus (tipo II) (Elosua R, et al., Obesity Modulates the Association among APOE Genotypes, 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, colelitíase (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.

"Distúrbio Psiquiátrico", como aqui utilizado, refere-se a esquizofrenia (Kampan 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)), distúrbio obsessivo-compulsivo (OCD), comportamento aditivo (dependência de tabaco, dependência de álcool, etc.), distúrbio bipolar (Dean B. et al., Plasma apolipoprotein E is decreased in schizophrenia spectrum and bipolar disorder. *Psychiatry Res.* **158**:75-78 (2008))

e outras doenças, distúrbios ou condições de uma natureza psiquiátrica.

“Desenvolvimento de uma condição”, como aqui utilizado, refere-se a um diagnóstico inicial de uma doença, distúrbio ou outra condição médica ou exacerbação de uma doença, distúrbio ou condição médica existente, para a qual o indivíduo já tenha sido diagnosticado.

“Diagnóstico” ou “prognóstico”, como aqui utilizado, refere-se à utilização de informação (e. g., informação ou dados genéticos de outros testes moleculares em amostras biológicas, sinais e sintomas, resultados de exames físicos, resultados de desempenho cognitivo, etc.) para antecipar os mais prováveis resultados, prazos e/ou respostas a um tratamento particular para uma dada doença, distúrbio ou condição, baseada em comparações com uma pluralidade de indivíduos partilhando sequências nucleotídicas comuns, sintomas, sinais, histórico familiar ou outros dados relevantes para consideração de um estado de saúde do doente.

“Amostra biológica”, como aqui utilizado, refere-se a um material suspeito de conter um ácido nucleico de interesse. Amostras biológicas contendo ADN incluem cabelo, pele, zaragatoa na bochecha e fluidos biológicos, tais como sangue, soro, plasma, saliva, fluido linfático, sémen, muco vaginal, fezes, urina, fluido espinal e semelhantes. O isolamento de ADN a partir de tais amostras é bem conhecido pelos especialistas na técnica.

Um “indivíduo”, de acordo com algumas formas de realização, é um indivíduo cujo(s) genótipo(s) ou haplótipo(s) deve(m) ser

determinado(s) e registado(s) em conjunto com a condição do indivíduo (*i. e.*, estado de doença ou distúrbio) e/ou resposta a um tratamento ou fármaco candidato. Sequências nucleotídicas de uma pluralidade de indivíduos são utilizadas para construir uma árvore filogenética e em seguida as sequências nucleotídicas análogas a partir de um único indivíduo podem ser comparadas com aquelas na árvore filogenética para fins de diagnóstico ou prognóstico.

“Gene”, aqui como utilizado, significa um segmento de ADN que contém toda a informação para a biossíntese regulada de um produto de ARN, incluindo promotores, exões, intrões e outras regiões não traduzidas que controlam a expressão.

“*Locus genético*” ou “*locus*”, como aqui utilizado, significa uma localização num cromossoma ou molécula de ADN, muitas vezes correspondente a um gene ou uma característica física ou fenotípica ou a um nucleótido ou extensão de nucleótidos particular. *Loci* é a forma plural de *locus*.

“Amplificação”, como aqui aplicado a ácidos nucleicos, refere-se a qualquer método que resulta na formação de uma ou mais cópias de um ácido nucleico, em que, de um modo preferido, a amplificação é exponencial. Um tal método para amplificação enzimática de sequências específicas de ADN é conhecida como a reacção em cadeia de polimerase (PCR), como descrito por Saiki *et al.*, 1986, *Science* 230:1350-1354. Os iniciadores utilizados em PCR variam, normalmente, em comprimento de cerca de 10 a 50 ou mais nucleótidos e são, tipicamente, seleccionados para ser, pelo menos, cerca de 15 nucleótidos para garantir especificidade suficiente. O fragmento de cadeia dupla que é

produzido é chamado um "amplicão" e pode variar em comprimento de tão pouco como cerca de 30 nucleótidos a 20000 ou mais.

Um "marcador" ou "marcador genético", como aqui utilizado, é uma variação conhecida de uma sequência de ADN num *locus* particular. A variação pode estar presente num indivíduo devido a mutação ou herança. Um marcador genético pode ser uma sequência curta de ADN, tal como uma sequência em torno de uma única alteração de pares de bases (polimorfismo de nucleótido único, SNP) ou uma longa, como minissatélites. Os marcadores podem ser utilizados para estudar a relação entre uma doença hereditária e a sua causa genética (por exemplo, uma mutação particular de um gene que resulta numa forma defeituosa ou de outro modo indesejável de proteína).

Um "factor de risco genético", como aqui utilizado, significa um marcador genético que está associado com susceptibilidade aumentada a uma condição, doença ou distúrbio. Este também pode referir-se a um marcador genético que está associado com uma resposta particular a um tratamento ou fármaco seleccionado de interesse.

"Associado com", como aqui utilizado, significa a ocorrência conjunta de duas ou mais características mais frequentemente do que seria esperado por acaso. Um exemplo de associação envolve uma característica na superfície de glóbulos brancos chamada HLA (HLA significa antigénio leucócito humano). Um tipo particular de HLA, HLA tipo B-27, está associado com um risco aumentado para um número de doenças incluindo espondilite anquilosante. A espondilite anquilosante é 87 vezes mais provável de ocorrer em pessoas com HLA B-27 do que na população em geral.

Um indivíduo “em risco aumentado de desenvolver uma condição” devido a um factor de risco genético é um que está predisposto à condição, tem susceptibilidade genética para a condição e/ou é mais provável que desenvolva a condição que indivíduos nos quais o factor de risco genético está ausente. Por exemplo, um indivíduo que está “em risco aumentado de desenvolver doença de Alzheimer” devido à presença de um ou dois alelos de ApoE 4 é mais provável que desenvolva doença de Alzheimer que um indivíduo que não transporta um alelo de ApoE 4.

“Polimorfismo”, como aqui utilizado, refere-se à existência de duas ou mais sequências nucleotídicas diferentes num *locus* particular no ADN no genoma. Os polimorfismos podem servir como marcadores genéticos e também podem ser referidos como variantes genéticas. Polimorfismos incluem substituições, inserções, deleções nucleotídicas e microssatelites e podem, mas não necessitam de, resultar em diferenças detectáveis em expressão génica ou função da proteína. Um sítio polimórfico está numa posição nucleotídica dentro de um *locus*, no qual a sequência nucleotídica varia de uma sequência de referência em, pelo menos, um indivíduo numa população.

Um “polimorfismo de deleção/inserção” ou “DIP”, como aqui utilizado, é uma inserção de um ou mais nucleótidos numa versão de uma sequência relativa a outra. Se for conhecido qual dos alelos representa alelos menores, o termo “deleção” é utilizado quando o alelo menor é uma deleção de um nucleótido e o termo “inserção” é utilizado quando o alelo menor é uma adição de um nucleótido. A expressão “polimorfismo de deleção/inserção” é também utilizada quando existem múltiplas formas ou comprimentos

e o alelo menor não é aparente. Por exemplo, para os polimorfismos poli-T aqui descritos, são observados múltiplos comprimentos de polimorfismos.

“Dados de polimorfismo”, como aqui utilizado, significa informação relativa a um ou mais dos seguintes para um gene específico: localização de sítios polimórficos; variação da sequência nestes sítios; frequência de polimorfismos numa ou mais populações; os diferentes genótipos e/ou haplótipos determinados para o gene; frequência de um ou mais destes genótipos e/ou haplótipos numa ou mais populações; e qualquer associação(ões) conhecida(s) entre uma característica e um genótipo ou a haplótipo para o gene.

“Haplótipo”, como aqui utilizado, refere-se a uma variante genética ou combinação de variantes transportadas em, pelo menos, um cromossoma num indivíduo. Um haplótipo inclui, muitas vezes, múltiplos loci polimórficos contíguos. Todas as partes de um haplótipo como aqui utilizado ocorrem na mesma cópia de um cromossoma ou molécula de ADN haplóide. A ausência evidente do contrário presume que um haplótipo representa uma combinação de múltiplos loci que são provavelmente transmitidos em conjunto durante a meiose. Cada humano transporta um par de haplótipos para qualquer dado locus genético, consistindo de sequências herdadas nos cromossomas homólogos de dois pais. Estes haplótipos podem ser idênticos ou podem representar duas variantes genéticas diferentes para o dado locus. Haplótipagem é um processo para determinar um ou mais haplótipos num indivíduo. Haplótipagem pode incluir utilização de genealogia de famílias, técnicas moleculares e/ou inferência estatística.

Uma "variante", "variância" ou "variante genética", como aqui utilizado, refere-se a uma isoforma específica de um haplótipo numa população, a forma específica diferindo de outras formas do mesmo haplótipo na sequência de, pelo menos, um e, frequentemente, mais que um, sítios variantes ou nucleótidos dentro da sequência do gene. As sequências destes sítios variantes que diferem entre diferentes alelos de um gene são designadas como "variantes da sequência génica", "alelos", "variâncias" ou "variantes". A expressão "forma alternativa" refere-se a um alelo que pode ser distinguido de outros alelos por possuir, pelo menos, um e, frequentemente, mais do que um sítios variantes dentro da sequência génica. Outros termos conhecidos na técnica para serem equivalente a "variâncias" ou "variantes" incluem mutações e polimorfismos de nucleótido único (SNP). A referência à presença de uma variância ou variâncias significa variâncias particulares, i. e., nucleótidos particulares em sítios polimórficos particulares, em vez de apenas a presença de qualquer variância no gene.

"Isoforma", como aqui utilizado, significa uma forma particular de um gene, ARNm, ADNc ou a proteína codificada desse modo, que se distingue de outras formas pela sua sequência e/ou estrutura particular. Por exemplo, a isoforma ApoE 4 da apolipoproteína E em oposto às isoformas ApoE 2 ou ApoE 3.

"Cistrão", como aqui utilizado, significa uma secção de ADN encontrada num único cromossoma que contém o código genético para um único polipeptídeo e funções como uma unidade hereditária. Um cistrão inclui exões, intrões e elementos regulatórios relacionados com uma unidade funcional única (i. e., um gene). O termo deriva do clássico teste *cis-trans* para determinar se os elementos genéticos foram capazes de

interagir funcionalmente independentemente de estarem localizados na mesma molécula de ADN (complementação "trans") ou apenas quando estavam localizados na mesma molécula de ADN (elementos de actuação "cis").

O termo "genótipo", no contexto desta invenção, refere-se à forma alélica particular de um gene, que pode ser definida pelo(s) nucleótido(s) particular(es) presente(s) numa sequência de ácido nucleico num sítio(s)particular(es). Genótipo também pode indicar o par de alelos presentes num ou mais *loci* polimórficos. Para organismos diplóides, tal como humanos, dois haplótipos formam um genótipo. Genotipagem é qualquer processo para determinar um genótipo de um indivíduo, e. g., por amplificação de ácidos nucleicos, ligação de anticorpo ou outra análise química. O genótipo resultante pode ser desfasado, o que significa que as sequências encontradas não são conhecidas para serem derivadas de um cromossoma parental ou o outro.

"Desequilíbrio de ligação", como aqui utilizado, significa a associação não aleatória de alelos em dois ou mais *loci*. Desequilíbrio de ligação descreve uma situação na qual algumas combinações de alelos ou marcadores genéticos ocorrem mais ou menos frequentemente numa população do que aquilo que seria de esperar numa formação aleatória de haplótipos a partir de alelos com base nas suas frequências. Associações não aleatórias entre polimorfismos em diferentes *loci* são medidas pelo grau de desequilíbrio de ligação.

"Alinhamento de sequências múltiplas" ou "MSA", como aqui utilizado, significa alinhamento de três ou mais sequências nucleotídicas a partir de ADN genómico derivado de uma pluralidade de indivíduos para determinar homologia e

heterologia entre as sequências. Em geral, assume-se que o conjunto de entrada de sequências de consulta tem uma relação evolutiva, pela qual estas partilham uma linhagem e são descendentes de um ancestral comum. Algoritmos computacionais são mais frequentemente utilizados para realizar a análise de sequências alinhadas.

Algumas formas de realização da presente invenção são descritas fazendo referência a diagramas de blocos ilustrando métodos (e. g., **Figura 1**), que podem incluir passos implementados por um computador e/ou produtos de programa de computador. Será entendido que cada bloco nos diagramas de blocos e/ou ilustrações operacionais e combinações de blocos nos diagramas de blocos e/ou ilustrações operacionais, podem ser implementadas por hardware análogo e/ou digital e/ou instruções de programa de computador. Estas instruções de programa de computador podem ser proporcionadas a um processador de um computador de utilização geral, computador de utilização específico, ASIC e/ou outro aparelho de processamento de dados programável, de tal modo que as instruções, que executam por meio do processador do computador e/ou outro aparelho de processamento de dados programável, criam meios para implementar as funções/acções especificadas nos diagramas de blocos e/ou ilustrações operacionais. Por conseguinte, será entendido que os diagramas de blocos e ilustrações operacionais suportam aparelhos, métodos e produtos de programa de computador.

Outro software, tal como um sistema operativo, também pode ser incluído. Será ainda apreciado que a funcionalidade do módulo de alinhamento da sequência múltipla, módulo de mapeamento e/ou outros módulos aqui descritos podem ser realizados, pelo menos em parte, utilizando componentes de

hardware discretos, um ou mais Circuitos Integrados de Aplicação Específica (ASIC) e/ou um ou mais processadores e/ou computadores de utilização digital específica.

"Mapeamento", como aqui utilizado, significa criar uma árvore filogenética atribuindo um nó a cada nova sequência nucleotídica variante observada, ligado esse nó a outro nó representando uma sequência conhecida transportada por um indivíduo no mesmo cromossoma ou cistrão e contando os números de cada tipo de indivíduo representado em cada nó. Ver **Figura 4** para um exemplo de uma árvore filogenética desenvolvida deste modo.

"Filogenética" significa relacionado com o estudo de ligações evolutivas entre vários grupos e organismos ou indivíduos dentro de uma espécie. Antes da informação genética estar prontamente disponível, a filogenia era baseada, principalmente, em observação fenotípica. "Mapeamento filogenético", como aqui utilizado, significa a utilização de sequência de dados de ADN para ligar sequências variantes relacionadas transportadas por uma pluralidade de indivíduos de modo a determinar ligações evolutivas e a cronologia de divergência. Uma "árvore filogenética" é o resultado de mapeamento das ligações entre variantes.

"Nó", como aqui utilizado, significa um ponto de dados de polimorfismo numa árvore filogenética representado uma sequência variante actual transportada por pelo menos um indivíduo. Um nó é ligado por uma ramificação a outro nó representando uma sequência variante transportada pelo mesmo indivíduo no mesmo cromossoma e no mesmo cistrão mas num diferente *locus* genético dentro do cistrão. A presença de um nó indica que pelo menos um

indivíduo transportando a sequência indicada pelo nó e a sequência representada pelo nó vizinho ao qual está ligado por uma ramificação.

“Ramificação”, como aqui utilizado, significa uma ligação entre dois nós representando duas sequências variantes distintas ou haplótipos, em que as duas variantes estão localizadas no mesmo cromossoma e no mesmo cistrão de um único indivíduo. “Ponto de ramificação” significa qualquer nó a partir do qual mais que duas ramificações se estendem, mas é aqui especialmente utilizado para referir um nó raiz a partir do qual três ou mais nós se estendem. Um “nó raiz” representa a sequência genética de um ancestral evolutivo comum a partir do qual a divergência genética gerou a variedade de sequência variantes próximas, representadas pelos nós conectados.

“Iterativamente” como aqui utilizado refere-se ao cálculo repetitivo de valores para cada carácter numa série. Por exemplo, cada nó numa árvore filogenética é analisado para calcular a proporção do número de indivíduos afectados com uma condição de interesse (tal como doença de Alzheimer) para controlar indivíduos não afectados; esta proporção é comparada com os nós conectados para localizar correlações com o risco aumentado ou diminuído de desenvolver uma doença, distúrbio ou condição de interesse. Uma alteração substancial nesta proporção entre um nó e o próximo indica a presença de uma variante que aumenta ou diminui o risco de aparecimento precoce de doença. “Examinar iterativamente as variantes genéticas” significa começar a análise com nós representando as sequências partilhadas pelo maior número de indivíduos únicos e sucessivamente analisar os nós ligados por ramificações estendendo-se a partir desse nó, seguido pelo segundo nível de

nós e assim por diante. A análise move-se, então, a partir das raízes da árvore para as ramificações exteriores e nós da árvore.

“Tratamento”, como aqui utilizado, inclui qualquer fármaco, procedimento, alteração de estilo de vida ou outro ajustamento introduzido na tentativa de produzir uma alteração num aspecto particular da saúde de um indivíduo (*i. e.*, dirigido a uma doença, distúrbio ou condição particular).

“Fármaco”, como aqui utilizado, refere-se a uma entidade química ou produto biológico ou combinação de entidades químicas ou produtos biológicos, administrados a uma pessoa para tratar ou prevenir ou controlar uma doença ou condição. O termo “fármaco”, como aqui utilizado, é sinónimo dos termos “remédio”, “medicamento”, “intervenção terapêutica” ou “produto farmacêutico”. De um modo muito preferido, o fármaco é aprovado por uma agência governamental para tratamento de, pelo menos, uma doença ou condição específica.

“Doença”, “distúrbio” e “condição” são geralmente reconhecidos na técnica e designam a presença de sinais e/ou sintomas num indivíduo ou doente que são, geralmente, reconhecidos como anormais e/ou indesejáveis. Doenças ou condições podem ser diagnosticadas e categorizadas com base em alterações patológicas. A doença ou condição pode ser seleccionada de tipos de doenças listadas em textos padrão, tal como Harrison's Principles of Intern Medicine, 1997 ou Robbins Pathologic Basis of Disease, 1998.

“Disfunção mitocondrial”, como aqui utilizado, significa quaisquer anormalidades prejudiciais da mitocôndria dentro da

célula ou células. Algumas doenças, distúrbios ou condições presentemente conhecidas na técnica para serem associadas com disfunção mitocondrial incluem doença de Alzheimer, doença de Parkinson e outras doenças neurodegenerativas, lesão de isquemia-reperfusão em acidente vascular cerebral e ataque cardíaco, epilepsia, diabetes e envelhecimento. Muitas outras doenças, distúrbios e condições têm sido associadas com disfunção mitocondrial na técnica. Com efeito, a mitocôndria é crítica para o funcionamento adequado da maioria dos tipos de células e o declínio mitocondrial leva, muitas vezes, à morte celular. Esta disfunção mitocondrial provoca danos celulares e morte por comprometer a produção de ATP, interrompendo a homeostase do cálcio e aumentando o stress oxidativo. Além disso, os danos mitocondriais podem levar a morte celular por apoptose por causar a libertação de citocromo c e outros factores pró-apoptóticos para o citoplasma (para revisão, ver Wallace, 1999; Schapira, 2006). Em relação a um exemplo específico aqui encontrado, é hipotetizado que as isoformas ApoE 3 e ApoE 4 causam disfunção mitocondrial através de interacções com TOMM40. Algumas variantes de TOMM40 podem actuar sinergicamente com isoforma ApoE 3 para acelerar o declínio mitocondrial. Pensa-se que este mecanismo mitocondrial contribui para muitas doenças, distúrbios e condições genéticas complexas.

“Indivíduos”, como aqui utilizado, são, de um modo preferido, mas não limitado a, indivíduos humanos. Os indivíduos podem ser masculinos ou femininos e podem ser de qualquer raça ou etnia, incluindo, mas não limitado a, Caucasianos, Afro-Americanos, Africanos, Asiáticos, Hispânicos, Indianos, etc. Os indivíduos podem ser de qualquer idade, incluindo recém-nascidos, neonatos, bebés, crianças, adolescentes, adultos e geriátricos. Indivíduos também podem incluir indivíduos

animais, particularmente indivíduos mamíferos, tais como caninos, felinos, bovinos, caprinos, equídeos, ovinos, porcinos, roedores (e. g., ratos e murganhos), lagomorfos, primatas (incluindo primatas não humanos), etc., seleccionados para medicina veterinária ou fins de desenvolvimento de fármacos farmacêuticos.

“Tratar”, “tratando” ou “tratamento”, como aqui utilizado, refere-se a qualquer tipo de medida que confere um benefício a um doente que sofre com uma doença, incluindo melhoria na condição do doente (e. g., em um ou mais sintomas), atraso no início ou progressão da doença, etc.

“Doença de Alzheimer de início tardio” ou “LOAD”, como aqui utilizado, é conhecida na técnica e é a classificação utilizada se a doença de Alzheimer tem um início ou é diagnosticada depois da idade de 65. É a forma mais comum da doença de Alzheimer.

2. Métodos para Identificar Variantes Genéticas

Enquanto listas de associações derivadas de varrimentos de genoma completo são úteis, estas são, geralmente, inadequadas para explicar a complexidade da doença. Famílias, vias e interacções de genes podem proporcionar especificidades. O mapeamento de variância de elevada-resolução pode revelar respostas para as interacções genéticas complexas. Isto é particularmente aplicável onde um conhecido factor de risco genético, que não explica ele próprio, inteiramente, uma associação com a doença, distúrbio ou condição de interesse pode apresentar um excelente locus genético candidato para investigações mais detalhadas. Além disso, farmacogenética,

enquanto útil para desenvolvimento de fármacos, pode também estender a relevância biológica. A análise de dados de sequência de um grande número de indivíduos para descobrir variâncias na sequência génica entre indivíduos numa população irá resultar em detecção de uma maior fracção de todas as variantes na população.

A informação da sequência inicial a ser analisada pelo método da presente invenção é derivada de ADN genómico de uma pluralidade de indivíduos. O organismo pode ser qualquer organismo para o qual estão disponíveis múltiplas sequências, mas é, de um modo preferido, de humano. Na identificação de novas variâncias, é muitas vezes útil seleccionar diferentes grupos populacionais com base na raça, etnia, género e/ou origem geográfica, pois as variâncias particulares podem diferir em frequência entre esses grupos. De um modo muito preferido, para doenças ou distúrbios que se pensa serem multigénicos (doenças/distúrbios geneticamente complexos), os fenótipos representados pela população submetida são dos extremos de um espectro. Amostras biológicas contendo ADN podem ser sangue, sémen, zaragatoa na bochecha, etc. O isolamento de ADN de tais amostras é bem conhecido na técnica.

Em algumas formas de realização, a invenção refere-se à análise de dados da sequência nucleotídica de uma pluralidade de indivíduos possuindo, pelo menos, um factor de risco conhecido para uma dada doença, distúrbio ou condição (genética ou outra). As sequências nucleotídicas são analisadas para gerar dados haplotípicos e os haplótipos ou variantes genéticas são então mapeadas para uma árvore filogenética para demonstrar a evolução das sequências representadas. Por comparação desta árvore com dados de fenótipo sobre a pluralidade de indivíduos, é possível

um prognóstico ou diagnóstico para um único indivíduo transportando haplótipos observados na árvore filogenética.

Em outras formas de realização, a invenção refere-se aos campos da farmacogenética e farmacogenómica e à utilização da informação haplotípica genética para prever a susceptibilidade um indivíduo à doença e/ou a sua resposta a um fármaco ou fármacos particular, de modo que fármacos, adaptados a diferenças genéticas de grupos populacionais, possam ser desenvolvidos e/ou administrados a indivíduos com o perfil genético apropriado.

A informação da sequência nucleotídica é derivada de ADN genómico. Os dados de sequências genómicas utilizadas podem ser obtidos de animais clínicos ou não humanos ou de células de cultura ou estudos de tecidos isolados. O organismo pode ser qualquer organismo para o qual estão disponíveis sequências múltiplas, mas é, de um modo preferido, de humano. Na identificação de novas variações é muitas vezes útil seleccionar diferentes grupos populacionais com base na raça, etnia, género, e/ou origem geográfica pois variações particulares podem diferir em frequência entre esses grupos. De um modo muito preferido, para doenças ou distúrbios que se pensa serem multigénicos (doenças/distúrbios geneticamente complexos), os fenótipos representados pela população submetida são extremos opostos.

Amostras biológicas contendo ADN podem ser sangue, sémen, zaragatoa na bochecha, etc. O isolamento de ADN de tais amostras é bem conhecido na técnica. Métodos para determinar a sequência de ADN num locus genético de interesse particular são também conhecidos na técnica. A sequenciação automática está agora amplamente disponível e requer apenas uma amostra de ADN isolada

e pelo menos um iniciador que é especificamente designado para reconhecer uma sequência altamente conservada dentro ou em proximidade estreia com o *locus* genético de interesse.

De acordo com algumas formas de realização, uma região ou *locus* de interesse genético definido (e. g., definido por um conjunto de iniciadores de PCR directo e reverso) é cuidadosamente sequenciado de um coorte de pessoas, inclusive de doentes que estão bem caracterizados para um distúrbio particular.

Uma sequência consenso é determinada e todas as sequências variantes observadas para um dado *locus* genético são compiladas numa lista. *Loci* possuindo o maior número de variantes observadas representa divergência evolutiva de um ancestral comum. Como tal, estes *loci* estão ligados em *cis* ao *loci* possuindo apenas um ou muito poucas variantes observadas. Durante as fases iniciais da investigação, pelo menos, é preferido que as populações sejam analisadas em grupos de indivíduos partilhando um fenótipo geral comum representando ancestralidade semelhante. De outro modo, a análise destes dados através da construção de árvores filogenéticas vai exigir um proibitivo grande número de indivíduos.

3. Alinhamento de sequências múltiplas.

A determinação da presença de uma variação particular ou pluralidade de variações num gene ou região génica numa população pode ser realizada numa variedade de modos, todos eles envolvendo localizar um *locus* genético particular por escolha de sequências dentro da região de interesse que são conhecidas por

serem altamente conservadas. A partir de *locus* altamente conservado, as sequências contíguas são facilmente obtidas através de uma de várias técnicas bem conhecidas na técnica.

O primeiro passo na análise de sequências de ADN em paralelo de uma pluralidade de indivíduos é alinhamento de sequências múltiplas ("MSA"). MSA é tipicamente utilizado para monitorizar alinhamento de sequências de amostras homólogas com diferenças polimórficas dentro de genes ou regiões génicas para mostrar áreas conservadas e sequências variantes. MSA da informação da sequência obtida no *locus* de interesse podem ser construídas utilizando uma ou mais de várias técnicas conhecidas e software disponível publicamente e são publicamente disponíveis de várias fontes incluindo a Internet. Métodos para analisar alinhamento de sequências múltiplas conhecidos na técnica incluem, e. g., aqueles descritos na Patente U.S. 6128587 de Sjolander; Patente U.S. 6291182 de Schork *et al.*; e Patente U.S. 6401043 de Stanton *et al.*.

4. Árvore filogenética e análise.

Vários métodos para a construção de "árvores filogenéticas" são conhecidos na técnica (Ver, e. g., Sanderson, 2008). Sun *et al.*, utilizaram análise de "bloco de haplótipo" para estudos de associação entre variantes de receptores tipo Toll (TLR) e cancro da próstata (2005) e Bardel *et al.*, (2005) utilizou uma abordagem de análise cladística para investigar associações entre variantes de gene CARD15 e a doença de Crohn. Contudo, nenhum dos loci genéticos utilizados anteriormente se associou com a doença para investigar ligações.

Podem ser construídas árvores filogenéticas, de acordo com algumas formas de realização, com uma topologia na qual variantes de sequência haplotípica observadas em indivíduos únicos humanos estudados formam nós (representando cada sequência observada nos dados) numa árvore. Aos nós podem ser juntos outros nós e o ancestral comum é encontrado no sítio da ramificação, raiz comum ou nó raiz da árvore. A árvore filogenética reflecte a relação evolutiva entre loci genéticos para os quais são analisados dados (ver Sanderson, 2008; Tzeng, 2005; Seltman, 2003). A **Figura 4** mostra uma árvore filogenética detalhada construída para a Região B do locus genético mostrado na **Figura 3**.

O ponto de partida para a estimação da árvore filogenética é geralmente um MSA (ver acima). Várias aplicações de software estão disponíveis para a construção de árvores filogenéticas com base em dados de sequência. Ver, e. g., Patente U.S. 7127466 e Patente U.S. 6532467 de Brocklebank, et al. A premissa básica é que um locus genético que exibe várias variantes é representado por estas variantes ligadas em *cis*. Os polimorfismos criam pontos de ramificação (nós) na árvore que definem grupos de sequências ou haplótipos relacionadas.

A árvore filogenética é utilizada para informação por examinar iterativamente proporções de indivíduos afectados com uma condição para indivíduos de controlo não afectados; os cálculos começam com nós observados no maior número de indivíduos e move-se para a periferia da árvore para nós observados em menos indivíduos. O objectivo é localizar um ponto de ramificação, ramo ou nó em que existe uma alteração substancial na proporção de indivíduos afectados com a condição de interesse para indivíduos de controlo não afectados. Tal

ponto de ramificação representa a divergência evolutiva de indivíduos com maior risco a partir de indivíduos com menor risco ou vice-versa.

A análise estatística da árvore filogenética gerada pode ser realizada de acordo com os métodos conhecidos na técnica. Um método reconhecido na técnica é o cálculo de intervalos de confiança "bootstrap" (ver Efron et al., Proc. Natl. Acad. Sci. USA 93, 13429-13434 (1996)).

5. Avaliação do doente.

Uma vez que uma árvore filogenética tenha sido gerada para um locus genético particular, um único indivíduo pode ser avaliado por comparação da sua sequência de ADN às sequências que compreendem a árvore filogenética. A presença de haplótipos ou sequências variantes correspondentes com regiões da árvore que representa indivíduos com maior incidência da condição de interesse (i. e., maiores proporções de indivíduos afectados com a doença ou distúrbio para indivíduos de controlo não afectados) significaria que os indivíduos únicos estão também com risco aumentado. De outro modo, proporções substancialmente menores correspondem a reduzido risco de desenvolver a condição de interesse.

As árvores filogenéticas podem também ser analisadas com base na responsividade da condição de interesse ao tratamento com um agente activo ou método de tratamento de interesse, de acordo com algumas formas de realização.

6. APOE e TOMM40.

Fenótipos e genótipos de ApoE são bem conhecidos na técnica. O sistema de nomenclatura estabelecido, bem como os fenótipos e genótipos para ApoE são aqui descritos, por exemplo, Zannis *et al.*, 1982.

Fenótipos e genótipos de TOMM40 (A Sub-Unidade de Canal da Membrana mitocondrial externa, 40 kDa) são também conhecidos. TOMM40 funciona como uma sub-unidade formadora de canal da translocase encontrada na membrana mitocondrial que é essencial para a importação de proteínas para dentro da mitocôndria. Dados de varrimento de associação de genoma completo de estudos de doentes da doença de Alzheimer têm identificado, inequivocamente, a região de desequilíbrio de ligação que contém o gene da apolipoproteína E (ApoE). A variante de ApoE 4 tem sido amplamente replicada como um gene de susceptibilidade confirmada desde as publicações iniciais em 1993 (ver, *e. g.*, Corder *et al.*,). Contudo, os dados de varrimento de associação de genoma completo resultaram numa "coincidência" notável observada em estudos de biologia celular envolvendo a co-localização de ApoE e TOMM40 para a membrana mitocondrial externa. Este outro gene, TOMM40, foi encontrado pela primeira vez durante estudos de modelação de desequilíbrio de ligação em torno de ApoE em 1998. Os polimorfismos foram localizados adjacentes a ApoE dentro de uma pequena região de desequilíbrio de ligação.

ApoE co-localiza para a membrana mitocondrial externa, sugerindo interacções específicas de isoforma, levando a um potencial papel para a apoptose mitocondrial induzida por ApoE como um passo inicial na expressão da doença de Alzheimer. Dados

biológicos têm demonstrado que a proporção de mitocôndrias móveis em cultura de células neuronais, bem como a velocidade com a qual estas se movem e a distância que atravessam, são factores que afectam apoptose mitocondrial aumentada. Dados filogenéticos sugerem um efeito genético independente no desenvolvimento de doença de Alzheimer para TOMM40.

ApoE liga-se especificamente à mitocôndria em culturas neuronais humanas (Chang, 2005) e a sequenciação desta região de desequilíbrio de ligação em centenas de doentes da doença de Alzheimer e respectivos controlos, combinada com mapeamento da evolução da variante genética de TOMM40, define uma região de particular interesse para interacções ApoE-TOMM40, como mostrado na **Figura 3**. Estes dados evolutivos suportam, ainda, a associação genética entre ApoE e TOMM40, e sugerem que a disfunção mitocondrial pode ser responsável por morte neuronal que ocorre lentamente ao longo de muitos anos. A idade da distribuição inicial da doença de Alzheimer (ver, e. g., Patente U.S. 6027896 de Roses *et al.*,) pode reflectir a herança de variantes intimamente ligadas de duas proteínas que interagem bioquimicamente que levam à expressão clínica da doença.

Como é aqui detalhado, a interacção entre múltiplos haplótipos de variantes de TOMM40 e alelos de ApoE contribui para a patogénese da doença de Alzheimer; em particular, haplótipos de TOMM40 em ligação com o alelo E 3 de ApoE contribui a patogénese da doença. Diversas das variantes génicas de TOMM40 evoluíram apenas ligadas a *cis* para ApoE 3. (De um modo semelhante, variantes específicas de TOMM40 podem ter evoluído ligadas a *cis* para ApoE 4 ou ApoE 2). Por este motivo, qualquer efeito genético adicional das variantes de TOMM40 segregam independentemente de ApoE 4 mas os dois produtos de

proteína variante podem interagir funcionalmente, em *trans*, para produzir um dado fenótipo ou característica observável. Por este motivo, qualquer efeito genético adicional das variantes de TOMM40 segregam independentemente de ApoE 4. Esta "coincidência" de genes de interacção adjacente pode explicar os dados estatísticos de associação extraordinariamente significativos encontrados em todos os estudos de varrimento de associação de genoma completo da doença de Alzheimer. É de interesse notar que as plataformas de varrimento de associação de genoma completo inicialmente comercializadas não continham quaisquer polimorfismos de ApoE, mas foram identificados com TOMM40 e ApoC1 SNP - mas a região é virtualmente sempre referida como a "região ApoE".

Estes dados, que combinam a genética de doença e mecanismos moleculares putativos de patogénese, podem também ser vistos dentro de um contexto farmacogenético. Devido ao forte efeito genético de herdar um alelo de ApoE 4, ApoE 4 tem sido referido como um gene de susceptibilidade complexa por mais que uma década. Replicações consistentes da idade de distribuição inicial como uma função de genótipo de ApoE confirmam que o papel da herança de ApoE 3 não é totalmente benigno, mas é um factor de risco mais baixo observado numa taxa mais lenta de início da doença. Existem variantes genéticas de TOMM40 que estão localizadas apenas em cadeias de ADN contendo ApoE 3 nas regiões de desequilíbrio de ligação (Roses et al., dados não publicados) e, por este motivo, não em equilíbrio de Hardy-Weinberg como foi requerido para SNP em painéis de associação de genoma completo. Alterações evolutivas em sequências de TOMM40 que estão ligadas em *cis* apenas a ApoE 3 actuam para aumentar o risco de doença de Alzheimer associado com ApoE 3, enquanto outras variantes de TOMM40 ligadas em *cis* a

ApoE 3 diminuem o risco associado com ApoE 3. Um teste genético independente seria para determinar se esses polimorfismos de TOMM40 associados com menor doença de Alzheimer segregam numa idade mais avançada em gráficos de distribuição de idade de início, para ApoE 3 contendo genótipos [ApoE 3/3 ou ApoE 4/3].

A detecção da presença ou ausência de ApoE 2, 3 ou 4, e/ou haplótipos de TOMM40 ou de ADN que codifica o mesmo (incluindo, em algumas formas de realização, o numero de alelos para cada) num indivíduo, pode ser realizado directa ou indirectamente por quaisquer meios adequados. São conhecidas uma variedade de técnicas pelos especialistas na técnica. Todas envolvem, geralmente, o passo de recolher uma amostra de material biológico contendo ADN ou proteína do indivíduo e, em seguida, detectar se o indivíduo possui ou não o haplótipo de interesse. Por exemplo, o passo de detecção em relação a ApoE pode ser realizado por recolha de uma amostra de ApoE do indivíduo (por exemplo, de fluido cerebrospinal ou qualquer outro fluido ou tecido contendo ApoE) e, então, determinar a presença ou ausência de uma isoforma de ApoE 2, 3 ou 4 na amostra de ApoE (e. g., por focagem isoeléctrica ou imunoensaio).

Determinar a presença ou ausência de ADN que codifica uma isoforma de ApoE e/ou TOMM40 pode ser realizado por sequenciação directa da região de interesse de ADN genómico, com uma sonda oligonucleotídica marcada com um grupo detectável adequado e/ou por meio de uma reacção de amplificação, tal como uma reacção em cadeia de polimerase ou reacção em cadeia de ligase (o produto de tal reacção de amplificação pode então ser detectado com uma sonda oligonucleotídica marcada ou um número de outras técnicas). Além disso, o passo de detecção pode incluir o passo de detectar se o indivíduo é heterozigótico ou homozigótico para

o gene que codifica um haplótipo de ApoE e/ou TOMM40. Numerosos formatos de ensaios com sonda oligonucleotídica diferentes são conhecidos que podem ser utilizados para realizar a presente invenção. Ver, e. g., documentos Pat. U.S. N° 4302204 de Wahl et al.,; Pat.U.S. N° 4358535 de Falkow et al.,; Pat. U.S. N° 4563419 de Ranki et al.,; e Pat. U.S. N° 4994373 de Stavrianopoulos et al.

Em algumas formas de realização, a detecção pode incluir amplificação múltipla do ADN (e. g., PCR fluorescente específico para alelo). Em algumas formas de realização, a detecção pode incluir hibridação a uma micromatriz (um chip, esferas, etc.). Em algumas formas de realização, a detecção pode incluir sequenciar porções apropriadas do gene contendo os haplótipos a ser detectado. Em algumas formas de realização, podem ser utilizados para detecção haplótipos que alteram a susceptibilidade à digestão por uma ou mais enzimas de restrição endonuclease. Por exemplo, pode ser utilizado polimorfismo de comprimento de fragmentos de restrição (RFLP), que se refere ao padrão de digestão quando várias enzimas de restrição são aplicadas ao ADN. Em algumas formas de realização, a presença de um ou mais haplótipos pode ser determinado pela amplificação específica do alelo. Em algumas formas de realização, a presença de haplótipos pode ser determinada pela extensão do iniciador. Em algumas formas de realização, a presença de haplótipos pode ser determinada por ligação oligonucleotídica. Em algumas formas de realização, a presença de haplótipos pode ser determinada por hibridação com uma sonda marcada de forma detectável. Ver, e. g., Pedido de de Patente Publicação U.S. N° 2008/0153088 de Sun et al.,; Kobler et al., Identification of an HT 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).

A amplificação de uma sequência de ácido nucleico, seleccionada ou alvo, pode ser realizada por qualquer meio adequado em ADN isolado de amostras biológicas. Ver geralmente D. Kwoh e T. Kwoh, 1990. Exemplos de técnicas de amplificação adequada incluem, mas não estão limitadas a, reacção em cadeia de polimerase, reacção em cadeia de ligase, amplificação por deslocamento de cadeia (ver geralmente Walker et al., 1992a; Walker et al., 1992b), amplificação baseada em transcrição (ver Kwoh et al., 1989), replicação de sequência auto-sustentada (ou "3SR") (ver Guatelli et al., 1990), o sistema replicase Q β (ver Lizardi et al., 1988), amplificação baseada na sequência de ácido nucleico (ou "NASBA") (ver Lewis, 1992), uma reacção em cadeia de reparação (ou "RCR") (ver Lewis, supra) e amplificação de ADN boomerang (ou "BDA") (ver Lewis, supra). Reacção em cadeia de polimerase é actualmente preferida.

As técnicas de amplificação do ADN, tais como as acima, podem envolver a utilização de uma sonda, um par de sondas ou dois pares de sondas que se ligam especificamente ao ADN que codifica ApoE 4, mas não se ligam ao ADN que codifica ApoE 2 ou ApoE 3 sob as mesmas condições de hibridação, e que serve como o iniciador ou iniciadores para a amplificação do ADN de ApoE 4 ou uma sua porção na reacção de amplificação. Do mesmo modo, pode utilizar-se uma sonda, um par de sondas ou dois pares de sondas

que se ligam especificamente ao ADN que codifica ApoE 2, mas não se ligam ao ADN que codifica ApoE 3 ou ApoE 4 sob as mesmas condições de hibridação e que servem como o iniciador ou iniciadores para a amplificação do ADN de ApoE 2 ou uma sua porção na reacção de amplificação; e pode utilizar-se uma sonda, um par de sondas ou dois pares de sondas que se ligam especificamente ao ADN que codifica ApoE 2, mas não se ligam ao ADN que codifica ApoE 2 ou ApoE 4 sob as mesmas condições de hibridação e que servem como o iniciador ou iniciadores para a amplificação do ADN de ApoE 3 ou uma sua porção na reacção de amplificação.

De um modo semelhante, pode utilizar-se uma sonda, um par de sondas ou dois pares de sondas que se ligam especificamente ao ADN que codifica um haplótipo de TOMM40 de interesse, mas não se ligam a outros haplótipos de TOMM40 sob as mesmas condições de hibridação e que servem como o iniciador ou iniciadores para a amplificação do ADN de TOMM40 ADN ou uma sua porção na reacção de amplificação.

Em geral, uma sonda oligonucleotídica que é utilizada para detectar ADN que codifica haplótipos de ApoE e/ou TOMM40 é uma sonda oligonucleotídica que se liga a ADN que codifica o haplótipo de interesse, mas não se liga a ADN que codifica outros haplótipos sob as mesmas condições de hibridação. A sonda oligonucleotídica é marcada com um grupo detectável adequado, tal como aquele estabelecido abaixo em associação com anticorpos.

A reacção em cadeia de polimerase (PCR) pode ser realizada de acordo com técnicas conhecidas. Ver, e. g., Pat. U.S. N° 4683195; 4683202; 4800159; e 4965188. Em geral, o PCR

envolve, primeiro, tratar uma amostra de ácido nucleico (e. g., na presença de uma polimerase de ADN termo-estável) com um iniciador oligonucleotídico para cada cadeia da sequência específica para ser detectada sob condições de hibridação, de modo que um produto da extensão de cada iniciador seja sintetizado que é complementar a cada cadeia de ácido nucleico, com os iniciadores suficientemente complementares a cada cadeia da sequência específica a hibridar respectivamente, de modo que o produto da extensão sintetizado a partir de cada iniciador, quando estiver separado do seu complemento, possa servir como um modelo para a síntese do produto de extensão do outro iniciador, e então tratar a amostra sob condições de desnaturação para separar os produtos de extensão do iniciador dos seus modelos, se a sequência ou sequências a serem detectadas estiverem presentes. Estes passos são repetidos ciclicamente até o grau desejado de amplificação ser obtido. A detecção da sequência amplificada pode ser realizada por adicionar ao produto da reacção uma sonda oligonucleotídica capaz de se hibridar com o produto da reacção (e. g., uma sonda oligonucleotídica da presente invenção), a sonda transportando uma etiqueta detectável e, em seguida, detectar a etiqueta de acordo com técnicas conhecidas ou por visualização directa num gel.

Quando as condições de PCR permitem amplificação de todos os tipos alélicos de ApoE, os tipos podem ser distinguidos por hibridação com sonda específica alélica, por digestão com endonuclease de restrição, por electroforese com géis de gradiente de desnaturação ou outras técnicas. Um protocolo de PCR para determinar o genótipo de ApoE é descrito em Wenham *et al.*, (1991). Exemplos de iniciadores eficazes para amplificação e identificação das isoformas de ApoE são ai descritos. Iniciadores específicos para a região polimórfica de

ApoE (seja ApoE 4, E3 ou E2) podem ser utilizados. Em Wenham, por exemplo, iniciadores de PCR são utilizados que amplificam uma região de 227 pb de ADN que abrange os sítios polimórficos de ApoE (codões 112 e 158, que contém nucleótidos 3745 e 3883). Os fragmentos amplificados são então submetidos a endonuclease de restrição CfoI que proporciona diferentes fragmentos de restrição a partir dos seis possíveis genótipos de ApoE que podem ser reconhecidos num gel de electroforese. Ver também, Hixon et al., (1990); Houlston et al., (1989) Wenham et al., (1991); e Konrula et al., (1990) para métodos adicionais.

Além da doença de Alzheimer, existem diversas outras doenças e distúrbios geneticamente complexos para os quais os métodos da presente invenção proporcionam vantagens sobre as análises existentes. Por exemplo, dados de múltiplos estudos genéticos de diabetes mellitus tipo 2 suportam a visão que serão necessárias grandes séries de casos clínicos/controlo para proporcionar significância estatística para loci definidos por estudos de associação de genoma completo.

7. Agente activos, composições e tratamento.

Como observado acima, árvores filogenéticas criadas utilizando os métodos aqui detalhados também podem ser analisadas com base na responsividade da condição de interesse ao tratamento com um agente activo ou método de tratamento de interesse, de acordo com algumas formas de realização e decisões de tratamento para um indivíduo ou doente podem ser com base em variantes genéticas específicas identificadas.

Agentes activos. Agentes activos incluem aqueles conhecidos para tratamento de uma condição de interesse, e são inclusive de agente activos de anti-doença de Alzheimer, incluindo, mas não limitados a, inibidores da acetilcolinesterase, antagonistas do receptor de NMDA, agonistas ou moduladores de PPAR, incluindo, mas não limitado aos fármacos nas classes de tiazolidinediona ou glitazar). O agente activo pode também ser um produto biofarmacêutico, por exemplo um anticorpo (e. g., monoclonal, policlonal, derivados de ou anticorpos modificados, tal como Domain Anticorpos™, Bapineuzumab, etc.), proteínas de fusão ou moléculas de ARN terapêuticas. O agente activo pode também ser uma combinação de qualquer destes produtos.

Exemplos de inibidores da acetilcolinesterase incluem, mas não são limitados a, donepezilo (comercializado como ARICEPT), galantamina (comercializada como RAZADYNE) e rivastigmina (comercializada como EXELON) e os seus sais farmaceuticamente aceitáveis. Exemplos adicionais incluem, mas não são limitados aos descritos nas Patentes U.S. Nº 6303633; 5965569; 5595883; 5574046; e 5171750.

Exemplos de antagonistas do receptor de NMDA, incluem, mas não são limitados a, memantina (comercializada como AKATINOL, AXURA, EBIXIA/ABIXIA, MEMOX e NAMENDA) e os seus sais farmaceuticamente aceitáveis. Exemplos adicionais incluem, mas não são limitados aos descritos nas Patentes US Nº 6956055; 6828462; 6642267; 6432985; e 5990126. Exemplos de tiazolidinedionas incluem, mas não são limitados a, rosiglitazona (comercializada como AVANDIA) e os seus sais farmaceuticamente aceitáveis. Exemplos adicionais incluem, mas não são limitados a: 5-(4-[2-(N-metil-N-(2-benzotiazolil)amino)etoxi]benzil)-2,4-tiazolididina diona;

5-(4-[2-(N-metil-N-(2-benzotiazolil)amino)etoxi]benzilideno)-
2,4-tiazol idinediona; 5-(4-[2-(N-metil-N-(2-
benzoxazolil)amino)etoxi]benzil)-2,4-tiazolidinediona;
5-(4-[2-(N-metil-N-(2-benzoxazolil)amino)etoxi]benzilideno)-2,4-
tiazolidinediona; 5-(4-[2-(N-metil-N-(2-
pirimidinil)amino)etoxi]benzil)-2,4-tiazolidinediona;
5-(4-[2-(N-metil-N-(2-pirimidinil)amino)etoxi]benzilideno)-2,4-
tiazolidinediona; 5-(4-(2-(N-metil-N-[2-(4,5-
dimetiltiazolil)]amino)etoxi]benzil)-2,4-tiazolidinediona; 5-(4-
[2-(N-metil-N-[2-(4,5-dimetiltiazolil)]amino)etoxi]benzilideno)-
2,4-tiazolidinediona; 5-(4-[2-(N-metil-N-(2-
tiazolil)amino)etoxi]benzil)-2,4-tiazolidinediona; 5-(4-[2-(N-
metil-N-(2-tiazolil)amino)etoxi]benzilideno)-2,4-
tiazolidinediona; 5-[4-(2-(N-metil-N-(2-(4-feniltiazolil))amino)
etoxi)benzil]-2,4-tiazolidinediona; 5-(4-[2-(N-metil-N-(2-(4-
feniltiazolil))amino)etoxi]benzilideno)-2,4-tiazolidinediona;
5-(4-[2-(N-metil-N-[2-(4-fenil-5-
metiltiazolil)]amino)etoxi]benzil)-2,4-tiazolidinediona;
5-(4-[2-(N-metil-N-[2-(4-fenil-5-metil tiazolil)]amino)etoxi]
benzilidene)-2,4-tiazolidinediona; 5-(4-[2-(N-metil-N-[2-(4-
metil-5-feniltiazolil)]amino)etoxi]benzil)-2,4-tiazolidinediona;
5-(4-[2-(N-metil-N-[2-(4-metil-5-
feniltiazolil)]amino)etoxi]benzilideno e)-2,4-tiazolidinediona;
5-(4-[2-(N-metil-N-[2-(4-metiltiazolil)]amino)etoxi]benzil)-2,4-
tiazolidinediona; 5-(4-[2-(N-metil-N-[2-(4-metiltiazolil)]amino)
etoxi]benzilideno)-2,4-tiazolidinediona; 5-[4-(2-(N-metil-N-[2-
(5-feniloxazolil)]amino)etoxi)benzil]-2,4-tiazolidinediona;
5-(4-[2-(N-metil-N-[2-(5-feniloxazolil)]amino)
etoxi]benzilideno)-2,4-tiazolidinediona;

5-(4-[2-(N-metil-N-[2-(4,5-dimetiloazolil)]amino)
etoxi]benzil)-2,4-tiazolidinediona; 5-(4-[2-(N-metil-N-[2-(4,5-

dimetiloxazolil)]amino)etoxi]benzilideno)-2,4-tiazolidinediona;
5-[4-(2-(2-pirimidinilamino)etoxi)benzil]-2,4-tiazolidinediona;
5-[4-(2-(2-pirimidinilamino)etoxi)benzilideno]-2,4-
tiazolidinediona; 5-(4-[2-(N-acetyl-N-(2-
pirimidinil)amino)etoxi]benzil)-2,4-tiazolidinediona;
5-(4-(2-(N-(2-benzotiazolil)-N-benzilamino)etoxi) benzilideno)-
2,4-tiazolidinediona; 5-(4-(2-(N-(2-benzotiazolil)-N-
benzilamino)etoxi) benzil)-2,4-tiazolidinediona;
5-(4-[3-(N-metil-N-(2-benzoxazolil)amino)propoxi]benzil)-2,4-
tiazolidinediona; 5-(4-[3-(N-metil-N-(2-
benzoxazolil)amino)propoxi]benzilideno)-2,4-tiazolidinediona;
5-(4-[2-(N-metil-N-(2-piridil)amino)etoxi]benzil)-2,4-
tiazolidinediona; 5-(4-[2-(N-metil-N-(2-
piridil)amino)etoxi]benzilideno)-2,4-tiazolidinediona;
5-(4-[4-(N-metil-N-(2-benzoxazolil)amino)butoxi]benzilideno)-
2,4-tiazolidinediona; 5-(4-[4-(N-metil-N-(2-
benzoxazolil)amino)butoxi]benzil)-2,4-tiazolidinediona;
5-(4-[2-(N-(2-benzoxazolil)amino)etoxi]benzilideno) 2,4-
tiazolidinediona; 5-(4-[2-(N-(2-
benzoxazolil)amino)etoxi]benzil)-2,4-tiazolidinediona;
5-(4-[2-(N-isopropil-N-(2-benzoxazolil)amino)etoxi]benzil)-2,4-
tiazolidin ediona, e seus sais farmaceuticamente aceitáveis.
Ver, e. g., Patente U.S. N° 5002953.

Os agentes activos aqui divulgados podem, como observado acima, ser preparados na forma dos seus sais farmaceuticamente aceitáveis. Sais farmaceuticamente aceitáveis são sais que retém a actividade biológica desejada do composto mãe e não conferem efeitos tóxicos indesejados. Exemplos de tais sais são (a) sais de adição de sais formados com ácidos inorgânicos, por exemplo, ácido clorídrico, ácido bromídrico, ácido sulfúrico, ácido fosfórico, ácido nítrico e semelhantes; e sais formados com

ácidos orgânicos, tal como, por exemplo, ácido acético, oxálico, ácido, ácido tartárico, succínico, ácido maleico, ácido fumárico, ácido glucónico, ácido cítrico, ácido málico, ácido ascórbico, ácido benzóico, ácido tânico, ácido palmítico, ácido algínico, ácido poliglutâmico, ácido naftalenossulfónico, ácido metanossulfónico, ácido p-toluenossulfónico, ácido naftalenodissulfónico, ácido poligalacturónico e semelhantes; (b) sais formados de anões elementares, cloro, bromo e iodo, e (c) sais derivados de bases, tal como sais de amónio, sais de metais alcalinos, tal como os de sódio e potássio, sais de metais alcalino-terrosos, tais como os de cálcio e magnésio, e sais com bases orgânicas, tais como diciclo-hexilamina e N-metil-D-glucamina.

Os agentes activos podem ser administrados como profármacos. "Profármacos", como aqui utilizado, refere-se aos profármacos dos compostos da presente invenção que são, dentro do âmbito do bom julgamento médico, adequados para utilização em contacto com os tecidos de humanos e animais inferiores sem indevida toxicidade, irritação, resposta alérgica e semelhantes, comensurável com uma relação risco/benefício, e eficaz para a sua utilização pretendida, bem como as suas formas zwitteriónicas, onde possível, dos compostos da invenção. O termo "profármaco" refere-se a compostos que são rapidamente transformados *in vivo* para originar o composto mãe das fórmulas acima, por exemplo, por hidrólise no sangue. É proporcionada uma discussão aprofundada em T. Higuchi e V. Stella, Prodrugs as Novel delivery Systems, Vol. 14 do A.C.S. Symposium Series e em Edward B. Roche, ed., Bioreversible Carriers in Drug Design, American Pharmaceutical Association e Pergamon Press, 1987. Exemplos incluem um profármaco que é metabolizado *in vivo* por um indivíduo a um fármaco activo possuindo uma actividade de

compostos activos como aqui descrito, em que o profármaco é um éster de um álcool ou grupo de ácido carboxílico, se um tal grupo está presente no composto; um acetal ou cetal de um grupo álcool, se um tal grupo está presente no composto; uma base N-Mannich ou uma imina de um grupo amina, se um tal grupo está presente no composto; ou uma base Schiff, oxima, acetal, éster enol, oxazolidina ou tiazolidina de um grupo carbonilo, se um tal grupo está presente no composto, tal como descrito na Patente U.S. Nº 6680324 e Patente U.S. Nº 6680322.

Composições. Os agentes activos descritos anteriormente podem ser formulados para administração num veículo farmacêutico, de acordo com técnicas conhecidas. Ver, e. g., Remington, The Science And Practice of Pharmacy (9^a Ed. 1995). Na produção de uma formulação farmacêutica, de acordo com a invenção, o composto activo (incluindo os seus sais fisiologicamente aceitáveis) é tipicamente misturado com, *inter alia*, um veículo aceitável. O veículo deve, naturalmente, ser aceitável no sentido de ser compatível com quaisquer outros ingredientes na formulação e não deve ser deletério para o doente. O veículo pode ser um sólido ou um líquido, ou ambos, e ser, de um modo preferido, formulado com o composto como uma formulação unidade-dose, por exemplo, um comprimido, que pode conter de 0,01 ou 0,5% a 95% ou 99% em peso do composto activo. Um ou mais compostos activos podem ser incorporados nas formulações da invenção, que pode ser preparada por qualquer das técnicas de farmácia bem conhecidas, compreendendo misturar os componentes, opcionalmente incluindo um ou mais ingredientes acessórios.

As formulações da invenção incluem aquelas adequadas para administração oral, rectal, tópica, bucal (e. g., sub-lingual),

vaginal, parentérica (e. g., subcutânea, intramuscular, intradermal ou intravenosa), tópica (i. e., ambas pele e superfícies mucosas, incluindo superfícies das vias respiratórias) e transdermal, embora a via adequada em qualquer caso dependa da natureza e gravidade da condição a ser tratada e da natureza do composto activo particular que está a ser utilizado.

Formulações adequadas para administração oral podem ser apresentadas em unidades discretas, tal como cápsulas, hóstias, pastilhas ou comprimidos, cada contendo uma quantidade predeterminada do composto activo; como um pó ou grânulos; como uma solução ou uma suspensão num líquido aquoso ou não aquoso; ou como uma emulsão óleo-em-água ou água-em-óleo. Tais formulações podem ser preparadas por qualquer método adequado de farmácia, que inclui o passo de colocar em associação o composto activo e um veículo adequado (que pode conter um ou mais ingredientes acessórios, como observado em cima). Em geral, as formulações da invenção são preparadas por misturar uniforme e intimamente o composto activo com um líquido ou veículo sólido finamente dividido ou ambos e, então, se necessário, moldar a mistura resultante. Por exemplo, um comprimido pode ser preparado por comprimir ou moldar um pó ou grânulos contendo o composto activo, opcionalmente com um ou mais ingredientes acessórios. Comprimidos compactados podem ser preparados por compressão, numa máquina adequada, o composto numa forma de fluolivre, tal como um pó ou granulados opcionalmente misturados com um aglutinante, lubrificante, diluente inerte e/ou agente(s) activos/de dispersão de superfície. Comprimidos moldados podem ser preparados por moldagem, numa máquina adequada, o composto triturado humedecido com um aglutinante líquido inerte.

Formulações adequadas para administração bucal (sub-lingual) incluem pastilhas compreendendo o composto activo numa base aromatizada, normalmente sacarose e acácia ou tragacanto; e pastilhas compreendendo o composto numa base inerte, tal como gelatina e glicerina ou sacarose e acácia.

Formulações da presente invenção adequadas para administração parentérica compreendem soluções de injecção estéril aquosas e não aquosas do(s) composto(s) activo(s), que são preparações, de um modo preferido, isotónicas com o sangue do receptor pretendido. Estas preparações podem conter antioxidantes, tampões, bacteriostáticos e solutos, que tornam a formulação isotónica com o sangue do receptor pretendido. Suspensões estéreis aquosas e não aquosas podem incluir agentes de suspensão e agentes espessantes. As formulações podem ser apresentadas em recipientes unidade\dose ou multi-dose, por exemplo, ampolas vedadas e frasquinhos e podem ser armazenadas numa condição seca por congelação (liofilizada) requerendo apenas a adição do veículo líquido estéril, por exemplo, soro fisiológico ou água para injecção, imediatamente antes de utilizar. Soluções e suspensões para injecção extemporânea podem ser preparadas a partir de pós estéreis, grânulos e comprimidos do tipo descrito anteriormente. Por exemplo, num aspecto da presente invenção, é proporcionado uma composição injectável, estável, estéril compreendendo um(uns) agentes activo(s) ou um seu sal, numa forma de dosagem unitária num recipiente vedado. O composto ou sal é proporcionado na forma de um liofilizado, que é capaz de ser reconstituído com um veículo adequado farmaceuticamente aceitável para formar uma composição líquida adequada para sua injecção num indivíduo. A forma da dosagem unitária comprehende, tipicamente, desde cerca de 10 mg a cerca de 10 gramas do composto ou sal. Quando o composto ou sal é

substancialmente insolúvel em água, uma quantidade suficiente de agente emulsionante, que é fisiologicamente aceitável, pode ser utilizado em quantidade suficiente para emulsionar o composto ou sal num veículo aquoso. Um tal agente emulsionante é fosfatidilcolina.

Formulações adequadas para aplicação tópica à pele, tomam à forma, de um modo preferido, de uma pomada, creme, loção, pasta, gel, spray, aerossol ou óleo. Veículos que podem ser utilizados incluem vaselina, lanolina, polietilenoglicóis, álcoois, potenciadores transdérmicos e combinações de dois ou mais destes.

Formulações adequadas para administração transdermal podem ser apresentadas como emplastros discretos adaptados para permanecer em contacto íntimo com a epiderme do destinatário por um prolongado período de tempo. Formulações adequadas para administração transdermal podem também ser administradas por iontopforese (ver, por exemplo, Pharmaceutical Research 3 (6):318 (1986)) e, tipicamente, tomam a forma de uma solução aquosa opcionalmente tamponada do composto activo. Formulações adequadas compreendem citrato ou tampão bis\tris (pH 6) ou etanol/água e contêm de 0,1 a 0,2 M de ingredientes activos.

Além do(s) composto(s) activo(s), as composições farmacêuticas podem conter outros aditivos, tal como aditivos de ajuste de pH. Em particular, agentes de ajuste de pH úteis incluem ácidos, tal como ácido clorídrico, bases ou tampões, tal como sódio lactato, acetato de sódio, fosfato de sódio, citrato de sódio, borato de sódio ou gluconato de sódio. Além disso, as composições podem conter conservantes microbianos. Conservantes microbianos úteis incluem metilparabeno, propilparabeno e álcool

benzílico. Os conservantes microbianos são, tipicamente, utilizados quando a formulação é colocada num frasquinho concebido para utilização multidose. Naturalmente, como indicado, as composições farmacêuticas da presente invenção podem ser liofilizadas utilizando técnicas bem conhecidas na técnica.

Dosagem. A dosagem terapeuticamente eficaz de qualquer agente activo específico, a utilização dos quais está no âmbito da presente invenção, irá variar um pouco de composto para composto e de doente para doente e irá depender da condição do doente e da via de administração. Para administração oral pode ser utilizada uma dosagem diária total de desde 1, 2 ou 3 mg até 30, 40 ou 50 mg, dado como uma dose diária única ou dividida em duas ou três doses diárias.

Tratamento. Variantes genéticas como aqui descritas ou descobertas utilizando os métodos como aqui ensinado, podem ser utilizadas para determinar o curso de tratamento de um doente afectado com uma condição (e. g., uma condição associada com ApoE e/ou TOMM40), por, e. g., determinar qual agente activo e/ou curso de tratamento administrar com base na presença ou ausência da variante ou variantes genética. A presença ou ausência da variantes genéticas podem indicar eficácia de um agente activo e/ou curso de tratamento para o doente, predizer a idade de início de uma condição, indicar regimes de dosagem preferidos, etc. Um perfil genético pode ser gerado para um doente e o perfil consultado para determinar se o doente esta entre um grupo de doentes que são susceptíveis de ser responsivos a um agente activo particular.

As instruções para utilização podem ser embaladas com, ou associadas de outro modo, com um agente activo, indicando recomendações para tratamento, tempo de tratamento, regimes de dose, etc., com base na presença ou ausência das variantes genéticas.

8. Métodos para determinar uma predição do risco de doença ou um prognóstico.

Para determinar uma predição de risco de doença para um indivíduo não sintomático ou um prognóstico (a perspectiva de aflição ou curso da doença como antecipado a partir do curso normal de doença ou particularidades do caso), de acordo com algumas formas de realização da presente invenção, dados de diagnóstico, incluindo o diagnóstico do doente ou historial médico e dados genéticos, tal como o genótipo do doente (e. g., genótipo ApoE e/ou TOMM40), pode ser realizado para proporcionar opções terapêuticas e previsões de resultados. O processamento pode incluir obter um "perfil do doente", tal como a colecção do historial médico do doente, incluindo idade e género, genotipando o loci de interesse (e. g., utilizando iniciadores concebidos apropriadamente e utilizando um passo de amplificação RT-PCR ou PCR e/ou fenotipagem, e. g., utilizando um método mediado por anticorpos ou testes enzimáticos) e estatística ou outras análises que convertem estes dados em bruto num prognóstico. O prognóstico pode incluir uma predição da idade de início da doença do paciente, resposta ao tratamento com fármaco, tempo de tratamento, eficácia do tratamento, etc. Em algumas formas de realização, o prognóstico pode incluir a utilização de um programa de software de computador para analisar dados do doente e executar controlos cruzados

estatísticos contra bases de dados relacionadas, de modo a converter os dados ou perfil do doente a um prognóstico.

Um "perfil do doente" inclui dados e/ou materiais que pertencem ao doente para os quais a análise preditiva e/ou prognóstica está a ser realizada. Os dados podem incluir informação do diagnóstico, idade, género e/ou genótipo do doente. O perfil do doente pode também incluir materiais do doente, tal como sangue, amostras de proteína de soro, fluido cerebrospinal ou ARN ou ADN purificado.

9. Estratificação do genótipo em ensaios clínicos.

A detecção de um genótipo aqui ensinado ou como determinado com os métodos aqui, pode ser utilizado na condução de um ensaio clínico, do mesmo modo que a outra informação gentípica é utilizada para conduzir um ensaio clínico, tal como descrito em, e. g., Patentes U.S. Nº 6573049, 6368797 e 6291175.

Em algumas formas de realização, tais métodos, de um modo vantajoso, estratificam ou permitem o aperfeiçoamento da população de doentes (e. g., por divisão da população num ou mais subgrupos), de modo que as vantagens de regimes de terapia particulares podem ser detectadas com maior precisão, particularmente em relação a sub-populações particulares de doentes com genótipos particulares. Em algumas formas de realização, tais métodos compreendem administrar um agente activo teste ou tratamento a uma pluralidade de indivíduos (um tratamento controlo ou placebo sendo tipicamente administrado a uma separada, mas de um modo semelhante caracterizada, pluralidade de indivíduos) e detectar a presença ou ausência de

um genótipo (e. g., ApoE e/ou TOMM40), como descrito acima na pluralidade de indivíduos. O genótipo pode ser detectado antes, depois ou concomitantemente com o passo de administrar o tratamento de teste. A influência de um ou mais alelos detectados no tratamento teste pode então ser determinado em qualquer parâmetro adequado ou tratamento potencial ou consequência, incluindo, mas não limitado, à eficácia do referido tratamento, falta de efeitos secundários do tratamento, etc.

Um ensaio clínico pode ser estabelecido para testar a eficácia de compostos de teste para tratar qualquer número de doenças para as quais foi determinado um genótipo particular para ser associado, para indivíduos que estão diagnosticados com a doença ou estão em risco de desenvolver a doença. Se os indivíduos são genotipados após a conclusão de um ensaio clínico, as análises podem ainda ser destinadas a determinar uma relação entre um tratamento para uma doença e o alelo a ser avaliado para eficácia. Em alternativa, se um indivíduo sintomático ou assintomático não foi ainda diagnosticado com a doença, mas foi determinado como em risco de desenvolver a doença, pode ser realizado um ensaio clínico semelhante ao ensaio clínico descrito anteriormente.

Os mecanismos biológicos subjacentes também podem ser considerados aquando da concepção dos grupos de tratamento. Por exemplo, o fragmento de ApoE 4 (1-272) liga-se à mitocôndria, diminui a dinâmica celular mitocondrial e diminui mais sinaptogénese do que o ApoE 3 (1-272). Rosiglitazona, um fármaco candidato para o tratamento da doença de Alzheimer, aumenta a mitogénese e aumenta a sinaptogénese - contrariando os efeitos da ligação do fragmento de ApoE - para ApoE 3 de uma maior forma

do que para ApoE 4. Por esse motivo, o fármaco ou tratamento candidato (e. g., rosiglitazona) pode ser seleccionado com base em mecanismos de acção subjacentes uma vez que se refere aos marcadores genéticos utilizados para as estratificações (e. g., variantes de ApoE 2, E 3, E 4 e/ou TOMM40).

A avaliação da eficácia de um fármaco escolhido para o ensaio pode incluir monitorizar o indivíduo durante um período de tempo e analisar o atraso do início da doença e a intensidade da doença no momento do início, bem como medindo o início dos sintomas que são associados com a doença. Um fármaco que, num ensaio clínico, elimina ou atrasa o início da doença ou reduz os sintomas da doença, pode ser um fármaco benéfico para utilização em doentes diagnosticados com a doença ou em risco de desenvolver a doença. Compostos de teste que podem ser utilizados em tais ensaios incluem os agentes como descrito acima, incluindo aqueles aprovados anteriormente para utilização clínica e novos compostos ainda não aprovados para utilização ou aprovados para tratar uma doença particular. Por este motivo, em algumas formas de realização, o ensaio clínico pode incluir a optimização da administração do fármaco, incluindo dosagem, tempo de administração, toxicidades ou efeitos secundários, via de administração e eficácia do tratamento.

10. Kits úteis para a detecção de variantes de genótipo em loci de interesse.

São aqui proporcionados kits para determinar se um indivíduo está em risco aumentado de desenvolver uma doença, desenvolver uma doença numa idade de início mais precoce e/ou um candidato para um tratamento particular, em que a doença é

associada com ApoE e/ou TOMM40 (e. g., início tardio da doença de Alzheimer). Os kits incluem, pelo menos, um reagente específico para detectar a presença ou ausência de uma variante de ApoE e/ou TOMM40, como aqui descrito, e pode incluir instruções para ajudar na determinação se o indivíduo está em risco aumentado de desenvolver a doença. O kit pode incluir, opcionalmente, um ácido nucleico para detecção de um gene ApoE (e. g., ApoE 2, ApoE 3 e/ou ApoE 4) ou instruções para métodos de focagem isoeléctrica para detecção do genótipo de ApoE; e/ou um ácido nucleico para detecção de uma variante de TOMM40, como aqui descrito. Em algumas formas de realização, o kit pode incluir, opcionalmente, um ou mais anticorpos que se ligam a ApoE 2, ApoE 3, ApoE 4, ou a isoformas de TOMM40. O kit de teste pode ser embalado de qualquer modo adequado, tipicamente com todos os elementos num único recipiente com uma folha de instruções impressa para realizar o teste.

Em algumas formas de realização, o kit pode conter, opcionalmente, tampões, enzimas e reagentes para amplificar os ácidos nucleicos genómicos, por meio de amplificação dirigida por iniciador. O kit também pode incluir um ou mais dispositivos para detectar a presença ou ausência de haplótipos particulares no ácido nucleico amplificado. Tais dispositivos podem incluir uma ou mais sondas, que hibridam com um haplótipo de ácido nucleico, que pode estar ligado a um bio-chip ou dispositivo de microarranjo, tal como qualquer dos descritos na Pat. U.S. Nº 6355429. O bio-chip ou dispositivo de microarranjo tem, opcionalmente, pelo menos, uma sonda de captura ligada a uma superfície que pode hibridar com uma sequência haplotípica. Em formas de realização preferidas, o bio-chip ou microarranjo contém múltiplas sondas e, de um modo muito preferido, contém pelo menos uma sonda para uma sequência haplotípica que, se

presente, irá ser amplificada por um conjunto de iniciadores flanqueadores. Por exemplo, se cinco pares de iniciadores flanqueadores são utilizados para amplificação, o dispositivo irá conter, pelo menos, uma sonda haplotípica para cada produto amplificado ou, pelo menos, cinco sondas. O kit, também de um modo preferido, inclui instruções para utilizar os componentes do kit.

A presente invenção é explicada com maior detalhe nos seguintes exemplos não limitativos.

EXEMPLO 1: Construção de Árvores Filogenéticas

Todos os conhecidos estudos de varrimento de genoma completo demonstram valores de p extremamente significantes em torno do locus da apolipoproteinC1 [ApoC1]. (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 importância é que cada série identificou um gene candidato significante "favorecido" limítrofe fora da área de desequilíbrio de ligação de ApoE, mas estes genes candidatos favorecidos eram diferentes em cada estudo. TOMM40 está próximo de ApoC1 e em desequilíbrio de ligação com ApoE. Pensa-se que interacções entre ApoE 3 ou ApoE 4 e diferentes isoformas de TOMM40 estão associadas com risco aumentado ou diminuído de desenvolver doença de Alzheimer numa gama de idades mais precoce. A idade de início das curvas para genótipos de Apo 4/4, 3/4, 3/3, 2/4 e 2/3 é mostrado na **Figura 2**, indicando uma gama de risco para desenvolvimento precoce da doença, dependendo do perfil de ApoE. O ApoE sozinho não parece, contudo, explicar todos os dados nestas idades das curvas início.

Vários métodos para a criação de perfil polimórfico de risco de doença de Alzheimer associado com os diferentes alelos de ApoE foram propostos (ver, e. g., Aplicação U.S. de Cox et al., N° 20060228728; Aplicação U.S. de Li e Grupe, N° 20080051318). Uma abordagem filogenética para o puzzle do ApoE 4 é aqui demonstrada.

Amostras biológicas, isolamento de ADN, amplificação de loci de interesse. Um total de 340 indivíduos, incluindo 135 casos de doença de Alzheimer e 99 controlos com igual idade no Grupo A bem como 57 casos e 49 controlos no Grupo B. Todos os indivíduos transportaram genótipos de ApoE previamente associados com maior risco para início precoce da doença (i. e., 3/3, 3/4 ou 4/4). Amostras biológicas contendo ADN foram recolhidas de todos os indivíduos. ADN genómico foi então isolado de acordo com métodos convencionais para sequenciação de loci genéticos no Cromossoma 19.

A **Figura 3** mostra as regiões genéticas no Cromossoma 19 destinado para estudo utilizando dados de varrimento de genoma completo de vários relatórios. A região é abrangida dentro da sequência de referência AF050154 do GenBank. Foi utilizado software para gerar múltiplos alinhamentos de sequência para loci variantes (e. g., ClustalW2, European Bioinformatics Institute). Subsequentemente, os múltiplos alinhamentos de sequência foram analisados utilizando software para desenvolver árvore filogenéticas (e. g., MEGA versão 2.1, Center for Evolutionary Functional Genomics, ÁRVOREVOLVE, Departamento de Zoologia, Universidade de Oxford ou software de construção com base em parcimónia, tal como PAUP, Sinauer Associates). Análises estatísticas podem ser realizadas com, e. g., Análise de Dados

Genéticos (GDA: Software para a Análise de Dados Genéticos Discretos, The Bioinformatics Research Center of North Carolina State University). Os resultados da análise da Região B são demonstrados na árvore filogenética da **Figura 4**.

Cada porção de dados na **Figura 4** representa uma sequência variante observada. Estas variantes podem ser substituições nucleotídicas, inserções, deleções ou microssatélites e podem, ou não, resultar em diferenças detectáveis na expressão génica ou função proteica. Cada nó representa uma variante (ou um número de variantes) que ocorre em mais que um cromossoma. Nós adjacentes definem os limites de sequências que estão em *cis* e, por esse motivo, mais propensas a serem herdadas como uma unidade, na região de interesse num cromossoma de um indivíduo. Nós que precedem o maior número de nós subsequentes representam variantes ancestrais evolucionárias, das quais ocorreu divergência genética ao longo do tempo.

A presença de haplótipos ou sequências de variantes correspondendo com regiões da árvore representando indivíduos com substancialmente maior incidência de doença de Alzheimer (*i. e.*, elevadas proporções de indivíduos afectados com a doença para indivíduos de controlo não afectados) significaria que o indivíduo único é também um risco aumentado. Por outro lado, substancialmente baixas proporções correspondem a risco reduzido de desenvolver doença de Alzheimer.

TOMM40 interage com ApoE directamente na regulação da importação da proteína mitocondrial e uma presente hipótese é que a expressão de uma(s) variante(s) particular(es) de TOMM40 exacerb(a)m o relativamente moderado risco de doença de Alzheimer associada com a presença dependente da dose do alelo

de ApoE 3. Uma tal variante de TOMM40 é descoberta dentro da Região B utilizando os métodos da presente invenção.

Testar novos fármacos em indivíduos humanos transporta imenso risco (ver Kenter e Cohen, Lancet, 368: 1387-91 (2006)). A utilização de árvores filogenéticas para antecipar respostas individuais a um fármaco ou tratamento de interesse tem potencial para aliviar, significativamente, esse risco. Estudos preliminares indicaram que rosiglitazona (Avandia) podem ter eficácia específica de perfil genético no tratamento de doença de Alzheimer (ver Risner et al., The Pharmacogenomics Journal 6, 246-254 (2006); Brodbeck et al., Proc. Nat. Acad. Sci. **105**, 1343-6 (2008)). Os dados do ensaio clínico de fase II indicam que doentes de doença de Alzheimer sem um alelo de ApoE 4 responderam melhor a rosiglitazona que doentes que transportam alelos 1 ou 2 de ApoE 4 (dados não mostrados). Isto suporta a hipótese que variantes identificadas com os métodos aqui ensinados podem ser utilizadas para antecipar resposta individual ao tratamento baseado em genótipo.

EXEMPLO 2: Identificação de Variantes de Interesse de TOMM40

174 sequências (2 de cada um dos 87 indivíduos) foram alinhadas utilizando o programa CLUSTAL X (versão 2.0.10, Larkin et al., Clustal W e Clustal X versão 2.0. Bioinformatics, 23:2947-2948 (2007)). O alinhamento de sequências múltiplas foi utilizado para construir uma árvore filogenética utilizando um algoritmo junção de vizinhos (Saitou e Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol.

Evol. Biol., **4**:406–425 (1987)) como implementado no sítio da internet do Instituto Europeu de Bioinformática (EBI).

A árvore filogenética resultante tem uma estrutura de dois grupos principais (A, B) na primeira divergência. As frequências genotípicas de ApoE para estes grupos estão tabuladas e mostradas na **Figura 5**. É claro que o grupo B contém haplótipos de indivíduos de genótipos primários $\epsilon 3/\epsilon 3$ e $\epsilon 3/\epsilon 4$ de ApoE e quase nenhum $\epsilon 4/\epsilon 4$. O grupo A contém quase todos os haplótipos de indivíduos com o genótipo $\epsilon 4/\epsilon 4$.

A lista de polimorfismos gerados pela plataforma de descoberta de SNP (Polimórfica) foi utilizada para identificar variantes específicas no gene TOMM40 que separavam os dados nos dois grupos. Um teste de razão de verosimilhança foi utilizado para identificar variantes significantes com um valor de p menor que 0,005.

A lista de variantes está sumarizada na **Tabela 1**. Na tabela, o termo "delecção" é utilizado quando o menor alelo é uma deleção de um nucleótido e o termo "inserção" é utilizado quando o menor alelo é uma adição de um nucleótido. A expressão "polimorfismo de deleção/inserção" é utilizada quando existem mais que duas possíveis formas e o alelo menor não é aparente. Por exemplo, para os polimorfismos poli-T, existem polimorfismos de múltiplos comprimentos observados. A segunda coluna da tabela proporciona informação sobre as identidades dos alelos específicos associados com a variante que divide as sequências em dois grupos. Por exemplo, T>A indica que o alelo T segregá sequências em grupo "A" na árvore filogenética. Quando dois alelos são listados, e. g., G>B; A>A, cada alelo segregá unicamente os dados de sequência em dois grupos, enquanto que

quando um alelo único é listado, é associado com a separação predominante dos dados, e o restante alelo não separa unicamente os dados num grupo homogéneo, mas ao invés uma mistura de ambos os grupos.

Tabela 1. Variantes de TOMM40 associadas com grupos na árvore filogenética que distribuem pelo genótipo de ApoE.

Variante	Alelo > grupo da árvore	Localização Genómica (NCBI Build 36.3)	Função	Classificação UCSC
50092565	T>A	50092565	Intrão 6	único
50092587	T>A	50092587	Intrão 6	único
Rs8106922	G>B; A>A	50093506	Intrão 6	único
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	50093609	Intrão 6	complexo
Rs34878901	T>B; C>A	50094317	Intrão 6	único
Rs35568738	C>B	50094558	Intrão 6	único
Rs10602329	T16, 17, 18>A T14, 15>B	50094716	Intrão 6	inserção/ delecção
50094733	->A	50094733	Intrão 7	inserção
Rs10524523	T12, 14, 15, 15, 17>B T21, 22, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36>A	50094889	Intrão 6	Inserção/ delecção
Rs1160985	T>B; C>A	50095252	Intrão 6	único
50095506	T>A	50095506	Intrão 6	único
Rs760136	A>A; G>B	50095698	Intrão 6	único
Rs1160984	T>B	50095764	Intrão 6	único
Rs741780	C>B; T>A	59996271	Intrão 8	único
Rs405697	A>A	50096531	Intrão 9	único
50096647 (DIP3)	->A	50096647	Intrão 9	delecção
50096697	C>A	50096697	Intrão 9	único
Rs1038025	C>B; T>A	50096812	Intrão 9	único
Rs1038026	C>B; A>A	50096902	Intrão 9	único
Rs1305062	C>B; G>A	50097361	Intrão 9	único
Rs34215622	G>B; ->A	50098378	Exão 10	inserção
Rs10119	A>A	50098513	Exão 10	único
Rs7259620	G>A; A>B	50099628	desconhecido	único

EXEMPLO 3: Duas formas distintas de ApoE 3: aquelas ligadas a haplótipos de TOMM40 que aumentam o risco e diminuem a idade de início e aqueles que diminuem o risco

A associação de genótipos de apolipoproteína E (ApoE), particularmente ApoE ε4 (ApoE 4), com o risco e idade de início de doença de Alzheimer (AD), continua a ser a associação genética mais confirmada para qualquer doença complexa. Estimativas da hereditariedade de ApoE 4 para início tardio de AD variam de 58% a 79% e o risco atribuível populacional devido ao alelo ApoE 4 está entre 20% e 70%. Esta estimativa sugere que outras variantes genéticas e/ou interacções entre variantes incorrem em risco de doença adicional e alteração da idade de início de distribuições.

Resultados de associação de varrimento de genoma completo para AD têm consistemente reproduzido a extraordinária associação da região de DL contendo ApoE. TOMM40, a proteína translocase da membrana mitocondrial externa, está em elevado DL com ApoE, e codifica para o canal de membrana através do qual atravessam péptidos citoplasmáticos e proteínas, de modo a sintetizar nova mitocôndria. Os objectivos da requerente foram identificar haplótipos adicionais dentro da região em DL que aumenta as estimativas de hereditariedade.

Métodos: A requerente examinou a região em DL contendo ApoE e TOMM40 utilizando sequenciação primária profunda (10x) em doentes com DA e controlos. Realizaram-se análises filogenéticas da região em DL abrangendo TOMM40 e ApoE em 66 doentes e 66 controlos com a mesma idade, em relação ao risco e idade de início de distribuição.

Conclusão: A requerente encontrou que famílias herdadas únicas e distintas de variantes de TOMM40 estão localizadas no mesmo intervalo genómico que ApoE 3, mas não no intervalo genómico contendo ApoE 4 e pode aumentar ou diminuir a idade de distribuição de risco da AD. Por esse motivo, a hereditariedade genética destas variantes de TOMM40 são independentes da hereditariedade de ApoE 4, proporcionando, efectivamente, uma diferenciação de duas formas distintas de ApoE 3: aquelas ligadas a haplótipos de TOMM40 que aumentam o risco e diminuem a idade de inicio, e aquelas que diminuem o risco. Estes dados aumentam a precisão da idade genética de inicio de risco, dependente da idade, genótipos de ApoE e TOMM40 e proporcionam a oportunidade para definir o risco elevado de AD durante os próximos 5-7 anos, *versus* baixo risco de AD.

EXEMPLO 4: Análise de três variantes DIP de TOMM40 identificadas

Três das variantes de TOMM40 identificadas neste pedido são polimorfismos deleção/inserção (DIP) localizados no intrão 6 ou intrão 9. Estes DIP são identificados como rs10524523 e rs10602329 na base de dados de dbSNP do National Center for Biotechnology Information, e um polimorfismo não descrito anteriormente, designado como DIP3. Estes polimorfismos estão localizados em crl9:50094889, crl9:50094731 e crl9:50096647, respectivamente, de acordo com NCBI build 36. Esta invenção descreve a identificação destes DIP utilizando análise filogenética do gene TOMM40, especificamente de um fragmento do gene de 10 Kb, e que os DIP estão associados com diferentes grupos evolucionários determinados por análise filogenética.

Esta invenção divulga, ainda, a utilidade destes DIP para (1) determinar o risco de uma pessoa de desenvolver doença de Alzheimer no futuro, e (2) para prever a idade de início de AD dentro de um espaço de tempo de, aproximadamente, 8 anos.

Os três polimorfismos DIP aqui caracterizados correspondem a diferentes comprimentos de repetições poli-T DIP no gene TOMM40. A associação de variantes poli-T DIP com risco de doença tem precedência. Por exemplo, uma variante poli-T no intrão 8 do gene regulador da condutância transmembranar da fibrose cística (CFTR) está associada com salto do exão 9 e o desenvolvimento da fibrose cística (Groman *et al.*, *Am J Hum Genet* **74**(1): 176–9 (2004)). É aqui divulgado: (1) utilização do novo método – análise de associação filogenética (descrito anteriormente) – para identificar DIP que são preditivos de risco de doença e/ou diferenças da idade de início de doença, (2) a identidade de três DIP específicos associados com diferenças na idade de início de AD e risco de AD, (3) a utilização destes SNP individualmente, em conjunto, ou com outras sequências variantes no TOMM40 ou ApoE para diagnosticar doença ou prever ou determinar características da doença, tal como idade de início da doença, prognóstico da doença, sub-tipos da doença, gravidade da doença, e também para analisar ou determinar a resposta a fármacos.

A análise filogenética revela a distribuição de DIPs rs10524523 e rs10602329 em dois clados diferentes. Esta análise revela que comprimentos poli-T mais pequenos neste mapa de loci para os clados identificadas filogeneticamente no grupo B, o grupo que também compreende maiores percentagens de indivíduos de genótipo ApoE e3/e3, efectivamente poucos (0%) indivíduos ApoE e4/e4 e menores proporções caso/controlo (*i. e.*, risco de

doença AD) (**Figura 5**). A associação entre o comprimento de DIP e grupo filogenético é estatisticamente significativa ($p < 0,0001$) pelo teste da razão de verosimilhança ou teste qui-quadrado de Pearson.

Devido à arquitectura genómica, o elevado desequilíbrio de ligação e as relações evolutivas como indicado na análise filogenética, entre os dois genes, e a putativa interacção física entre os dois produtos génicos, a influencia do genótipo de TOMM40 é provável que se extenda a outras doenças que são influenciadas pelo genótipo de ApoE. Estas doenças incluem, mas não estão limitadas a, doença de Parkinson, esclerose Múltipla, doença cardiovascular, dislipidemia, recuperação de lesão cerebral traumática, recuperação de episódios isquémicos cerebrais, resposta a anestésicos e resposta a fármacos utilizados para tratar AD e as doenças aqui listadas.

Estes polimorfismos podem também ser utilizados em esforços para descobrir fármacos para a pesquisa de compostos úteis para tratar doenças influenciadas por variações na proteína TOMM40 ou ApoE ou variantes genicas.

Além disso, as variantes podem influenciar ou determinar tratamentos baseados em biofarmacêuticos direcionados especificamente, como exemplificado por anticorpos monoclonais e moléculas ARNsi.

Os polimorfismos DIP em TOMM40 que são aqui divulgados podem ser identificados a partir de uma amostra de ADN de um indivíduo utilizando diversas metodologias de análise molecular de nucleótidos diferentes, incluindo, mas não limitado a,

sequenciação de ADN com os iniciadores indicados na **Tabela 4** listada abaixo.

EXEMPLO 5: Extensões de poli-T mais longos em rs10524523 são significativamente correlacionadas com idade de início precoce de LOAD

A análise filogenética tem sido utilizada para identificar relações genómicas entre variante genéticas de baixa frequência e para agrupar haplótipos 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 metodologia foi utilizada para explorar o bloco DL ApoE-TOMM40 para a existência de novas determinantes de risco para LOAD. Num estudo exploratório, 23 Kb de ADN contendo os genes TOMM40 e ApoE foram amplificados e sequenciados e foram determinados haplótipos resolvidos em fase, para 72 casos de LOAD e 60 controlos com a mesma idade (Li et al., Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch Neurol* **65**, 45-53 (2008)). Foi possível construir uma árvore filogenética distinta para 10 Kb, exões 2-10 de codificação, desta região. Dois clados (A e B) foram distinguidos com forte suporte "bootstrap" (98%, 1000 replicados). Houve uma diferença significativa na distribuição dos genótipos de ApoE entre os dois clados de haplótipos de TOMM40 nesta árvore filogenética, sugerindo que esta região pode ser funcionalmente significativa. Ambos os clados continham indivíduos com o genótipo ε3/ε3, mas 98% de todos os haplótipos do clado B ocorreram em *cis* com o alelo ApoE ε3 ($P = 1,2 \times 10^{-18}$, teste exacto de Fisher, de duas caudas).

A estrutura filogenética desta região of TOMM40 de 10 Kb, a hereditariedade específica de ApoE ε3 de haplótipos particulares, e a identificação dos polimorfismos do clado foram subsequentemente confirmados em dois coortes caso/controlo independentes de LOAD, incluindo um coorte com estado AD confirmado por autopsia e dados da idade de início da doença. A associação entre os dois clados e risco de doença e idade de início da doença, em que os dados estão disponíveis, foi também explorado para estes dois coortes. O primeiro coorte (AS) composto por casos de AD ($n = 74$) e controlos ($n = 31$) apurados no Arizona Alzheimer's Disease Research Center (ADRC). O segundo coorte (DS) foi montado no ADRC Duke Bryan e composto apenas por indivíduos ApoE ε3/ε4 (40 casos confirmados por autópsia com conhecida idade de início de doença e 33 controlos) (**Tabela 2**). Embora a sequenciação de ADN tenha sido bem sucedida, para um sub-conjunto do coorte DS que tinha início de doença dos 50 aos 68 anos de idade, análises de associação foram limitadas a um sub-conjunto de doentes que desenvolveram AD depois da idade de 60.

Tabela 2. Composições do coorte. O número de casos e controlos, média de idade e percentagem que são do sexo feminino são mostrados para cada série. A média de idade é dada como idade no diagnóstico para AD para casos e idade no exame para controlos. O desvio padrão da média é dado em parêntesis.

Series	<i>n</i>		Média de Idade (SD)		% do Sexo Feminino	
	Casos	Controlos	Casos	Controlos	Casos	Controlos
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

Uma árvore filogenética de estrutura semelhante àquela gerada no estudo exploratório foi desenvolvida com forte suporte “bootstrap” (97%, 1000 replicados) para o coorte AS. Indivíduos ApoE ϵ 4/ ϵ 4 ocorrem apenas no clado A (98% de separação entre grupos, $P = 2,0 \times 10^{-4}$ teste exacto de Fisher, de duas caudas), enquanto os restantes genótipos de ApoE foram distribuídos entre os clados A e B (**Figura 6**). Ou seja, ApoE ϵ 4 estava sempre em LD com variantes do clado A enquanto ApoE ϵ 3 ocorreu em ambos os haplótipos do clado A e clado B. A análise da distribuição dos poucos indivíduos ApoE ϵ 2/ ϵ 4 na árvore filogenética sugere que haplótipos ApoE ϵ 2-TOMM40 partilham uma história evolutiva semelhante com haplótipos ApoE ϵ 3-TOMM40 (dados não mostrados). Para verificar a estrutura filogenética utilizando um método separado e para garantir que a recombinação dentro do intervalo genético não confunde a estrutura da árvore filogenética desenvolvida para o coorte AS, redes haplotípicas foram também construídas utilizando parcimónia estatística (TCS versão 1.21 (Clement et al., TCS: a computer program to estimate gene genealogies. *Mol Ecol* **9**, 1657-9 (2000))). Os principais grupos indivíduo-haplótipo derivados dos dois métodos (máxima parcimónia e TCS) eram congruentes.

O clado A foi mais frequentemente associada com casos de AD que o clado B (OR = 1,44, 95% IC = 0,76 - 2,70). Heterozigóticos ApoE ϵ 3/ ϵ 4 (n = 36) foram analisados para estimar o risco de doença associado com haplótipos do clado A enquanto controlando o efeito de ApoE ϵ 4. Houve uma maior tendência para maior incidência de LOAD para o sub-conjunto que era homozigótico para TOMM40 do clado A, relativo ao sub-conjunto que era heterozigótico para clado A e clado B (OR = 1,36, 95% IC = 0,40 - 4,61) e, por este motivo, foi postulado que, pelo

menos, algumas das variantes de TOMM40, que definem clado A conferem risco de LOAD independente de ApoE ϵ 4.

A análise dos dados de sequência do coorte AS identificou 39 sítios polimórficos na região de 10 Kb de TOMM40, dos quais houve 30 sítios informativos por parcimónia (pelo menos dois nucleótidos diferentes, cada representado em, pelo menos, duas sequências). Dos 30 sítios informativos por parcimónia, 18 tinham uma frequência de alelo menor (MAF) $> 0,10$ e seis SNP estavam fora do limite do gene TOMM40. 10 SNP ocorreram exclusivamente no contexto de ApoE ϵ 3 ($P = 6,07 \times 10^{-50}$, Teste exacto de Fisher, de duas caudas, $n = 210$) e nunca foram observados em indivíduos homozigóticos ApoE ϵ 4/ ϵ 4 ($n = 16$). A maioria das variantes de TOMM40 específicas para ϵ 3 foi localizada em regiões intrónicas.

A **Figura 7** ilustra os 10 SNP e 6 polimorfismos de inserção/delecção que distinguem os clados A e B de TOMM40 (a $P < 0,001$) para os indivíduos ApoE ϵ 3/ ϵ 3 do coorte AS. Estes polimorfismos foram testados individualmente e como haplótipos para associação com risco de LOAD (**Tabela 3**). A razão de probabilidades para risco de doença para cada alelo do clado B, em todos os casos o menor alelo, sugere que os alelos do clado B são protectivos do risco de AD no coorte AS, contudo, em cada caso, a associação perdeu significância por pouco. Para ter em conta o efeito de ApoE ϵ 4 nas razões de probabilidades apresentadas na **Tabela 3**, foi construído um conjunto equilibrado de 48 casos de AD e 48 controlos de AD por selecção aleatória de sequências de indivíduos ApoE ϵ 3/ ϵ 4 do grupo dos coortes AS e DS. SNP únicos foram, novamente, associados não significativamente com LOAD neste conjunto equilibrado de dados. Contudo, os alelos menores de quatro dos SNP (rs8106922,

rs1160985, rs760136, rs741780) que distinguem TOMM40 do clado B foram analisados anteriormente em três estudos de associação de genoma completo caso/controlo de LOAD e verificou-se serem protectivos de risco de doença ($OR < 1$ em cada caso), que é consistente com a tendência observada no presente estudo (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's Disease in Japanese. *Genomics* **93**, 441–448 (2009)).

Tabela 3. Estatística descritiva e resultados de associação genotípica e alélica para os SNP individuais.

Todos os genótipos ApoE																				
SNP ID	Posição	Alelo	Alelo		MAF (todos)	MAF (casos)	MAF (controlos)	LOAD (M)	LOAD (m)	Controlo (M)	Controlo (m)	LOAD (MM)	LOAD (Mm)	LOAD (mm)	Controlo (MM)	Controlo (Mm)	Controlo (mm)	OR	Cl inferior 95%	Cl superior 95%
			do Clado	B																
rs1038025	50096812	T/c	c	0,31	0,28	0,37	106	41	39	23	40	27	7	1	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	1	17	3	0,66	0,35	1,23	
rs1160985	50095252	C/t	t	0,30	0,28	0,37	106	41	39	23	40	27	7	1	17	3	0,66	0,35	1,23	
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	0,31	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																				
SNP ID	Posição	Alelo	Alelo		MAF (todos)	MAF (casos)	MAF (controlos)	LOAD (M)	LOAD (m)	Controlo (M)	Controlo (m)	LOAD (MM)	LOAD (Mm)	LOAD (mm)	Controlo (MM)	Controlo (Mm)	Controlo (mm)	OR	Cl inferior 95%	Cl superior 95%
			do Clado	B																
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	

Outro polimorfismo que distinguiu os dois clados e, por esse motivo, dois grupos de haplótipos ApoE ϵ 3, foi uma variante poli-T (rs 10524523) localizada no intrão 6 de TOMM40. Em cromossomas de ApoE ϵ 4, a variante era relativamente longa, com uma distribuição de comprimentos estreita, unimodal de (resíduos T 21 - 30, média = 26,78, s.d. = 2,60, n = 32), enquanto que em cromossomas de ApoE ϵ 3, foi evidente uma distribuição de comprimentos com picos de resíduos T a 15,17 (s.d. = 0,85, n = 36) e 33,15 (s.d. = 2,09, n = 55) (**Figura 8**). Comprimentos de poli-T mais longos ($T \geq 27$) segregaram quase exclusivamente em clado A, o clado de maior risco, no coorte AS ($P = 7,6 \times 10^{-46}$, n = 210, teste exacto de Fisher, de duas caudas). A proporção de caso/controlo para a categoria que contem os dois, mais comuns, comprimentos mais curtos (resíduos T 15 ou 16) foi 1,46 (95% IC = 1,25 - 1,75), e a proporção de caso/controlo para a de comprimentos mais compridos (resíduos T 28, 29, 33 e 34) foi 2,02 (95% IC = 1,13 - 2,87). Estes dados mostraram uma tendência para uma associação entre o comprimento de poli-T mais longo rs10524523 e AD (OR = 1,38, 95% IC = 0,80 - 2,39).

Enquanto havia apenas tendências de associação de haplótipos de TOMM40 ou polimorfismos individuais com LOAD para o coorte AS, havia uma associação significativa entre a categoria de comprimento poli-T de rs 10524523 e idade de início de LOAD. Isso foi testado utilizando o coorte DS de indivíduos ApoE ϵ 3/ ϵ 4 confirmados por autópsia para os quais havia dados de início de doença. Alelos poli-T mais longos (≥ 27 resíduos T) foram significativamente associados com início de doença a uma idade muito mais jovem (70,5 anos \pm 1,2 versus 77,6 anos \pm 2,1, $P = 0,02$, n = 34) (Fig. 5).

Este polimorfismo, por esse motivo, teve impacto, significativamente, na idade de início de doença para indivíduos que transportavam um alelo ApoE ϵ 3. Três outros polimorfismos de comprimento poli-T localizados no intrão 6 (rs34896370, rs56290633 e rs10602329) também distinguiram clados A e B, mas estes polimorfismos não estavam associados com idade de início da doença. De um modo semelhante, não houve relação entre haplótipos de SNP distinguidos por clado e idade de LOAD, ou para o SNP único, rs8106922, que tinha sido, significativamente, associado com risco de DA em três estudos de associação de 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's Disease in Japanese. *Genomics* **93**, 441-448 (2009)) (dados não mostrados).

Concluiu-se que extensões de poli-T mais longas em rs 10524523 eram correlacionadas significativamente com idade de início de LOAD precoce. O comprimento desta variante é relativamente homogéneo e relativamente longo, em cromossomos de ApoE ϵ 4, enquanto que existem duas categorias de comprimentos de poli-T ligados a ApoE ϵ 3. Cromossomas de ApoE ϵ 2 também parecem transportar repetições de poli-T de comprimento variável semelhante a cromossomas ϵ 3, mas é necessária investigação adicional para verificar esta conclusão preliminar e para determinar se a repetição de poli-T tem impacto na idade de início da doença muito avançada para portadores de ApoE ϵ 2.

Enquanto é possível que existam outras variantes que influenciam a idade de início de LOAD para indivíduos que não são homozigóticos para ApoE ε4, o comprimento do polimorfismo poli-T no intrão 6 de TOMM40 parece ser o preditor genético mais potente nesta região de ligação e deve ser validado prospectivamente. Estes dados sugerem que a idade estratificada por genótipo de ApoE de curvas de início (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's Disease. *Arch Neurol* **65**, 45-53 (2008)) são, na realidade, grupos de curvas, com cada curva reflectindo uma interacção específica de polimorfismos ligados em ApoE e TOMM40. Por esse motivo, estes dados adicionam resolução à previsão da idade de início de LOAD, dentro de uma janela de 5-7 anos, para indivíduos com mais de 60 anos de idade. O estudo para validar a associação de genótipos de ApoE e haplótipos de TOMM40 ou rs 10524523 com a idade de início da doença a ser actualmente planeada. Este estudo será um estudo prospectivo, de 5 anos, de base populacional conduzido em diversos grupos étnicos e será combinado com uma prevenção ou atraso de início da doença do ensaio do fármaco.

MÉTODOS

Os dois coortes analisados neste estudo foram do Arizona Alzheimer's Disease Research Center (ADRC), Phoenix, Arizona e do Duke Bryan ADRC, Durham, Carolina do Norte. Todos os indivíduos eram de descendência Europeia. Os estudos do Arizona e Duke foram aprovados por conselhos de revisão institucionais e foram obtidos consentimentos informados apropriados de todos os

participantes. Os dados da idade e género para os casos e controlos em cada coorte são mostrados na **Tabela 2**. Para o coorte Duke, a idade de início da doença foi determinada retrospectivamente e o diagnóstico da doença foi confirmado por autópsia.

As amostras foram plaqueadas em placas de 96 poços para PCR de longo alcance e sequenciação de ADN em Polymorphic DNA Technologies (Alameda, CA).

O PCR de longo alcance foi realizado utilizando Polimerase Takara LA Taq (Takara Mirus Bio). A mistura de reacção e condições de PCR foram as mesmas que as recomendadas pelo fabricante. O PCR foi conduzido num volume de 50 µL com 2,5 U de LA Taq e 200–400 ng de ADN genómico humano. A termociclagem foi realizada com as seguintes condições: 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. Iniciadores para PCR de longo alcance são mostrados na **Tabela 4**.

Tabela 4. Iniciadores de sequenciação directos e reversos estão listados. A linha sombreada indica os iniciadores directo e reverso utilizados para PCR de longo alcance de R2 (Figura 2).

Iniciadores Directos				Iniciadores Reversos			
Sequência	Coordenada do Iniciador UCSC (da extremidade 3' do iniciador)	Posição do Iniciador em Produto de PCR Clonado (da extremidade 3' do iniciador)	SEQ ID Nº	Sequência	Coordenada do Iniciador UCSC (da extremidade 3' do iniciador)	Posição do Iniciador em Produto de PCR Clonado (da extremidade 3' do iniciador)	SEQ ID Nº
CAGGAACAGCTATGAC	50092292	243	1	CCCCACTGGTTGTGA	50093034	2456	2
GTTGTGATGGTATTCAAC	50093038	634	3	GAATAGGGGCTTCA	50093282	878	6
CTGCAGGTATGAAAG	50093287	883	5	CAATCTCTAGGTGTC	50093512	1108	8
GTCTCTCGAGATGTG	50093601	1197	7	CGGAAGTTGCAGTAG	50093706	1302	10
TACTGCAACTTCCGC	50093722	1318	9	AAGGTCAAGGTTACACT	50094318	1914	12
TCTCTGTGCCACG	50094289	1885	11	ACAACCTCTAGGTACAT	50094790	2386	14
CCCAAVTAATTTGTATTG	50094609	2205	13	CCTGTAATCCCAGCTAT	50095002	2598	16
ACATTGTGGCCTGTAC	50095129	2725	15	TCATCTCTGTGGAACCTAA	50095324	2920	18
CCCATGGCTTGTTG	50095603	3199	17	GGCAAAATGACGATCAGT	50095804	3400	20
CCCAGATGCCAAATC	50096082	3678	19	GCAGCACCAAGCTAGT	50096218	3814	22
AACTCTGAGTGGATGTG	50096471	4067	21	GATGGGTCTAACCTCCTTA	50096620	4216	24
CTATAGTCCCAAATCTGA	50096730	4326	23	TTTTTCCAAGCATAAAACATAGTA	50096863	4459	26
AGTCCCGTACTTA	50097080	4676	25	GGGGATGGACAAAGCT	50097268	4864	28
ACCACAGGTATGCC	50097451	5047	27	TGAAAAGCCCTCTAGAC	50097898	5494	30
GAACAGATTCCATCCGA	50097864	5460	29	CACCCACGATCCAGTT	50098141	5737	32
TGTGGATACCAACTGGAT	50098148	5744	31	CAAGCCACACTGAAACTT	50098231	5827	34
GGGATCTGAGTACCA	50098469	6065	33	CAGAATCCCTGCGT	50098526	6122	36
TGCTGCCCTAACGTCG	50098937	6533	35	ACACTTGAGAAAACGG	50098797	6393	38
CTGGGGTCAGCTGAT	50099350	6946	37	ACAAAGTCTCTATAGCC	50098077	6673	40
TGAAACATCTGGGATTATAAC	50099679	7275	39	TAACCTGGGGTTGGTT	50099429	7025	42
CTGGAACCAACAATACC	50099990	7586	41	AAGTTCTTGCTATCAG	50099829	7425	44
ATCTCGCTCACTGTG	50100261	7857	43	GCAAGAGGGAGACTGT	50100207	7803	46
GTCAAAAGACCTATGTC	50100739	8335	45	TGTGCCTGGATGAATGTA	50100567	8163	48
AGGACTCCACGAGT	50101197	8793	47	TGAGCTCATCCCCGT	50100960	8556	50
			49	CCGTGTTCATTATGAG	50101328	8924	51
				GTAAAACGACGGCCAG	50101681	9277	52

Os produtos de PCR foram corridos num gel de agarose a 0,8%, visualizado por corante violeta de cristal comparados com padrões de tamanho, cortados do gel e extraídos com materiais de purificação incluídos no kit de Clonagem TOPO XL PCR (Invitrogen). Os produtos do PCR de longo alcance foram clonados num vector de clonagem TOPO XL PCR. Este sistema utiliza um vector de clonagem TA e é recomendado para inserções de até 10 kb. Segundo as instruções do fabricante, células electrocompetentes (do mesmo kit) foram transformadas pelo vector, plaqueadas na presença de antibiótico e incubadas. Dez clones de cada placa foram escolhidos e cultivados num formato de 96 poços.

Foram transferidas culturas diluídas para um tampão de desnaturação que fazia parte do kit de Amplificação de Sequenciação de ADN TempliPhi (GE HealthCare/Amersham Biosciences). Este tampão provoca a libertação de ADN plasmídico mas não ADN bacteriano. As culturas foram aquecidas, arrefecidas, centrifugadas e transferidas para placas contendo a enzima TempliPhi e outros componentes. Esta mistura foi incubada, a 30 °C, durante 18 horas para promover a amplificação dos modelos de plasmídeos. Estes produtos foram então centrifugados e aquecidos até 65 °C para destruir a enzima.

Foram utilizados modelos de plasmídeos em reacções de sequenciação de ADN utilizando o Corante Big, versão 3.1 do kit de sequenciação (Applied Biosystems). Para cada reacção, foi utilizado um iniciador de sequenciação apropriado (**Tabela 4**) que foi concebido para emparelhar com uma única localização do modelo. A sequenciação do ciclo foi realizada com uma temperatura de emparelhamento de 50 °C, uma temperatura de elongamento de 60 °C, e uma temperatura de desnaturação de

96 °C, durante um total de 30 ciclos. Os produtos de reacção de sequenciação foram corridos num sequenciador de ADN ABI 3730XL com uma matriz capilar de 50 cm, utilizando o modo de corrida padrão.

Um programa de análise de sequenciação de propriedade chamado "Agent" (desenvolvido por Celera) foi utilizado para alinhar leituras de sequenciação à sequência de referência apropriada e produzir 'contíguos' associados com cada clone. O sistema proporciona índices de qualidade estimados para todas as bases para as quais existe qualquer variação para qualquer das amostras. O relatório da sequenciação para cada amostra foi analisado para a presença de SNP que foram correlacionados num haplótipo padrão para um sub-conjunto de clones e num diferente haplótipo padrão para os restantes clones. Um arquivo de referência para a região de interesse foi preparado por listagem das variações conhecidas para aquela região, publicamente disponíveis de dbSNP NCBI. Um arquivo de genótipo para a região de interesse foi criada para pesquisar cada relatório do haplótipo de cada indivíduo para todas as variações entre a sequência de referência conhecida e as sequências haplotípicas consensos.

A magnitude do erro de leitura de comprimento para as variantes poli-T (e. g., rs10524523) foi estimada por exame da variação dos comprimentos observados dos 10 clones que foram preparados para amostras que tinham um único haplótipo. Para uma amostra típica com comprimento curto de poli-T de 16, o desvio padrão para os 10 clones foi 0,97. Para uma amostra típica com comprimento longo de poli-T, e. g., 27, o desvio padrão foi 1,58.

A análise filogenética foi conduzida. Um alinhamento de sequências múltiplas das sequências foi realizado utilizando o programa ClustalW2 (versão 2.0.10), utilizando parâmetros padrão. O ajustamento manual dos alinhamentos foi concluído utilizando Genedoc (versão 2.7.000). Foram construídas árvores filogenéticas utilizando reconstruções Bayesianas, máxima verosimilhança e baseadas na distância. O software de construção da árvore filogenética utilizado foi Paup* (versão 4.0b10), ClustalX2 (métodos de junção de vizinhos, versão 2.0.10) e Mr. Bayes (versão 3.1.2).

Foi utilizada bissecção de árvore e reconexão de troca de em todos os métodos. O melhor modelo de ajustamento da evolução da sequência foi estimado utilizando o programa Modeltest (versão 3.7) que proporcionou estimativas para os seguintes factores chave: matriz de taxa, forma da distribuição gama e proporção de sítios não variáveis. Foi realizada análise “bootstrap” utilizando 1000 replicados para determinar suporte estatístico para a morfologia da árvore específica.

Redes haplotípicas foram também construídas a partir dos dados de sequência utilizando o programa TCS (versão 1.21 (Clement et al., TCS: a computer program to estimate gene genealogies. *Mol Ecol* **9**, 1657–9 (2000))) para comparar as árvores filogenéticas com cladogramas estimados utilizando parcimónia estatística. As árvores filogenéticas e redes haplotípicas foram construídas duas vezes, com lacunas tratadas como dados perdidos para a primeira instância e como um quinto carácter para a segunda instância. A diversidade nucleotídica na região de interesse foi calculada utilizando DnaSP (versão 5.00.02 (Librado et al., DnaSP v5: a software for

comprehensive analysis of ADN polymorphism data. *Bioinformatics* **25**, 1451–2 (2009))).

Após a construção das árvores filogenéticas, a rede haplotípica e conclusão da análise da diversidade nucleotídica na região de interesse, os resultados dos diferentes métodos foram comparados e reconciliados com uma árvore consenso. Foi presumido que grupos de sequências partilhando uma mutação da doença recente segregavam mais proximamente na árvore filogenética, contudo, casos esporádicos devido a fenocópias, dominância e epistasia, pode introduzir interferência na relação fenótipo-haplótipo (Tachmazidou *et al.*, Genetic association mapping via evolution-based clustering of haplótipos. *PLoS Genet* **3**, e111 (2007)).

Contudo, casos esporádicos devido a fenocópias, dominância e epistasia pode introduzir interferência na relação fenótipo-haplótipo. Esta análise filogenética focou-se numa agregação de alto nível de clados, de modo a minimizar estes efeitos. Os clados determinados na primeira divisão na árvore filogenética foram utilizados para testar a hipótese que haplótipos de indivíduos de TOMM40 do clado "B" foram associados com início de AD numa idade mais avançada que haplótipos de indivíduos do clado "A", (cada indivíduo contribuiu com dois haplótipos para a idade de início do sinal de associação de AD). O número de testes de associação que são realizados utilizando esta abordagem foi ordens de magnitude menores que em estudos de associação de genoma completo típicos, dado que a análise filogenética identificou categorias de haplótipos de indivíduos relacionados evolutivamente. Se os testes de associação confirmaram que os diferentes clados classificaram os dados de haplótipo de indivíduos por idade de início, análise estatística

adicional foi realizada para identificar as variantes que separaram as sequências em cada clado. Efectivamente, esta análise confirmou a significância de cada variante como um factor que influencia a idade de início utilizando uma série de testes de um grau de liberdade guiados pela estrutura da árvore. As análises filogenéticas foram realizadas utilizando nucleótidos únicos e polimorfismos de inserção/deleção. Os testes estatísticos de associação foram ajustados com uma correcção de Bonferroni para o número de sítios polimórficos incluídos na análise.

Relatórios de haplótipo do software de análise Polimórfica e relatórios do software DnaSP (versão 5.00.02 (Librado et al., DnaSP v5: a software for comprehensive analysis of ADN polymorphism data. *Bioinformatics* **25**, 1451-2 (2009))) foram utilizados para análises estatísticas subsequentes. Foram analisadas variantes de TOMM40 SNP individuais, haplótipos de TOMM40 e comprimento de repetições de poli-T para associação com risco de LOAD para o coorte AS e idade de início de LOAD para o coorte DS. Diferenças nas proporções de alelos específicos de TOMM40 associados com cada alelo de ApoE ou genótipo de ApoE foram comparados utilizando o teste exacto de Fisher (de duas caudas). Começando com 30 sítios informativos por parcimónia e $\alpha = 0,05$, uma correcção de Bonferroni para a significância de uma associação alélica específica iria requerer um valor de P de 0,001. Foram calculadas razões de probabilidade (OR) como o (numero de alelos menores em casos/numero de alelos menores em controlos)/numero de alelos maiores em casos/numero de alelos maiores em controlos) e referidos com intervalo de confiança de 95%. Médias para idade de LOAD definida de grupos de início foram comparadas por testes *t*, de duas caudas. Um teste F padrão em variações de grupo foi realizado para determinar se o teste *t*

foi calculado assumindo variâncias iguais ou diferentes. A análise estatística foi concluída utilizando software JMP (versão 8, SAS Institute, Cary, NC).

Códigos de Acesso: GenBank: TOMM40, translocase da membrana mitocondrial externa 40 homologa, 10452; ApoE, apolipoproteína E, 348

O acima mencionado é ilustrativo da presente invenção e não deve ser interpretado como limitativo da mesma. A invenção é definida pelas seguintes reivindicações, com equivalentes das reivindicações a ser aqui incluídas.

LISTAGEM DE SEQUÊNCIAS

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

<120> MÉTODO PARA IDENTIFICAR FACTORES DE RISCO DE DOENÇA

<130> 9719-2WO

<150> US 61/224,647
<151> 2009-07-10

<150> US 61/186,673
<151> 2009-06-12

<150> US 61/088,203
<151> 2008-08-12

<160> 52

<170> PatentIn versão 3.3

<210> 1

<211> 25

<212> ADN

<213> Artificial

<220>

<223> Iniciador de PCR de longo alcance

<400> 1

aactcagagg ccagagattc taagt 25

<210> 2

<211> 25

<212> ADN

<213> Artificial

<220>

<223> Iniciador de PCR de longo alcance

<400> 2

aacagcctaa tcccagcaca ttac 25

<210> 3

<211> 17

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 3

cagggaaacag ctatgac 17

<210> 4

<211> 15

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 4

cccaactgggtt gttga 15

<210> 5

<211> 18

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 5

gtgtgatgggt gattcaac 18

<210> 6

<211> 16

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 6

gaataggggc ctttca 16

<210> 7

<211> 15

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 7

ctgcaggtat gaaag 15

<210> 8

<211> 16

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 8

caatctccata gggtgc 16

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

<220>
<223> Iniciador de sequenciação

<400> 9
gtctctgcag atgtg 15

<210> 10
<211> 16
<212> ADN
<213> Artificial

<220>
<223> Iniciador de sequenciação

<400> 10
cggaaagtgc agtaag 16

<210> 11
<211> 15
<212> ADN
<213> Artificial

<220>
<223> Iniciador de sequenciação

<400> 11
tactgcaact tccgc 15

<210> 12
<211> 17
<212> ADN
<213> Artificial

<220>
<223> Iniciador de sequenciação

<400> 12
aaggtcaagg ttacact 17

<210> 13
<211> 15
<212> ADN
<213> Artificial

<220>
<223> Iniciador de sequenciação

<400> 13
tctctgttgc ccacg 15

<210> 14
<211> 17
<212> ADN
<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 14

acaaggcttag gtgacat 17

<210> 15

<211> 21

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 15

cccaactaat ttttgtattc g 21

<210> 16

<211> 17

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 16

cctgttatcc cagctat 17

<210> 17

<211> 17

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 17

acatttgtgg cctgtac 17

<210> 18

<211> 20

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 18

tcatctctgtgaaacctaa 20

<210> 19

<211> 16

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 19

ccacatgggc ttgtgt 16

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

<220>
<223> Iniciador de sequenciação

<400> 20
ggcaaaatga cgatcagt 18

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

<220>
<223> Iniciador de sequenciação

<400> 21
cccagatgcc caaatc 16

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

<220>
<223> Iniciador de sequenciação

<400> 22
gcagcaccag ctagt 15

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

<220>
<223> Iniciador de sequenciação

<400> 23
aactctgagt ggatgtg 17

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

<220>
<223> Iniciador de sequenciação

<400> 24
gatggctca atccctta 19

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

<220>

<223> Iniciador de sequenciação

<400> 25

ctatagtc
ccc aactactga 19

<210> 26

<211> 25

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 26

ttttccaa gcataaaaca tagta 25

<210> 27

<211> 15

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 27

agtccccgct actta 15

<210> 28

<211> 16

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 28

ggggatggac aaagct 16

<210> 29

<211> 16

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 29

accacaggtg tatgcc 16

<210> 30

<211> 17

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 30

tgaaaagccc tctagac 17

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

<220>
<223> Iniciador de sequenciação

<400> 31
gaacagattc atccgca 17

<210> 32
<211> 16
<212> ADN
<213> Artificial

<220>
<223> Iniciador de sequenciação

<400> 32
cacccacgat ccagtt 16

<210> 33
<211> 18
<212> ADN
<213> Artificial

<220>
<223> Iniciador de sequenciação

<400> 33
tgtggatagc aactggat 18

<210> 34
<211> 19
<212> ADN
<213> Artificial

<220>
<223> Iniciador de sequenciação

<400> 34
caaagccaca ctgaaaactt 19

<210> 35
<211> 16
<212> ADN
<213> Artificial

<220>
<223> Iniciador de sequenciação

<400> 35
gggattctga gtagca 16

<210> 36
<211> 13
<212> ADN
<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 36

cagaatccctg cgt 13

<210> 37

<211> 16

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 37

tgctgccta agtccg 16

<210> 38

<211> 16

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 38

acacttgaga aaacgg 16

<210> 39

<211> 15

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 39

ctgggtcag ctgat 15

<210> 40

<211> 18

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 40

acaaagtccct ctatagcc 18

<210> 41

<211> 22

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 41

tgaaacatct gggattata ac 22

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

<220>
<223> Iniciador de sequenciação

<400> 42

taacctgggg ttggtt 16

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

<220>
<223> Iniciador de sequenciação

<400> 43
ctggaaaacca caataacc 17

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

<220>
<223> Iniciador de sequenciação

<400> 44

aagttccctt gctcatcag 19

<210> 45

<211> 19

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 45

atctcggttc actgtt 16

<210> 46

<211> 16

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 46

gcaagaggga gactgt 16

<210> 47

<211> 18

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 47

gtcaaaagac ctctatgc 18

<210> 48

<211> 18

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 48

tgtgcctgga tgaatgta 18

<210> 49

<211> 14

<212> ADN

<213> Artificial

<220>

<223> Iniciador de sequenciação

<400> 49

aggactccac gagt 14

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

<220>
<223> Iniciador de sequenciação

<400> 50
tgagctcatc cccgt 15

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

<220>
<223> Iniciador de sequenciação

<400> 51
ccgtgttcca ttatgag 18

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

<220>

<223> Iniciador de sequenciação

<400> 52

gtaaaacgac ggccag 16

Lisboa, 15 de Maio de 2014

REIVINDICAÇÕES

1. Método para determinar o risco de desenvolvimento da doença de Alzheimer num indivíduo, compreendendo:
 - (a) detectar, a partir de uma amostra biológica contendo ADN retirado do referido indivíduo, a presença ou ausência de uma variante genética do gene TOMM40 associado com risco aumentado ou diminuído da doença de Alzheimer, em que a referida variante é um polimorfismo de deleção/inserção no intrão 6 ou intrão 9 do gene TOMM40; e
 - (b) determinar se o referido indivíduo está em risco aumentado ou diminuído da Doença de Alzheimer quando a referida variante genética está presente ou ausente.
2. Método da reivindicação 1, compreendendo, ainda, detectar se o referido indivíduo é um indivíduo Apo E2/E2, E2/E3, E2/E4, E3/E3, E3/E4, ou E4/E4.
3. Método da reivindicação 1, compreendendo ainda detectar se o referido indivíduo é um indivíduo Apo E3/E3 ou E3/E4.
4. Quantidade eficaz de tratamento de um agente activo anti-doença de Alzheimer seleccionado do grupo consistindo de inibidores da acetilcolinesterase, antagonistas do receptor de NMDA, agonistas ou moduladores de receptor activado por proliferadores de peroxissoma (PPAR), anticorpos, proteínas de fusão, moléculas de ARN terapêuticas e suas combinações, para utilização num método

de tratamento da doença de Alzheimer, caracterizada pelo facto do método compreender o passo de determinar se um indivíduo está ou não em risco de desenvolver doença de Alzheimer, em que o risco é determinado de acordo com o método de qualquer uma das reivindicações 1-3.

5. Quantidade eficaz de tratamento de um agente activo anti-doença de Alzheimer para utilização como na reivindicação 4, em que o referido agente é um agonista ou modulador de receptor activado por proliferador de peroxissoma.
6. Quantidade eficaz de tratamento de um agente activo anti-doença de Alzheimer para utilização como na reivindicação 5, em que o agonista ou modulador é uma tiazolidinediona.
7. Quantidade eficaz de tratamento de um agente activo anti-doença de Alzheimer para utilização como na reivindicação 5, em que o agonista ou modulador é um fármaco tiazolidinediona.
8. Quantidade eficaz de tratamento de um agente activo anti-doença de Alzheimer para utilização em qualquer uma das reivindicações 4-7, em que o referido tratamento é para atrasar o início da doença de Alzheimer.
9. Quantidade eficaz de tratamento de um agente activo anti-doença de Alzheimer para utilização como na reivindicação 4 ou reivindicação 5, em que a referida variante genética do gene TOMM40 é uma variante de inserção

poli-T na localização genómica 50094889 de NCBI Build 36.3 (rs10524523).

10. Quantidade eficaz de tratamento de um agente activo anti-doença de Alzheimer para utilização como na reivindicação 6 ou reivindicação 7, em que a referida variante genética do gene TOMM40 é uma variante de inserção poli-T na localização genómica 50094889 de NCBI Build 36.3 (rs10524523).
11. Quantidade eficaz de tratamento de um agente activo anti-doença de Alzheimer para utilização de acordo com qualquer das reivindicações 4-8, em que o referido polimorfismo de inserção/delecção (DIP) é um polimorfismo de inserção/delecção poli-T.
12. Quantidade eficaz de tratamento de um agente activo anti-doença de Alzheimer para utilização de acordo com qualquer das reivindicações 4-8, em que a referida variante genética do gene TOMM40 é rs10524523.
13. Quantidade eficaz de tratamento de um agente activo anti-doença de Alzheimer para utilização de acordo com a reivindicação 4 ou reivindicação 5, em que o referido DIP é um polimorfismo de inserção/delecção poli-T tendo um poli-T de entre 20 e 50 pares de base contíguos.
14. Quantidade eficaz de tratamento de um agente activo anti-doença de Alzheimer para utilização de acordo com a reivindicação 6 ou reivindicação 7, em que o referido DIP é um polimorfismo de inserção/delecção poli-T tendo um poli-T de entre 20 e 50 pares de base contíguos em rs10524523.

15. Quantidade eficaz de tratamento de um agente activo anti-doença de Alzheimer para utilização de acordo com a reivindicação 10, em que o referido tratamento comprehende atrasar o inicio da doença de Alzheimer.
16. Método para determinar um prognóstico ou o risco de desenvolver doença de Alzheimer num doente comprehendendo obter um perfil do doente, em que a referida obtenção de um perfil do doente comprehende:

detectar a presença ou ausência de, pelo menos, um alelo ApoE 2, ApoE 3 ou ApoE 4 numa amostra biológica do referido doente, e

detectar a presença ou ausência de, pelo menos, um polimorfismo de deleção/inserção (DIP) de TOMM40 localizado no intrão 6 ou intrão 9 do gene TOMM40, e depois converter o referido perfil do doente no referido prognóstico, em que a presença do referido alelo ApoE 2, ApoE 3 ou ApoE 4 e a presença do referido pelo menos um polimorfismo DIP de TOMM40 identifica o referido doente como um doente em risco de desenvolver doença de Alzheimer.
17. Método da reivindicação 16, em que o referido DIP é um polimorfismo de inserção/deleção poli-T.
18. Método da reivindicação 16, em que o referido DIP é rs10524523.

19. Método da reivindicação 16, em que a variante genética do gene TOMM40 é o polimorfismo de inserção/delecção rs10524523 tendo um poli-T de entre 20 e 50 pares de base contíguos.

20. Método da reivindicação 16, compreendendo ainda:

detectar, a partir de uma amostra biológica contendo ADN retirado do referido indivíduo, a presença ou ausência de uma variante genética do gene TOMM40 associado com risco aumentado ou diminuído de doença de Alzheimer, em que a referida variante é um polimorfismo de deleção/inserção no intrão 6 ou intrão 9 do gene TOMM40; e

determinar o referido indivíduo está em risco aumentado ou diminuído da doença de Alzheimer quando a referida variante genética está presente ou ausente.

Lisboa, 15 de Maio de 2014

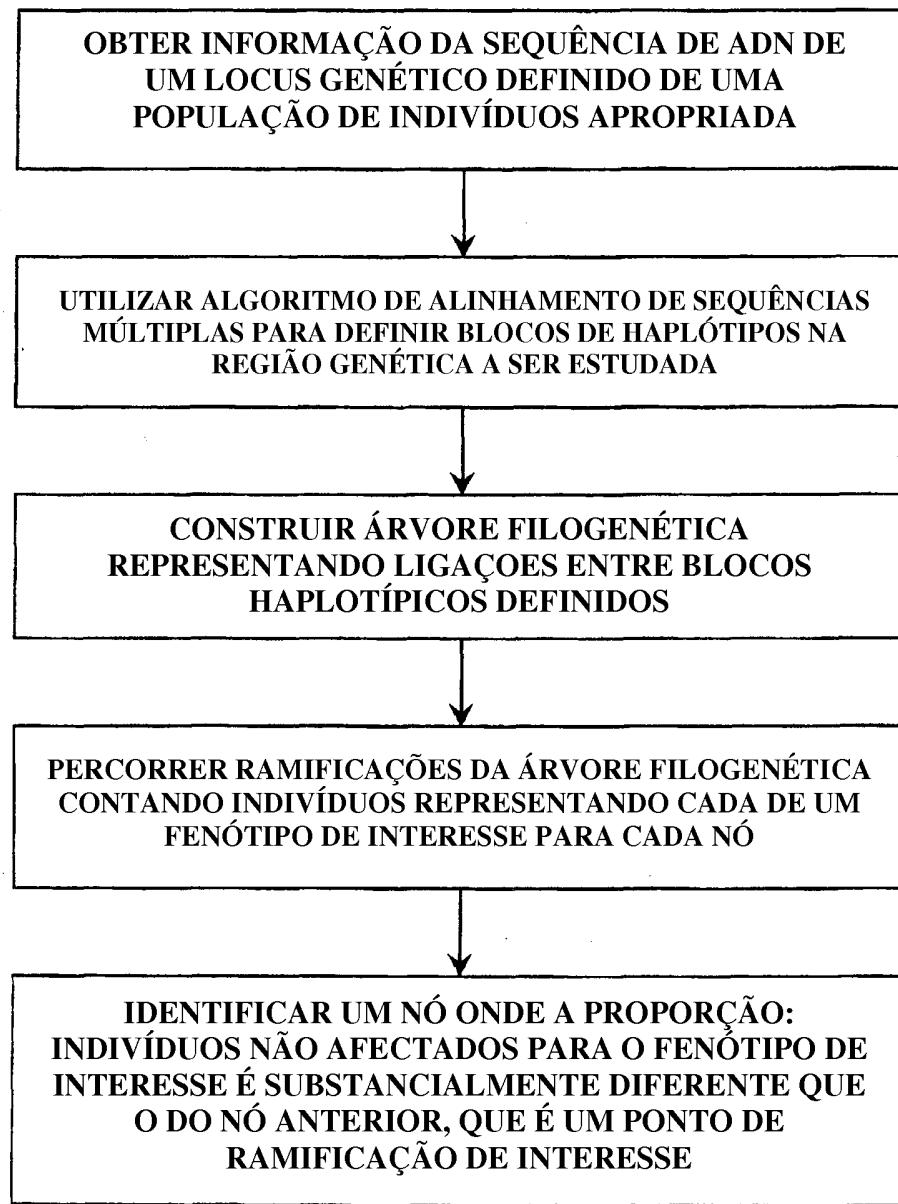
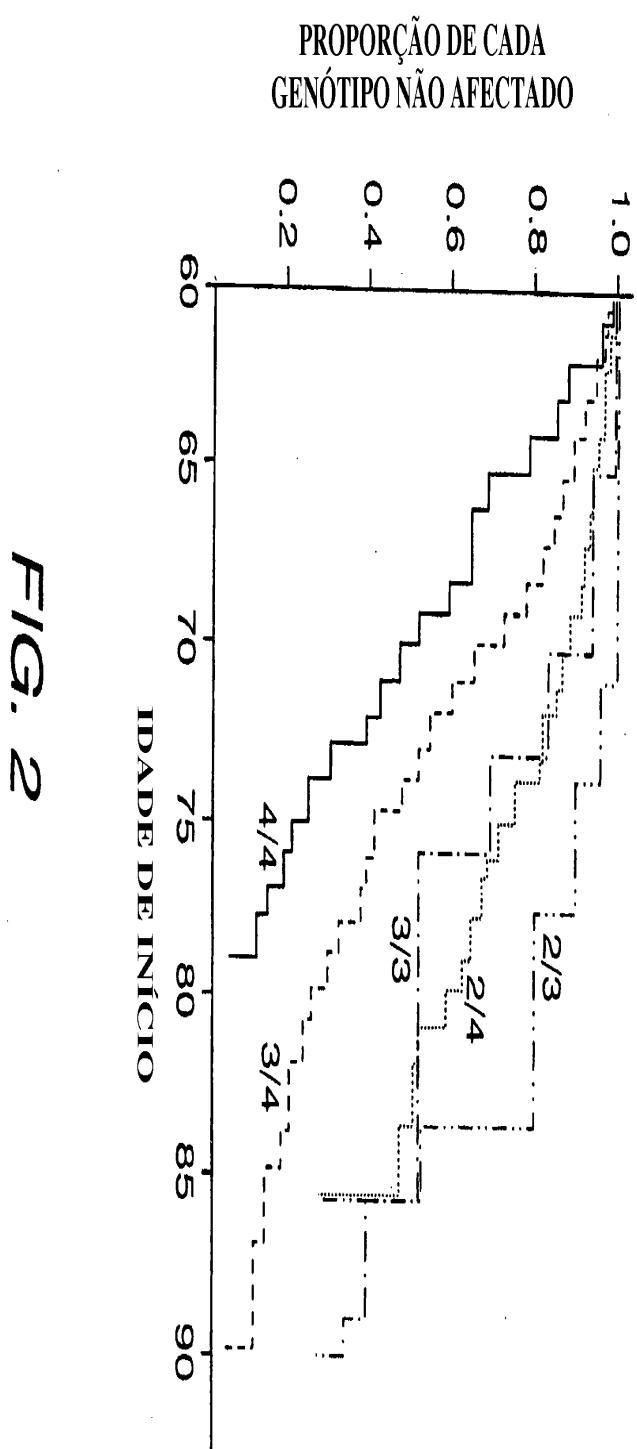


FIG. 1



CR 19

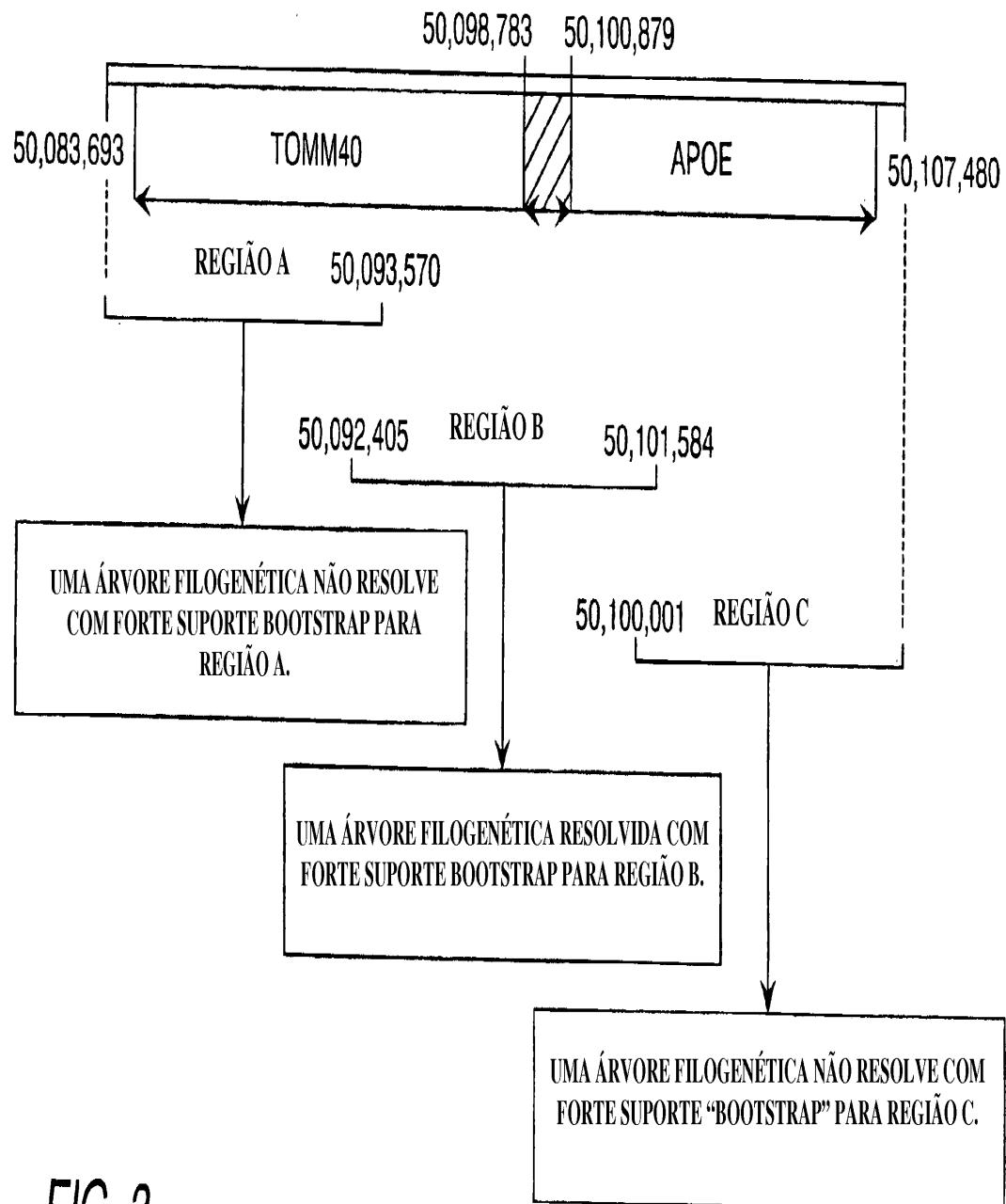


FIG. 3

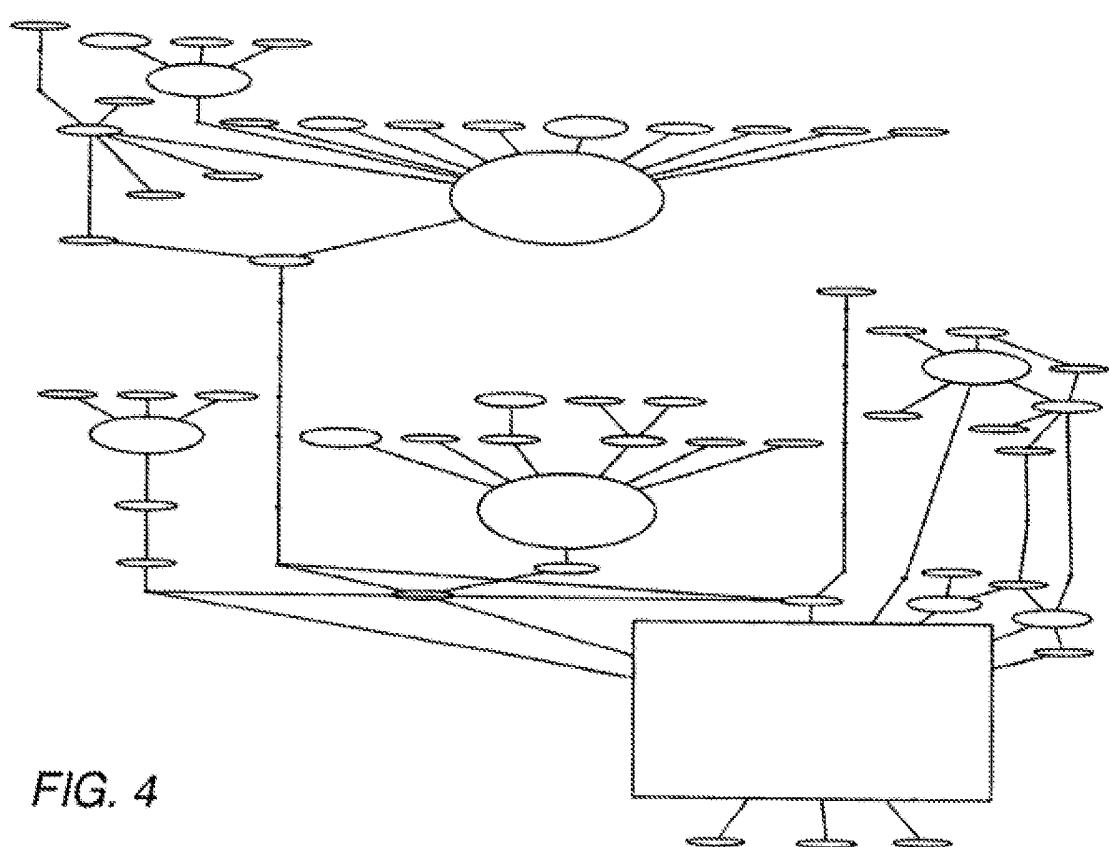


FIG. 4

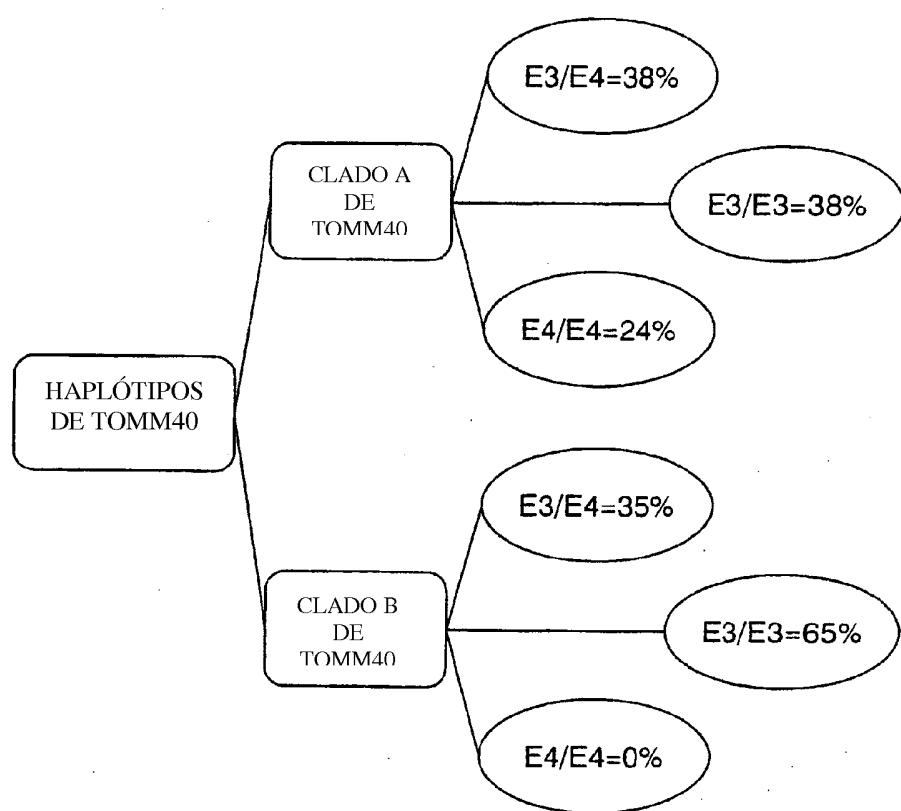
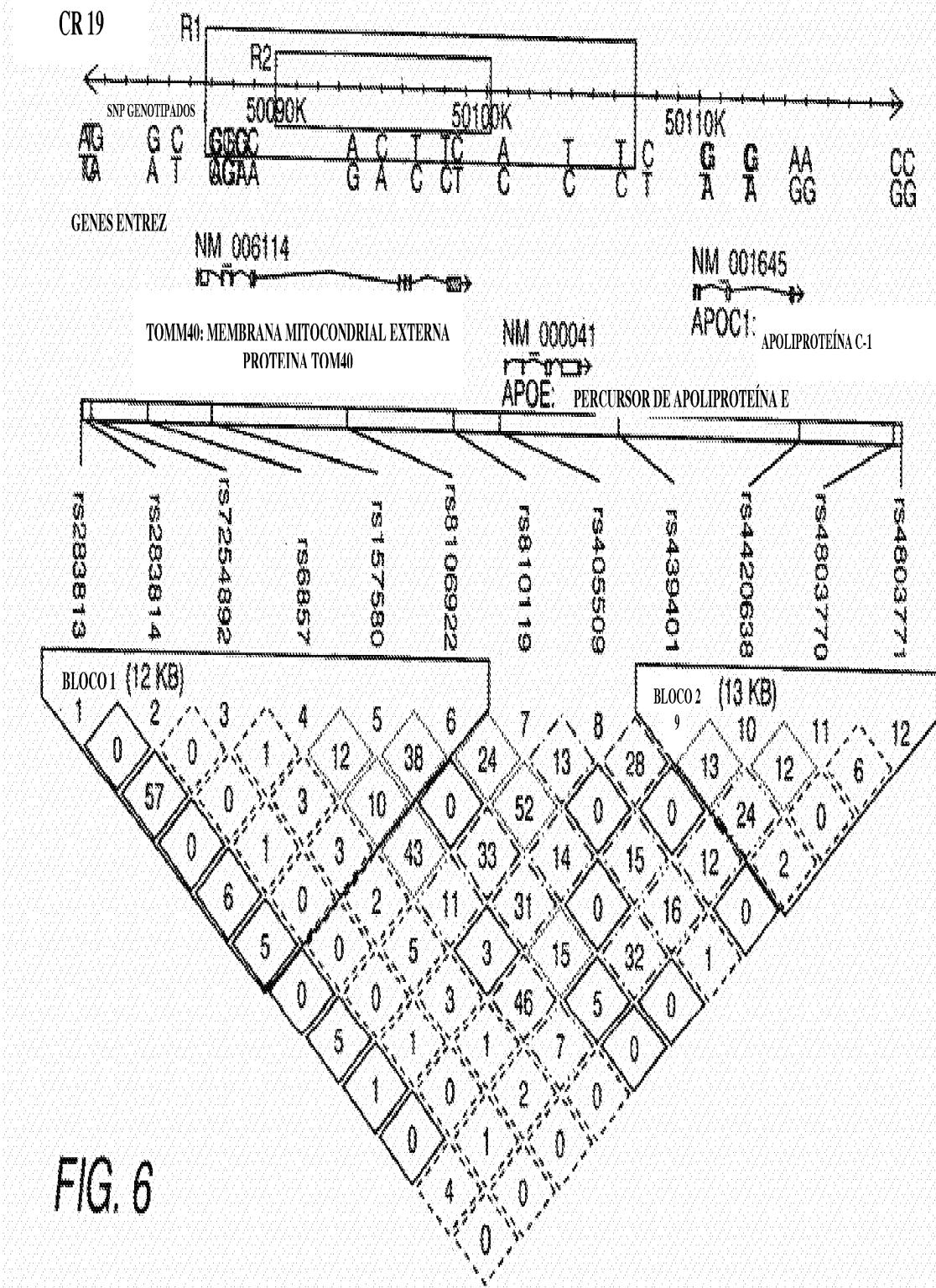
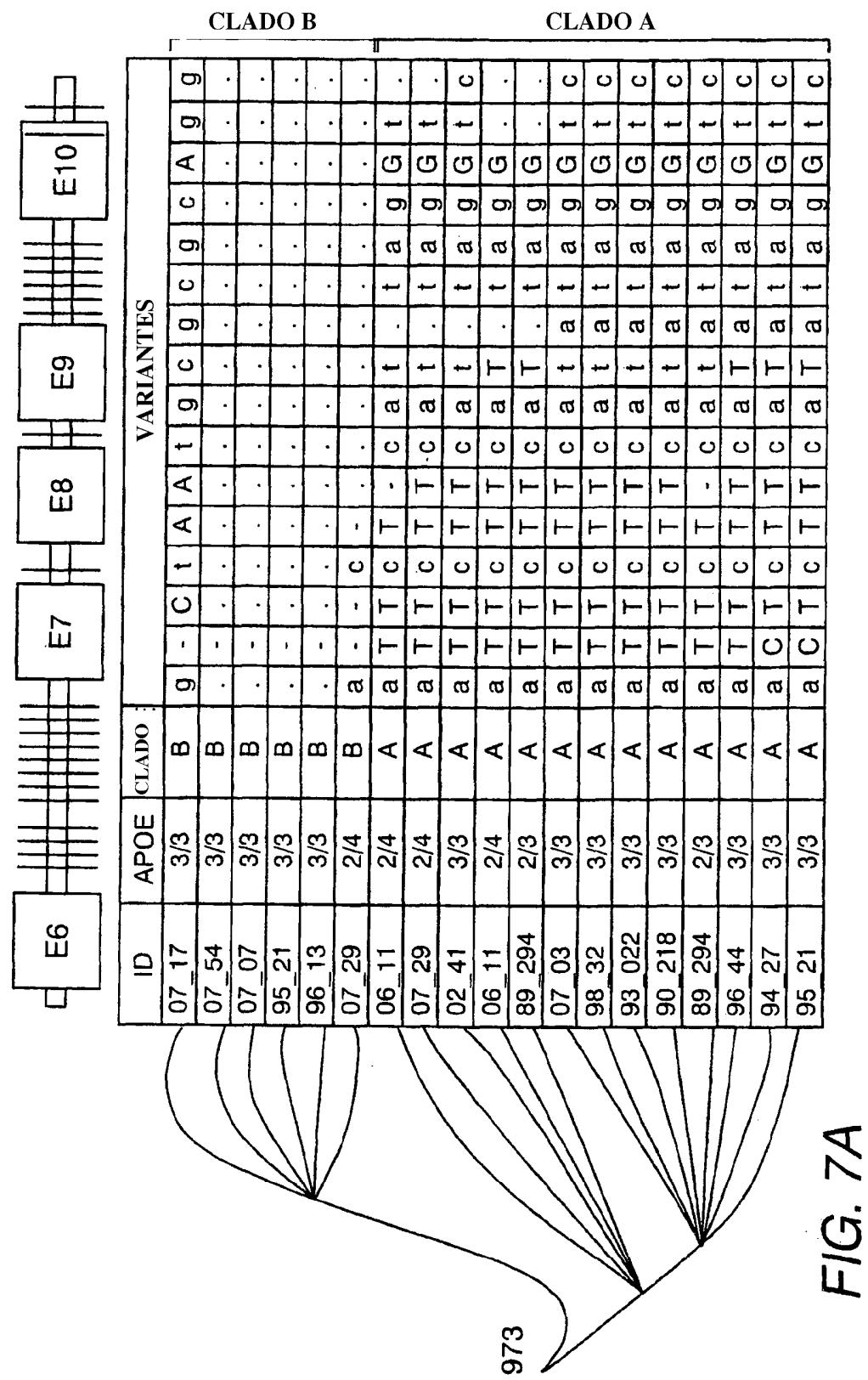
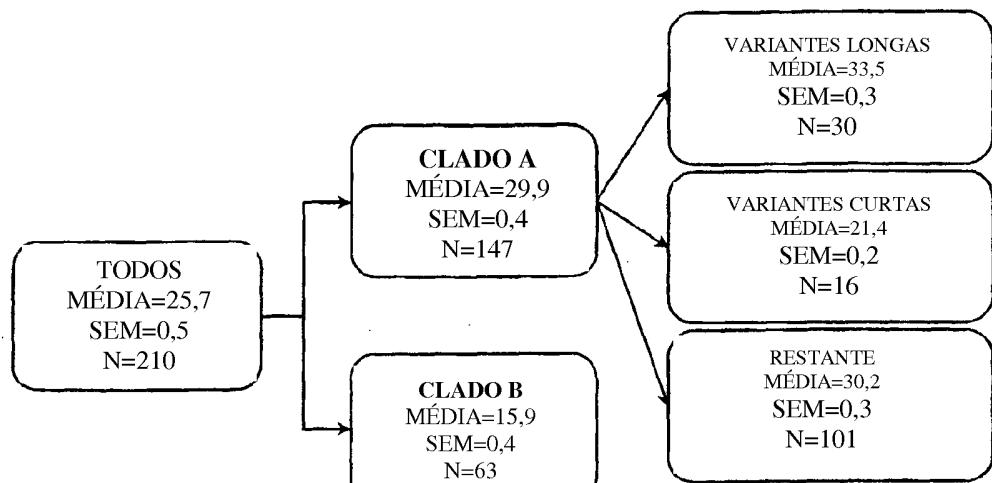
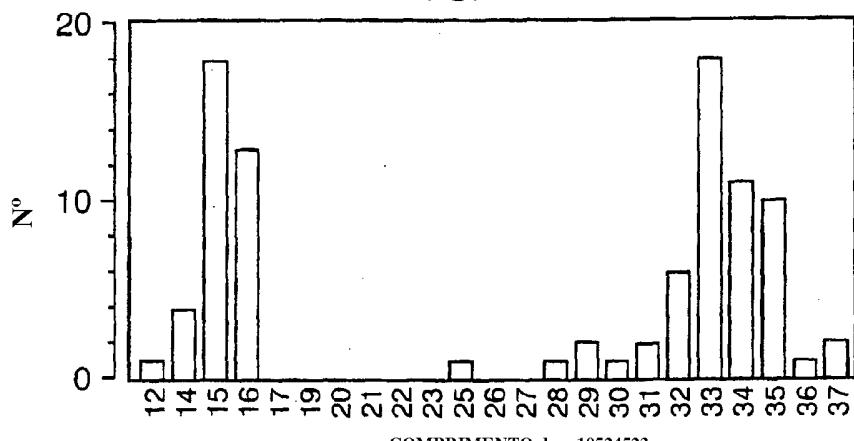
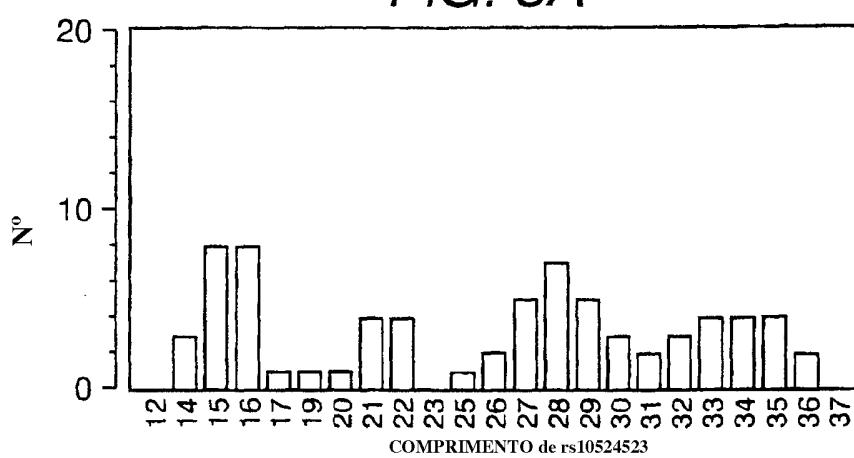
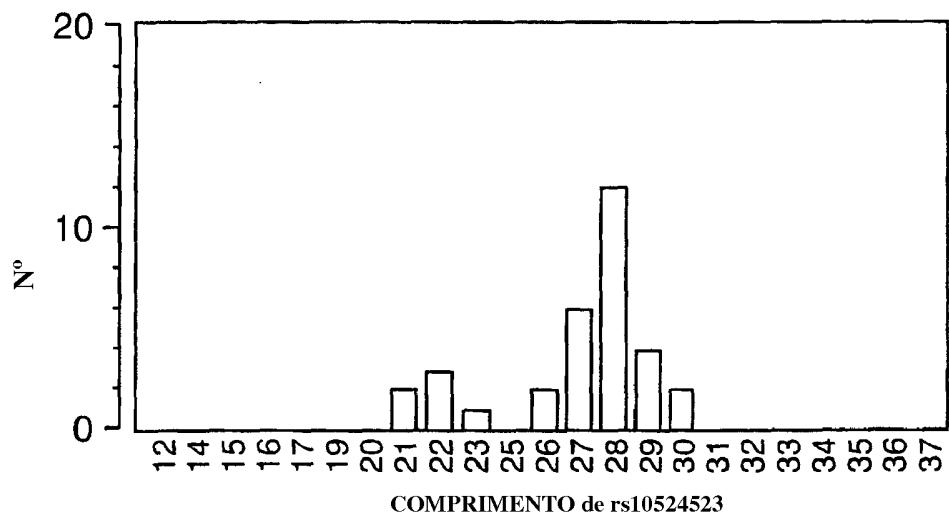
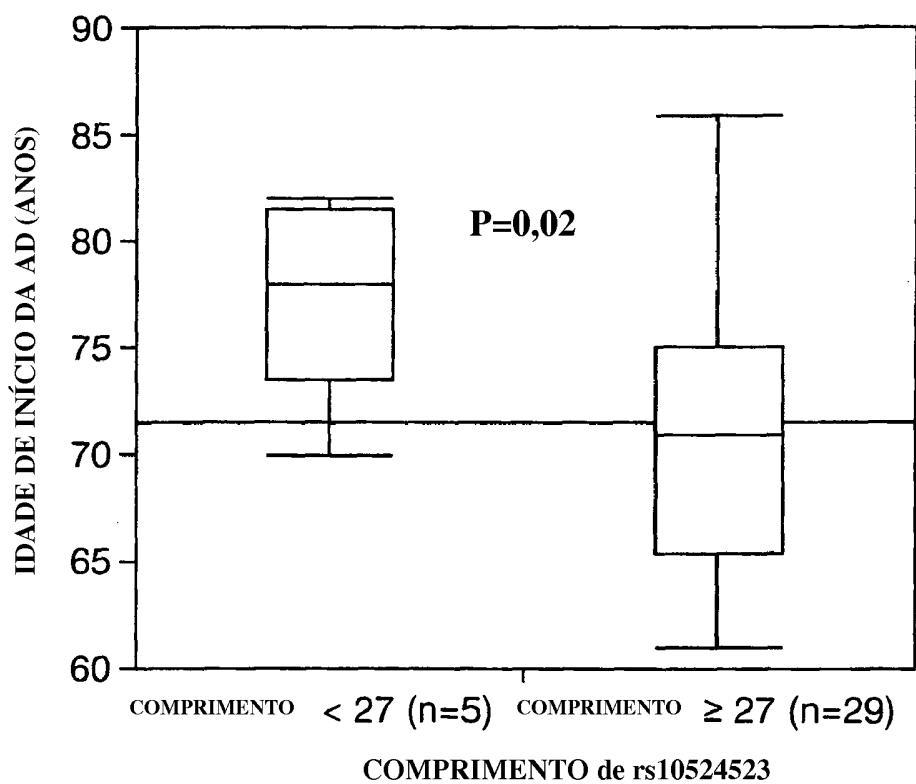


FIG. 5





**FIG. 7B****FIG. 8A****FIG. 8B**

**FIG. 8C****FIG. 9**