



República Federativa do Brasil
Ministério da Indústria, Comércio Exterior
e Serviços
Instituto Nacional da Propriedade Industrial

(11) PI 0720116-8 B1

(22) Data do Depósito: 05/12/2007

(45) Data de Concessão: 02/05/2018



* B R P I 0 7 2 0 1 1 6 B 1 *

(54) Título: MÉTODO PARA DETERMINAR UMA SUSCETIBILIDADE À ARRITMIA CARDÍACA OU ACIDENTE VASCULAR CEREBRAL EM UM INDIVÍDUO HUMANO, USO DE UMA Sonda DE OLIGONUCLEOTÍDEO, E, APARELHO PARA DETERMINAR UM INDICADOR GENÉTICO PARA A ARRITMIA CARDÍACA E/OU ACIDENTE VASCULAR CEREBRAL EM UM INDIVÍDUO HUMANO

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

(30) Prioridade Unionista: 05/12/2006 IS 8576, 29/06/2007 IS 8658

(73) Titular(es): DECODE GENETICS EHF

(72) Inventor(es): ANNA HELGADOTTIR; DANIEL GUDBJARTSSON

“MÉTODO PARA DETERMINAR UMA SUSCETIBILIDADE A ARRITMIA CARDÍACA OU ACIDENTE VASCULAR CEREBRAL EM UM INDIVÍDUO HUMANO, E, APARELHO PARA DETERMINAR UM INDICADOR GENÉTICO PARA A ARRITMIA CARDÍACA E/OU ACIDENTE VASCULAR CEREBRAL EM UM INDIVÍDUO HUMANO”

FUNDAMENTOS DA INVENÇÃO

A arritmia cardíaca é um grupo de condições médicas, em que a atividade elétrica do coração é irregular ou é mais lenta ou mais rápida do que o normal. Algumas arritmias são ameaçadoras da vida e podem causar parada cardíaca ou morte súbita. Outras causam ou predisõem a outros sintomas ou doença agravantes, incluindo o acidente vascular cerebral. A fibrilação é uma forma séria de arritmia, em que o músculo cardíaco se apresenta com movimento irregular ou tremor devido à falta de unidade na função das células contráteis. A fibrilação pode afetar o átrio (Fibrilação atrial (AF) ou Flúter atrial (AFI)) ou o ventrículo (Fibrilação Ventricular (VF)).

A fibrilação atrial (AF) é um ritmo cardíaco anormal (arritmia cardíaca) que envolve as duas câmaras cardíacas pequenas, superiores (os átrios). As batidas cardíacas em um coração normal começam depois que a eletricidade gerada nos átrios pelo nodo sinusal se espalha através do coração e causa a contração do músculo cardíaco e bombeamento do sangue.

Na AF, os impulsos elétricos regulares do nodo sinoatrial são substituídos pelos impulsos elétricos desorganizados, rápidos que resultam em batimentos cardíacos irregulares.

A fibrilação atrial é a arritmia cardíaca mais comum. O risco de desenvolver a fibrilação atrial aumenta com a idade - a AF afeta quatro por cento de indivíduos nos anos 80. Um indivíduo pode espontaneamente alternar entre AF e um ritmo normal (fibrilação atrial paroxismal) ou pode continuar com AF como o ritmo cardíaco dominante sem reversão para o ritmo normal (fibrilação atrial crônica). A fibrilação atrial é frequentemente assintomática, mas pode resultar em sintomas de palpitações, desmaio, dor no peito ou mesmo insuficiência cardíaca. Estes sintomas são especialmente comuns quando a fibrilação atrial resulta em uma frequência cardíaca que é muito rápida ou muito lenta. Além disso, o movimento errático dos átrios leva à estagnação do sangue (estase) que aumenta o risco de coágulos sanguíneos que podem viajar do coração para o cérebro e outras áreas. Assim, a AF é um fator de risco importante para o acidente vascular cerebral, a complicação mais temida da fibrilação atrial.

Os sintomas da fibrilação atrial podem ser tratados com medicações que diminuem a frequência cardíaca. Diversas medicações assim como a cardioversão elétrica podem ser usados para converter a AF em um ritmo cardíaco normal. As terapias com base em cirurgia e cateter também podem ser usadas para prevenir a fibrilação atrial em certos indivíduos. Pessoas com AF são frequentemente administradas afinadores do sangue tais como varfarina para protegê-los de acidentes vasculares cerebrais.

Qualquer paciente com 2 ou mais episódios identificados de fibrilação atrial é dito ter fibrilação atrial recorrente. Isto é ainda classificado em paroxísmico e persistente com base em quando o episódio termina sem terapia. A fibrilação atrial é dita ser paroxísmica quando ela termina espontaneamente dentro de 7 dias, mais habitualmente dentro de 24 horas. A

fibrilação atrial persistente ou crônica é a AF estabelecida por mais do que sete dias. A diferenciação de AF paroxísmica de crônica ou estabelecida está fundamentada na história de episódios recorrentes e da duração do episódio corrente de AF (Levy S., J Cardiovasc Electrophysiol. 8 Supl, S78-82 (1998)).

5 A fibrilação atrial isolada (LAF) é definida como a fibrilação atrial na ausência de descobertas clínicas ou ecocardiográficas de doença cardiopulmonar.

 A fibrilação atrial é usualmente acompanhada por sintomas relacionados com a frequência cardíaca rápida ou embolização. As
10 frequências cardíacas rápidas e irregulares podem ser percebidas como palpitações, intolerância a exercício e ocasionalmente produz angina e sintomas congestivos de encurtamento da respiração ou edema. Algumas vezes a arritmia será identificada com o início de um acidente vascular cerebral ou um ataque isquêmico transitório (TIA). Não é incomum
15 identificar a fibrilação atrial em um exame físico de rotina ou eletrocardiograma (ECG/EKG), visto que a mesma pode ser assintomática em alguns casos. A fibrilação atrial paroxísmica é a ocorrência episódica da arritmia e pode ser difícil de diagnosticar. Os episódios podem ocorrer com o sono ou com exercícios e a sua natureza episódica pode requerer
20 monitoramento de ECG prolongados (por exemplo, um monitor Holter) para diagnóstico.

 A fibrilação atrial é diagnosticada em um eletrocardiograma, uma investigação realizada rotineiramente sempre que a batida cardíaca irregular é suspeitada. As descobertas características incluem a ausência de
25 ondas P, atividade elétrica desorganizada em seu lugar e irregularidade do intervalo R-R devido à condução irregular de impulsos para os ventrículos. Se AF paroxísmica é suspeitada, episódios podem ser documentados com o uso de monitoramento de Holter (registros de ECG contínuos por 24 horas ou mais longos).

Embora muitos casos de AF não têm nenhuma causa definida, estes podem ser o resultado de vários outros problemas (ver abaixo). Por este motivo, a função renal e eletrólitos são rotineiramente determinados, assim como hormônio que estimula a tireóide e uma contagem sanguínea. Um raio X do tórax é em uma maneira geral realizado. Na AF de início agudo associada com dor no peito, troponinas cardíacas ou outros marcadores de dano ao músculo cardíaco podem ser receitados. Estudos de coagulação (INR/aPTT) são usualmente realizados, visto que medicação anticoagulante pode ser começada. Um ecocardiograma transesofágico pode ser indicado para identificar qualquer trombo intracardíaco (Fuster V., *et al.*, Circulation.; 104, 2118-2150 (2001)).

Flúter atrial (AFI) é caracterizado por um ritmo cardíaco rápido anormal nos átrios. Os pacientes que se apresentam com flúter atrial habitualmente também experienciam fibrilação atrial e vice versa (Waldo, A., Progr Cardiovasc Disease, 48: 41-56 (2005)). Mecanística e biologicamente, AF e AFI são assim prováveis de serem altamente relacionados.

AF (e AFI) é ligada a diversas causas cardíacas, mas podem ocorrer de outro modo em corações normais. As associações conhecidas incluem: Pressão sanguínea alta, Estenose mitral (por exemplo, devido à doença cardíaca reumática ou prolapso de válvula mitral), Regurgitação mitral, Cirurgia cardíaca, Doença da artéria coronária, Cardiomiopatia hipertrófica, Consumo de álcool excessivo (“consumo desenfreado de álcool” ou “coração em festa”), Hipertireoidismo, Hiperestimulação do nervo vago, usualmente por ter refeições grandes (“compulsão por comida”), Patologia pulmonar (tal como pneumonia, câncer pulmonar, embolismo pulmonar, Sarcoidose), Pericardite, Distúrbio emocional intenso e Doença cardíaca congênita.

O sistema de condução elétrica normal do coração permite que o impulso que é gerado pelo nodo sinoatrial (nodo SA) do coração seja

propagado e estimula o miocárdio (músculo do coração). Quando o miocárdio é estimulado, este se contrai. É a estimulação ordenada do miocárdio que permite a contração eficiente do coração, permitindo deste modo que o sangue seja bombeado para o corpo. Na fibrilação atrial, os impulsos regulares produzidos pelo nó sinusal para fornecer a contração rítmica do coração são subjugados pelas descargas rápidas aleatoriamente geradas produzidas pelas áreas maiores do tecido atrial. Um impulso elétrico organizado no átrio produz a contração atrial; a falta de um tal impulso, como na fibrilação atrial, produz fluxo sanguíneo estagnado, especialmente no apêndice atrial e predispõe à coagulação. O deslocamento de um coágulo do átrio resulta em um êmbolo e o dano produzido está relacionado a onde a circulação o leva. Um êmbolo para o cérebro produz a complicação mais temida da fibrilação atrial, o acidente vascular cerebral, enquanto que um êmbolo também pode alojar-se na circulação mesentérica (a circulação que abastece os órgãos abdominais) ou dedo, produzindo o dano específico de órgão.

O tratamento da fibrilação atrial é direcionado por dois objetivos principais: (i) prevenir a instabilidade circulatória temporária; (ii) prevenir o acidente vascular cerebral. Os métodos mais comuns para se obter o primeiro inclui o controle da frequência e ritmo, enquanto que a anticoagulação é usualmente o método desejado para o último (Prystowsky E. N., *Am J Cardiol.*; 85, 3D-11D (2000); van Walraven C, *et al.*, *Jama.* 288, 2441-2448 (2002)). Os métodos comuns para o controle da frequência, isto é, para reduzir a frequência cardíaca ao normal, incluem bloqueadores beta (por exemplo, metoprolol), glicosídeos cardíacos (por exemplo, digoxina) e bloqueadores do canal de cálcio (por exemplo, verapamil). Todas estas medicações trabalham pela diminuição da geração de pulsos a partir dos átrios e da condução a partir dos átrios para os ventrículos. Outros medicamentos habitualmente usados incluem quinidina, flecainida, propafenona, disopiramida, sotalol e amiodarona. O controle do ritmo pode ser obtido pela

cardioversão elétrica, isto é, pela aplicação de choque elétrico DC ou pela cardioversão química, usando medicamentos tais como amiodarona, propafenona e flecainida.

Medidas preventivas para o acidente vascular cerebral incluem anticoagulantes. Os exemplos representativos de agentes anticoagulantes são Dalteparina (por exemplo, Fragmin), Danaparoid (por exemplo, Orgaran), Enoxaparina (por exemplo, Lovenox), Heparina (várias), Tinzaparina (por exemplo, Innohep), Varfarina (por exemplo, Coumadin). Alguns pacientes com fibrilação atrial isolada são algumas vezes tratados com aspirina ou clopidogrel. Existe evidência de que a aspirina e o clopidogrel sejam eficazes quando usados juntos, mas a combinação é ainda inferior à varfarina (Connolly S., *et al.* Lancet.; 367, 1903-1912 (2006)).(2) O novo anticoagulante ximelagatran mostrou prevenir o acidente vascular cerebral com eficácia igual como a varfarina, sem o processo de monitoração difícil associado com a varfarina e possivelmente com menos eventos hemorrágicos adversos. Infelizmente, ximegalatran e outros medicamentos anticoagulantes similares (habitualmente aludido como inibidores diretos de trombina), têm que ser ainda amplamente licenciados.

Determinar quem deve receber anti-coagulação ou não com varfarina não está óbvio. A contagem de CHADS2 é o melhor método validado de determinar o risco de acidente vascular cerebral (e portanto quem deve ser anticoagulado). A diretrizes da UK NICE têm ao contrário optado por um método de algoritmo. O problema subjacente é que se um paciente tem um risco anual de acidente vascular cerebral que é menor do que 2%, então os riscos associados com o uso de varfarina sobrepuja o risco de se ter um acidente vascular cerebral (Gage B. F. *et al.* Stroke 29, 1083-1091 (1998))

A fibrilação atrial pode algumas vezes ser controlada com tratamento. A tendência natural da fibrilação atrial, entretanto, é tornar-se uma condição crônica. A AF crônica leva a um risco aumentado de morte.

Pacientes com fibrilação atrial estão em chance significativamente aumentada de acidente vascular cerebral.

A fibrilação atrial é comum entre adultos mais velhos. Nos países desenvolvidos, o número de pacientes com fibrilação atrial provavelmente deve aumentar durante os 50 anos seguintes, devido à proporção crescente de indivíduos idosos (Go A.S. *et al.*, *Jama.*, 285, 2370-2375 (2001))(3). No estudo de Framingham o risco durante a vida para o desenvolvimento de AF é de 1 em 4 para homens e mulheres de 40 anos de idade e mais velhos. Os riscos durante a vida quanto a AF são altos (1 em 6). DE acordo com os dados do National Hospital Discharge Survey (1996-2001) em casos que incluíram AF como um diagnóstico de alta primário descoberto que 45% dos pacientes são do sexo masculino e que a idade média para os homens foi de 66,8 anos e 74,6 para as mulheres. A relação racial para as admissões foi descoberto ser de 71,2% de brancos, 5,6% de negros, 2% de outras raças e 20% não especificado. Além disso, os pacientes afro-americanos foram, em média, muito mais jovens do que as outras raças. A incidência nos homens variou de 20,58/100.000 pessoas por ano para as idades dos pacientes de 15 a 44 anos para 1203/100.000 pessoas por ano para aquelas idades de 85 e mais velhos. De 1996 a 2001, as hospitalizações com AF como o primeiro diagnóstico listado, aumentaram em 34%.

O acidente vascular cerebral é uma doença comum e séria. A cada ano nos Estados Unidos mais do que 600.000 indivíduos sofrem de um acidente vascular cerebral e mais do que 160.000 morrem de causas relacionadas com o acidente vascular cerebral (Sacco, R. L. *et al.*, *Stroke* 28, 1507-17 (1997)). Além disso, mais de 300.000 indivíduos se apresentam com Ataque Isquêmico Transitório, uma forma branda de acidente vascular cerebral, a cada ano nos US. Nos países ocidentais o acidente vascular cerebral brando é a causa principal de incapacidade grave e a terceira causa principal de morte (Bonita, R., *Lancet* 339, 342-4 (1992)). O risco por toda a

vida daqueles que atingem a idade de 40 anos excede 10%.

O fenótipo clínico de acidente vascular cerebral é complexo mas é amplamente dividido em isquêmico (responsável por 80 a 90%) e acidente vascular cerebral hemorrágico (10 a 20%) (Caplan, L.R. Caplan's
 5 Stroke: A Clinical Approach, 1-556 (Butterworth-Heinemann, 2000)). O acidente vascular cerebral isquêmico é ainda subdividido em doença oclusiva de vaso grande (aqui aludida como acidente vascular cerebral da carótida), usualmente devido ao envolvimento aterosclerótico das artérias carótidas comuns e internas, doença oclusivas de vaso pequeno, considerada ser um
 10 estreitamento não aterosclerótico de artérias finais pequenas dentro do cérebro e o acidente vascular cerebral cardiogênico devido aos coágulos sanguíneos que surgem do coração usualmente no cenário da doença cardíaca da fibrilação atrial ou isquêmica (aterosclerótica) (Adams, H. P., Jr. *et al.*, Stroke
 24, 35-41 (1993)). Portanto, parece que o acidente vascular cerebral não é
 15 uma doença mas um grupo heterogêneo de distúrbios que refletem diferenças nos mecanismos patogênicos (Alberts, M. J. Genetics of Cerebrovascular Disease, 386 (Futura Publishing Company, Inc., Nova Iorque, 1999); Hassan, A. & Markus, H. S. Brain 123, 1784-812 (2000)). Entretanto, todas as formas de acidente vascular cerebral compartilham fatores de risco tais como
 20 hipertensão, diabetes, hiperlipidemia e o ato de fumar (Sacco, R. L. *et al.*, Stroke 28, 1507-17 (1997); Leys, D. *et al.*, J. Neurol. 249, 507-17 (2002)). A história familiar de acidente vascular cerebral é também um fator de risco independente que sugere a existência de fatores genéticos que podem interagir com fatores ambientais (Hassan, A. & Markus, H. S. Brain 123, 1784-812
 25 (2000); Brass, L. M. & Alberts, M. J. Baillieres Clin. Neurol. 4, 221-45 (1995)).

Os determinantes genéticos das formas comuns de acidente vascular cerebral são ainda amplamente desconhecidos. Existem exemplos de mutações em genes específicos que causam formas Mendelianas raras de

acidente vascular cerebral tais como o gene Notch3 em CADASIL (arteriopatia dominante autossômica cerebral com infartações subcorticais e leucoencefalopatia) (Tournier-Lasserre, E. *et al.*, Nat. Genet. 3, 256-9 (1993); Joutel, A. *et al.*, Nature 383, 707-10 (1996)), A Cistatina C no tipo islandês de hemorragia cerebral hereditária com amiloidose (Palsdottir, A. *et al.*, Lancet 2, 603-4 (1988)), APP no tipo holandês de hemorragia cerebral hereditária (Levy, E. *et al.*, Science 248, 1124-6 (1990)) e o gene KRIT1 em pacientes com angioma cavernoso hereditário (Gunel, M. *et al.*, Proc. Natl. Acad. Sci. USA 92, 6620-4 (1995); Sahoo, T. *et al.*, Hum. Mol. Genet. 8, 2325-33 (1999)). Nenhuma destas formas raras de acidente vascular cerebral ocorre no cenário da aterosclerose e portanto, os genes correspondentes não são prováveis de desempenhar papéis nas formas comuns de acidente vascular cerebral que mais frequentemente ocorrem com a aterosclerose.

É muito importante para o sistema de cuidado de saúde desenvolver estratégias para prevenir o acidente vascular cerebral. Uma vez que um acidente vascular cerebral acontece, a morte celular irreversível ocorre em uma porção significativa do cérebro abastecido pelos vasos sanguíneos afetados pelo acidente vascular cerebral. Infelizmente, os neurônios que morrem não podem ser revividos ou substituídos a partir de uma população de célula tronco. Portanto, existe uma necessidade para prevenir os acidentes vasculares cerebrais de acontecer em primeiro lugar. Embora já conheçamos certos fatores de risco clínicos que aumentam o risco de acidente vascular cerebral (listados acima), existe uma necessidade médica não atingida para definir os fatores genéticos envolvidos no acidente vascular cerebral para definir mais precisamente o risco de acidente vascular cerebral. Além disso, se alelos predispostos são comuns na população geral e a especificidade de prognosticar uma doença com base na sua presença é baixa, locais adicionais tais como locais protetores são necessários para o prognóstico significativo da disposição do estado de doença. Existe também

uma necessidade enorme quanto a agentes terapêuticos para prevenir o primeiro acidente vascular cerebral ou acidentes vasculares cerebrais adicionais em indivíduos que sofreram de um acidente vascular cerebral anterior ou ataque isquêmico transitório.

5 A AF é um fator de risco independente para o acidente vascular cerebral, aumentando o risco em cerca de 5 vezes. O risco para o acidente vascular cerebral atribuível ao AF aumenta com a idade. A AF é responsável por cerca de 15 a 20% de todos os acidentes vasculares cerebrais. A AF também é um fator de risco independente para a recorrência do acidente vascular cerebral e severidade do acidente vascular cerebral. Um relato recente mostrou pessoas que tiveram AF e não foram tratados com anticoagulantes tiveram um aumento de 2,1 vezes no risco para acidente vascular cerebral recorrente e um aumento de 2,4 vezes no risco para acidente vascular cerebral grave recorrente. Pessoas que têm acidente vascular cerebral causado pela AF foram relatados como 2,23 vezes mais prováveis de estar acamados comparado com aqueles que tiveram acidentes vasculares cerebrais a partir de outras causas.

10

15

Existe uma necessidade quanto a um entendimento dos fatores de suscetibilidade que levam à predisposição aumentada para AF e acidente vascular cerebral. A identificação de variantes em risco para AF pode, por exemplo, ser útil para avaliar quais indivíduos estão particularmente em alto risco para AF e acidente vascular cerebral subsequente. Além disso, o tratamento preventivo pode ser administrado aos indivíduos que sofrem de AF e que são portadores de variantes de suscetibilidade em risco para AF e/ou acidente vascular cerebral. Finalmente, a identificação de variantes em risco para AF e/ou acidente vascular cerebral pode levar à identificação de novos alvos para a terapia medicamentosa, assim como o desenvolvimento de novas medidas terapêuticas.

20

25

SUMÁRIO DA INVENÇÃO

A presente invenção diz respeito à descoberta de que certos marcadores genéticos têm sido mostrado estar associados com a arritmia cardíaca, em particular a fibrilação atrial e flúter atrial e acidente vascular cerebral. Esta descoberta pode ser utilizada em uma variedade de métodos, procedimentos, aparelhos, meios e kits, como aqui descrito, com respeito aos métodos e procedimentos de diagnose e/ou determinação de uma suscetibilidade, métodos de genotipar variantes associadas, métodos de prognosticar resposta aos agentes terapêuticos, métodos de prognosticar prognóstico, métodos de monitorar o progresso de tratamento e sistemas e kits para o uso em tais métodos.

Um aspecto da invenção diz respeito a um método de determinar uma suscetibilidade para a arritmia cardíaca ou acidente vascular cerebral em um indivíduo humano, o método compreendendo determinar a presença ou ausência de pelo menos um alelo de pelo menos um marcador polimórfico em uma amostra de ácido nucleico do indivíduo, em que o pelo menos um marcador polimórfico é selecionado dos marcadores polimórficos apresentados na Tabela 5 e marcadores em desequilíbrio de ligação com estes, em que a determinação da presença ou ausência do pelo menos um alelo é indicativa de uma suscetibilidade para a arritmia cardíaca ou acidente vascular cerebral no indivíduo. Em uma forma de realização, o pelo menos um marcador polimórfico está localizado dentro do bloco LD C04, apresentado na SEQ ID NO: 50 aqui contida. Em uma outra forma de realização, o pelo menos um marcador polimórfico é selecionado dos marcadores apresentados na Tabela 9 e marcadores em desequilíbrio de ligação com estes. Em uma forma de realização, o pelo menos um marcador é selecionado do marcador rs2220427 (SEQ ID NO: 1) e marcador rs10033464 (SEQ ID NO: 41) e marcadores em desequilíbrio de ligação com estes. Em uma outra forma de realização, o pelo menos um marcador polimórfico é selecionado dos

marcadores apresentados na Tabela 19. Em uma forma de realização, o método compreende ainda uma etapa de avaliar pelo menos um haplótipo que compreende pelo menos dois marcadores polimórficos no indivíduo.

5 Em um outro aspecto, a invenção diz respeito a um método de determinar uma suscetibilidade para a arritmia cardíaca ou acidente vascular cerebral em um indivíduo humano, que compreende determinar se pelo menos um alelo em risco em pelo menos um marcador polimórfico está presente em um conjunto de dados de genótipo derivado do indivíduo, em que o pelo menos um marcador polimórfico é selecionado dos marcadores apresentados na Tabela 5 e marcadores em desequilíbrio de ligação com estes e em que a determinação da presença do pelo menos um alelo em risco é indicativa de suscetibilidade aumentada para a arritmia cardíaca ou acidente vascular cerebral no indivíduo.

15 O conjunto de dados de genótipo compreende em uma forma de realização informação a cerca da identidade de marcador e o status alélico do indivíduo para o pelo menos um marcador polimórfico, isto é, informação a cerca da identidade dos dois alelos carregados pelo indivíduo para o marcador e/ou informação a cerca de se um indivíduo é um carregador de um alelo em risco particular para o pelo menos um marcador polimórfico. O conjunto de dados de genótipo pode compreender informação alélica a cerca de um ou mais marcadores, incluindo dois ou mais marcadores, três ou mais marcadores, cinco ou mais marcadores, cem ou mais marcadores, etc. Em algumas formas de realização, o conjunto de dados de genótipo compreende informação de genótipo de uma avaliação genômica inteira do indivíduo

20 incluindo centenas de milhares de marcadores ou ainda um milhão ou mais de marcadores.

A invenção, em um outro aspecto, diz respeito a um procedimento que compreende uma etapa de analisar um ácido nucleico de um indivíduo humano para determinar a presença ou ausência de pelo menos

um alelo de pelo menos um marcador polimórfico ou haplótipo associados com a sequência genômica com a sequência como apresentada na SEQ ID NO: 50; e uma etapa de determinar o status de um indicador genético de arritmia cardíaca ou acidente vascular cerebral no indivíduo a partir da presença ou ausência do pelo menos um marcador ou haplótipo. Assim o status do genótipo e/ou haplótipo do indivíduo é usado como um indicador de arritmia cardíaca, incluindo fibrilação atrial e flúter atrial, assim como acidente vascular cerebral, no indivíduo.

A invenção também diz respeito a um método de avaliar uma suscetibilidade para a arritmia cardíaca ou acidente vascular cerebral em um indivíduo humano, que compreende triar um ácido nucleico do indivíduo quanto a pelo menos um marcador polimórfico ou haplótipo na SEQ ID NO: 50 que se correlacione com a ocorrência aumentada de arritmia cardíaca ou acidente vascular cerebral em uma população humana; em que a determinação da presença de um alelo marcador em risco no pelo menos um polimorfismo ou um haplótipo em risco no ácido nucleico identifica o indivíduo como tendo suscetibilidade elevada para a arritmia cardíaca e/ou acidente vascular cerebral e em que a ausência do pelo menos um alelo marcador em risco ou haplótipo em risco no ácido nucleico identifica o indivíduo como não tendo a suscetibilidade elevada.

O procedimento ou métodos da invenção em uma forma de realização envolve pelo menos um marcador polimórfico ou haplótipo que compreendem um fragmento de ácido nucleico contíguo de bloco LD C04 como apresentado na SEQ ID NO: 50 ou o complemento deste, em que o fragmento é menor do que 500 nucleotídeos no tamanho e especificamente hibridiza a um segmento complementar de bloco LD C04. Em uma forma de realização, o fragmento é maior do que 15 nucleotídeos e menor do que 400 nucleotídeos no tamanho e em que o fragmento especificamente hibridiza a um segmento complementar do bloco LD C04 como apresentado na SEQ ID

NO: 50.

Em formas de realização alternativas, a suscetibilidade conferida pelos marcadores polimórficos ou haplótipos é a suscetibilidade diminuída, isto é, os marcadores e haplótipos da invenção conferem risco diminuído de um indivíduo desenvolver arritmia cardíaca, incluindo fibrilação atrial e flúter atrial, e/ou acidente vascular cerebral. Em uma tal forma de realização, a suscetibilidade diminuída é caracterizada por uma razão de chances (OR) ou Risco Relativo (RR) de menos do que 0,8.

Em uma outra forma de realização, a suscetibilidade diminuída é caracterizada por uma razão de chances (OR) de menos do que 0,7. Em uma outra forma de realização, a suscetibilidade diminuída é caracterizada por uma OR ou RR de menos do que 0,6. Em uma outra forma de realização, a suscetibilidade diminuída é caracterizada pela OR ou RR de menos do que 0,5. Outras formas de realização dizem respeito a outros valores para OR ou RR incluindo valores de 0,9, 0,85, 0,75, 0,65, 0,55, etc.

Um outro aspecto da invenção diz respeito a um método de identificação de um marcador para o uso na avaliação da suscetibilidade aos sintomas associados com a arritmia cardíaca e/ou acidente vascular cerebral em um indivíduo humano, o método compreendendo pelo menos um marcador polimórfico dentro da SEQ ID NO: 50 ou pelo menos um marcador polimórfico em desequilíbrio de ligação com pelo menos um marcador dentro da SEQ ID NO: 50, determinar o status genotípico de uma amostra de indivíduos diagnosticados com a arritmia cardíaca e/ou acidente vascular cerebral e o status genotípico de uma amostra de indivíduos de controle, em que uma diferença significativa na frequência de pelo menos um alelo em pelo menos um polimorfismo em indivíduos diagnosticados com a arritmia cardíaca e/ou acidente vascular cerebral quando comparada com a frequência do pelo menos um alelo na amostra de controle é indicativa do pelo menos um polimorfismo sendo útil para avaliar a suscetibilidade para a arritmia cardíaca

e/ou acidente vascular cerebral. Em uma forma de realização, um aumento na frequência do pelo menos um alelo no pelo menos um polimorfismo em indivíduos diagnosticados com a arritmia cardíaca e/ou acidente vascular cerebral, quando comparada com a frequência do pelo menos um alelo na amostra de controle, é indicativa do pelo menos um polimorfismo sendo útil para avaliar a suscetibilidade aumentada para a arritmia cardíaca. Em uma outra forma de realização, uma diminuição na frequência do pelo menos um alelo no pelo menos um polimorfismo em indivíduos diagnosticados com a arritmia cardíaca e/ou acidente vascular cerebral, quando comparada com a frequência do pelo menos um alelo na amostra de controle, é indicativa do pelo menos um polimorfismo sendo útil para avaliar a suscetibilidade diminuída para ou proteção contra, a arritmia cardíaca e/ou acidente vascular cerebral. Em formas de realização preferidas, a diferença significativa na frequência é caracterizada por uma medida estatística. Em uma forma de realização, a medida estatística é um valor P. em formas de realização particulares, um valor P significativo é menor do que 0,05, menor do que 0,01, menor do que 0,001, menor do que 0,0001, menor do que 0,00001, menor do que 0,000001, menor do que 0,0000001 ou menor do que 0,00000001. Em outras formas de realização, a diferença significativa é caracterizada por uma razão de chances (OR) ou Risco Relativo (RR) com valores de intervalo de confiança particular (CE).

Em um outro aspecto, a invenção diz respeito a um método de genotipar uma amostra de ácido nucleico obtida de um indivíduo humano, que compreende determinar a presença ou ausência de pelo menos um alelo de pelo menos um marcador polimórfico preditivo de risco aumentado de arritmia cardíaca e/ou acidente vascular cerebral na amostra, em que o pelo menos um marcador é selecionado dos marcadores apresentados na Tabela 5 e marcadores em desequilíbrio de ligação com estes e em que a determinação da presença ou ausência do pelo menos um alelo do pelo menos um marcador

polimórfico é preditiva de risco aumentado de arritmia cardíaca e/ou acidente vascular cerebral no indivíduo. Em uma forma de realização, a genotipagem é realizada usando um processo selecionado de hibridização de sonda específica de alelo, extensão de iniciador específico de alelo, amplificação específica de alelo, sequenciamento de ácido nucleico, digestão de 5'-exonuclease, ensaio de marca molecular, ensaio de ligação de oligonucleotídeo, análise de tamanho e análise conformacional de filamento simples. Em uma forma de realização preferida, o processo compreende a hibridização de sonda específica de alelo. O processo de genotipagem preferivelmente compreende amplificar um segmento de um ácido nucleico que compreende o pelo menos um marcador polimórfico, pela Reação em Cadeia de Polimerase (PCR), usando um par iniciador de nucleotídeo que flanqueia o pelo menos um marcador polimórfico. Em um método de genotipagem preferido, as seguintes etapas são realizadas:

1. contatar cópias do ácido nucleico com uma sonda de oligonucleotídeo de detecção e uma sonda de oligonucleotídeo realçadora sob condições para a hibridização específica da sonda de oligonucleotídeo com o ácido nucleico; em que

a) a sonda de oligonucleotídeo de detecção tem de 5 a 100 nucleotídeos no comprimento e especificamente hibridiza a um primeiro segmento do ácido nucleico cuja sequência de nucleotídeo é dada pela SEQ ID NO: 50 que compreende pelo menos um sítio polimórfico;

b) a sonda de oligonucleotídeo de detecção compreende um rótulo detectável no seu terminal 3' e uma porção de extinção no seu terminal 5';

c) o oligonucleotídeo realçador tem de 5 a 100 nucleotídeos no comprimento e é complementar a um segundo segmento da sequência de nucleotídeo que está a 5' em relação à sonda de oligonucleotídeo, tal que o oligonucleotídeo realçador está localizado 3' em relação à sonda de

oligonucleotídeo de detecção quando ambos oligonucleotídeos são hibridizados ao ácido nucleico; e

d) um intervalo de base única existe entre o primeiro segmento e o segundo segmento, tal que quando a sonda de oligonucleotídeo e a sonda de oligonucleotídeo realçador são ambas hibridizadas ao ácido nucleico, um intervalo de base única existe entre os oligonucleotídeos;

2. tratar o ácido nucleico com uma endonuclease que clivará o rótulo detectável do terminal 3' da sonda de detecção para liberar rótulo detectável livre quando a sonda de detecção é hibridizada ao ácido nucleico; e

medir o rótulo detectável livre, em que a presença do rótulo detectável livre indica que a sonda de detecção especificamente hibridiza ao primeiro segmento do ácido nucleico e indica a sequência do sítio polimórfico como o complemento da sonda de detecção.

Um outro aspecto da invenção diz respeito a um método de determinar uma suscetibilidade para a arritmia cardíaca ou acidente vascular cerebral em um indivíduo humano, o método compreendendo determinar a identidade de pelo menos um alelo de pelo menos um marcador polimórfico em uma amostra de ácido nucleico obtida do indivíduo, em que o pelo menos um marcador é selecionado do grupo de marcadores associados com o gene PITX2, em que a presença do pelo menos um alelo é indicativa de uma suscetibilidade para a arritmia cardíaca ou acidente vascular cerebral no indivíduo.

Algumas formas de realização da invenção dizem respeito a uma outra etapa de avaliar pelo menos um biomarcador adicional para a fibrilação atrial, flúter atrial ou acidente vascular cerebral, em que combinar a informação genética dos marcadores fornece avaliação de risco para a fibrilação atrial, flúter atrial ou acidente vascular cerebral. Em algumas destas formas de realização, o biomarcador é um marcador ou haplótipo genéticos, isto é, fatores de risco genético mostram estar ou considerados estar,

relacionados com o risco aumentado ou diminuído de fibrilação atrial, flúter atrial ou acidente vascular cerebral. Em outras formas de realização o biomarcador é um biomarcador de proteína. O biomarcador de proteína é em algumas formas de realização selecionado de dímero D da fibrina, fragmento de ativação da protrombina 1,2 (F1,2), complexos de trombina-antitrombina III (TAT), fibrinopeptídeo A (FPA), fosfolipase associada à lipoproteína A2 (Ip-PLA2), beta-tromboglobulina, fator de plaqueta 4, P-selectina, Fator de von Willebrand, peptídeo pró-natriurético (BNP), metaloproteinase de matriz-9 (MMP-9), PARK7, difosfato cinase de nucleosídeo (NDKA), tau, enolase específica de neurônio, fator de crescimento neurotrófico tipo B, proteína astrogial S-100b, proteína ácida fibrilar glial, proteína reativa C, amiloide A sérico, metaloproteinase de matriz-9, moléculas de adesão celular vascular e intracelular, fator alfa de necrose de tumor e interleucinas, incluindo interleucina-1, -6 e -8). Em uma forma de realização, o pelo menos um biomarcador inclui células progenitoras. Em formas de realização particulares, mais do que um biomarcador é determinado. Em uma forma de realização preferida, o biomarcador é medido no plasma do indivíduo. Outras formas de realização dizem respeito ainda a combinar informação não genética para fazer a avaliação de risco, diagnose ou prognóstico de fibrilação atrial, flúter atrial ou acidente vascular cerebral no indivíduo. A informação não genética pode compreender idade, idade no início da doença, gênero, etnicidade, diagnose de doença anterior, por exemplo, diagnose de arritmia cardíaca (por exemplo, fibrilação atrial) e acidente vascular cerebral, história médica do indivíduo, história familiar de doença, medições bioquímicas e medições clínicas (por exemplo, pressão sanguínea, níveis lipídicos séricos). A análise de tal informação combinada a partir de vários marcadores genéticos ou marcadores genéticos mais marcadores não genéticos é possível pelos métodos conhecidos por aqueles habilitados na técnica. Em uma forma de realização, a análise é realizada calculando o risco global pela regressão

logística.

A invenção diz respeito ainda a um método de diagnosticar a suscetibilidade aumentada de acidente vascular cerebral em um indivíduo humano, que compreende as etapas de (a) determinar se o indivíduo tem
5 experienciado sintomas associados com a Fibrilação atrial, Flúter atrial ou um Ataque Isquêmico Transitório; (b) determinar se uma amostra de ácido nucleico do indivíduo compreende pelo menos uma cópia de um alelo em risco de pelo menos um marcador polimórfico selecionado dos marcadores apresentados na Tabela 5 e marcadores em desequilíbrio de ligação com estes;
10 em que a presença de sintomas associados com a Fibrilação atrial, Flúter atrial e/ou Ataque Isquêmico Transitório e a presença da pelo menos uma cópia do alelo em risco é indicativa de suscetibilidade aumentada de acidente vascular cerebral.

A invenção em um outro aspecto diz respeito a um método de
15 avaliar um indivíduo quanto a probabilidade de resposta a um agente terapêutico para prevenir e/ou melhorar sintomas associados com a arritmia cardíaca e/ou acidente vascular cerebral, que compreende: determinar a presença ou ausência de pelo menos um alelo de pelo menos um marcador polimórfico em uma amostra de ácido nucleico obtida do indivíduo, em que o
20 pelo menos um marcador polimórfico é selecionado dos marcadores apresentados na Tabela 9 e marcadores em desequilíbrio de ligação com estes, em que a determinação da presença do pelo menos um alelo do pelo menos um marcador é indicativa de uma probabilidade de uma resposta positiva ao agente terapêutico para a arritmia cardíaca e/ou acidente vascular cerebral.

25 Em uma forma de realização, o agente terapêutico é um anticoagulante, um agente anti-arrítmico, um agente de controle da frequência cardíaca, um agente de cardioversão ou um agente de controle do ritmo cardíaco. Em uma outra forma de realização, o agente terapêutico é selecionado de varfarina, heparina, heparinas de peso molecular baixo,

inibidores do fator Xa e inibidores de trombina, bloqueadores do canal de sódio, bloqueadores beta, bloqueadores do canal de potássio e bloqueadores do canal de cálcio.

Em uma outra forma de realização, o agente terapêutico é selecionado de varfarinag, ximelagatran, heparina, enoxaparina, dalteparina, tinzaparina, ardeparina, nadroparina, reviparina, fondaparinux, idraparinux, lepirudina, bivalirudina, argatroban, danaparóide, disopiramida, moricizina, procainamida, quinidina, lidocaína, mexiletino, tocainida, fenitoína, encainida, flecainida, propafenona, aimalina, cibenzolina, detaímio, esmolol, propranolol, metoprolol, alprenolol, atenolol, carvedilol, bisoprolol, acebutolol, nadolol, pindololol, labetalol, oxprenotol, penbutolol, timolol, betaxolol, cartelol, sotalol, levobunolol, amiodarono, azimilida, bretílio, dofetilida, tedisamil, ibutilida, sematilida, N-acetil procainamida, cloridreto de nifecalant, vernacalant, ambasilida, verpamil, mibefradil, diltiazem, digoxin, adenosina, ibutilida, amiodarona, procainamida, profafenona e flecainida.

Já um outro aspecto da invenção diz respeito a um método de prognosticar prognóstico de um indivíduo diagnosticado com, arritmia cardíaca e/ou acidente vascular cerebral, o método compreendendo determinar a presença ou ausência de pelo menos um alelo de pelo menos um marcador polimórfico em uma amostra de ácido nucleico obtida do indivíduo, em que o pelo menos um marcador polimórfico é selecionado dos marcadores apresentados na Tabela 9 e marcadores em desequilíbrio de ligação com estes, em que a determinação da presença do pelo menos um alelo é indicativa de um prognóstico pior da arritmia cardíaca e/ou acidente vascular cerebral no indivíduo.

Os métodos de monitorar o progresso de um tratamento de um indivíduo que passa por tratamento para a arritmia cardíaca e/ou acidente vascular cerebral também estão dentro do escopo da invenção, os métodos

compreendendo determinar a presença ou ausência de pelo menos um alelo de pelo menos um marcador polimórfico em uma amostra de ácido nucleico obtida do indivíduo, em que o pelo menos um marcador polimórfico é selecionado dos marcadores apresentados na Tabela 9 e marcadores em desequilíbrio de ligação com estes, em que a determinação da presença do

Em formas de realização particulares da invenção, por exemplo nos vários métodos, usos, procedimentos, aparelho e kits da invenção, o fenótipo de arritmia cardíaca é caracterizado ainda como sendo fibrilação atrial ou flúter atrial. Os inventores determinaram que o risco conferido pelas variantes em risco de AF aqui descritas é maior para indivíduo com idade precoce no início do que para os indivíduos com idade avançada no início. Assim em uma forma de realização, a fibrilação atrial ou flúter atrial são caracterizados ainda por uma idade de início no indivíduo de menos do que 80 anos. Em uma outra forma de realização, a fibrilação atrial ou flúter atrial são caracterizados ainda por uma idade de início no indivíduo de menos do que 70 anos. Ainda em uma outra forma de realização, a fibrilação atrial ou flúter atrial são caracterizados ainda por uma idade de início no indivíduo de menos do que 60 anos. Outros cortes de idade são possíveis em formas de realização alternativas da invenção e também são consideradas, incluindo, mas não limitadas a, corte de idade de menos do que 75 anos, menos do que 65 anos e menos do que 55 anos. Além disso, a idade no início ou diagnose acima da idade de 55, 60, 65, 70, 75 ou 80 também são consideradas e dentro do escopo da invenção, como o são as faixas de idade dentro das quais a diagnose ou sintomas ou início da doença ocorrem, incluindo, mas não limitadas a, idade de 50 a 80, idade de 55 a 75, idade de 60 a 80, idade de 65 a 75, etc.

Em certas formas de realização da invenção, o acidente vascular cerebral é caracterizado ainda como acidente vascular cerebral

isquêmico. Em outras formas de realização, o fenótipo de acidente vascular cerebral pode ser caracterizado como um ou mais dos sub-fenótipos de acidente vascular cerebral isquêmico a aterosclerose de artéria grande (LAA), acidente vascular cerebral cardioembólico (CES) e doença de vaso pequeno (SVD).

Em formas de realização particulares da invenção, o desequilíbrio de ligação (LD) é definido por um corte quantitativo específico. Como descrito aqui em detalhes, o desequilíbrio de ligação pode ser quantitativamente determinado pelas medidas tais como r^2 e $|D'|$. Como uma consequência, certas formas de realização da invenção dizem respeito a marcadores em desequilíbrio de ligação por uma medida dentro de uma certa faixa especificada pelos valores particulares de r^2 e/ou $|D'|$. Em uma tal forma de realização, o LD é caracterizado pelos valores numéricos para r^2 maiores do que 0,1. Em uma outra forma de realização, o LD é caracterizado pelos valores numéricos para r^2 maiores do que 0,1. Em uma outra forma de realização, o LD é caracterizado pelos valores numéricos para r^2 maiores do que 0,5. Ainda em uma outra forma de realização, o LD é caracterizado pelos valores numéricos para r^2 maiores do que 0,8. Outros valores de corte para r^2 também são considerados, como aqui descrito em mais detalhes. Em certas formas de realização, o LD é caracterizado por certos valores de corte para r^2 e/ou $|D'|$. Em uma tal forma de realização, o LD é caracterizado pelos valores para r^2 e/ou $|D'|$ de mais do que 0,2 e 0,8, respectivamente. Outra combinação e permutas destas ou outras medidas de LD são possíveis para praticar a invenção e também são consideradas dentro do escopo da invenção.

Os procedimentos, usos ou métodos da invenção em algumas formas de realização compreendem ainda uma etapa de administrar a um indivíduo determinado estar em risco aumentado para desenvolver arritmia cardíaca ou acidente vascular cerebral uma composição que compreende pelo menos um agente terapêutico eficaz para tratar ou prevenir a arritmia cardíaca

ou acidente vascular cerebral ou prevenir sintomas associados com a arritmia cardíaca ou acidente vascular cerebral. Assim, a invenção pode ser usada para determinar se um indivíduo é adequado para um módulo de tratamento particular.

5 Kits para o uso nos vários métodos e procedimentos aqui descritos também estão dentro do escopo da invenção. Assim, em um aspecto, a invenção diz respeito a um kit para avaliar a suscetibilidade para a arritmia cardíaca e/ou acidente vascular cerebral em um indivíduo humano, o kit compreendendo reagentes para detectar seletivamente pelo menos um alelo de
10 pelo menos um marcador polimórfico no genoma do indivíduo, em que o marcador polimórfico é selecionado do grupo que consiste dos marcadores polimórficos dentro do segmento cuja sequência é apresentada na SEQ ID NO: 50 e marcadores em desequilíbrio de ligação com estes e em que a
- presença do pelo menos um alelo é indicativa de uma suscetibilidade para a
-
15 arritmia cardíaca e/ou acidente vascular cerebral.

 Em uma forma de realização, o pelo menos um marcador polimórfico é selecionado dos marcadores apresentados na Tabela 5. Em uma outra forma de realização, o pelo menos um marcador polimórfico é selecionado do grupo de marcadores apresentados na Tabela 9 e marcadores
20 em desequilíbrio de ligação com estes. Em uma outra forma de realização, o pelo menos um marcador polimórfico é selecionado do marcador rs2220427 (SEQ ID NO: 1) e rs10033464 (SEQ ID NO: 41) e marcadores em desequilíbrio de ligação com estes. Em uma forma de realização preferida, o pelo menos um marcador polimórfico é selecionado dos marcadores
25 apresentados na Tabela 19. Em uma outra forma de realização preferida, o pelo menos um marcador polimórfico é selecionado de D4S406 (SEQ ID NO: 45), rs2634073 (SEQ ID NO: 33), rs2200733 (SEQ ID NO: 28), rs2220427 (SEQ ID NO: 1), rs10033464 (SEQ ID NO: 41) e rs13143308 (SEQ ID NO: 51). Em uma forma de realização, os reagentes compreendem pelo menos um

oligonucleotídeo contíguo que hibridiza a um fragmento do genoma do indivíduo que compreende o pelo menos um marcador polimórfico, um tampão e um rótulo detectável.

Em uma outra forma de realização, os reagentes compreendem pelo menos um par de oligonucleotídeos que hibridizam aos filamentos opostos de um segmento de ácido nucleico genômico obtido do paciente, em que cada par iniciador de oligonucleotídeo é planejado para amplificar seletivamente um fragmento do genoma do indivíduo que inclui um marcador polimórfico e em que o fragmento tem pelo menos 30 pares de base no tamanho. O pelo menos um oligonucleotídeo é, em formas de realização preferidas, completamente complementar ao genoma do indivíduo. Em uma forma de realização, o oligonucleotídeo tem cerca de 18 a cerca de 50 nucleotídeos no comprimento. Em uma outra forma de realização, o oligonucleotídeo tem de 20 a 30 nucleotídeos no comprimento. Em uma forma de realização preferida, o kit compreende:

- a. uma sonda de oligonucleotídeo de detecção que tem de 5 a 100 nucleotídeos no comprimento;
- b. uma sonda de oligonucleotídeo realçadora que tem de 5 a 100 nucleotídeos no comprimento; e
- c. uma enzima de endonuclease;

em que a sonda de oligonucleotídeo de detecção especificamente hibridiza a um primeiro segmento do ácido nucleico cuja sequência de nucleotídeo é dada pela SEQ ID NO: 2 que compreende pelo menos um sítio polimórfico; em que a sonda de oligonucleotídeo de detecção compreende um rótulo detectável no seu terminal 3' e uma porção de extinção no seu terminal 5'; em que o oligonucleotídeo realçador tem de 5 a 100 nucleotídeos no comprimento e é complementar a um segundo segmento da sequência de nucleotídeo que é 5' em relação à sonda de oligonucleotídeo, tal que o oligonucleotídeo realçador está localizado 3' em relação à sonda de

oligonucleotídeo de detecção quando ambos oligonucleotídeos são hibridizados ao ácido nucleico; em que um intervalo de base única existe entre o primeiro segmento e o segundo segmento, tal que quando a sonda de oligonucleotídeo e a sonda de oligonucleotídeo realçador são ambas hibridizadas ao ácido nucleico, um intervalo de base única existe entre os oligonucleotídeos; e em que tratar o ácido nucleico com a endonuclease clivará o rótulo detectável do terminal 3' da sonda de detecção para liberar o rótulo detectável livre quando a sonda de detecção é hibridizada ao ácido nucleico.

Os marcadores polimórficos aqui descritos como preditivo de risco de arritmia cardíaca (por exemplo, AF e Flúter atrial) e acidente vascular cerebral são úteis como marcadores de diagnóstico. No aspecto, a invenção portanto diz respeito ao uso de uma sonda de oligonucleotídeo na fabricação de um reagente de diagnóstico para diagnosticar e/ou avaliar a suscetibilidade para a arritmia cardíaca e/ou acidente vascular cerebral em um indivíduo humano, em que a sonda hibridiza a um segmento de um ácido nucleico cuja sequência de nucleotídeo é dada pela SEQ ID NO: 50 que compreende pelo menos um sítio polimórfico, em que o fragmento tem 15 a 500 nucleotídeos no comprimento.

Em uma tal forma de realização, o sítio polimórfico é selecionado dos marcadores polimórficos apresentados na Tabela 5 e polimorfismos em desequilíbrio de ligação com estes. Em uma outra forma de realização, o pelo menos um marcador polimórfico é selecionado de D4S406 (SEQ ID NO: 45), rs2634073 (SEQ ID NO: 33), rs2200733 (SEQ ID NO: 28), rs2220427 (SEQ ID NO: 1), rs10033464 (SEQ ID NO: 41) e rs13143308 (SEQ ID NO: 51),

Meio legível por computador para armazenar informação a cerca de marcadores associados com doença como aqui descrito também estão dentro do escopo da presente invenção. Em um tal aspecto, a invenção diz

respeito a um meio legível por computador no qual é armazenado um identificador quanto a pelo menos um marcador polimórfico; um indicador da frequência de pelo menos um alelo do dito pelo menos um marcador polimórfico em uma pluralidade de indivíduos diagnosticados com fibrilação atrial, flúter atrial e/ou acidente vascular cerebral; e um indicador da frequência do pelo menos um alelo do dito pelo menos um dos marcadores polimórficos em uma pluralidade de indivíduos de referência; em que o pelo menos um marcador polimórfico é selecionado dos marcadores polimórficos apresentados na Tabela 5 e polimorfismos em desequilíbrio de ligação com estes. Em uma forma de realização preferida, o pelo menos um marcador polimórfico é selecionado de D4S406 (SEQ ID NO: 45), rs2634073 (SEQ ID NO: 33), rs2200733 (SEQ ID NO: 28), rs2220427 (SEQ ID NO: 1), rs10033464 (SEQ ID NO: 41) e rs13143308 (SEQ ID NO: 51).

A invenção também está relacionada com o um aparelho para determinar um indicador genético para a arritmia cardíaca e/ou acidente vascular cerebral em um indivíduo humano, que compreende: uma memória legível por computador; e uma rotina armazenada na memória legível por computador; em que a rotina é adaptada para ser executada em um processador para analisar dados de genótipo e/ou haplótipo quanto a pelo menos um indivíduo humano com respeito a pelo menos um marcador polimórfico selecionado dos marcadores apresentados na Tabela 5 e marcadores em desequilíbrio de ligação com estes e gerar um produto com base nos dados de marcador ou haplótipo, em que o produto compreende uma medição de risco do pelo menos um marcador ou haplótipo como um indicador genético de arritmia cardíaca e/ou acidente vascular cerebral para o indivíduo humano. Em uma forma de realização preferida, a rotina compreende ainda determinar um indicador da frequência de pelo menos um alelo de pelo menos um marcador polimórfico e/ou pelo menos um haplótipo em uma pluralidade de indivíduos diagnosticados com a arritmia cardíaca e/ou

acidente vascular cerebral e um indicador da frequência de pelo menos um alelo de pelo menos um marcador polimórfico ou pelo menos um haplótipo em uma pluralidade de indivíduos de referência e calcular uma medição de risco para o pelo menos um alelo e/ou haplótipo com base neste; e em que

5 uma medição de risco para o indivíduo é calculada com base em uma comparação do pelo menos um status de marcador e/ou haplótipo para o indivíduo para o risco calculado para a pelo menos uma informação de marcador e/ou haplótipo para a pluralidade de indivíduos diagnosticados com a fibrilação atrial, flúter atrial e/ou acidente vascular cerebral. Em certas

10 formas de realização, a medição de risco é caracterizada por uma razão de chances (OR) ou um Risco Relativo (RR), como aqui descrito em mais detalhes.

Os marcadores polimórficos descobertos na presente invenção como preditivos de uma suscetibilidade de arritmia cardíaca e acidente

15 vascular cerebral, como descrito, assim como marcadores em desequilíbrio de ligação com estes, são todos úteis para praticar os vários aspectos da presente invenção. Assim, embora marcadores polimórficos particulares fossem usados pelos presentes inventores para detectar uma associação de uma região particular do cromossoma 4 com a arritmia cardíaca (por exemplo, fibrilação

20 atrial e flúter atrial) e acidente vascular cerebral, é igualmente útil avaliar marcadores em desequilíbrio de ligação forte com estes marcadores. Como uma consequência, em uma forma de realização dos métodos, usos, kits, procedimentos, aparelho e meios da invenção, o pelo menos um marcador polimórfico ou haplótipo útil nos métodos ou procedimentos da invenção

25 compreende pelo menos um dos marcadores apresentados na Tabela 5 (por exemplo, Tabela 5A e Tabela 5B) e marcadores em desequilíbrio de ligação com estes. Em uma outra forma de realização, o pelo menos um marcador polimórfico ou haplótipo compreende pelo menos um dos marcadores apresentados na Tabela 9 e marcadores em desequilíbrio de ligação com estes.

Em uma forma de realização, o pelo menos um marcador polimórfico ou haplótipo compreende pelo menos um dos marcadores apresentados na Tabela 5. Em uma outra forma de realização, o pelo menos um marcador polimórfico ou haplótipo compreende pelo menos um dos marcadores apresentados na Tabela 9. Em uma outra forma de realização, o pelo menos um marcador polimórfico é selecionado dos marcadores apresentados na Tabela 4. Em uma forma de realização, o pelo menos um marcador é selecionado do marcador rs2220427 (SEQ ID NO: 1) e marcador rs10033464 (SEQ ID NO: 41) e marcadores em desequilíbrio de ligação com estes.

10 Em uma outra forma de realização, o pelo menos um marcador polimórfico é selecionado dos marcadores apresentados na Tabela 19.

Em uma forma de realização, o pelo menos um marcador ou haplótipo compreende pelo menos um dos marcadores D4S406 (SEQ ID NO: 45), rs2723296 (SEQ ID NO: 35), rs16997168 (SEQ ID NO: 36), rs2723316 (SEQ ID NO: 37), rs6419178 (SEQ ID NO: 38), rs1448817 (SEQ ID NO: 39), rs2634073 (SEQ ID NO: 33), rs2200733 (SEQ ID NO: 28), rs2220427 (SEQ ID NO: 1), rs13105878 (SEQ ID NO: 40), rs10033464 (SEQ ID NO: 41), rs13141190 (SEQ ID NO: 42), rs3853444 (SEQ ID NO: 43) e rs4576077 (SEQ ID NO: 44). Em uma outra forma de realização, o pelo menos um marcador ou haplótipo compreende pelo menos um dos marcadores D4S406 (SEQ ID NO: 45), rs2634073 (SEQ ID NO: 33), rs2200733 (SEQ ID NO: 28), rs2220427 (SEQ ID NO: 1), rs10033464 (SEQ ID NO: 41) e rs13143308 (SEQ ID NO: 51), Ainda em uma outra forma de realização, o pelo menos um marcador é selecionado de rs10033464, rs2200733, rs13143308 e rs2220427 e marcadores em desequilíbrio de ligação com estes.

Em uma outra forma de realização, a presença de alelos -2, -4 e/ou -8 do marcador D4S406, alelo G do marcador rs2723296, alelo T do marcador rs16997168, alelo T do marcador rs2723316, alelo A do marcador rs6419178, alelo G do marcador rs1448817, alelo A do marcador rs2634073,

alelo T do marcador rs2200733, alelo T do marcador rs2220427, alelo C do marcador rs13105878, alelo T do marcador rs10033464, alelo A do marcador rs13141190, alelo A do marcador rs3853444, e/ou alelo T do marcador rs4576077 é indicativa de suscetibilidade aumentada de arritmia cardíaca ou acidente vascular cerebral no indivíduo.

Em formas de realização particulares da invenção, a suscetibilidade conferida pela variante de risco (isto é, um alelo particular em um marcador polimórfico (por exemplo, um SNP) ou um haplótipo particular) é a suscetibilidade aumentada, isto é, os marcadores e haplótipos da invenção conferem risco aumentado de um indivíduo desenvolver arritmia cardíaca, incluindo fibrilação atrial e flúter atrial e acidente vascular cerebral. A suscetibilidade é tipicamente caracterizada pela medida da Razão de Chances (OR) ou, alternativamente, por um Risco Relativo (RR). Em uma forma de realização, a suscetibilidade aumentada é caracterizada por uma razão de chances (OR) de pelo menos 1,3. Em uma outra forma de realização, a suscetibilidade aumentada é caracterizada por uma razão de chances (OR) de pelo menos 1,4. Em uma outra forma de realização, a suscetibilidade aumentada é caracterizada por uma razão de chances (OR) de pelo menos 1,5. Em uma outra forma de realização, a suscetibilidade aumentada é caracterizada por uma razão de chances (OR) ou Risco Relativo (RR) de pelo menos 1,6. Ainda em uma outra forma de realização, a suscetibilidade aumentada é caracterizada por uma razão de chances (OR) ou Risco Relativo (RR) de pelo menos 1,8. Outras formas de realização dizem respeito a outros valores para OR ou valores comparáveis para RR incluindo valores de 1,25, 1,35, 1,45, 1,55, etc.

Certas formas de realização da invenção dizem respeito a indivíduos de uma etnicidade ou ancestralidade particular. Em uma tal forma de realização, o indivíduo humano tem ancestralidade selecionada da etnicidade negra, etnicidade asiática, etnicidade caucasiana, etnicidade

hispânica e etnicidade árabe. Em formas de realização particulares, a etnicidade é auto-relatada. Em outras formas de realização, a ancestralidade é determinada pela avaliação de marcadores genéticos específicos de etnicidade particulares.

5 DESCRIÇÃO RESUMIDA DAS FIGURAS

O precedente e outros objetivos, características e vantagens da invenção estarão evidentes a partir da seguinte descrição mais particular de formas de realização preferidas da invenção.

A FIG. 1 mostra uma plotagem de desequilíbrio de ligação (LD) na região que compreende variantes da presente invenção para a população CEPH (dados HapMap). O bloco LD C04 (111.954.811 - 112.104.250 no Cromossoma 4, posições da Construção 35 da NCBI) é indicado na Figura por uma caixa preta. A plotagem mostra duas medidas de LD, isto é, D' na parte superior e esquerda da Figure 1 e r^2 na parte inferior e direita da figura.

A FIG. 2 Mostra um esquemático da estrutura de haplótipo na região associada dentro do bloco LD. As áreas dos círculos escuros (esquerda) são proporcionais às frequências de haplótipo dos haplótipos na Islândia e as áreas dos círculos claros (direita) são proporcionais às frequências de haplótipo em Hong Kong. O haplótipo intermediário, mostrado no centro do gráfico, certamente não mais existe em cada uma das duas populações (a sua frequência estimada é menor do que 0,2% que é indistinguível de erros de genotipagem).

A FIG. 3 É uma vista geral de uma vizinhança genômica de 200 kb de rs2200733 e rs10033464. Isto inclui ESTs prognosticados, as localizações das três classes principais de SNPs equivalentes nas amostras de CEU HapMap e uma vista geral da estrutura LD da região nas várias amostras HapMap étnicas.

A FIG 4. Mostra a análise de Northern Blot da expressão de

PITX2 em coração e aorta humanos. O clone de cDNA de PITX2 HU3_p983E0327D foi usado como uma sonda e detectou transcritos de 1,8, 2 e 3 kb e transcritos de PITX2 de 2,2 e 3 kb no átrio esquerdo e aorta respectivamente. Linha 1: Coração fetal, linha 2: Coração inteiro, linha 3: Aorta, linha 4: Ápice do coração, linha 5: Átrio esquerdo, linha 6: Átrio direito, linha 7: Ventrículo esquerdo linha 8: Ventrículo direito. Mancha sondada com clone de cDNA PITX2 (HU3_p983E0327D).

DESCRIÇÃO DETALHADA DA INVENÇÃO

Uma descrição de formas de realização preferidas da invenção segue.

Definições

Os seguintes termos, no presente contexto, devem ter os significados como indicados:

Fibrilação atrial (AF), como aqui descrito, refere-se à AF como habitualmente definida de acordo com critérios médicos estabelecidos. A AF é classificada por ICD-10 na classe 148 e por ICD-9 na classe 427,3

Flúter atrial (AFI), como aqui descrito, refere-se à AFI como habitualmente definida de acordo com critérios médicos estabelecidos. A AFI é classificada por ICD-10 classe 148 e por ICD-9 na classe 427,32.

Um “marcador polimórfico”, algumas vezes aludido como um “marcador”, como aqui descrito, refere-se a um sítio polimórfico genômico. Cada marcador polimórfico tem pelo menos duas variações de sequência características de alelos particulares no sítio polimórfico. Assim, a associação genética a um marcador polimórfico implica que existe associação a pelo menos um alelo específico daquele marcador polimórfico particular. O marcador pode compreender qualquer alelo de qualquer tipo de variante encontrado no genoma, incluindo SNPs, microssatélites, inserções, deleções, duplicações e translocações.

Um “alelo” refere-se à sequência de nucleotídeo de um dado

local (posição) em um cromossoma. Um alelo marcador polimórfico assim refere-se à composição (isto é, sequência) do marcador em um cromossoma. O DNA genômico de um indivíduo contém dois alelos para qualquer marcador polimórfico dado, representativo de cada cópia do marcador em cada cromossoma.

Uma posição de nucleotídeo na qual mais do que uma sequência é possível em uma população (uma população natural ou uma população sintética, por exemplo, uma biblioteca de moléculas sintéticas) é aqui aludida como um “sítio polimórfico”.

Um “Polimorfismo de Nucleotídeo Único” ou “SNP” é uma variação de sequência de DNA que ocorre quando um nucleotídeo único em um local específico no genoma difere entre membros de uma espécie ou entre cromossomas emparelhados em um indivíduo. A maioria dos polimorfismos SNP tem dois alelos. Cada indivíduo é neste caso homozigoto para um alelo do polimorfismo (isto é, ambas as cópias cromossômicas do indivíduo têm o mesmo nucleotídeo na localização SNP) ou o indivíduo é heterozigoto (isto é, os duas cromátides irmãs do indivíduo contêm nucleotídeos diferentes). A nomenclatura SNP como aqui relatada refere-se ao rótulo de identificação ID da Referência oficial SNP (rs) como designado a cada SNP único pelo National Center for Biotechnological Information (NCBI).

Uma “variante”, como aqui descrita, refere-se a um segmento de DNA que difere do DNA de referência. Um “marcador” ou um “marcador polimórfico”, como aqui definido, é uma variante. Alelos que diferem da referência são aludidos como alelos “variante”.

Um “microsatélite” é um marcador polimórfico que tem repetições pequenas múltiplas de bases que têm 2 a 8 nucleotídeos no comprimento (tais como repetições CA) em um sítio particular, em que o número de comprimentos de repetições varia na população geral. Um “indel” é uma forma comum de polimorfismo que compreende uma inserção ou

deleção pequenas que têm tipicamente apenas uns poucos nucleotídeos de comprimento.

Um “haplótipo,” como aqui descrito, refere-se a um segmento de DNA genômico que é caracterizado por uma combinação específica de alelos dispostos ao longo do segmento. Para os organismos diplóides tais como humanos, um haplótipo compreende um membro do par de alelos para cada marcador ou local polimórfico. Em uma certa forma de realização, o haplótipo pode compreender dois ou mais alelos, três ou mais alelos, quatro ou mais alelos ou cinco ou mais alelos.

O termo “suscetibilidade”, como aqui descrito, abrange tanto a suscetibilidade aumentada quanto a suscetibilidade diminuída. Assim, marcadores polimórficos e/ou haplótipos particulares da invenção podem ser característicos de suscetibilidade aumentada (isto é, risco aumentado) de fibrilação atrial ou acidente vascular cerebral, como caracterizado por um Risco Relativo (RR) ou Razão de Chances (OR) de mais do que um. Alternativamente, os marcadores e/ou haplótipos da invenção são característicos de suscetibilidade diminuída (isto é, risco diminuído) de fibrilação atrial ou acidente vascular cerebral, como caracterizado por um Risco Relativo de menos do que um.

Uma “amostra de ácido nucleico” é uma amostra obtida de um indivíduo que contém ácido nucleico. Em certas formas de realização, isto é, a detecção de marcadores polimórficos e/ou haplótipos específicos, a amostra de ácido nucleico compreende DNA genômico. Uma tal amostra de ácido nucleico pode ser obtida a partir de qualquer fonte que contenha DNA genômico, incluindo como uma amostra de sangue, amostra de fluido amniótico, amostra de fluido cerebrospinal ou amostra de tecido da pele, músculo, mucosa bucal ou conjuntiva, placenta, trato gastrointestinal ou outros órgãos.

O termo “agente terapêutico da fibrilação atrial e/ou acidente

vascular cerebral” refere-se a um agente que pode ser usado para melhorar ou prevenir sintomas associados com a fibrilação atrial (AF), flúter atrial (AFI) ou acidente vascular cerebral, como aqui descrito em mais detalhes.

O termo “ácido nucleico associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral”, como aqui descrito, refere-se a um ácido nucleico que foi descoberto estar associado com a arritmia cardíaca, por exemplo, fibrilação atrial (AF), flúter atrial (AFI) ou acidente vascular cerebral. Isto inclui, mas não é limitado aos marcadores e haplótipos aqui descritos e marcadores e haplótipos em forte desequilíbrio de ligação (LD) com estes. Em uma forma de realização, um ácido nucleico associado com a fibrilação atrial, flúter atrial ou acidente vascular cerebral refere-se ao bloco LD C04 descoberto estar associado com a fibrilação atrial e acidente vascular cerebral. Em uma outra forma de realização, ácido nucleico associado com a fibrilação atrial, flúter atrial ou acidente vascular cerebral refere-se ao gene PITX2.

O termo “bloco LD C04”, como aqui descrito, refere-se ao bloco de Desequilíbrio de Ligação (LD) no Cromossoma 4 entre as posições 111.954.811 e 112.104.250 da Construção 35 da NCBI (National Center for Biotechnology Information), com a sequência genômica como apresentada na SEQ ID NO: 50.

O termo “fragmento”, como aqui descrito, refere-se a um segmento de um ácido nucleico ou sequência de proteína. Os fragmentos são de tamanho menor do que o seu ponto de referência, isto é, um fragmento de uma molécula de ácido nucleico de referência que tem 1000 nucleotídeos no tamanho é menor do que 1000 nucleotídeos no tamanho. Os fragmentos de ácido nucleico da invenção são habitualmente maiores do que 5 nucleotídeos no tamanho e tipicamente maiores do que 15 nucleotídeos no tamanho, com um limite superior como definido pelo seu nucleotídeo de referência ou pela utilidade prática do fragmento de nucleotídeo. Por exemplo, fragmentos de

nucleotídeo úteis como sondas de hibridização em algumas formas de realização da invenção são maiores do que 15 nucleotídeos e menores do que cerca de 500 nucleotídeos no tamanho. Outras faixas de tamanho aplicar-se-ão para outros fragmentos de nucleotídeo e proteína ou fragmentos de peptídeo da invenção.

O termo "PITX2", como aqui descrito, refere-se ao gene do fator de transcrição 2 de homeodomínio emparelhado no cromossoma 4q25. Este gene também é aludido como homeobox 2 pituitário (PTX2), fator de transcrição 1 de homeobox relacionado a *rieg bicoid* (RIEG1), *solurshin* e gene 1 responsivo todo-1 (ARP1).

A presente invenção diz respeito à observação de que certos marcadores polimórficos no cromossoma 4q25 do genoma humano foram descobertos estar associados com a arritmia cardíaca e acidente vascular cerebral. Em formas de realização particulares da invenção, marcadores polimórficos no cromossoma 4q25 estão associados com as arritmia cardíacas Fibrilação atrial (AF) e Flúter atrial (AFI) e acidente vascular cerebral. Esta observação tem implicações importantes e inesperadas para o desenvolvimento de métodos, usos, kits e sistemas de diagnóstico e terapêuticos, como aqui descrito em mais detalhes.

Em uma varredura ampla do genoma quanto a variantes genéticas que conferem suscetibilidade à AF, diversos marcadores no cromossoma 4q25 foram descobertos estar associados com AF. A associação mais significativa foi descoberta para os marcadores rs2220427 e rs2220733, ambos os quais deram valores p próximos a 10^{-9} (Tabela 2) para AF e menores, mas associação nominalmente significativa ao acidente vascular cerebral (Tabela 3). Um número grande de marcadores foram identificados como substitutos perfeitos para estes marcadores, incluindo o marcador de microssatélite D4S406 (Tabela 1) e vários marcadores de SNP (Tabela 4).

Outro refinamento dos resultados revelaram que o sinal de

associação parece centralizar, em termos genéticos, aos marcadores de rs2200733 e rs10033464 (Tabela 7) e marcadores em desequilíbrio de ligação com estes marcadores (incluindo, mas não limitados aos marcadores de SNP listados na Tabela 9).

A observação original na população islandesa foi replicada em um grupo AF/AFI islandês independente, em um grupo AF sueco e em um grupo AF US (Tabela 7). Quando combinado com as amostras islandesas, a associação ao rs2200733 foi inequívoca ($OR = 1,72$, $P = 3,3 \times 10^{-41}$) e a significância de rs10033464 foi bem além do limiar de significância da amplitude do genoma ($OR = 1,39$, $P = 6,9 \times 10^{-11}$). Assumindo o modelo multiplicativo, o risco atribuível à população (PAR) das duas variantes combinadas é de aproximadamente 20% em populações de ancestralidade européia. Além disso, a associação replicada em um grupo AF chinês de Hong Kong (Tabela 7).

Os inventores também descobriram que a idade no diagnóstico de AF/AFI para as amostras islandesas correlacionam-se com os dois SNPs rs2200733 e rs10033464. Assim, o diagnóstico ocorre 2,28 anos mais cedo para o alelo T de rs2200733 e 1,10 ano mais cedo para o alelo T de rs10033464 (junta $P = 1,29 \times 10^{-6}$). Este efeito é manifestado pela associação das duas variantes sendo mais forte naqueles diagnosticados em uma idade mais jovem, embora o risco permaneça significativo mesmo naqueles diagnosticados depois de atingir 80 anos de idade (Tabela 8). Uma idade similar no início do efeito é observado no grupo US (Tabela 8).

Os inventores também observaram uma forte associação entre as variantes e AFI, que parecem ser ainda mais forte do que para AF. Assim é revelado pela associação ao subconjunto ($N = 116$) dos pacientes islandeses que têm um diagnóstico de AFI ($OR = 2,60$, 95% de intervalo de confiança (CI) = 1,83-3,68, $P = 7,5 \times 10^{-8}$ para rs2200733 ou = 1,94, 95% de CI = 1,26-3,00, $P = 0,0028$ para rs10033464). De fato, para rs2200733, a OR para estes

casos de AFI definidos é significativamente mais alta do que aquela para os casos com um fenótipo AF ($P = 0,0026$) e próximo a significativamente mais alto para rs10033464 ($P = 0,084$). Estes resultam que tanto AF quanto AFI têm fatores de risco genético significantes que são ilustrados pela associação a SNPs rs2200733 e rs10033464.

Os inventores além disso estabeleceram que as variantes que se associam com AF/AFI também estão associados com acidente vascular cerebral, em particular acidente vascular cerebral isquêmico (Tabela 21). O marcador rs2200733 replicou significativamente no acidente vascular cerebral isquêmico e no subfenótipo do acidente vascular cerebral isquêmico (IS), o acidente vascular cerebral cardioembólico (CES). Tanto este marcador quanto o marcador rs10033464 foram descobertos, depois de genotipar casos de IS islandês adicionais e controles (total 1.943 casos/25.708 controles) e quatro conjuntos de replicação de caso IS/controle grandes (4.294 casos/3.709 controles), associarem-se mais firmemente com o CES, do que AF é a causa primária, (rs2200733: OR = 1,53, $P = 1,5 \times 10^{-12}$; rs10033464: OR = 1,27, $P = 5,9 \times 10^{-4}$) (Tabela 21).

Não existe nenhum gene conhecido presente no bloco LD contendo rs2200733 e rs10033464 (Figura 3). O bloco LD contém um EST unido (DA725631) e dois exons EST isolados (DB324364 e AF017091). A RT-PCR de bibliotecas de cDNA de vários tecidos não detectou a expressão destes ESTs (Tabela 16). O gene PITX2 localizado no bloco LD a montante adjacente é o gene mais próximo das variantes de risco. Diversos marcadores dentro do bloco LD contendo o gene PITX2 são correlacionados com os marcadores que mostram a associação ao AF e AfI, como mostrado na Tabela 18. É portanto possível que as variantes dentro do gene PITX2 sejam as variantes causativas subjacentes. Alternativamente, é possível que as variantes da presente invenção, como aqui descritas, afetem a função, estabilidade, expressão, modificação pós-traducional, junção, estabilidade de mensagem de

PITX2 ou por outros meios afetam o gene de modo a predispor aos sintomas associados com a fibrilação atrial, flúter atrial e/ou acidente vascular cerebral. A proteína codificada por este gene, o fator de transcrição de homeodomínio no modo emparelhado 2, é um candidato interessante para AF/AFI visto que é conhecido desempenhar um papel importante no desenvolvimento cardíaco pelo direcionamento da morfogênese assimétrica do coração (Franco, D., Trends Cardiovasc Med 13: 157-63 (2003)). Além disso, em um modelo de camundongo silenciado Pitx2 mostrou suprimir um caminho default para a formação do nodo sinoatrial no átrio esquerdo. Existe muito pouca expressão de mRNA de PITX2 em todos os tecidos facilmente acessíveis, tais como sangue e tecido adiposo, dificultando o estudo de correlação entre genótipos e níveis de expressão. O gene seguinte a montante de PITX2 é ENPEP, uma aminopeptidase responsável pela ruptura da angiotensina II no endotélio vascular. Este gene é expressado mais amplamente, mas as variantes associadas com AF não mostraram nenhuma correlação para a sua expressão no sangue ou tecido adiposo. Nenhum outro gene anotado está localizado dentro de uma região de 400 kb a montante e regiões 1,5 Mb a jusante das variantes associadas.

Avaliação quanto a marcadores e haplótipos

A sequência genômica dentro de populações não é idêntica quando os indivíduos são comparados. Ao invés, o genoma exibe variabilidade de sequência entre indivíduos em muitas localizações no genoma. Tais variações na sequência são habitualmente aludidas como polimorfismos e existem muitos de tais sítios dentro de cada genoma. Por exemplo, o genoma humano exibe variações de sequência que ocorrem em média a cada 500 pares de base. A variante de sequência mais comum consiste de variações de base em uma única posição de base no genoma e tais variantes ou polimorfismos de sequência, são habitualmente chamados Polimorfismos de Nucleotídeo Único (“SNPs”). Acredita-se que estes SNPs

tenham ocorrido em um único evento mutacional e portanto existem usualmente dois alelos possíveis em cada sítio de SNP; o alelo original e o alelo mutado. Devido ao curso genético natural e possivelmente também pressão seletiva, a mutação original tem resultado em um polimorfismo caracterizado por uma frequência particular de seus alelos em qualquer população dada. Muitos outros tipos de variantes de sequência são encontrados no genoma humano, incluindo microssatélites, inserções, deleções, inversões e variações de número de cópia. Um microssatélite polimórfico tem repetições pequenas múltiplas de bases (tais como repetições de CA, TG no filamento complementar) em um sítio particular em que o número de comprimentos de repetição varia na população geral. Em termos gerais, cada versão da sequência com respeito ao sítio polimórfico representa um alelo específico do sítio polimórfico. Estas variantes de sequência podem ser todas aludidas como polimorfismos, que ocorrem em sítios polimórficos específicos característicos da variante de sequência em questão. Em termos gerais, os polimorfismos podem compreender qualquer número de alelos específicos. Assim em uma forma de realização da invenção, o polimorfismo é caracterizado pela presença de dois ou mais alelos em qualquer população dada. Em uma outra forma de realização, o polimorfismo é caracterizado pela presença de três ou mais alelos. Em outras formas de realização, o polimorfismo é caracterizado por quatro ou mais alelos, cinco ou mais alelos, seis ou mais alelos, sete ou mais alelos, nove ou mais alelos ou dez ou mais alelos. Todos de tais polimorfismos podem ser utilizados nos métodos e kits da presente invenção e estão assim dentro do escopo da invenção.

Em alguns casos, referência é feita a alelos diferentes em um sítio polimórfico sem escolher um alelo de referência. Alternativamente, uma sequência de referência pode ser aludida para um sítio polimórfico particular. O alelo de referência é algumas vezes aludido como o alelo do “tipo selvagem” e usualmente é escolhido como o primeiro alelo sequenciado ou

como o alelo de um indivíduo “não afetado” (por exemplo, um indivíduo que não demonstra um traço ou fenótipo de doença).

Alelos para marcadores de SNP como aqui aludidos referem-se às bases A, C, G ou T visto que eles ocorrem no sítio polimórfico no ensaio de SNP utilizado. O alelo que codifica as SNPs aqui usadas são como segue: 1 = A, 2 = C, 3 = G, 4 = T. A pessoa habilitada na técnica entretanto constatará que ensaiando-se ou lendo-se o filamento de DNA oposto, o alelo complementar pode em cada caso ser medido. Assim, para um sítio polimórfico (marcador polimórfico) caracterizado por um polimorfismo A/G, o ensaio utilizado pode ser planejado para detectar especificamente a presença de um ou ambas das duas bases possíveis, isto é, A e G. Alternativamente, pelo planejamento de um ensaio que é planejado para detectar o filamento oposto no padrão de DNA, a presença das bases complementares T e C pode ser medida. Quantitativamente (por exemplo, em termos de Risco Relativo), resultados idênticos podem ser obtidos da medição do filamento de DNA (filamento + ou filamento -).

Tipicamente, uma sequência de referência é aludida para uma sequência particular. Os alelos que diferem da referência são algumas vezes aludidos como alelos “variante”. Uma sequência variante, como aqui usada, refere-se a uma sequência que difere da sequência de referência mas é de outro modo substancialmente similar. Os alelos nos marcadores genéticos polimórficos aqui descritos são variantes. As variantes adicionais podem incluir mudanças que afetam um polipeptídeo. As diferenças de sequência, quando comparadas a uma sequência de nucleotídeo de referência, pode incluir a inserção ou deleção de um único nucleotídeo ou de mais do que um nucleotídeo, resultando em uma mudança de matriz; a mudança de pelo menos um nucleotídeo, resultando em uma mudança no aminoácido codificado; a mudança de pelo menos um nucleotídeo, que resulta na geração de um códon de parada prematuro; a deleção de diversos nucleotídeos, que

resultam em uma deleção de um ou mais aminoácidos codificados pelos nucleotídeos; a inserção de um ou vários nucleotídeos, tal como pela recombinação desigual ou conversão de gene, que resulta em uma interrupção da sequência codificadora de uma matriz de leitura; duplicação de toda ou uma parte de uma sequência; transposição; ou um rearranjo de uma sequência de nucleotídeo,. Tais mudanças de sequência podem alterar o polipeptídeo codificado pelo ácido nucleico. Por exemplo, se a mudança na sequência de ácido nucleico causa uma mudança de matriz, a mudança de matriz pode resultar em uma mudança nos aminoácidos codificados, e/ou pode resultar na geração de um códon de parada prematura, causando a geração de um polipeptídeo truncado. Alternativamente, um polimorfismo associado com uma doença ou traço pode ser uma mudança sinônima em um ou mais nucleotídeos (isto é, uma mudança que não resulta em uma mudança na sequência de aminoácido). Um tal polimorfismo pode, por exemplo, alterar sítios de junção, afetar a estabilidade ou transporte de mRNA ou de outro modo afetar a transcrição ou tradução de um polipeptídeo codificado. O mesmo também pode alterar o DNA para aumentar a possibilidade de que as mudanças estruturais, tais como amplificações ou deleções, ocorram ao nível somático. O polipeptídeo codificado pela sequência de nucleotídeo de referência é o polipeptídeo de “referência” com uma sequência de aminoácido de referência particular e os polipeptídeos codificados pelos alelos variantes são aludidos como polipeptídeos “variantes” com sequências de aminoácido variantes. Uma sequência ou uma sequência de referência pode representar a direção (+) ou (-) de DNA de filamento duplo. Tais sequências estão relacionadas como sendo o complemento reverso um do outro, como bem conhecido pela pessoa habilitada.

Um haplótipo refere-se a um segmento de DNA que é caracterizado por uma combinação específica de alelos dispostos ao longo do segmento. Para organismos diplóides tais como os seres humanos, um

haplótipo compreende um membro do par de alelos para cada marcador ou local polimórficos. Em uma certa forma de realização, o haplótipo pode compreender dois ou mais alelos, três ou mais alelos, quatro ou mais alelos ou cinco ou mais alelos, cada alelo correspondendo a um marcador polimórfico específico ao longo do segmento. Os haplótipos podem compreender uma combinação de vários marcadores polimórficos, por exemplo, SNPs e microssatélites, tendo alelos particulares nos sítios polimórficos. Os haplótipos assim compreendem uma combinação de alelos em vários marcadores genéticos.

A detecção de marcadores polimórficos e/ou haplótipos específicos pode ser realizada pelos métodos conhecidos na técnica para detectar sequências nos sítios polimórficos. Por exemplo, as técnicas padrão para genotipar quanto a presença de SNPs e/ou marcadores de microssatélite podem ser usadas, tais como as técnicas com base na fluorescência (Chen, X. *et al.*, Genome Res. 9(5): 492-98 (1999)), utilizando PCR, LCR, PCR Abrigada e outras técnicas para a amplificação de ácido nucleico. As metodologias específicas disponíveis para genotipar SNP incluem, mas não são limitadas aos ensaios de genotipagem TaqMan e plataformas SNPlex (Applied Biosystems), espectrometria de massa (por exemplo, o sistema MassARRAY da Sequenom), métodos de mini-sequenciamento, PCR em tempo real, sistema Bio-Plex (BioRad), sistemas CEQ e SNPstream (Beckman), tecnologia de série de Sonda de Inversão Molecular (por exemplo, GeneChip da Affymetrix) e BeadArray Technologies (por exemplo, os ensaios Illumina GoldenGate e Infinium). Por estes ou outros métodos disponíveis para a pessoa habilitada na técnica, um ou mais alelos em marcadores polimórficos, incluindo microssatélites, SNPs ou outros tipos de marcadores polimórficos, podem ser identificados.

Em certos métodos aqui descritos, um indivíduo que está em uma suscetibilidade aumentada (isto é, risco aumentado) para qualquer

doença ou traço específicos sob estudo, é um indivíduo em quem pelo menos um alelo específico em um ou mais marcadores polimórficos ou haplótipo que conferem suscetibilidade aumentada para a doença ou traço são identificados (isto é, alelos marcadores ou haplótipos em risco). Em um aspecto, o

5 marcador ou haplótipo em risco é um que confere um risco aumentado (ou suscetibilidade) significativa da doença ou traço. Em uma forma de realização, a significância associada com um marcador ou haplótipo é medida por um Risco Relativo (RR). Em uma outra forma de realização, a significância associada com um marcador ou haplotipo é medida por uma razão de chances

10 (OR). Em uma outra forma de realização, a significância é medida por uma porcentagem. Em uma forma de realização, um risco aumentado significativo é medido como um risco (Risco Relativo e/ou Razão de Chances) de pelo menos 1,2, incluindo mas não limitado a: pelo menos 1,2, pelo menos 1,3, pelo menos 1,4, pelo menos 1,5, pelo menos 1,6, pelo menos 1,7, 1,8, pelo

15 menos 1,9, pelo menos 2,0, pelo menos 2,5, pelo menos 3,0, pelo menos 4,0 e pelo menos 5,0. Em uma forma de realização particular, um risco (Risco Relativo e/ou Razão de Chances) de pelo menos 1,2 é significativo. Em uma outra forma de realização particular, um risco de pelo menos 1,3 é significativo. Ainda em uma outra forma de realização, um risco de pelo

20 menos 1,4 é significativo. Em uma outra forma de realização, um Risco Relativo de pelo menos cerca de 1,5 é significativo. Em mais uma outra forma de realização, um aumento significativo no risco de pelo menos cerca de 1,7 é significativo. Entretanto, outros cortes também são considerados, por exemplo pelo menos 1,15, 1,25, 1,35 e assim por diante e tais cortes também estão

25 dentro do escopo da presente invenção. Em outras formas de realização, um aumento significativo no risco é de pelo menos cerca de 20%, incluindo mas não limitado a cerca de 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 150%, 200%, 300% e 500%. Em uma forma de realização particular, um aumento significativo no risco é de

pelo menos 20%. Em outras formas de realização, um aumento significativo no risco é de pelo menos 30%, pelo menos 40%, pelo menos 50%, pelo menos 60%, pelo menos 70%, pelo menos 80%, pelo menos 90% e pelo menos 100%. Outros cortes ou faixas como julgados adequados pela pessoa habilitada na técnica para caracterizar a invenção são entretanto também considerados e estes também estão dentro do escopo da presente invenção.

Um marcador polimórfico ou haplótipo em risco da presente invenção é um onde pelo menos um alelo de pelo menos um marcador ou haplótipo está mais frequentemente presente em um indivíduo em risco para a doença ou traço (por exemplo, arritmia cardíaca ou acidente vascular cerebral) (afetado), comparado com a frequência da sua presença em um grupo de comparação (controle) e em que a presença do marcador ou haplótipo é indicativa de suscetibilidade para a doença ou traço. O grupo de controle em uma forma de realização pode ser uma amostra de população, isto é, uma amostra aleatória da população geral. Em uma outra forma de realização, o grupo de controle é representado por um grupo de indivíduos que são isentos de doença. Tal controle isento de doença em uma forma de realização pode ser caracterizado pela ausência de um ou mais sintomas associados a doença específicos. Em uma outra forma de realização, o grupo de controle isento de doença é caracterizado pela ausência de um ou mais fatores de risco específicos de doença. Tais fatores de risco são em uma forma de realização pelo menos um fator de risco ambiental. Os fatores ambientais representativos são produtos naturais, minerais ou outros produtos químicos que são conhecidos por afetar ou considerados afetar, o risco de desenvolver a doença ou traço específicos. Outros fatores de risco ambientais são fatores de risco relacionados com o estilo de vida, incluindo mas não limitados aos hábitos alimentares e bebidas, localização geográfica do habitat principais e fatores de risco ocupacionais. Em uma outra forma de realização, os fatores de risco são pelo menos um fator de risco genético.

Como um exemplo de um teste simples para a correlação seria um teste exato de Fisher em uma tabela de dois por dois. Dado um grupo de cromossomas, a tabela de dois por dois é construída do número de cromossomas que incluem ambos dos marcadores ou haplótipos, um dos marcadores ou haplótipos mas não o outro e nenhum dos marcadores ou haplótipos.

Em outras formas de realização da invenção, um indivíduo que está em uma suscetibilidade diminuída (isto é, em uma risco diminuído) para a doença ou traço é um indivíduo no qual pelo menos um alelo específico em um ou mais marcadores polimórficos ou haplótipos que conferem suscetibilidade diminuída para a doença ou traço é identificado. Os alelos e/ou haplótipos marcadores que conferem risco diminuído também são ditos serem protetores. Em um aspecto, o marcador ou haplótipo protetores é um que confere um risco diminuído (ou suscetibilidade) significativa da doença ou traço. Em uma forma de realização, o risco diminuído significativa é medido como um Risco Relativo de menos do que 0,9, incluindo mas não limitado a menor do que 0,9, menor do que 0,8, menor do que 0,7, menor do que 0,6, menor do que 0,5, menor do que 0,4, menor do que 0,3, menor do que 0,2 e menor do que 0,1. Em uma forma de realização particular, o risco diminuído significativa é menor do que 0,7. Em uma outra forma de realização, o risco diminuído significativa é menor do que 0,5. Ainda em uma outra forma de realização, o risco diminuído significativa é menor do que 0,3. Em uma outra forma de realização, a diminuição no risco (ou suscetibilidade) é de pelo menos 20%, incluindo mas não limitada a pelo menos 25%, pelo menos 30%, pelo menos 35%, pelo menos 40%, pelo menos 45%, pelo menos 50%, pelo menos 55%, pelo menos 60%, pelo menos 65%, pelo menos 70%, pelo menos 75%, pelo menos 80%, pelo menos 85%, pelo menos 90%, pelo menos 95% e pelo menos 98%. Em uma forma de realização particular, uma diminuição significativa no risco é de pelo menos cerca de 30%. Em uma outra forma de

realização, uma diminuição significativa no risco é de pelo menos cerca de 50%. Em uma outra forma de realização, a diminuição no risco é de pelo menos cerca de 70%. Outros cortes ou faixas como julgados adequados pela pessoa habilitada na técnica para caracterizar a invenção são entretanto também considerados e estes também estão dentro do escopo da presente invenção.

A pessoa habilitada na técnica avaliará que para os marcadores com dois alelos presentes na população que é estudada e em que um alelo é encontrado em frequência aumentada em um grupo de indivíduos com um traço ou doença na população, comparado com controles, o outro alelo do marcador será descoberto na frequência diminuída no grupo de indivíduos com o traço ou doença, comparado com controles. Em um tal caso, um alelo do marcador (aquele encontrado em frequência aumentada em indivíduos com o traço ou doença) será o alelo em risco, enquanto que o outro alelo será um alelo protetor.

Desequilíbrio de ligação

O fenômeno natural de recombinação, que ocorre em média uma vez para cada par de cromossoma durante cada evento meiótico, representa um modo em que a natureza fornece variações na sequência (e função biológica por consequência). foi descoberto que a recombinação não ocorre aleatoriamente no genoma; ao invés, existem variações grandes na frequência de taxas de recombinação, que resulta em regiões pequenas de alta frequência de recombinação (também chamada de ponto de recombinação) e regiões grandes de baixa frequência de recombinação, que são habitualmente aludidas como blocos de desequilíbrio de ligação (LD) (Myers, S. *et al.*, Biochem Soc Trans 34: 526-530 (2006); Jeffreys, A. J., *et al.*, Nature Genet 29: 217-222 (2001); May, C. A., *et al.*, Nature Genet 31: 272-275(2002)).

O desequilíbrio de ligação (LD) refere-se a uma seleção não aleatória de dois elementos genéticos. Por exemplo, se um elemento genético

particular (por exemplo, um alelo de um marcador polimórfico ou um haplótipo) ocorre em uma população em uma frequência de 0,25 (25%) e um outro elemento ocorre em uma frequência de 0,25 (25%), então a ocorrência prognosticada de uma pessoa tendo ambos os elementos é de 0,125 (12,5%), assumindo uma distribuição aleatória dos elementos. Entretanto, se é descoberto que os dois elementos ocorrem juntos em uma frequência mais alta do que 0,125, então os elementos são ditos estar em desequilíbrio de ligação, visto que eles tendem a ser herdados juntos em uma taxa mais alta do que aquela que as suas frequências independentes de ocorrência (por exemplo, frequências de alelo ou haplótipo) prognosticariam. Em termos gerais, LD está em uma maneira geral correlacionado com a frequência de eventos de recombinação entre os dois elementos. As frequências de alelo ou haplótipo podem ser determinadas em uma população pela genotipagem de indivíduos em uma população e determinando a frequência da ocorrência de cada alelo ou haplótipo na população. Para as populações de diplóides, por exemplo, populações humanas, indivíduos tipicamente terão dois alelos para cada elemento genético (por exemplo, um marcador, haplótipo ou gene).

Muitas medidas diferentes foram propostas para avaliar a força de desequilíbrio de ligação (LD). A maioria captura a força de associação entre pares de sítios bialélicos. Duas medidas aos pares importantes de LD são r^2 (algumas vezes denotada Δ^2) e $|D'|$. Ambas as medidas variam de 0 (nenhum desequilíbrio) a 1 (desequilíbrio ‘completo’), mas a sua interpretação é levemente diferente. $|D'|$ é definido em um tal modo que seja igual a 1 se apenas dois ou três dos haplótipos possíveis estão presentes e é < 1 se todos os quatro haplótipos possíveis estão presentes. Portanto, um valor de $|D'|$ que é < 1 indica que a recombinação histórica pode ter ocorrido entre dois sítios (a mutação recorrente também pode fazer com que $|D'|$ seja < 1 , mas para os polimorfismos de nucleotídeo único (SNPs) isto é usualmente considerado como sendo menos provável do que a recombinação). A medida

r^2 representa a correlação estatística entre dois sítios e assume o valor de 1 se apenas dois haplótipos estão presentes.

A medida r^2 é discutivelmente a medida mais relevante para o mapeamento de associação, porque existe uma relação inversa simples entre r^2 e o tamanho da amostra requerido para detectar a associação entre o local de suscetibilidade e SNPs. Estas medidas são definidas para pares de sítios, mas para algumas aplicações uma determinação de quão forte LD é através de uma região inteira que contém muitos sítios polimórficos seria desejável (por exemplo, testando se a força de LD difere significativamente entre populações locais ou cruzadas ou se existe mais ou menos LD em uma região do que prognosticado sob um modelo particular). Medir LD através de uma região não é direto, mas um método é usar a medida r , que foi desenvolvida nas populações genéticas. Em termos gerais, r mede quanta recombinação seria requerida sob um modelo de população particular para gerar o LD que é observado nos dados. Este tipo de método potencialmente também pode fornecer um método estatisticamente rigoroso para o problema de determinar se os dados de LD fornecem evidência quanto a presença de pontos de recombinação. Para os métodos e procedimentos aqui descritos, um valor r^2 significativo pode ser de pelo menos 0,1 tal como pelo menos 0,1, 0,15, 0,2, 0,25, 0,3, 0,35, 0,4, 0,45, 0,5, 0,55, 0,6, 0,65, 0,7, 0,75, 0,8, 0,85, 0,9, 0,91, 0,92, 0,93, 0,94, 0,95, 0,96, 0,97, 0,98, 0,99 ou 1,0. Em uma forma de realização preferida, o valor r^2 significativo pode ser de pelo menos 0,2. Alternativamente, o desequilíbrio de ligação como aqui descrito, refere-se ao desequilíbrio de ligação caracterizado pelos valores de $|D'|$ de pelo menos 0,2, tal como 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,85, 0,9, 0,95, 0,96, 0,97, 0,98, 0,99. Assim, o desequilíbrio de ligação representa uma correlação entre alelos de marcadores distintos. O mesmo é medido pelo coeficiente de correlação ou $|D'|$ (r^2 até 1,0 e $|D'|$ até 1,0). O desequilíbrio de ligação pode ser determinado em uma única população humana, como aqui definido ou pode ser

determinado em uma coleção de amostras que compreendem indivíduos de mais do que uma população humana. Em uma forma de realização da invenção, LD é determinado em uma amostra de uma ou mais das populações HapMap (caucasiana, africana, japonesa, chinesa), como definido (<http://www.hapmap.org>). Em uma tal forma de realização, LD é determinado na população CEU das amostras de HapMap. Em uma outra forma de realização, LD é determinado na população YRI. Ainda em uma outra forma de realização, LD é determinado em amostras da população islandesa.

Se todos os polimorfismos no genoma fossem idênticos ao nível de população, então cada um deles necessitaria ser investigado em estudos de associação. Entretanto, devido ao desequilíbrio de ligação entre polimorfismos, polimorfismos firmemente ligados são fortemente correlacionados, o que reduz o número de polimorfismos que precisam ser investigados em um estudo de associação para observar uma associação significativa. Uma outra consequência de LD é que muitos polimorfismos podem dar um sinal de associação devido ao fato de que estes polimorfismos são fortemente correlacionados.

Os mapas LD genômicos foram gerados através do genoma e tais mapas LD foram propostos para servir como matriz para o mapeamento de doença-genes (Risch, N. & Merkiangas, K, *Science* 273: 1516-1517 (1996); Maniatis, N., *et al.*, *Proc Natl Acad Sci USA* 99: 2228-2233 (2002); Reich, DE *et al.*, *Nature* 411: 199-204 (2001)).

É agora estabelecido que muitas porções do genoma humano podem ser quebradas em séries de blocos de haplótipo separados contendo uns poucos haplótipos comuns; para este blocos, os dados de desequilíbrio de ligação fornecem pouca evidência indicando recombinação (ver, por exemplo, Wall, J. D. e Pritchard, J. K., *Nature Reviews Genetics* 4: 587-597 (2003); Daly, M. *et al.*, *Nature Genet.* 29: 229-232 (2001); Gabriel, S. B. *et al.*, *Science* 296: 2225-2229 (2002); Patil, N. *et al.*, *Science* 294: 1719-1723

(2001); Dawson, E. *et al.*, Nature 418: 544-548 (2002); Phillips, M. S. *et al.*, Nature Genet. 33: 382-387 (2003)).

Existem dois métodos principais para definir estes blocos de haplótipo: Blocos podem ser definidos como regiões de DNA que têm diversidade de haplótipo limitada (ver, por exemplo, Daly, M. *et al.*, Nature Genet. 29: 229-232 (2001); Patil, N. *et al.*, Science 294: 1719-1723 (2001); Dawson, E. *et al.*, Nature 418: 544-548 (2002); Zhang, K. *et al.*, Proc. Natl. Acad. Sci. USA 99: 7335-7339 (2002)) ou como regiões entre zonas de transição tendo recombinação histórica extensiva, identificado usando desequilíbrio de ligação (ver, por exemplo, Gabriel, S. B. *et al.*, Science 296: 2225-2229 (2002); Phillips, M. S. *et al.*, Nature Genet. 33: 382-387 (2003); Wang, N. *et al.*, Am. J. Hum. Genet. 71: 1227-1234 (2002); Stumpf, M. P. e Goldstein, D. B., Curr. Biol. 13: 1-8 (2003)). Mais recentemente, um mapa de escala fina de taxas de recombinação e pontos correspondentes através do genoma humano foi gerado (Myers, S., *et al.*, Science 310: 321-32324 (2005); Myers, S. *et al.*, Biochem Soc Trans 34: 526-530 (2006)). O mapa revela a enorme variação na recombinação através do genoma, com taxas de recombinação tão altas quanto 10 a 60 cM/Mb nos pontos, embora mais próximo a 0 nas regiões intervenientes, que assim representam regiões de diversidade de haplótipo limitada e LD alta. O mapa portanto pode ser usado para definir blocos de haplótipo/Blocos LD como regiões flanqueadas pelos pontos de recombinação. Como aqui usado, os termos “bloco de haplótipo” ou “bloco LD” incluem blocos definidos por qualquer uma das características descritas acima ou outros métodos alternativos usados pelas pessoas habilitadas na técnica para definir tais regiões.

Os blocos de haplótipo podem ser usados para mapear associações entre a situação de fenótipo e haplótipo, usando marcadores únicos ou haplótipos que compreendem uma pluralidade de marcadores. Os haplótipos principais podem ser identificados em cada bloco de haplótipo e

depois um conjunto de SNPs “rotuladores” ou marcadores (o conjunto menor de SNPs ou marcadores necessários para distinguir entre os haplótipos) podem ser depois identificados. Estes SNPs rotuladores ou marcadores podem ser depois usados na avaliação de amostras de grupos de indivíduos, de modo a identificar a associação entre fenótipo e haplótipo. Se desejado, blocos de haplótipo vizinhos podem ser avaliados concorrentemente, visto que também pode existir desequilíbrio de ligação entre os blocos de haplótipo.

Assim tem se tornado evidente que para qualquer associação observada dada a um marcador polimórfico no genoma, é provável que marcadores adicionais no genoma também mostrem associação. Isto é uma consequência natural da distribuição não uniforme de LD através do genoma, como observado pela grande variação em taxas de recombinação. Os marcadores usados para detectar a associação assim em um sentido representa “rótulos” para uma região genômica (isto é, um bloco de haplótipo ou bloco LD) que esteja em associação com uma dada doença ou traço e como tais são úteis para o uso nos métodos e kits da presente invenção. Uma ou mais variantes ou mutações causativas (funcionais) podem residir dentro da região descoberta estar em associação com a doença ou traço. Tais variantes podem conferir um Risco Relativo (RR) ou Razão de Chances (OR) mais altos do que os observados para os marcadores de rotulação usados para detectar a associação. A presente invenção refere-se assim ao marcadores usados para detectar a associação com a doença, como aqui descrito, assim como marcadores em desequilíbrio de ligação com os marcadores. Assim, em certas formas de realização da invenção, marcadores que estão em LD com os marcadores e/ou haplótipos da invenção, como aqui descritos, podem ser usados como marcadores substitutos. Os marcadores substitutos têm em uma forma de realização valores de Risco Relativo (RR) e/ou Razão de Chances (OR) menores do que para os marcadores ou haplótipos inicialmente descobertos estar em associação com a doença, como aqui descrito. Em outras

formas de realização, os marcadores substitutos têm valores RR ou OR maiores do que aqueles inicialmente determinados para os marcadores inicialmente descobertos estar em associação com a doença, como aqui descrito. Um exemplo de uma tal forma de realização seria uma variante rara ou em relativamente rara ($< 10\%$ de frequência de população alélica) em LD com uma variante mais comum ($> 10\%$ de frequência de população) inicialmente descoberto estar em associação com a doença, tais como as variantes aqui descritas. A identificação e uso tais marcadores para detectar a associação descoberta pelos inventores como aqui descrita podem ser realizados pelos métodos de rotina bem conhecidos pela pessoa habilitada na técnica e estão portanto dentro do escopo da presente invenção.

É possível que certos marcadores polimórficos em desequilíbrio de ligação com os marcadores aqui mostrados estar associados com a arritmia cardíaca (por exemplo, fibrilação atrial e flúter atrial) e acidente vascular cerebral estão localizados fora dos limites físicos do bloco LD C04 como aqui definido pela sequência apresentada na SEQ ID NO: 50. Isto é uma consequência das taxas de recombinação histórica na região em questão, que pode ter levado a uma região de LD forte (o bloco LD), com marcadores residuais fora do bloco em LD com marcadores dentro do bloco. Tais marcadores também estão dentro do escopo da presente invenção, visto que também são úteis para praticar a invenção em virtude de sua relação genética com os marcadores aqui mostrados estar associados com a arritmia cardíaca e acidente vascular cerebral. Os exemplos de tais marcadores são mostrados na Tabela 18 (rs7668322 (SEQ ID NO: 46), rs2197815 (SEQ ID NO: 47), rs6831623 (SEQ ID NO: 48), rs2595110 (SEQ ID NO: 49))

Determinação da frequência de haplótipo

As frequências de haplótipos em grupos de paciente e controle podem ser estimadas usando um algoritmo de expectativa-maximização (Dempster A. *et al.*, J. R. Stat. Soc. 8, 39: 1-38 (1977)). Uma implementação

deste algoritmo que pode manusear genótipos perdidos e inconstantes com a fase pode ser usada. Sob a hipótese nula, os pacientes e os controles são assumidos ter frequências idênticas. Usando um método de probabilidades, uma hipótese alternativa é testada, onde um haplótipo em risco candidato, que pode incluir os marcadores aqui descritos, é deixado ter uma frequência mais alta em pacientes do que nos controles, enquanto que as razões das frequências de outros haplótipos são assumidos ser as mesmas em ambos os grupos. As probabilidades são maximizadas separadamente sob ambas as hipóteses e uma estatística da razão de probabilidade 1-df correspondente é usada para avaliar a significância estatística.

Para examinar quanto aos marcadores e haplótipos em risco e protetores dentro de uma região de ligação, por exemplo, a associação de todas as combinações possíveis de marcadores genotipados é estudada, contanto que estes marcadores transponham uma região prática. Os grupos de paciente e controle combinados podem ser aleatoriamente divididos em dois conjuntos, iguais no tamanho ao grupo original de pacientes e controles. A análise de marcador e haplótipo é depois repetida e o valor p mais significativo registrado é determinado. Este esquema de randomização pode ser repetido, por exemplo, mais de 100 vezes para construir uma distribuição empírica de valores p. Em uma forma de realização preferida, um valor p de $< 0,05$ é indicativo de uma associação de marcador e/ou haplótipo significativa.

Análise de Haplótipo

Um método geral para a análise de haplótipo envolve o uso da inferência com base na probabilidade aplicada ao Modelos Abridados (Gretarsdottir S., *et al.*, Nat. Genet. 35: 131-38 (2003)). O método é implementado no programa NEMO, que permite muitos marcadores polimórficos, SNPs e microsatélites. O método e software são especificamente planejados para os estudos de controle de caso onde o propósito é identificar grupos de haplótipo que conferem riscos diferentes.

Este também é uma ferramenta para estudar estruturas LD. Em NEMO, as estimativas de probabilidade máxima, as razões de probabilidade e os valores p são calculados diretamente, com a ajuda do algoritmo EM, para os dados observados tratando-os como um problema de dados perdidos.

Embora testes da razão de probabilidade com base nas probabilidades computadas diretamente para os dados observados, que têm capturado a perda de informação devido à incerteza na fase e genótipos perdidos, podem ser confiados para dar valores p válidos, seria ainda de interesse saber quanta informação foi perdida devido à informação que está incompleta. A medida de informação para a análise de haplótipo é descrita em Nicolae e Kong (Technical Report 537, Department of Statistics, University of Statistics, University of Chicago; Biometrics, 60(2): 368-75 (2004)) como uma extensão natural de medidas de informação definidas para a análise de ligação e é implementada em NEMO.

Para a associação de marcador único a uma doença, o teste exato de Fisher pode ser usado para calcular valores p de dois lados para cada alelo individual. Usualmente, todos os valores p são apresentados não ajustados para comparações múltiplas a menos que especificamente indicado. As frequências apresentadas (para microssatélites, SNPs e haplótipos) são frequências alélicas como oposto às frequências do carregador. Para minimizar qualquer tendência devido à relacionabilidade dos pacientes que foram recrutados como famílias para a análise de ligação, as relações de primeiro e segundo grau podem ser eliminadas da lista de paciente. Além disso, o teste pode ser repetido para associação corrigindo quanto a qualquer relacionabilidade remanescente entre os pacientes, pela extensão de um procedimento de ajuste de variação descrito em Risch, N. & Teng, J. (Genome Res., 8: 1273-1288 (1998)), reunião de DNA (*ibid*) para *sibships* de modo que possa ser aplicado às relações familiares gerais e apresenta valores p tanto ajustados quanto não ajustados para comparação. As diferenças são

em uma maneira geral muito pequenas como esperado. Para avaliar a significância da associação de marcador único corrigido para teste múltiplo nós podemos realizar um teste de randomização usando os mesmos dados de genótipo. Os grupos de pacientes e controles podem ser randomizados e a análise de associação refeita múltiplas vezes (por exemplo, até 500.000 vezes) e o valor p é a fração de replicações que produziu um valor p para algum alelo marcador que é mais baixo do que ou igual ao valor p que nós observamos usando os grupos de paciente e controle originais.

Para as análises tanto de marcador quanto de haplótipo únicos, o Risco Relativo (RR) e o risco atribuível à população (PAR) podem ser calculados assumindo um modelo multiplicativo (modelo Risco Relativo de haplótipo) (Terwilliger, J. D. & Ott, J., Hum. Hered. 42: 337-46 (1992) e Falk, C. T. & Rubinstein, P, Ann. Hum. Genet. 51 (Pt 3): 227-33 (1987)), isto é, que os riscos dos dois alelos/haplótipos que uma pessoa carrega multiplicam-se. Por exemplo, se RR é o risco de um em relação a um, então o risco de uma pessoa homozigota AA será RR vezes aquele de um Aa heterozigoto e RR^2 vezes aquele de um aa homozigoto. O modelo multiplicativo tem uma boa propriedade que simplifica a análise e computações — os haplótipos são independentes, isto é, em equilíbrio de Hardy-Weinberg, dentro da população afetada assim como dentro da população de controle. Como uma consequência, as contagens de haplótipo dos afetados e controles cada um tem distribuições multinomiais, mas com frequências de haplótipo diferentes sob a hipótese alternativa. Especificamente, para dois haplótipos, h_i e h_j , $\text{risco}(h_i)/\text{risco}(h_j) = (f_i/p_i)/(f_j/p_j)$, onde f e p denotam, respectivamente, frequências na população afetada e na população de controle. Enquanto existe alguma perda de energia se o modelo verdadeiro não é multiplicativo, a perda tende a ser branda exceto para casos extremos. Mais importantemente, os valores p são sempre válidos visto que eles são computados com respeito à hipótese nula.

Desequilíbrio de ligação Usando NEMO

LD entre pares de marcadores pode ser calculado usando a definição padrão de D' e r^2 (Lewontin, R., Genetics 49: 49-67 (1964); Hill, W. G. & Robertson, A. Theor. Appl. Genet. 22: 226-231 (1968)). Usando NEMO, as frequências das duas combinações de alelo marcador são estimadas pela probabilidade máxima e o desvio de equilíbrio de ligação é avaliado por um teste de razão de probabilidade. As definições de D' e r^2 são prolongadas para incluir microssatélites pelo cálculo em média em relação aos valores para todas as combinações de alelo possíveis dos dois marcadores pesados pelas probabilidades de alelo marginal. Quando da plotagem de todas as combinações de marcador para elucidar a estrutura LD em uma região particular, nós plotamos D' no canto superior esquerdo e o valor p no canto inferior direito. Nas plotagens de LD os marcadores podem ser plotados equidistantes ao invés de acordo com a sua localização física, se desejado.

Avaliação de Risco e Diagnósticos

Como aqui descrito, certos marcadores polimórficos e haplótipos que compreendem tais marcadores são descobertos ser úteis para a avaliação de risco da arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) ou acidente vascular cerebral. A avaliação de risco pode envolver o uso dos marcadores para diagnosticar uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) ou acidente vascular cerebral. Os alelos particulares de marcadores polimórficos são encontrados mais frequentemente em indivíduos com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) ou acidente vascular cerebral, do que em indivíduos sem diagnose de arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) ou acidente vascular cerebral. Portanto, estes alelos marcadores têm valor preditivo para detectar a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral ou uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou

flúter atrial) ou acidente vascular cerebral, em um indivíduo. Os marcadores de rotulação dentro dos blocos de haplótipo ou Blocos LD que compreendem marcadores em risco, tais como os marcadores da presente invenção, podem ser usados como substitutos para outros marcadores e/ou haplótipos dentro do

5 bloco de haplótipo ou bloco LD. Marcadores com valores de r^2 iguais a 1 são substitutos perfeitos para as variantes em risco, isto é, genótipos para um marcador perfeitamente prognostica genótipos para o outro. Marcadores com valores de r^2 menores do que 1 também podem ser substitutos para a variante de risco ou alternativamente representam variantes com valores de Risco

10 Relativo tão alto ou possivelmente ainda mais alto do que a variante de risco. A variante de risco identificada pode não ser a própria variante funcional, mas está neste caso em desequilíbrio de ligação com a variante funcional verdadeira. A presente invenção abrange a avaliação de tais marcadores substitutos para os marcadores como aqui divulgados. Tais marcadores são

15 anotados, mapeados e listados em bases de dados públicas, como bem conhecido pela pessoa habilitada ou alternativamente podem ser facilmente identificados pelo sequenciamento da região ou uma parte da região identificada pelos marcadores da presente invenção em um grupo de indivíduos e identificam polimorfismos no grupo resultante de sequências.

20 Como uma consequência, a pessoa habilitada na técnica pode facilmente e sem experimentação indevida genotipar marcadores substitutos em desequilíbrio de ligação com os marcadores e/ou haplótipos como aqui descritos. A rotulação ou marcadores substitutos em LD com as variantes em risco detectadas, também têm valor preditivo para detectar a associação para a

25 arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral ou uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) ou acidente vascular cerebral, em um indivíduo. Esta rotulação ou marcadores substitutos que estão em LD com os marcadores da presente invenção também podem incluir outros marcadores

que distinguem entre haplótipos, visto que estes similarmente têm valor preditivo para detectar a suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral.

Os marcadores e haplótipos da invenção, por exemplo, os marcadores apresentados nas Tabelas 5 e 9, assim como marcadores em desequilíbrio de ligação com estes podem ser úteis para a avaliação de risco e propósitos de diagnóstico para, sozinho ou em combinação. Assim, mesmo em casos onde o aumento no risco pelas marcadores individuais é relativamente modesto, isto é, da ordem de 10 a 30%, a associação pode ter implicações significantes. Assim, variantes relativamente comuns podem ter contribuição significativa para o risco global (o Risco Atribuível à População é alto) ou a combinação de marcadores pode ser usada para definir grupos de indivíduos que, com base no risco combinado dos marcadores, está em risco combinado significativo de desenvolver a doença.

Assim, em uma forma de realização da invenção, uma pluralidade de variantes (marcadores e/ou haplótipos) é usada para a avaliação global de risco. Estas variantes estão em uma forma de realização selecionada das variantes como aqui divulgadas. Outras formas de realização incluem o uso das variantes da presente invenção em combinação com outras variantes conhecidas serem úteis para diagnosticar uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Em tais formas de realização, a situação genotípica de uma pluralidade de marcadores e/ou haplótipos é determinada em um indivíduo e a situação do indivíduo comparada com a frequência da população das variantes associadas ou a frequência das variantes em pacientes clinicamente saudáveis, tais como pacientes igualados em idade e igualados em sexo. Métodos conhecidos na técnica, tais como a análise multivariada ou análise de risco em comum, podem ser subsequentemente usados para determinar o risco global conferido com base na situação genotípica nos locais múltiplos. A avaliação

de risco com base em tal análise pode ser subsequentemente usada nos métodos e kits da invenção, como aqui descrito.

Como descrito no acima, a estrutura de bloco de haplótipo do genoma humano tem o efeito que um número grande de variantes (marcadores e/ou haplótipos) em desequilíbrio de ligação com a variante originalmente associada com uma doença ou traço pode ser usado como marcadores substitutos para avaliar a associação com a doença ou traço. O número de tais marcadores substitutos dependerão de fatores tais como a taxa de recombinação histórica na região, a frequência mutacional na região (isto é, o número de sítios polimórficos ou marcadores na região) e o grau de LD (tamanho do bloco LD) na região. Estes marcadores estão usualmente localizados dentro dos limites físicos do bloco LD ou bloco de haplótipo em questão como definido usando os métodos aqui descritos ou pelos outros métodos conhecidos pela pessoa habilitada na técnica. Entretanto, algumas vezes a associação de marcador e haplótipo é descoberta prolongar-se além dos limites físicos do bloco de haplótipo como definido. Tais marcadores e/ou haplótipos nestes casos também podem ser usados como marcadores e/ou haplótipos substitutos para os marcadores e/ou haplótipos que fisicamente residem dentro do bloco de haplótipo como definido. Como uma consequência, marcadores e haplótipos em LD (tipicamente caracterizados por r^2 maior do que 0,1, tal com r^2 maior do que 0,2, incluindo r^2 maior do que 0,3, também incluindo r^2 maior do que 0,4) com os marcadores e haplótipos da presente invenção também estão dentro do escopo da invenção, mesmo se eles estão fisicamente localizados além dos limites do bloco de haplótipo como definido. Isto inclui marcadores que são aqui descritos (por exemplo, Tabelas 5 e 9), mas também podem incluir outros marcadores que estão em LD forte (caracterizados por r^2 maior do que 0,1 ou 0,2 e/ou $|D'| > 0,8$) com um ou mais dos marcadores listados nas Tabelas 5 e 9.

Para os marcadores de SNP aqui descritos, o alelo oposto ao

alelo descoberto estar em excesso em pacientes (alelo em risco) é descoberto em frequência diminuída em pacientes com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Estes marcadores e haplótipos em LD e/ou que compreendem tais marcadores, são assim protetores para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Isto é, eles conferem um risco ou suscetibilidade diminuídos de indivíduos que carregam estes marcadores e/ou haplótipos desenvolverem arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral.

Certas variantes da presente invenção, incluindo certos haplótipos compreendem, em alguns casos, uma combinação de vários marcadores genéticos, por exemplo, SNPs e microssatélites. A detecção de haplótipos pode ser realizada pelos métodos conhecidos na técnica e/ou aqui descritos para detectar sequências em sítios polimórficos. Além disso, a correlação entre certos haplótipos ou conjuntos de marcadores e fenótipo de doença pode ser verificada usando técnicas padrão. Um exemplo representativo de um teste simples para a correlação seria um teste exato de Fisher em uma tabela dois por dois.

Em formas de realização específicas, um marcador ou haplótipo descoberto estar associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral, (por exemplo, marcadores como listados na Tabela 5 (Tabelas 5A e 5B), Tabela 9 e/ou Tabela 19 e marcadores em desequilíbrio de ligação com estes) é um em que o alelo ou haplótipo marcadores estão mais frequentemente presentes em um indivíduo em risco para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral (afetado), comparado com a frequência da sua presença em um indivíduo saudável (controle), em que a presença do alelo ou haplótipo marcadores é indicativa de arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral

ou uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Em outras formas de realização, marcadores em desequilíbrio de ligação em risco com um ou mais marcadores descobertos estar associados com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral (por exemplo, alelos marcadores como listados nas Tabelas 5A e 5B e marcadores em desequilíbrio de ligação com estes) são marcadores de rotulação que estão mais frequentemente presentes em um indivíduo em risco para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral (afetado), comparado com a frequência da sua presença em um indivíduo saudável (controle), em que a presença dos marcadores de rotulação é indicativa de suscetibilidade aumentada para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Em uma outra forma de realização, alelos marcadores em risco (isto é, que conferem suscetibilidade aumentada) em desequilíbrio de ligação com um ou mais marcadores descobertos estar associados com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. (por exemplo, alelos marcadores como listados nas Tabelas 5A e 5B e marcadores em desequilíbrio de ligação com estes), são marcadores que compreendem um ou mais alelos que estão mais frequentemente presentes em um indivíduo em risco para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral comparado com a frequência da sua presença em um indivíduo saudável (controle), em que a presença dos marcadores é indicativa de suscetibilidade aumentada para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral.

População de estudo

Em um sentido geral, os métodos e kits da invenção podem ser utilizados de amostras contendo DNA genômico de qualquer fonte, isto é, qualquer indivíduo. Em formas de realização preferidas, o indivíduo é um

indivíduo humano. O indivíduo pode ser um adulto, criança ou feto. A presente invenção também fornece a avaliação de marcadores e/ou haplótipos em indivíduos que são membros de uma população alvo. Uma tal população alvo é em uma forma de realização uma população ou grupo de indivíduos em

5 risco de desenvolver a doença, com base em outros fatores genéticos, biomarcadores, parâmetros biofísicos (por exemplo, peso, BMD, pressão sanguínea) ou saúde geral e/ou estilo de vida (por exemplo, história de doença ou doenças relacionadas, diagnóstico anterior de doença, história familiar de doença).

10 A invenção fornece formas de realização que incluem indivíduos de subgrupos de idade específica, tais com aqueles acima da idade de 40, acima da idade de 45 ou acima da idade de 50, 55, 60, 65, 70, 75, 80 ou 85. Outras formas de realização da invenção dizem respeito a outros grupos de idade, tais com indivíduos de idade menor do que 85, tal com menor do

15 que a idade de 80, menor do que a idade de 75 ou menor do que idade de 70, 65, 60, 55, 50, 45, 40, 35 ou idade de 30. Outras formas de realização dizem respeito a indivíduos com idade no início da doença em qualquer uma das faixas de idade descritas acima. Também é considerado que uma faixa de idades pode ser relevante em certas formas de realização, tal com a idade no

20 início maior do que a idade de 45 porém menor do que a idade de 60. Outras faixas de idade entretanto também são consideradas, incluindo todas as faixas de idade agrupadas pelos valores de idade listados acima.

Outras formas de realização relacionadas com os indivíduos com idade no início da doença na idade particular ou faixa de idade. Assim, é

25 conhecido que fatores de predisposição, genéticos e não genéticos, podem afetar em qual idade um indivíduo desenvolve uma doença. Para distúrbios cardiovasculares, incluindo arritmias cardíacas e acidente vascular cerebral, os fatores de risco comuns podem influenciar se e em qual idade, um indivíduo desenvolve a doença. Algumas formas de realização da invenção

portanto dizem respeito à idade no início ou idade no diagnóstico de arritmia cardíaca (por exemplo, fibrilação atrial e/ou flúter atrial) ou acidente vascular cerebral em uma certa faixa de idade. Em uma forma de realização, os indivíduos em risco para desenvolver arritmia cardíaca (por exemplo, fibrilação atrial e/ou flúter atrial) ou acidente vascular cerebral têm a idade no início ou idade no diagnóstico acima da idade de 40. Em outras formas de realização, os indivíduos têm idade no início ou idade no diagnóstico acima da idade de 45 ou acima da idade de 50, 55, 60, 65, 70, 75, 80 ou 85. Outras formas de realização da invenção dizem respeito aos indivíduos que têm uma idade no início ou idade no diagnóstico em idade menor do que 85, tal com menor do que a idade de 80, menor do que a idade de 75 ou menor do que a idade de 70, 65, 60, 55, 50, 45, 40, 35 ou a idade de 30. Uma forma de realização preferida inclui indivíduos diagnosticados com fibrilação atrial ou flúter atrial ou acidente vascular cerebral abaixo da idade de 80. Uma outra forma de realização preferida diz respeito a indivíduos diagnosticados com fibrilação atrial ou flúter atrial ou acidente vascular cerebral abaixo da idade de 70. Uma outra forma de realização preferida, diz respeito a indivíduos diagnosticados com fibrilação atrial ou flúter atrial ou acidente vascular cerebral abaixo da idade de 60. Ainda uma outra forma de realização preferida diz respeito a indivíduos diagnosticados com fibrilação atrial ou flúter atrial ou acidente vascular cerebral abaixo da idade de 50. Outras formas de realização dizem respeito a indivíduos com idade no início da doença em faixas de idade específicas, descritas acima. Também é considerado que uma faixa de idades pode ser relevante em certas formas de realização, tal com a idade no início maior do que a idade de 45 porém menor do que a idade de 60, idade no início na idade maior do que 60 e menor do que a idade de 70, idade no início na idade maior do que 70 e menor do que 80 ou idade no início na idade maior do que 60 e menor do que 80. Outras faixas de idade entretanto também são consideradas, incluindo todas as faixas

de idade agrupadas pelos valores de idade listados acima.

A invenção além disso diz respeito a indivíduos de qualquer sexo, masculino ou feminino. Esta também fornece formas de realização que dizem respeito a pacientes humanos que são de uma ou mais populações humanas incluindo, mas não limitadas à, Banto, Mandenk, Yoruba, San, Mbuti, Pigmeia, Órcades, Adygel, Russa, Sarda, Toscana, Mozabite, Beduína, Drusa, Palestina, Balochi, Brahui, Makrani, Sindi, Pathan, Burusho, Hazara, Uigur, Kalash, Han, Dai, Daur, Hezhen, Lahu, Miao, Orogen, She, Tujia, Tu, Xibo, Yi, Mongoliana, Naxi, Cambojana, Japonesa, Yakut, Melanésia, Papuásia, Karitinan, Surui, Colombiana, Maia e Pima. A invenção também diz respeito às populações Européias, populações Americanas, populações Eurasianas, populações Asiáticas, populações Asiáticas Central/Meridional, populações Asiáticas Orientais, populações do oriente Médio, populações Africanas, populações Hispânicas e populações Oceânicas. As populações Européias incluem, mas não são limitadas às populações Sueca, Norueguesa, Filandesa, Russa, Dinamarquesa, Islandesa, Irlandesa, Celta, Inglesa, Escocesa, Holandesa, Belga, Francesa, Alemã, Espanhola, Portuguesa, Italiana, Polonesa, Búlgara, Eslava, Sérvia, Bósnia, Checa, Grega e Turca.

Em uma forma de realização preferida, a invenção diz respeito às populações que incluem a ancestralidade africana negra tais com populações que compreendem pessoas de descendência ou linhagem africana. A ancestralidade africana negra pode ser determinada pela auto denominação como Africano-Americanos, Afro-Americanos, Americanos Pretos, sendo um membro da raça preta ou sendo um membro da raça negra. Por exemplo, Americanos Africanos ou Americanos Pretos são aquelas pessoas que vivem na América do Norte e tendo origens em qualquer um dos grupos raciais pretos da África. Em um outro exemplo, as pessoas auto-denominadas de ancestralidade africana preta podem ter pelo menos um parente de ancestralidade africana preta ou pelo menos um avô de ancestralidade africana

preta.

A contribuição racial em pacientes individuais também pode ser determinada pela análise genética. A análise genética de ancestralidade pode ser realizada usando marcadores de microssatélite não ligados tais com aqueles demonstrado em Smith *et al.* (Am J Hum Genet 74, 1001-13 (2004)).

Em certas formas de realização, a invenção diz respeito a marcadores e/ou haplótipos identificados em populações específicas, como descritos acima. A pessoa habilitada na técnica avaliará que as medidas de desequilíbrio de ligação (LD) podem dar resultados diferentes quando aplicadas em populações diferentes. Isto é devido à história de população diferente de populações humanas diferentes assim como pressões seletivas diferenciais que podem ter levado às diferenças no LD em regiões genômicas específicas. Também é bem conhecido para a pessoa habilitada na técnica que certos marcadores, por exemplo marcadores de SNP, são polimórficos em uma população mas não em uma outra. A pessoa habilitada na técnica entretanto aplicará os métodos disponíveis e como aqui divulgado ?? para praticar a presente invenção em qualquer população humana dada. Isto pode incluir a avaliação de marcadores polimórficos na região LD da presente invenção, de modo a identificar aqueles marcadores que dão associação mais forte dentro da população específica. Assim, as variantes em risco da presente invenção podem residir em fundos de haplótipo diferentes e em frequências diferentes em várias populações humanas. Entretanto, utilizando os métodos conhecidos na técnica e os marcadores da presente invenção, a invenção pode ser praticada em qualquer população humana dada.

Utilidade do Teste Genético

A pessoa habilitada na técnica avaliará e entenderá que as variantes aqui descritas em uma maneira geral, por elas mesmas, não fornecem uma identificação absoluta de indivíduos que desenvolverão a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente

vascular cerebral. As variantes aqui descritas entretanto indicam a probabilidade aumentada e/ou diminuída de que os indivíduos carreguem as variantes em risco ou protetoras da invenção desenvolverão sintomas associados com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Esta informação é entretanto extremamente valiosa por si só, como esboçado em mais detalhes abaixo, visto poder ser usada, por exemplo, para iniciar medidas preventivas em um estágio inicial, realizar exames físicos e/ou mentais regulares para monitorar a progressão e/ou aparecimento de sintomas ou para programar exames em um intervalo regular para identificar a condição em questão, de modo a ser capaz de aplicar tratamento em um estágio inicial.

O conhecimento em torno de uma variante genética que confere um risco de desenvolver arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral oferece a oportunidade de aplicar um teste genético para distinguir entre indivíduos com risco aumentado de desenvolver a doença (isto é, carregadores da variante de risco) e aqueles com risco diminuído de desenvolver a doença (isto é, carregadores da variante protetor). Os valores de núcleo de teste genético, para indivíduos que pertencem a ambos dos grupos mencionados acima, são as possibilidades de ser capaz de diagnosticar a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral ou uma predisposição para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral em um estágio inicial e fornecer informação para o médico acerca do prognóstico de arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral de modo a ser capaz de aplicar o tratamento mais apropriado.

Os indivíduos com uma história familiar de arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral e portadores de variantes em risco podem se beneficiar do teste genético visto

que o conhecimento da presença de um fator de risco genético ou evidência para risco aumentado de ser um portador de um ou mais fatores de risco, pode fornecer incentivo aumentado para implementar um estilo de vida mais saudável, evitando-se ou minimizando-se fatores de risco ambientais conhecidos para as doenças cardiovasculares relacionadas com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. O teste genético de pacientes com arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral além disso pode dar informação valiosa a cerca da causa primária da doença e pode ajudar ao médico na seleção das melhores opções de tratamento e medicação para cada indivíduo.

A presente invenção além disso diz respeito à avaliação de risco para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral, incluindo determinar se um indivíduo está em risco para desenvolver arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Os marcadores polimórficos da presente invenção podem ser usados sozinhos ou em combinação, assim como em combinação com outros fatores, incluindo outros fatores de risco genético ou biomarcadores, para a avaliação de risco de um indivíduo para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Muitos fatores conhecidos afetar a predisposição de um indivíduo para desenvolver risco de doença cardiovascular são fatores de suscetibilidade para as arritmias cardíacas (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral e são conhecidos pela pessoa habilitada na técnica e podem ser utilizados em tal avaliação. Estes incluem, mas não são limitados a, idade, gênero, a situação do ato de fumar, atividade física, razão da circunferência da cintura para quadril, história familiar de arritmia cardíaca (em particular fibrilação atrial e/ou flúter atrial) e/ou acidente vascular cerebral, arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou

acidente vascular cerebral anteriormente diagnosticado, obesidade, hipertrigliceridemia, colesterol HDL baixo, hipertensão, pressão sanguínea elevada, níveis de colesterol, colesterol HDL, colesterol LDL, triglicerídeos, níveis da apolipoproteína A1 e B, níveis de fibrinogênio, ferritina, proteína reativa C e leucotrieno. Os biomarcadores particulares que foram associados com a Fibrilação atrial/Flúter atrial e acidente vascular cerebral são debatidos em Allard *et al.* (Clin Chem 51: 2043-2051 (2005) e Becker (J Thromb Trombolys 19: 71-75 (2005)). Estes incluem, mas não são limitados a, dímero D da fibrina, fragmento de ativação da protrombina 1,2 (F1,2), complexos de trombina-antitrombina III (TAT), fibrinopeptídeo A (FPA), fosfolipase associada à lipoproteína A2 (lp-PLA2), beta-tromboglobulina, fator de plaqueta 4, P-selectina, Fator de von Willebrand, peptídeo pró-natriurético (BNP), metaloproteinase de matriz 9 (MMP-9), PARK7, difosfato cinase de nucleosídeo (NDKA), tau, enolase específica de neurônio, fator de crescimento neurotrófico tipo B, proteína astrogial S-100b, proteína ácida fibrilar glial, proteína reativa C, amiloide A sérico, metaloproteinase de matriz 9, moléculas de adesão celular vascular e intracelular, fator alfa de necrose de tumor e interleucinas, incluindo interleucina 1, 6 e 8). As células progenitoras circulantes também foram implicadas como sendo biomarcadores úteis para o AF. Em formas de realização particulares, mais do que um biomarcador é determinado para um indivíduo e combinado com resultados de uma determinação de pelo menos um marcador polimórfico como aqui descrito. Preferivelmente, o biomarcador é medido no plasma ou soro do indivíduo. Alternativamente, o biomarcador é determinado em outros tecidos adequados contendo quantidades mensuráveis do biomarcador e tais formas de realização também estão dentro do escopo da invenção.

Os métodos conhecidos na técnica podem ser usados para a avaliação de risco global, incluindo análises multivariadas ou logística de regressão.

A fibrilação atrial é uma doença de enorme significância tanto para o paciente individual quanto para o sistema de saúde pública como um todo. Esta pode ser uma condição permanente mas também pode ser paroxísmica e recorrente caso este em que a mesma pode ser muito desafiadora para diagnosticar. A complicação mais devastadora da fibrilação atrial e flúter atrial é a ocorrência do acidente vascular cerebral debilitante. Importantemente o risco de acidente vascular cerebral é igual na fibrilação atrial permanente e paroxísmica. Foi repetidamente mostrado que a terapia com a anticoagulação com varfarina pode significativamente reduzir o risco do primeiro ou episódios adicionais de acidente vascular cerebral no cenário da fibrilação atrial. Por essa razão, a anticoagulação com varfarina é a terapia padrão para quase todos os pacientes com fibrilação atrial para a prevenção de acidente vascular cerebral, tenham eles a do tipo permanente ou paroxísmico. Os únicos pacientes para os quais a varfarina não é fortemente recomendada são aqueles mais jovens do que 65 anos de idade que são considerados de baixo risco, isto é, eles não têm nenhuma doença cardíaca orgânica, incluindo, nem hipertensão nem doença da artéria coronária, nenhuma história anterior de acidente vascular cerebral ou Ataques Isquêmicos Transitórios e nenhuma diabetes. Este grupo tem um risco mais baixo de acidente vascular cerebral e a prevenção do acidente vascular cerebral com aspirina é recomendada.

Devido à natureza da fibrilação atrial paroxísmica pode ser muito difícil de diagnosticá-la. Quando o paciente procura atendimento médico devido aos sintomas relacionados com a doença, tais com palpitações, dor no peito, encurtamento da respiração, tontura, insuficiência cardíaca, Ataques Isquêmicos Transitórios ou mesmo acidente vascular cerebral, o ritmo cardíaco normal pode ser prontamente restabelecido impedindo o diagnóstico da arritmia. Nestes casos o monitoramento do ritmo cardíaco é frequentemente aplicado na tentativa de diagnosticar a condição. O ritmo cardíaco é habitualmente monitorado continuamente por 24 a 48 horas.

Infelizmente os episódios de fibrilação atrial são imprevisíveis e frequentemente perdidos por este método. A oportunidade para diagnosticar a arritmia, estabelecer a terapia recomendada e possivelmente prevenir um enfraquecer o primeiro acidente vascular cerebral ou recorrente pode ser perdida com resultados devastadores para o paciente. Medidas de monitoramento do ritmo cardíaco prolongado e mais complexo são disponíveis e aplicados ocasionalmente quando a suspeita de fibrilação atrial é muito forte. Estes testes são caros, o rendimento de diagnóstico com método corrente é frequentemente baixo e eles são usados escassamente para esta indicação. Nestas circunstâncias o risco adicional de estratificação com teste genético pode ser extremamente útil. Entendendo que o indivíduo em questão carrega uma variante genética em risco ou uma protetora pode ser uma contribuição inestimável para a motivação da decisão de diagnóstico e/ou tratamento. Deste modo, em alguns casos, teste e terapia desnecessárias podem ser evitadas e em outros casos, com a ajuda de método de diagnóstico mais agressivo, a arritmia pode ser diagnosticada e/ou terapia apropriada iniciada e complicações posteriores de doença diminuídas.

MÉTODOS DA INVENÇÃO

Métodos para a avaliação de risco de arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral são aqui descritos e são abrangidos pela invenção. A invenção também abrange métodos de avaliar um indivíduo quanto a probabilidade de resposta a um agente terapêutico para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral, assim como métodos para prognosticar a eficácia de um agente terapêutico para tratar pacientes com arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Kits para ensaiar uma amostra de um paciente para detectar a suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral também são abrangidos pela

invenção.

Diagnóstico e ensaios de triagem da invenção

Em certas formas de realização, a presente invenção diz respeito aos métodos de diagnosticar ou ajudar no diagnóstico de, arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral ou uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral, pela detecção de alelos particulares em marcadores genéticos que aparecem mais frequentemente na arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou pacientes de acidente vascular cerebral ou pacientes que são suscetíveis para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Em uma forma de realização particular, a invenção é um método de diagnosticar uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral pela detecção de pelo menos um alelo de pelo menos um marcador polimórfico (por exemplo, o marcadores aqui descrito). A presente invenção descreve métodos por meio dos quais a detecção de alelos particulares de marcadores ou haplótipos particulares é indicativa de uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Tais ensaios de prognóstico ou preditivos também podem ser usados para determinar tratamento profilático de um paciente antes do início dos sintomas de arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral.

A presente invenção diz respeito em algumas formas de realização aos métodos de aplicações clínicas de diagnóstico, por exemplo, diagnóstico realizado por um profissional médico, que pode incluir uma avaliação ou determinação de variantes de risco genético e a sua interpretação. Em outras formas de realização, a invenção diz respeito aos métodos de avaliação de risco (ou diagnóstico) realizados por um leigo ou um

profissional não médico. Avanços tecnológicos recentes em tecnologias de genotipagem, incluindo a genotipagem de alto rendimento de marcadores de SNP, tal como a tecnologia de arranjo de Sonda de Inversão Molecular (por exemplo, Affymetrix GeneChip) e BeadArray Technologies (por exemplo, 5 ensaios Illumina GoldenGate e Infinium) têm tornado possível para os indivíduos ter seu próprio genoma avaliado por grande número de variações simultaneamente ou até um milhão de SNPs. A informação de genótipo resultante, tornada disponível para o indivíduo, pode ser comparada à informação da literatura científica pública a cerca dos riscos associados com a 10 doença ou traço com vários SNPs. A aplicação em diagnóstico de alelos associados com doença como aqui descrita, pode ser assim realizada por um profissional de saúde com base nos resultados de um teste clínico ou por um leigo ou profissional não médico, incluindo um indivíduo que fornece serviço para realizar uma avaliação de SNPs através da genotipagem de SNP, em uma 15 base de SNP individual ou pelos métodos de alto rendimento em escala grande tal com as tecnologias de série. Em outras palavras, o diagnóstico ou avaliação de uma suscetibilidade com base no risco genético pode ser feito pelos profissionais de saúde, conselheiro genético, fornecedores de serviços de genótipo ou pelo leigo, com base na informação a cerca de seu genótipo e 20 publicações em vários fatores de risco. No presente contexto, o termo “diagnosticar” e “diagnóstico de uma suscetibilidade”, é intencionado a se referir a qualquer método de diagnóstico disponível, incluindo aqueles mencionados acima.

Além disso, em certas outras formas de realização, a presente 25 invenção diz respeito aos métodos de diagnosticar ou auxiliar no diagnóstico de, uma suscetibilidade diminuída para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral pela detecção de alelos marcadores ou haplótipos genéticos particulares que aparecem menos frequentemente em pacientes com arritmia cardíaca (por exemplo,

fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral do que no indivíduo não diagnosticado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral ou na população geral.

Como aqui descrito e exemplificado, alelos marcadores ou
5 haplótipos particulares (por exemplo os marcadores e haplótipos como listados na Tabela 5 (Tabelas 5A e 5B) e marcadores em desequilíbrio de ligação com estes, por exemplo, os marcadores listados nas Tabelas 4 e/ou 9 marcadores em desequilíbrio de ligação com estes, por exemplo, os marcadores como apresentados na Tabela 19) são associados com a arritmia
10 cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Em uma forma de realização, o alelo marcador ou haplótipo é um que confere um risco ou suscetibilidade significantes para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Em uma outra forma de realização, a invenção diz respeito a um método de
15 diagnosticar uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral em um indivíduo humano, o método compreendendo determinar a presença ou ausência de pelo menos um alelo de pelo menos um marcador polimórfico em uma amostra de ácido nucleico obtida do indivíduo, em que o pelo menos um
20 marcador polimórfico é selecionado do grupo que consiste dos marcadores polimórficos listados nas Tabelas 5A e 5B e marcadores em desequilíbrio de ligação com estes. Em uma outra forma de realização, a invenção diz respeito aos métodos de diagnosticar uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral em
25 um indivíduo humano, pela triagem quanto a pelo menos um alelo marcador ou haplótipo como listados nas Tabelas 5A e 5B ou marcadores em desequilíbrio de ligação com estes. Em uma outra forma de realização, o alelo marcador ou haplótipo é mais frequentemente presentes em um paciente tendo ou que é suscetível à, arritmia cardíaca (por exemplo, fibrilação atrial ou

flúter atrial) e/ou acidente vascular cerebral (afetado), quando comparado com a frequência da sua presença em um paciente saudável (controle, tal com controle da população). Em certas formas de realização, a significância da associação do pelo menos um alelo marcador ou haplótipo é caracterizada por um valor $p < 0,05$. Em outras formas de realização, a significância da associação é caracterizada pelos valores p menores, tais como $< 0,01$, $< 0,001$, $< 0,0001$, $< 0,00001$, $< 0,000001$, $< 0,0000001$, $< 0,00000001$ ou $< 0,000000001$.

Nestas formas de realização, a presença do pelo menos um alelo marcador ou haplótipo é indicativa de uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Estes métodos de diagnóstico envolvem detectar a presença ou ausência de pelo menos um alelo marcador ou haplótipo que estejam associados com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Os haplótipos aqui descritos incluem combinações de alelos em vários marcadores genéticos (por exemplo, SNPs, microssatélites). A detecção dos alelos marcadores genéticos particulares que compõem os haplótipos particulares pode ser realizada por uma variedade de métodos aqui descritos e/ou conhecidos na técnica. Por exemplo, marcadores genéticos podem ser detectados ao nível de ácido nucleico (por exemplo, pelo sequenciamento de nucleotídeo direto ou por outros meios conhecidos pelos habilitados na técnica) ou ao nível de aminoácido se o marcador genético afeta a sequência codificadora de uma proteína codificada por um ácido nucleico associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral (por exemplo, pelo sequenciamento de proteína ou pelos imunoenaios usando anticorpos que reconhecem uma tal proteína). Os alelos marcadores ou haplótipos da presente invenção correspondem aos fragmentos de uma sequência de DNA genômico associada com a arritmia cardíaca (por exemplo, fibrilação atrial ou

flúter atrial) e/ou acidente vascular cerebral. Tais fragmentos abrangem a sequência de DNA do marcador polimórfico ou haplótipo em questão, mas também podem incluir segmentos de DNA em forte LD (desequilíbrio de ligação) com o marcador ou haplótipo. Em uma forma de realização, tais
5 segmentos compreendem segmentos em LD com o marcador ou haplótipo como determinado por um valor de r^2 maior do que 0,2 e/ou $|D'| > 0,8$.

Em uma forma de realização, o diagnóstico de uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral pode ser realizado usando
10 métodos de hibridização, tais como a análise de Southern, análise de Northern, e/ou hibridizações *in situ* (ver Current Protocols in Molecular Biology, Ausubel, F. *et al.*, eds., John Wiley & Sons, incluindo todos os suplementos). Uma amostra biológica de um paciente ou indivíduo de teste (uma “amostra de teste”) de DNA genômico, RNA ou cDNA é obtida de um
15 paciente suspeito de ter, estando suscetível a ou predisposto para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral (o “paciente de teste”). O paciente pode ser um adulto, criança ou feto. A amostra de teste pode ser de qualquer fonte que contenha DNA genômico, tal com uma amostra de sangue, amostra de fluido amniótico,
20 amostra de fluido cerebrospinal ou amostra de tecido da pele, músculo, mucosa bucal ou conjuntiva, placenta, trato gastrointestinal ou outros órgãos. Uma amostra de teste de DNA de células ou tecido fetais pode ser obtida pelos métodos apropriados, tal com pela amniocentese ou amostragem de vilo coriônico. As amostras de DNA, RNA ou cDNA são depois examinadas. A
25 presença de um alelo marcador específico pode ser indicada pela Hibridização específica de sequência de uma sonda de ácido nucleico específica para o alelo particular. A presença de mais do que um alelo marcador específico ou um haplótipo específico pode ser indicada pelo uso de diversas sondas de ácido nucleico específicas de sequência, cada uma sendo específica para um

alelo particular. Em uma forma de realização, um haplótipo pode ser indicado por uma única sonda de ácido nucleico que é específica para o haplótipo específico (isto é, hibridiza especificamente a um filamento de DNA que compreende os alelos marcadores específicos característicos do haplótipo).

5 Uma sonda específica de sequência pode ser direcionada para hibridizar ao DNA genômico, RNA ou cDNA. Uma “sonda de ácido nucleico”, como aqui usada, pode ser uma sonda de DNA ou uma sonda de RNA que hibridiza a uma sequência complementar. Uma pessoa de habilidade na técnica saberia como planejar uma tal sonda de modo que a hibridização específica de
10 sequência ocorrerá apenas se um alelo particular está presente em uma sequência genômica de uma amostra de teste.

Para diagnosticar uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral, uma amostra de hibridização é formada contatando-se a amostra de teste
15 contendo um ácido nucleico associado com a fibrilação atrial e/ou acidente vascular cerebral, tal com uma amostra de DNA genômico, com pelo menos uma sonda de ácido nucleico. Um exemplo não limitante de uma sonda para detectar mRNA ou DNA genômico é uma sonda de ácido nucleico rotulada que é capaz de hibridizar às sequências de mRNA ou DNA genômico aqui
20 descritas. A sonda de ácido nucleico pode ser, por exemplo, uma molécula de ácido nucleico de tamanho natural ou uma porção desta, tal com um oligonucleotídeo de pelo menos 15, 30, 50, 100, 250 ou 500 nucleotídeos no comprimento que é suficiente para hibridizam especificamente sob condições severas ao mRNA ou DNA genômico apropriados. A sonda de ácido nucleico
25 pode ser de até 1000 ou mais nucleotídeos no comprimento, incluindo até 500 nucleotídeos, 400 nucleotídeo, 300 nucleotídeos, 200 nucleotídeos ou 100 nucleotídeos. Certas formas de realização incluem sondas de nucleotídeo que são de 15 a 1000 nucleotídeos no comprimento. Outras formas de realização dizem respeito ao uso de sondas de nucleotídeo que são de 15 a 500

nucleotídeos no comprimento ou de 15 a 400 nucleotídeos no comprimento ou de 20 a 400 nucleotídeos no comprimento. Outras faixas de tamanho das sondas de nucleotídeo da invenção são consideradas, como bem conhecidos pela pessoa habilitada. Em uma forma de realização, a sonda de ácido nucleico pode compreender toda ou uma porção da sequência de nucleotídeo de bloco LD C04, como aqui descrita, opcionalmente compreendendo pelo menos um alelo de um marcador aqui descrito ou pelo menos um haplótipo aqui descrito ou a sonda pode ser a sequência complementar de uma tal sequência. Em uma forma de realização particular, a sonda de ácido nucleico é uma porção da sequência de nucleotídeo de bloco LD C04 como apresentada na SEQ ID NO: 50 ou, como aqui descrita, opcionalmente compreendendo pelo menos um alelo de um marcador aqui descrito ou pelo menos um alelo de um marcador polimórfico ou haplótipo que compreende pelo menos um marcador polimórfico aqui descrito ou a sonda pode ser a sequência complementar de uma tal sequência. Outras sondas adequadas para o uso nos ensaios de diagnóstico da invenção são aqui descritas. A hibridização pode ser realizada pelos métodos bem conhecidos pela pessoa habilitada na técnica (ver, por exemplo, *Current Protocols in Molecular Biology*, Ausubel, F. *et al.*, eds., John Wiley & Sons, incluindo todos os suplementos). Em uma forma de realização, a hibridização refere-se à hibridização específica, isto é, hibridização sem nenhum desemparelhamento (hibridização exata). Em uma forma de realização, as condições de hibridização para a hibridização específica são severidade alta.

A hibridização específica, se presente, é detectada usando métodos padrão. Se a hibridização específica ocorre entre a sonda de ácido nucleico e o ácido nucleico na amostra de teste, então a amostra contém o alelo que é complementar ao nucleotídeo que está presente na sonda de ácido nucleico. O processo pode ser repetido para quaisquer marcadores da presente invenção ou marcadores que compõem um haplótipo da presente invenção ou

sondas múltiplas podem ser usadas concorrentemente para detectar mais do que um dos alelos marcadores ao mesmo tempo. Também é possível planejar uma única sonda contendo mais do que um dos alelos marcadores de um haplótipo particular (por exemplo, uma sonda contendo alelos complementares a 2, 3, 4, 5 ou a todos os marcadores que compõem um haplótipo particular). A detecção dos marcadores particulares do haplótipo na amostra é indicativa de que a fonte da amostra tem o haplótipo particular (por exemplo, um haplótipo) e portanto é suscetível para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral.

Em um outro método de hibridização, a análise de Northern (ver *Current Protocols in Molecular Biology*, Ausubel, F. *et al.*, eds., John Wiley & Sons, supra) é usada para identificar a presença de um polimorfismo associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Para a análise de Northern, uma amostra de teste de RNA é obtida do paciente por meios apropriados. Como aqui descrito, a hibridização específica de uma sonda de ácido nucleico ao RNA do paciente é indicativa de um alelo particular complementar à sonda. Para exemplos representativos de uso de sondas de ácido nucleico, ver, por exemplo, Patente U.S. N^{os} 5.288.611 e 4.851.330.

Adicional ou alternativamente, uma sonda de ácido nucleico peptídico (PNA) pode ser usada além ou invés de, uma sonda de ácido nucleico nos métodos de hibridização aqui descritos. Um PNA é uma imitação de DNA tendo uma cadeia principal como peptídeo, inorgânica, tal com unidades de N-(2-aminoetil)glicina, com uma base orgânica (A, G, C, T ou U) ligada ao nitrogênio da glicina por intermédio de um ligador de metileno carbonila (ver, por exemplo, Nielsen, P., *et al.*, *Bioconjug. Chem.* 5: 3-7 (1994)). A sonda de PNA pode ser planejada para hibridizar especificamente a uma molécula em uma amostra suspeita de conter um ou mais dos alelos marcadores ou haplótipos que estão associados com a arritmia

cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. A hibridização da sonda de PNA é assim diagnóstico para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral ou uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral.

Em uma forma de realização da invenção, uma amostra de teste contendo DNA genômico obtida do paciente é coletada e a Reação em Cadeia de Polimerase (PCR) é usada para amplificar um fragmento que compreende um ou mais marcadores ou haplótipos da presente invenção.

Como aqui descrito, a identificação de um alelo marcador ou haplótipo particulares associados com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral, pode ser realizada usando uma variedade de métodos (por exemplo, análise de sequência, análise pela digestão de restrição, hibridização específica, ensaios de polimorfismo de conformação de filamento simples (SSCP), análise eletroforética, etc.). Em uma outra forma de realização, o diagnóstico é realizado pela análise de expressão usando a PCR quantitativa (ciclagem térmica cinética). Esta técnica, por exemplo, pode utilizar tecnologias comercialmente disponíveis, tais com TaqMan® (Applied Biosystems, Foster City, CA). A técnica pode avaliar a presença de uma alteração na expressão ou composição de um polipeptídeo ou variante(s) de junção que é/são codificada(s) por um ácido nucleico associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Além disso, a expressão da(s) variante(s) pode ser quantificada como física ou funcionalmente diferente.

Em um outro método da invenção, a análise pela digestão de restrição pode ser usada para detectar um alelo particular se o alelo resulta na criação ou eliminação de um sítio de restrição em relação a uma sequência de referência. A análise de polimorfismo de comprimento de fragmento de restrição (RFLP) pode ser conduzida, por exemplo, como descrito em Current

Protocols in Molecular Biology, supra. O padrão de digestão do fragmento de DNA relevante indica a presença ou ausência do alelo particular na amostra.

A análise de sequência também pode ser usada para detectar alelos ou haplótipos específicos associados com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral (por exemplo os marcadores polimórficos da Tabela 5 (Tabelas 5A e 5B), Tabela 9 e/ou Tabela 19). Portanto, em uma forma de realização, a determinação da presença ou ausência de um alelo marcador ou haplótipo particulares compreende a análise de sequência de uma amostra de teste de DNA ou RNA obtida de um paciente ou indivíduo. A PCR ou outro métodos apropriados podem ser usados para amplificar uma porção de um ácido nucleico associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral e a presença de um alelo específico pode ser depois detectada diretamente pelo sequenciamento do sítio polimórfico (ou sítios polimórficos múltiplos em um haplótipo) do DNA genômico na amostra.

Os oligonucleotídeos específicos de alelo também podem ser usados para detectar a presença de um alelo particular em um ácido nucleico associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral, (por exemplo os marcadores polimórficos da Tabela 5 (Tabelas 5A e 5B), Tabela 9 e/ou Tabela 19), através do uso de hibridização de ponto-mancha de oligonucleotídeos amplificados com sondas de oligonucleotídeo específicas de alelo (ASO) (ver, por exemplo, Saiki, R. *et al.*, Nature, 324: 163-166 (1986)). Um “oligonucleotídeo específico de alelo” (também aqui aludido como uma “sonda de oligonucleotídeo específica de alelo”) é um oligonucleotídeo de aproximadamente 10 a 500 pares de base, aproximadamente 15 a 400 pares de base, aproximadamente 15 a 200 pares de base, aproximadamente 15 a 100 pares de base, aproximadamente 15 a 50 pares de base ou aproximadamente 15 a 30 pares de base, que especificamente hibridizam a um ácido nucleico

associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral e que contém um alelo específico em um sítio polimórfico (por exemplo, um marcador ou haplótipo como aqui descrito). Uma sonda de oligonucleotídeo específica de alelo que é específica para um ou mais ácidos nucleicos particulares associados com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral pode ser preparada usando métodos padrão (ver, por exemplo, Current Protocols in Molecular Biology, supra). A PCR pode ser usada para amplificar a região desejada. O DNA contendo a região amplificada pode ser submetida a ponto-mancha usando métodos padrão (ver, por exemplo, Current Protocols in Molecular Biology, supra) e a mancha pode ser contatada com a sonda de oligonucleotídeo. A presença de hibridização específica da sonda para a região amplificada pode ser depois detectada. A hibridização específica de uma sonda de oligonucleotídeo específica de alelo para DNA do paciente é indicativa de um alelo específico em um sítio polimórfico associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral (ver, por exemplo, Gibbs, R. *et al.*, Nucleic Acids Res., 17: 2437-2448 (1989) e WO 93/22456).

Com a adição de tais análogos como ácido nucleicos bloqueados (LNAs), o tamanho de iniciadores e sondas pode ser reduzido a tanto quanto 8 bases. LNAs são uma nova classe de análogos de DNA bicíclicos em que as posições 2' e 4' no anel de furanose são unidos por intermédio de uma porção de O-metileno (óxi-LNA), S-metileno (tio-LNA) ou amino metileno (amino-LNA). Comum a todas destas variantes de LNA é uma afinidade aos ácidos nucleicos complementares, que é de longe a mais alta relatada para um análogo de DNA. Por exemplo, todos os nonâmeros óxi-LNA particulares foram mostrados ter temperaturas de fusão (T_m) de 64°C e 74°C quando em complexo com DNA ou RNA complementares, respectivamente, como oposto a 28°C tanto para DNA quanto RNA para o

nonúmero de DNA correspondente. Aumentos substanciais na T_m também são obtidos quando monômeros de LNA são usados em combinação com monômeros de DNA ou RNA padrão. Para iniciadores e sondas, dependendo de onde os monômeros de LNA são incluídos (por exemplo, na extremidade 3', na extremidade 5' ou no centro), a T_m pode ser consideravelmente aumentada.

Em uma outra forma de realização, arranjos de sondas de oligonucleotídeo que são complementares aos segmentos de sequência de ácido nucleico alvo de um paciente, podem ser usados para identificar polimorfismos em um ácido nucleico associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral (por exemplo os marcadores polimórficos das Tabelas 5A e 5B e marcadores em desequilíbrio de ligação com estes). Por exemplo, um arranjo de oligonucleotídeo pode ser usado. Arranjos de oligonucleotídeo tipicamente compreendendo uma pluralidade de sondas de oligonucleotídeo diferentes que são ligadas a uma superfície de um substrato em locais conhecidos diferentes. Estes arranjos de oligonucleotídeo, também descritos como “Genechips[®],” foram em uma maneira geral descritos na técnica (ver, por exemplo, a Patente U.S. Nº 5.143.854, Publicações de Patente PCT Nºs WO 90/15070 e 92/10092). Estes arranjos em uma maneira geral podem ser produzidos usando métodos de síntese mecânicas ou métodos de síntese direcionadas por luz que incorporam uma combinação de métodos fotolitográficos e métodos de síntese de oligonucleotídeo de fase sólida ou por outros métodos conhecidos pela pessoa habilitada na técnica (ver, por exemplo, Fodor, S. *et al.*, Science, 251: 767-773 (1991); Pirrung *et al.*, Patente U.S. Nº 5.143.854 (ver também Pedido PCT publicado Nº WO 90/15070); e Fodor, S. *et al.*, Pedido PCT publicado Nº WO 92/10092 e Patente U.S. Nº 5.424.186, as divulgações inteiras de cada um dos quais são aqui incorporadas por referência). As técnicas para a síntese destes arranjos usando métodos de

síntese mecânicas são descritas, por exemplo, na Patente U.S. Nº 5.384.261; as divulgações inteiras das quais são aqui incorporadas por referência. Em um outro exemplo, arranjos lineares podem ser utilizados. Descrições adicionais de uso de arranjos de oligonucleotídeo para a detecção de polimorfismos

5 podem ser encontradas, por exemplo, nas Patentes U.S. Nºs 5.858.659 e 5.837.832, as divulgações inteiras de ambas das quais são aqui incorporadas por referência.

Outros métodos de análise de ácido nucleico que são disponíveis para aqueles habilitados na técnica podem ser usados para

10 detectar um alelo particular em um sítio polimórfico associado com a fibrilação atrial e/ou acidente vascular cerebral (por exemplo os marcadores polimórficos da Tabela 5 (Tabelas 5A e 5B), Tabela 9 e/ou Tabela 19). Os métodos representativos incluem, por exemplo, o sequenciamento manual direto (Church e Gilbert, Proc. Natl. Acad. Sci. USA, 81: 1991-1995 (1988);

15 Sanger, F., *et al.*, Proc. Natl. Acad. Sci, USA, 74: 5463-5467 (1977); Beavis, *et al.*, Patente U.S. Nº 5.288.644); sequenciamento fluorescente automatizado; ensaios de polimorfismo de conformação de filamento simples (SSCP); eletroforese em gel desnaturante clampeado (CDGE); eletroforese em gel de gradiente desnaturante (DGGE) (Sheffield, V., *et al.*, Proc. Natl. Acad. Sci.

20 USA, 86: 232-236 (1989)), análise de mudança de mobilidade (Orita, M., *et al.*, Proc. Natl. Acad. Sci. USA, 86: 2766-2770 (1989)), análise de enzima de restrição (Flavell, R., *et al.*, Cell, 15: 25-41 (1978); Geever, R., *et al.*, Proc. Natl. Acad. Sci. USA, 78: 5081-5085 (1981)); análise de heteroduplex; clivagem de desemparelhamento químico (CMC) (Cotton, R., *et al.*, Proc.

25 Natl. Acad. Sci. USA, 85: 4397-4401 (1985)); ensaios de proteção da RNase (Myers, R., *et al.*, Science, 230: 1242-1246 (1985); uso de polipeptídeos que reconhecem desemparelhamentos de nucleotídeo, tais com a proteína mutS da *E. coli*; e PCR específica de alelo.

Em uma outra forma de realização da invenção, o diagnóstico

de arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral ou uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral pode ser feita pelo exame da expressão e/ou composição de um polipeptídeo codificado por um ácido nucleico associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral nestes casos onde o(s) marcador(es) genético(s) ou haplótipo(s) da presente invenção resultam em uma mudança na composição ou expressão do polipeptídeo. Assim, o diagnóstico de uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral pode ser feito examinando-se a expressão e/ou composição de um destes polipeptídeos ou um outro polipeptídeo codificado por um ácido nucleico associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral, nestes casos onde o marcador ou haplótipo genéticos da presente invenção resulta em uma mudança na composição ou expressão do polipeptídeo. Os haplótipos e marcadores da presente invenção que mostram associação para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral podem desempenhar um papel através do seu efeito sobre um ou mais destes genes vizinhos (por exemplo, o gene PITX2). Mecanismos possíveis que afetam estes genes incluem, por exemplo, efeitos sobre a transcrição, efeitos sobre a união de RNA, alterações nas quantidades de relação de formas unidas alternativas de mRNA, efeitos sobre a estabilidade do RNA, efeitos sobre o transporte do núcleo para o citoplasma e efeitos sobre a eficiência e precisão da tradução.

Assim, em uma outra forma de realização, as variantes (marcadores ou haplótipos) da invenção que mostra a associação para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral afetam a expressão de um gene vizinho. É bem conhecido

que a expressão de gene que afeta elemento regulatório pode estar localizado a dezenas ou mesmo centenas de quilobases longe da região promotora de um gene. Ensaando-se quanto a presença ou ausência de pelo menos um alelo de pelo menos um marcador polimórfico da presente invenção, é assim possível

5 avaliar o nível de expressão de tais genes vizinhos. É assim considerado que a detecção dos marcadores ou haplótipos da presente invenção pode ser usada para avaliar a expressão quanto a um ou mais genes que são ligados à arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral.

10 Uma variedade de métodos podem ser usados para detectar os níveis de expressão de proteína, incluindo ensaios imunossorventes ligados à enzima (ELISA), Western blots, imunoprecipitações e imunofluorescência. Uma amostra de teste de um paciente é avaliada quanto a presença de uma alteração na expressão e/ou uma alteração na composição do polipeptídeo

15 codificada por um ácido nucleico associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Uma alteração na expressão de um polipeptídeo codificado por um ácido nucleico associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral pode ser, por exemplo, uma

20 alteração na expressão de polipeptídeo quantitativa (isto é, a quantidade de polipeptídeo produzida). Uma alteração na composição de um polipeptídeo codificado por um ácido nucleico associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral é uma alteração na expressão de polipeptídeo qualitativa (por exemplo,

25 expressão de um polipeptídeo mutante ou de uma variante de união diferente). Em uma forma de realização, o diagnóstico de uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral é feita pela detecção de uma variante de junção particular codificada por um ácido nucleico associado com a arritmia cardíaca (por

exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral ou um padrão particular de variantes de união.

Ambas de tais alterações (quantitativa e qualitativa) também podem estar presentes. Uma “alteração” na expressão ou composição de polipeptídeo, como aqui usada, refere-se a uma alteração na expressão ou composição em uma amostra de teste, quando comparada à expressão ou composição do polipeptídeo em uma amostra de controle. Uma amostra de controle é uma amostra que corresponde à amostra de teste (por exemplo, é do mesmo tipo de células) e é de um paciente que não é afetado por, e/ou que não tem uma suscetibilidade para, arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Em uma forma de realização, a amostra de controle é de um paciente que não possui um alelo marcador ou haplótipo como aqui descrito. Similarmente, a presença de uma ou mais variantes de união diferentes na amostra de teste ou a presença de quantidades significativamente diferentes de variantes de união diferentes na amostra de teste, quando comparado com a amostra de controle, pode ser indicativa de uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Uma alteração na expressão ou composição do polipeptídeo na amostra de teste, quando comparada com a amostra de controle, pode ser indicativa de um alelo específico no caso onde o alelo altera um sítio de união em relação à referência na amostra de controle. Vários meios de examinar a expressão ou composição de um polipeptídeo codificado por um ácido nucleico são conhecidos pela pessoa habilitada na técnica e podem ser usados, incluindo espectroscopia, colorimetria, eletroforese, focalização isoeletrica e imunoenaios (por exemplo, David *et al.*, Pat. U.S. Nº 4.376.110) tal com imunomanchamento (ver, por exemplo, Current Protocols in Molecular Biology, particularmente o capítulo 10, *supra*).

Por exemplo, em uma forma de realização, um anticorpo (por

exemplo, um anticorpo com um rótulo detectável) que é capaz de ligar a um polipeptídeo codificado por um ácido nucleico associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral podem ser usados. Anticorpos podem ser policlonais ou monoclonais. Um anticorpo intacto ou um fragmento deste (por exemplo, Fv, Fab, Fab', F(ab')₂) podem ser usados. O termo "rotulado", com respeito à sonda ou anticorpo, é intencionado abranger a rotulação direta da sonda ou anticorpo pela ligação (isto é, ligando fisicamente) uma substância detectável à sonda ou anticorpo, assim como rotulação indireta da sonda ou anticorpo pela reatividade com um outro reagente que é diretamente rotulado. Os exemplos de rotulação indireta incluem detecção de um anticorpo primário usando um anticorpo secundário rotulado (por exemplo, um anticorpo secundário fluorescentemente rotulado) e rotulação de extremidade de uma sonda de DNA com biotina tal que a mesma possa ser detectada com estreptavidina fluorescentemente rotulada.

Em uma forma de realização deste método, o nível ou a quantidade de polipeptídeo codificado por um ácido nucleico associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral em uma amostra de teste são comparados com o nível ou quantidade do polipeptídeo em uma amostra de controle. Um nível ou quantidade do polipeptídeo na amostra de teste que é mais alta ou mais baixa do que o nível ou quantidade do polipeptídeo na amostra de controle, tal que a diferença é estatisticamente significativa, é indicativa de uma alteração na expressão do polipeptídeo codificada pelo ácido nucleico e é diagnóstico para um alelo ou haplótipo particulares responsáveis por causar a diferença na expressão. Alternativamente, a composição do polipeptídeo em uma amostra de teste é comparada com a composição do polipeptídeo em uma amostra de controle. Em uma outra forma de realização, tanto o nível ou quantidade quanto a composição do polipeptídeo podem ser avaliados na amostra de teste

e na amostra de controle.

Em uma outra forma de realização, o diagnóstico de uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral é feita pela detecção de pelo menos um marcador ou haplótipos da presente invenção (por exemplo, alelos associados dos marcadores listados nas Tabelas 5A e 5B e marcadores em desequilíbrio de ligação com estes), em combinação com um ensaio com base em proteína, com base em RNA ou com base em DNA adicionais. Os métodos da invenção também podem ser usados em combinação com uma análise de uma história da família do paciente e fatores de risco (por exemplo, fatores de risco ambientais, fatores de risco do estilo de vida).

Kits

Os kits úteis nos métodos e procedimentos da invenção compreendem componentes úteis em qualquer um dos métodos aqui descritos, incluindo por exemplo, sondas de hibridização, enzimas de restrição (por exemplo, para a análise de RFLP), oligonucleotídeos específico de alelos, anticorpos que se ligam a um polipeptídeo alterado codificado por um ácido nucleico da invenção como aqui descrita (por exemplo, um segmento genômico que compreende pelo menos um marcador polimórfico e/ou haplótipo da presente invenção) ou a um polipeptídeo não alterado (nativo) codificado por um ácido nucleico da invenção como aqui descrita, meios para a amplificação de um ácido nucleico associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral, meios para analisar a sequência de ácido nucleico de um ácido nucleico associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral, meios para analisar a sequência de aminoácido de um polipeptídeo codificado por um ácido nucleico associado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral, etc. Os kits podem por exemplo incluir tampões

necessários, iniciadores de ácido nucleico para amplificar ácidos nucleicos da invenção (por exemplo, um ou mais dos marcadores polimórficos como aqui descritos) e reagentes para a detecção específica de alelo dos fragmentos amplificados usando tais iniciadores e enzimas necessárias (por exemplo, DNA polimerase). Adicionalmente, os kits podem fornecer reagentes para os ensaios a serem usados em combinação com os métodos da presente invenção, por exemplo, reagentes para o uso com ensaios de diagnóstico para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral.

Em uma forma de realização, a invenção é um kit para ensaiar uma amostra de um paciente para detectar a presença de arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral ou uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral em um paciente, em que o kit compreende reagentes necessários para detectar seletivamente pelo menos um alelo de pelo menos um polimorfismo da presente invenção no genoma do indivíduo. Em uma forma de realização particular, os reagentes compreendem pelo menos um oligonucleotídeo contíguo que hibridiza a um fragmento do genoma do indivíduo que compreende pelo menos um polimorfismo da presente invenção. Em uma outra forma de realização, os reagentes compreendem pelo menos um par de oligonucleotídeos que hibridizam aos filamentos opostos de um segmento genômico obtido de um paciente, em que cada par de iniciador de oligonucleotídeo é planejado para amplificar seletivamente um fragmento do genoma do indivíduo que inclui pelo menos um polimorfismo, em que o polimorfismo é selecionado do grupo que consiste dos polimorfismos como listados nas Tabelas 5A e 5B e marcadores polimórficos em desequilíbrio de ligação com estes. Ainda em uma outra forma de realização o fragmento tem pelo menos 20 pares de base no tamanho. Tais oligonucleotídeos ou ácidos nucleicos (por exemplo,

iniciadores de oligonucleotídeo) podem ser planejados usando porções da sequência de ácido nucleico que flanqueiam polimorfismos (por exemplo, SNPs ou microssatélites) que são indicativos de arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Em uma outra forma de realização, o kit compreende um ou mais ácidos nucleicos rotulados capazes de detecção específica de alelo de um ou mais marcadores polimórficos ou haplótipos específicos associados com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral e reagentes para detecção do rótulo. Os rótulos adequados incluem, por exemplo, um radioisótopo, um rótulo fluorescente, um rótulo de enzima, um rótulo de co-fator de enzima, um rótulo magnético, um rótulo de spin, um rótulo de epítipo.

Em formas de realização particulares, o marcador polimórfico ou haplótipo a serem detectados pelos reagentes do kit compreende um ou mais marcadores, dois ou mais marcadores, três ou mais marcadores, quatro ou mais marcadores ou cinco ou mais marcadores selecionados do grupo que consiste dos marcadores nas Tabelas 5A e 5B. Em uma outra forma de realização, o marcador ou haplótipo a serem detectados compreendem os marcadores listados nas Tabelas 5A e 5B. Em uma outra forma de realização, o marcador ou haplótipo a serem detectados compreende os marcadores listados nas Tabelas 4 e 9. Em uma outra forma de realização, o marcador ou haplótipo a serem detectados compreendem pelo menos um marcador do grupo de marcadores em forte desequilíbrio de ligação, como definido pelos valores de r^2 maiores do que 0,2, a pelo menos um do grupo de marcadores que consiste dos marcadores listados nas Tabelas 5A e 5B. Em uma outra forma de realização, o marcador ou haplótipo a serem detectados compreende pelo menos um marcador dos marcadores em forte desequilíbrio de ligação, como definido pelos valores de r^2 maiores do que 0,2, a pelo menos um do grupo de marcadores que consiste dos marcadores listados nas Tabelas 4 e 9.

Em uma outra forma de realização, o marcador ou haplótipo a serem detectados compreendem o marcador rs2220427 (SEQ ID NO: 1) ou marcador rs1033464 (SEQ ID NO: 41) ou marcadores em desequilíbrio de ligação com estes. Em uma outra forma de realização, o marcador ou haplótipo a serem detectados compreendem pelo menos um dos marcadores apresentados na Tabela 19. Em uma outra forma de realização, o marcador ou haplótipo a serem detectados compreendem os marcadores D4S406 (SEQ ID NO: 45), rs2634073 (SEQ ID NO: 33), rs2200733 (SEQ ID NO: 28), rs2220427 (SEQ ID NO: 1), rs10033464 (SEQ ID NO: 41) e rs13143308 (SEQ ID NO: 51) e marcadores em desequilíbrio de ligação com estes. Ainda em uma outra forma de realização, o marcador ou haplótipo compreendem os alelos em risco 2, 4 e/ou 8 no marcador D4S406, alelo A do marcador rs2634073, alelo T do marcador rs2200733, alelo T do marcador rs2220427, alelo T do marcador rs10033464, e/ou alelo G do marcador rs13143308. Em uma tal forma de realização, o desequilíbrio de ligação é definido pelos valores de r^2 maiores do que 0,1. Em uma outra de tal forma de realização, o desequilíbrio de ligação é definido pelos valores de r^2 maiores do que 0,2.

Em uma forma de realização preferida, o kit para detectar os marcadores da invenção compreende uma sonda de oligonucleotídeo de detecção, que hibridiza a um segmento de DNA padrão contendo um polimorfismo de SNP a ser detectado, uma sonda de oligonucleotídeo realçadora e uma endonuclease. Como explicado no acima, a sonda de oligonucleotídeo de detecção compreende uma porção fluorescente ou grupo no seu terminal 3' e um arrefecedor no seu término 5' e um oligonucleotídeo realçador, é utilizado, como descrito por Kuttyavin *et al.* (Nucleic Acid Res. 34: e128 (2006)). A porção fluorescente pode ser Gig Harbor Green ou Yakima Yellow ou outras porções fluorescentes adequadas. A sonda de detecção é planejada para hibridizar a uma sequência de nucleotídeo curta que inclui o polimorfismo SNP a ser detectado. Preferivelmente, o SNP está em

qualquer lugar do resíduo terminal até 6 resíduos da extremidade 3' da sonda de detecção. O realçador é uma sonda de oligonucleotídeo curta que hibridiza ao padrão de DNA 3' em relação à sonda de detecção. As sondas são planejadas tal que um único intervalo de nucleotídeo existe entre a sonda de
5 detecção e a sonda de nucleotídeo realçadora quando ambos são ligados ao padrão. O intervalo cria um sítio abásico sintético que é reconhecido por uma endonuclease, tal com Endonuclease IV. A enzima cliva o pigmento fora da sonda de detecção totalmente complementar, mas não pode clivar uma sonda de detecção contendo um desemparelhamento. Assim, medindo-se a
10 fluorescência da porção fluorescente liberada, a avaliação da presença de um alelo particular definido pela sequência de nucleotídeo da sonda de detecção pode ser realizada.

A sonda de detecção pode ser de qualquer tamanho adequado, embora preferivelmente a sonda seja relativamente curta. Em uma forma de
15 realização, a sonda tem de 5 a 100 nucleotídeos no comprimento. Em uma outra forma de realização, a sonda é de 10 a 50 nucleotídeos no comprimento e em uma outra forma de realização, a sonda é de 12 a 30 nucleotídeos no comprimento. Outros comprimentos da sonda são possíveis e dentro do escopo da habilidade da pessoa média habilitada na técnica.

20 Em uma forma de realização preferida, o padrão de DNA contendo o polimorfismo de SNP é amplificado pela Reação em Cadeia de Polimerase (PCR) antes da detecção e iniciadores para tal amplificação são incluídos no kit de reagente. Em uma tal forma de realização, o DNA amplificado serve como o padrão para a sonda de detecção e a sonda
25 realçadora.

Certas formas de realização da sonda de detecção, a sonda realçadora e/ou os iniciadores usados para a amplificação do padrão pela PCR incluem o uso de bases modificadas, incluindo A modificada e G modificada. O uso de bases modificadas pode ser útil para ajustar a temperatura de fusão

da molécula de nucleotídeo (sonda e/ou iniciador) para o DNA padrão, por exemplo para aumentar a temperatura de fusão em regiões contendo uma porcentagem baixa de bases G ou C, em que A modificada com a capacidade de formar três ligações de hidrogênio ao seu T complementar pode ser usada ou para diminuir a temperatura de fusão em regiões contendo uma alta porcentagem de bases G ou C, por exemplo pelo uso de bases G modificadas que formam apenas duas ligações de hidrogênio à sua base C complementar em uma molécula de DNA de filamento duplo. Em uma forma de realização preferida, as bases modificadas são usadas no planejamento da sonda de detecção nucleotídica. Qualquer base modificada conhecida pela pessoa habilitada pode ser selecionada nestes métodos e a seleção de bases adequadas está bem dentro do escopo da pessoa habilitada com base nas divulgações aqui contidas e bases conhecidas disponíveis de fontes comerciais como conhecido pela pessoa habilitada.

Em uma de tais formas de realização, a presença do marcador ou haplótipo é indicativa de uma suscetibilidade (suscetibilidade aumentada ou suscetibilidade diminuída) para fibrilação atrial e/ou acidente vascular cerebral. Em uma outra forma de realização, a presença do marcador ou haplótipo é indicativa de resposta ao agente terapêutico para a fibrilação atrial e/ou acidente vascular cerebral. Em uma outra forma de realização, a presença do marcador ou haplótipo é indicativa do prognóstico de fibrilação atrial e/ou acidente vascular cerebral. Ainda em uma outra forma de realização, a presença do marcador ou haplótipo é indicativa de progresso do tratamento da fibrilação atrial e/ou acidente vascular cerebral. Tal tratamento pode incluir a intervenção pela cirurgia, medicação ou por outros meios (por exemplo, mudanças no estilo de vida).

Agentes terapêuticos

Variantes da presente invenção (por exemplo, os marcadores e/ou haplótipos da invenção, por exemplo, os marcadores listados nas Tabelas

5A e 5B e/ou Tabela 19) podem ser usados para identificar novos alvos terapêuticos para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Por exemplo, genes contendo ou em desequilíbrio de ligação com, variantes (marcadores e/ou haplótipos) associados com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral ou seus produtos, assim como genes ou seus produtos que são direta ou indiretamente regulados pela OR interação com estes genes variantes ou seus produtos, podem ser alvejados para o desenvolvimento de agentes terapêuticos para tratar arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral ou prevenir ou retardar o início de sintomas associados com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Os agentes terapêuticos podem compreender um ou mais de, por exemplo, moléculas pequenas que não de proteína e não de ácido nucleico, proteínas, peptídeos, fragmentos de proteína, ácidos nucleicos (DNA, RNA), PNA (ácidos nucleicos peptídicos) ou seus derivados ou miméticos que podem modular a função e/ou níveis dos genes alvo ou seus produtos de gene.

Os ácidos nucleicos e/ou variantes da invenção ou ácidos nucleicos que compreendem a sua sequência complementar, podem ser usados como construções de anti-sentido para controlar a expressão de gene em células, tecidos ou órgãos. A metodologia associada com técnicas de anti-sentido é bem conhecida para o técnico habilitado e é descrito e revisado em *Antisense Drug Technology: Principles, Strategies and Applications*, Crooke, ed., Marcel Dekker Inc., Nova Iorque (2001). Em uma maneira geral, as moléculas de ácido nucleico de anti-sentido são planejadas para serem complementares a uma região de mRNA expressada por um gene, de modo que a molécula de anti-sentido hibridiza ao mRNA, bloqueando assim a tradução do mRNA em proteína. Diversas classes de oligonucleotídeo de anti-sentido são conhecidos por aqueles habilitados na técnica, incluindo

clivadores e bloqueadores. O primeiro liga-se aos sítios de RNA alvo, ativam nucleases intracelulares (por exemplo, RnaseH ou Rnase L), que clivam o RNA alvo. Os bloqueadores se ligam ao RNA alvo, inibem a tradução da proteína pelo impedimento estérico dos ribossomas. Os exemplos de bloqueadores incluem ácidos nucleicos, compostos de morfolino, ácido nucleicos bloqueados e metilfosfonatos (Thompson, *Drug Discovery Today*, 7: 912-917 (2002)). Oligonucleotídeos de anti-sentido são úteis diretamente como agentes terapêuticos e também são úteis para determinar e validar a função de gene, por exemplo pelo silenciamento de gene ou eliminação de gene. A tecnologia de anti-sentido é ainda descrita em Lavery *et al.*, *Curr. Opin. Drug Discov. Devel.* 6: 561-569 (2003), Stephens *et al.*, *Curr. Opin. Mol. Ther.* 5: 118-122 (2003), Kurreck, *Eur. J. Biochem.* 270: 1628-44 (2003), Dias *et al.*, *Mol. Cancer Ter.* 1: 347-55 (2002), Chem. Methods Mol. Med. 75: 621-636 (2003), Wang *et al.*, *Curr. Cancer Drug Targets* 1: 177-96 (2001) e Bennett, *Anti-sense Nucleic Acid Drug. Dev.* 12: 215-24 (2002).

As variantes aqui descritas podem ser usadas para a seleção e planejamento de reagentes de anti-sentido que são específicos para as variantes particulares. Usando informação a cerca das variantes aqui descritas, oligonucleotídeos de anti-sentido ou outras moléculas de anti-sentido que especificamente alvejam moléculas de mRNA que contêm uma ou mais variantes da invenção podem ser planejados. Desta maneira, a expressão de moléculas de mRNA que contêm uma ou mais variantes da presente invenção (marcadores e/ou haplótipos) podem ser inibidos ou bloqueados. Em uma forma de realização, as moléculas de anti-sentido são planejadas para ligar especificamente uma forma alélica particular (isto é, uma ou diversas variantes (alelos e/ou haplótipos)) do ácido nucleico alvo, inibindo deste modo a tradução de um produto que se origina deste alelo ou haplótipo específicos, mas que não se ligam a outras variantes ou variantes alternativas nos sítios polimórficos específicos da molécula de ácido nucleico alvo.

Visto que as moléculas de anti-sentido podem ser usadas para inativar mRNA de modo a inibir a expressão de gene e assim a expressão de proteína, as moléculas podem ser usadas para tratar uma doença ou distúrbio, tal com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. A metodologia pode envolver a clivagem por meio de sequências de nucleotídeo complementares contendo ribozimas a uma ou mais regiões no mRNA que atenuam a capacidade do mRNA para ser traduzido. Tais regiões de mRNA incluem, por exemplo, regiões que codificam proteína, em particular regiões que codificam proteína que correspondem à atividade catalítica, substrato e/ou sítios de ligação de ligando ou outros domínios funcionais de uma proteína.

O fenômeno da interferência de RNA (RNAi) foi ativamente estudado na última década, visto que a sua descoberta original em *C. elegans* (Fire *et al.*, Nature 391: 806-11 (1998)) e em anos recentes o seu uso potencial no tratamento de doença humana tem sido ativamente procurado (revisado em Kim & Rossi, Nature Rev. Genet. 8: 173-204 (2007)). A interferência de RNA (RNAi), também chamada de silenciamento de gene, está fundamentada no uso de moléculas de RNA de filamento duplo (dsRNA) para desligar genes específicos. Na célula, as moléculas de RNA de filamento duplo citoplásmicas (dsRNA) são processadas pelos complexos celulares em RNA interferente pequeno (siRNA). O siRNA guia o alvejamento de um complexo de RNA proteína aos sítios específicos em um mRNA alvo, levando à clivagem do mRNA (Thompson, Drug Discovery Today, 7: 912-917 (2002)). As moléculas de siRNA são tipicamente de cerca de 20, 21, 22 ou 23 nucleotídeos no comprimento. Assim, um aspecto da invenção diz respeito a moléculas de ácido nucleico isoladas e ao uso destas moléculas para a interferência de RNA, isto é, como moléculas de RNA interferentes pequenas (siRNA). Em uma forma de realização, as moléculas de ácido nucleico isoladas têm 18 a 26 nucleotídeos no comprimento, preferivelmente

19 a 25 nucleotídeos no comprimento, mais preferivelmente 20 a 24 nucleotídeos no comprimento e mais preferivelmente 21, 22 ou 23 nucleotídeos no comprimento.

Um outro caminho para o silenciamento de gene mediado pelo RNAi origina-se nos transcritos de microRNA primários endogenamente codificados (pri-miRNA), que são processados na célula para gerar miRNA precursor (pré-miRNA). Estas moléculas de miRNA são exportadas do núcleo para o citoplasma, onde elas passam pelo processamento para gerar moléculas de miRNA maduras (miRNA), que direcionam a inibição traducional pelo reconhecimento de sítios alvo nas regiões 3' não traduzidas de mRNAs e subsequente degradação de mRNA pelo processamento de corpos P (revisado em Kim & Rossi, *Nature Rev. Genet.* 8: 173-204 (2007)).

As aplicações clínicas de RNAi incluem a incorporação de duplexes de siRNA sintéticos, que preferivelmente têm aproximadamente 20 a 23 nucleotídeos no tamanho e preferivelmente têm sobreposições em 3' de 2 nucleotídeos. A eliminação da expressão de gene é estabelecida pelo planejamento específico de sequência para o mRNA alvo. Diversos sítios comerciais para o planejamento ótimo e síntese de tais moléculas são conhecidos por aqueles habilitados na técnica.

Outras aplicações fornecem moléculas de siRNA mais longas (tipicamente de 25 a 30 nucleotídeos no comprimento, preferivelmente cerca de 27 nucleotídeos), assim como RNAs de grampo de cabelo pequenos (shRNAs; tipicamente de cerca de 29 nucleotídeos no comprimento). Os últimos são naturalmente expressados, como descrito em Amarzguioui *et al.* (*FEBS Lett.* 579: 5974-81 (2005)). siRNAs e shRNAs quimicamente sintéticos são substratos para o processamento *in vivo* e em alguns casos fornecem silenciamento de gene mais potente do que os projetos mais curtos (Kim *et al.*, *Nature Biotechnol.* 23: 222-226 (2005); Siolas *et al.*, *Nature Biotechnol.* 23: 227-231 (2005)). Em uma maneira geral siRNAs fornecem o

silenciamento transitório de expressão de gene, porque a sua concentração intracelular é diluída pelas divisões celulares subsequentes. Ao contrário, shRNAs expressados medeiam a eliminação de longa duração, estável de transcritos alvo, contanto que a transcrição do shRNA ocorra (Marques *et al.*, Nature Biotechnol. 23: 559-565 (2006); Brummelkamp *et al.*, Science 296: 550-553 (2002)).

Visto que as moléculas de RNAi, incluindo siRNA, miRNA e shRNA, atuam em uma maneira dependente de sequência, as variantes da presente invenção (por exemplo, os marcadores e haplótipos associados com bloco LD C04, por exemplo, os marcadores listados nas Tabelas 5A e 5B) podem ser usados para planejar reagentes de RNAi que reconhecem moléculas de ácido nucleico específicas que compreendem alelos e/ou haplótipos específicos (por exemplo, os alelos e/ou haplótipos da presente invenção), enquanto que não reconhece as moléculas de ácido nucleico que compreendem outros alelos ou haplótipos. Estes reagentes de RNAi podem assim reconhecer e destruir as moléculas de ácido nucleico alvo. Como com reagentes de anti-sentido, os reagentes de RNAi podem ser úteis como agentes terapêuticos (isto é, para desligar genes associados com doença ou variantes de gene associados com doença), mas também podem ser úteis para caracterizar e validar a função de gene (por exemplo, pelos experimentos de silenciamento de gene ou eliminação de gene).

A liberação de RNAi pode ser realizada por uma faixa de metodologias conhecidas por aqueles habilitados na técnica. Os métodos de utilizar a liberação não viral incluem colesterol, partícula de ácido nucleico-lipídeo estável (SNALP), fragmento de anticorpo de cadeia pesada (Fab), aptâmeros e nanopartículas. Os métodos de liberação viral incluem o uso de lentivírus, adenovírus e vírus adeno-associados. As moléculas de siRNA são em algumas formas de realização quimicamente modificadas para aumentar a sua estabilidade. Isto pode incluir modificações na posição 2' da ribose,

incluindo 2'-O-metilpurinas e 2'-fluoropirimidinas, que fornecem resistência à atividade de Rnase. Outras modificações químicas são possíveis e conhecidas por aqueles habilitados na técnica.

As seguintes referências fornecem um outro resumo de RNAi e possibilidades para alvejar genes específicos usando RNAi: Kim & Rossi, Nat. Rev. Genet. 8: 173-184 (2007), Chen & Rajewsky, Nat. Rev. Genet. 8: 93-103 (2007), Reynolds, *et al.*, Nat. Biotechnol. 22: 326-330 (2004), Chi *et al.*, Proc. Natl. Acad. Sci. USA 100: 6343-6346 (2003), Vickers *et al.*, J. Biol. Chem. 278: 7108-7118 (2003), Agami, Curr. Opin. Chem. Biol. 6: 829-834 (2002), Lavery, *et al.*, Curr. Opin. Drug Discov. Devel. 6: 561-569 (2003), Shi, Trends Genet. 19: 9-12 (2003), Shuey *et al.*, Drug Discov. Today 7: 1040-46 (2002), McManus *et al.*, Nat. Rev. Genet. 3: 737-747 (2002), Xia *et al.*, Nat. Biotechnol. 20: 1006-10 (2002), Plasterk *et al.*, Curr. Opin. Genet. Dev. 10: 562-7 (2000), Bosher *et al.*, Nat. Cell Biol. 2: E31-6 (2000) e Hunter, Curr. Biol. 9: R440-442 (1999).

Um defeito genético que leva à predisposição ou risco aumentados para o desenvolvimento de uma doença, incluindo arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral ou um defeito que causa a doença, podem ser corrigidos permanentemente pela administração a um paciente que carrega o defeito um fragmento de ácido nucleico que incorpora uma sequência de reparo que fornece o(s) nucleotídeo(s) normal(is)/tipo selvagem no sítio do defeito genético. Tal sequência de reparo específica de sítio pode *concompass* um oligonucleotídeo de RNA/DNA que opera para promover o reparo endógeno de um DNA genômico do paciente. A administração da sequência de reparo pode ser realizada por um veículo apropriado, tal com um complexo com polietilenimina, encapsulado em lipossomas aniônicos, um vetor viral tal com um vetor de adenovírus ou outras composições farmacêuticas adequadas para promover a captação intracelular do ácido nucleico administrado. O defeito

genético pode ser depois superado, visto que os oligonucleotídeos quiméricos induzem a incorporação da sequência normal no genoma do paciente, levando à expressão do produto de gene normal/tipo selvagem. A substituição é propagada, restituindo assim um reparo permanente e alívio dos sintomas associados com a doença ou condição.

A presente invenção fornece métodos para identificar compostos ou agentes que podem ser usados para tratar arritmia cardíaca, por exemplo fibrilação atrial e flúter atrial e acidente vascular cerebral. Assim, as variantes da invenção são úteis como alvos para a identificação e/ou desenvolvimento de agentes terapêuticos. Tais métodos podem incluir ensaiar a capacidade de um agente ou composto para modular a atividade e/ou expressão de um ácido nucleico que inclui pelo menos uma das variantes (marcadores e/ou haplótipos) da presente invenção ou o produto codificado do ácido nucleico. Este por sua vez pode ser usado para identificar agentes ou compostos que inibem ou alteram a atividade ou expressão indesejadas do produto de ácido nucleico codificado. Os ensaios para realizar tais experimentos podem ser realizados em sistemas com base em célula ou em sistemas isentos de célula, como conhecidos pela pessoa habilitada. Os sistemas com base em célula incluem células que expressam naturalmente as moléculas de ácido nucleico de interesse ou células recombinantes que foram geneticamente modificadas de modo a expressar uma certa molécula de ácido nucleico desejada.

A expressão de gene variante em um paciente pode ser avaliada pela expressão de uma sequência de ácido nucleico contendo variante (por exemplo, um gene contendo pelo menos uma variante da presente invenção, que pode ser transcrita no RNA contendo a pelo menos uma variante e por sua vez traduzida em proteína) ou pela expressão alterada de uma sequência de ácido nucleico normal/tipo selvagem devido às variantes que afetam o nível ou padrão de expressão dos transcritos normais, por

exemplo variantes na região regulatória ou de controle do gene. Ensaios para a expressão de gene incluem ensaios de ácido nucleico diretos (mRNA), ensaios quanto ao nível de proteína expressada ou ensaios de compostos colaterais envolvidos em um caminho, por exemplo um caminho de sinal.

- 5 Além disso, a expressão de genes que são supra- ou infra-regulados em resposta ao caminho de sinal também podem ser ensaiados. Uma forma de realização inclui ligar operavelmente um gene repórter, tal com luciferase, à região reguladora do(s) gene(s) de interesse.

Moduladores da expressão de gene em uma forma de
10 realização podem ser identificados quando uma célula é contatada com um composto candidato ou agente e a expressão de mRNA é determinada. O nível de expressão de mRNA na presença do composto ou agente candidatos é comparado com o nível de expressão na ausência do composto ou agente. Com base nesta comparação, compostos ou agentes candidatos para tratar
15 distúrbios tais com a fibrilação atrial, flúter atrial e acidente vascular cerebral podem ser identificados como aqueles que modulam a expressão de gene do gene variante. Quando a expressão de mRNA ou a proteína codificada é de modo estatístico significativamente maior na presença do composto ou agente candidatos do que na sua ausência, então o composto ou agente candidatos
20 são identificados como um estimulador ou supra-regulador da expressão do ácido nucleico. Quando a expressão de ácido nucleico ou nível de proteína é de modo estatístico significativamente menor na presença do composto ou agente candidatos do que na sua ausência, então o composto candidato é identificado como um inibidor ou infra-regulador da expressão de ácido
25 nucleico.

A invenção fornece ainda métodos de tratamento usando um composto identificado através da triagem de medicamento (composto e/ou agente) como um gene modulador (isto é, estimulador e/ou inibidor da expressão de gene).

Em um outro aspecto da presente invenção, um pacote farmacêutico (kit) é fornecido, o pacote compreendendo um agente terapêutico e um conjunto de instruções para a administração do agente terapêutico aos seres humanos diagnosticamente testado quanto a uma ou mais variantes da presente invenção, como aqui divulgadas. O agente terapêutico pode ser um medicamento de molécula pequena, um anticorpo, um peptídeo, uma molécula de anti-sentido ou RNAi ou outras moléculas terapêuticas. Em uma forma de realização, um indivíduo identificado como um portador de pelo menos uma variante da presente invenção é instruído tomar uma dose prescrita do agente terapêutico. Em uma tal forma de realização, um indivíduo identificado como um portador homozigoto de pelo menos uma variante da presente invenção é instruído tomar uma dose prescrita do agente terapêutico. Em uma outra forma de realização, um indivíduo identificado como um não portador de pelo menos uma variante da presente invenção é instruído tomar uma dose prescrita do agente terapêutico.

Métodos de avaliar a probabilidade de resposta aos agentes terapêuticos, métodos de monitorar o progresso de tratamento e métodos de tratamento

Como é conhecido na técnica, indivíduos podem ter respostas diferenciais a uma terapia particular (por exemplo, um agente terapêutico ou método terapêutico). Farmacogenômicos tratam o problema de como variações genéticas (por exemplo, as variantes (marcadores e/ou haplótipos) da presente invenção) afetam a resposta ao medicamento, devido à disposição de medicamento alterada e/ou ação anormal ou alterada do medicamento. Assim, a base da resposta diferencial pode ser geneticamente determinada em parte. Resultados clínicos devido às variações genéticas que afetam a resposta a medicamento podem resultar em toxicidade do medicamento em certos indivíduos (por exemplo, carregadores ou não carregadores das variantes genéticas da presente invenção) ou insuficiência terapêutica do medicamento. Portanto, as variantes da presente invenção podem determinar a maneira em

que um agente terapêutico e/ou método atuam sobre o corpo ou o modo em que o corpo metaboliza o agente terapêutico.

Consequentemente, em uma forma de realização, a presença de um alelo particular em um sítio polimórfico ou haplótipo é indicativa de um diferente, por exemplo uma taxa de resposta diferente, a uma modalidade de tratamento particular. Isto significa que um paciente diagnosticado com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral e que carrega um certo alelo em um polimórfico ou haplótipo da presente invenção (por exemplo, os alelos em risco e protetores e/ou haplótipos da invenção) responderiam melhor ou pior a um produto terapêutico, medicamento e/ou outra terapia específicos usados para tratar a doença. Portanto, a presença ou ausência do alelo marcador ou haplótipo poderia ajudar na decisão de qual tratamento deva ser usado para um paciente. Por exemplo, para um paciente recém diagnosticado, a presença de um marcador ou haplótipo da presente invenção pode ser avaliada (por exemplo, através de teste de derivado de DNA a partir de uma amostra de sangue, como aqui descrito). Se o paciente é positivo para um alelo marcador ou haplótipo (isto é, pelo menos um alelo específico do marcador ou haplótipo, está presente), então o médico recomenda uma terapia particular, enquanto se o paciente for negativo para o pelo menos um alelo de um marcador ou um haplótipo, então um curso diferente de terapia pode ser recomendada (que pode incluir recomendar que nenhuma terapia imediata, outra que não o monitoramento em série quanto a progressão da doença, seja realizada). Assim, a situação de portador do paciente seria usada para ajudar a determinar se uma modalidade de tratamento particular deve ser administrada. O valor reside dentro das possibilidades de ser capaz de diagnosticar a doença em um estágio inicial, para selecionar o tratamento mais apropriado e fornecer informação ao médico a cerca do prognóstico/agressividade da doença de modo a ser capaz de aplicar o tratamento mais apropriado.

Tratamento da Fibrilação atrial e Flúter atrial é em uma maneira geral direcionado por dois objetivos principais: (i) prevenir o acidente vascular cerebral e (ii) tratar os sintomas.

(i) Prevenção do Acidente Vascular Cerebral

5 Anticoagulação é a terapia de escolha para a prevenção do acidente vascular cerebral na fibrilação atrial e é indicada para a maioria dos pacientes com esta arritmia. Os únicos pacientes para os quais a anticoagulação não é fortemente recomendada são aqueles mais jovens do que 65 anos de idade que são considerados de risco baixo, isto é, eles não têm
10 nenhuma doença de coração orgânica, nenhuma hipertensão, nenhuma história anterior de acidente vascular cerebral ou Ataques Isquêmicos Transitórios e nenhuma diabete. Este grupo como um todo tem um risco mais baixo de acidente vascular cerebral e a prevenção do acidente vascular cerebral com aspirina é em uma maneira geral recomendada. Para todos os
15 outros pacientes, a anticoagulação é indicada se a fibrilação atrial é permanente, paroxísmicos recorrentes ou recorrente persistente. Não pode ser generalizado como os pacientes que se apresentam com o seu primeiro episódio de fibrilação atrial paroxísmica deve ser tratado e a decisão precisa ser individualizado para cada paciente. A anticoagulação também é indicada
20 mesmo quando o paciente com fibrilação atrial está sentindo ser mantido em ritmo sinusal com a terapia antiarrítmica (ritmo controlado) visto que este tipo de terapia não afeta o risco de acidente vascular cerebral.

 Anticoagulantes. A anticoagulação é recomendada na fibrilação atrial, como detalhado acima, para a prevenção de cardioembolismo
25 e acidente vascular cerebral. O anticoagulante oral mais amplamente estudado é a varfarina e esta medicação é universalmente recomendada para a anticoagulação oral crônica na fibrilação atrial. A varfarina tem poucos efeitos colaterais com a exceção do risco de hemorragia mas requer monitoramento regular e cuidadoso de valores sanguíneos durante a terapia

(para medir o efeito da anticoagulação). O anticoagulante oral ximelagatran mostrou-se promissor na prevenção do acidente vascular cerebral em pacientes com fibrilação atrial e teve a vantagem de não requerer o monitoramento regular como a varfarina. Ximelagatran foi descoberto
 5 entretanto causar lesão hepática inexplicada e foi retirado do mercado em 2006. Vários agentes estão disponíveis para a terapia intravenosa e/ou subcutânea, incluindo heparina e as heparinas de peso molecular baixo (por exemplo enoxaparina, dalteparina, tinzaparina, ardeparina, nadroparina e reviparina). Estas medicações são recomendadas quando o início rápido da
 10 anticoagulação é necessário ou se a terapia de anticoagulação oral tem que ser interrompida em pacientes de alto risco ou por tempo mais longo do que uma semana em outros pacientes por exemplo devido a uma série de procedimentos. Outros anticoagulantes parenterais são disponíveis mas não especificamente recomendados como terapia na fibrilação atrial; por exemplo,
 15 os inibidores do fator Xa fondaparinux e idraparinux, os inibidores de trombina lepirudin, bivalirudin e argatroban assim como danaparoid.

*(ii) Controle de Sintoma. A terapia médica e cirúrgica aplicada ao controle dos sintomas de fibrilação atrial é feita de encomenda para o paciente individual e consiste do controle da frequência e/ou ritmos cardíacos com
 20 medicações, remoção por radiofrequência e/ou cirurgia.*

Medicações Antiarrítmicas. Em termos gerais, os agentes antiarrítmicos são usados para suprimir ritmos anormais do coração que são característicos de arritmias cardíacas, incluindo fibrilação atrial e flúter atrial. Uma classificação de agentes antiarrítmicos é a classificação de Vaughan
 25 Williams, em que cinco categorias principais de agentes antiarrítmicos são definidas. Os agentes de Classe I são bloqueadores rápidos do canal de sódio e são subclassificados com base nas cinéticas e força de bloqueio assim como seu efeito na repolarização. A Classe Ia inclui disopiramida, moricizina, procainamida e quinidina. Os agentes da Classe Ib são lidocaína, mexiletina,

tocainida e fenitoína. Os agentes da Classe Ic são encainida, flecainida, propafenona, aimalina, cibenzolina e detaímio. Os agentes de Classe II são bloqueadores beta, eles bloqueiam os efeitos de catecolaminas nos receptores beta-adrenérgicos. Os exemplos de bloqueadores beta são esmolol, propranolol, metoprolol, alprenolol, atenolol, carvedilol, bisoprolol, acebutolol, nadolol, pindolol, labetalol, oxprenotol, penbutolol, timolol, betaxolol, cartelol, sotalol e levobunolol. Os agentes de Classe III têm propriedades mistas mas são coletivamente bloqueadores do canal de potássio e prolongam a repolarização. As medicações nesta categoria são amiodarona, azimilida, bretílio, dofetilida, tedisamil, ibutilida, sematilida, sotalol, N-acetil procainamida, cloridreto de nifecalante, vernacalante e ambasilida. Os agentes de Classe IV são bloqueadores do canal de cálcio e incluem verapamil, mibefradil e diltiazem.

Finalmente, a classe V consiste de antiarrítmicos mistos e inclui digoxina e adenosina.

Controle da Frequência Cardíaca. Medidas farmacológicas para a manutenção do controle da frequência cardíaca incluem bloqueadores beta, bloqueadores do canal de cálcio e digoxina. Todas estas medicações diminuem a condução elétrica através do nodo atrioventricular e diminuem a resposta da frequência ventricular para a fibrilação atrial rápida. Alguns antiarrítmicos usados primariamente para controle do ritmo (ver abaixo) também diminuem a frequência de condução do nodo atrioventricular e assim a resposta cardíaca à frequência ventricular. Estes incluem algumas medicações da classe III e Ic tais com amiodarona, sotalol e flecainida.

Cardioversão. A cardioversão do ritmo cardíaco da fibrilação atrial ou flúter atrial para ritmo sinusal pode ser obtido eletricamente, com cardioversão de corrente direta sincronizada ou com medicações tais como ibutilida, amiodarona, procainamida, propafenona e flecainida.

Controle do ritmo cardíaco

As medicações usadas para a manutenção do ritmo sinusal, isto é, controle de ritmo, incluem principalmente medicações antiarrítmicas das classes III, Ia e Ic. Os exemplos são sotalol, amiodarona e dofetilida da classe III, disopiramida, procainamida e quinidina da classe Ia e flecínida e propafenona da classe Ic. O tratamento com estas medicações antiarrítmicas é complicada, pode ser perigosa e deve ser direcionado pelos médicos especificamente treinados para o uso destas medicações. Muitos dos antiarrítmicos têm efeitos colaterais sérios e devem ser usados apenas em populações específicas. Por exemplo, as medicações de classe Ic não devem ser usadas em pacientes com doença da artéria coronária e mesmo se eles pudessem suprimir a fibrilação atrial, eles poderiam realmente promover a resposta ventricular rápida no flúter atrial. As medicações de Classe Ia podem ser usados como último recurso em pacientes sem doenças do coração estruturais. Sotalol (como a maioria dos antiarrítmicos de classe III) pode causar prolongação significativa do intervalo QT, especificamente em pacientes com insuficiência renal e promover arritmias ventriculares sérias. Tanto o sotalol quanto a dofetilida assim como as medicações Ia precisam ser iniciadas em uma base de paciente internado para monitorar o intervalo QT. Embora a amiodarona seja usualmente bem tolerada e seja amplamente usada, a amiodarona tem muitos efeitos colaterais sérios com a terapia de longa duração.

Como o teste genético pode afetar diretamente a escolha de tratamento

Quando indivíduos presentes com o seu primeiro episódio (diagnosticado) de fibrilação atrial paroxísmica e converte espontaneamente para ritmo sinusal ou passa pela cardioversão elétrica ou química em menos do que 48 horas no episódio, a decisão para iniciar ou não, a terapia de anticoagulação, é individualizada com base no perfil de risco do paciente em questão e a preferência dos médicos administradores. Esta pode ser uma

escolha difícil a fazer visto que comprometer o paciente com a terapia de anticoagulação tem um impacto maior sobre a vida do paciente. Frequentemente a escolha é feita para impedir a anticoagulação em uma tal situação e isto pode ser de nenhuma consequência significativa para o paciente. Por outro lado o paciente pode mais tarde desenvolver um acidente vascular cerebral e a oportunidade de prevenção pode assim ter sido perdida. Em tais circunstâncias, saber que o paciente é um portador da variante de risco pode ser de enorme significância e sustenta o início do tratamento de anticoagulação.

Indivíduos que são diagnosticados com fibrilação atrial sob a idade de 65 e são de outro modo considerados de risco baixo para o acidente vascular cerebral, isto é, não tem nenhuma doença cardíaca orgânica, nenhuma hipertensão, nenhuma diabetes e nenhuma história anterior de acidente vascular cerebral, são em uma maneira geral tratados com apenas com aspirina para a prevenção do acidente vascular cerebral e não anticoagulação. Se um tal paciente é descoberto ser portador para as variantes de risco aqui descritas, isto pode ser considerado suporte para iniciar a anticoagulação mais cedo do que de outro modo recomendado. Isto seria uma consideração razoável visto que os resultados de acidente vascular cerebral a partir da fibrilação atrial pode ser devastadora.

O acidente vascular cerebral isquêmico é em uma maneira geral classificado em cinco subtipos com base na causa suspeita; aterosclerose de artéria grande, oclusão de artéria pequena, cardioembolismo (a maioria devido à fibrilação atrial), acidente vascular cerebral de outra causa determinada e acidente vascular cerebral de causa não determinada (nenhuma causa descoberta ou mais do que 1 causa plausível). De maneira importante, os acidentes vasculares cerebrais devido ao cardioembolismo têm a recorrência mais alta, são mais incapacitantes e estão associados com a sobrevivência mais baixa. É portanto imperativo não examinar a fibrilação

atrial como a causa principal do acidente vascular cerebral, particularmente visto que as medidas de tratamento variam com base no subtipo. Portanto, se um indivíduo é diagnosticado com acidente vascular cerebral ou um Ataque Isquêmico Transitório e uma causa plausível não é identificada a despeito do trabalho padrão, saber que o paciente é um portador da variante de risco pode ser de enorme valor e sustenta o início do tratamento de anticoagulação ou teste de diagnóstico mais agressivo na tentativa de diagnosticar a fibrilação atrial.

Além disso, os marcadores da presente invenção podem ser usados para aumentar o poder e a eficácia de testes clínicos. Assim, indivíduos que são portadores de pelo menos uma variante de risco da presente invenção, isto é, indivíduos que são portadores de pelo menos um alelo de pelo menos um marcador polimórfico que confere risco aumentado de desenvolver arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral podem ser mais prováveis de responder a uma modalidade de tratamento particular, por exemplo, como descrito acima. Em uma forma de realização, indivíduos que portam variantes de risco para o(s) gene(s) em um caminho e/ou rede metabólica para os quais um tratamento particular (por exemplo, medicamento de molécula pequena) é alvejado, são mais prováveis de serem respondedores para o tratamento. Em uma outra forma de realização, indivíduos que portam variantes de risco para um gene, cuja expressão e/ou função são alteradas pela variante de risco, são mais prováveis de serem respondedores a uma modalidade de tratamento que alveja este gene, a sua expressão ou seu produto de gene. Esta aplicação pode melhorar a segurança de testes clínicos, mas também pode realçar a chance de que um teste clínico demonstrará eficácia estatisticamente significativa, que pode ser limitada a um certo subgrupo da população. Assim, um resultado possível de um tal teste é que carregadores de certas variantes genéticas, por exemplo, os marcadores e haplótipos da presente invenção, são de modo

estatisticamente significativa prováveis de mostrar resposta positiva ao agente terapêutico, isto é, experimentar alívio de sintomas associados com a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral quando tomar o agente terapêutico ou medicamento como prescritos.

Em um outro aspecto, os marcadores e haplótipos da presente invenção podem ser usados para alvejar a seleção de agentes farmacêuticos para indivíduos específicos. A seleção personalizada de modalidades de tratamento, mudanças de estilo de vida ou combinação dos dois, pode ser realizada pela utilização das variantes de risco da presente invenção. Assim, o conhecimento de uma situação do indivíduo quanto aos marcadores particulares da presente invenção, pode ser útil para a seleção de opções de tratamento que alvejam genes ou produtos de gene afetados pelas variantes de risco da invenção. Certas combinações de variantes podem ser adequadas para uma seleção de opções de tratamento, enquanto que outras combinações de variante de gene podem alvejar outras opções de tratamento. Tal combinação de variante pode incluir uma variante, duas variantes, três variantes ou quatro ou mais variantes, como necessário para determinar com precisão clinicamente confiável a seleção do módulo de tratamento.

Aplicações implementadas por computador

A presente invenção também diz respeito a aplicações implementadas por computador dos marcadores polimórficos e haplótipos aqui descrito a serem associados com a arritmia cardíaca (por exemplo, fibrilação atrial e flúter atrial) e acidente vascular cerebral. Tais aplicações podem ser úteis para armazenar, manipular ou de outro modo analisar dados genotípicos que são úteis nos métodos da invenção. Um exemplo diz respeito a armazenar informação de derivado de genótipo de um indivíduo em meio legível, de modo a ser capaz de fornecer a informação de genótipo a um terceiro interessado (por exemplo, o indivíduo) ou para derivar informação a partir dos dados genotípicos, por exemplo, pela comparação dos dados

genotípicos para informação a cerca de fatores de risco genético que contribuem para a suscetibilidade aumentada para a arritmia cardíaca (por exemplo, fibrilação atrial e flúter atrial) e acidente vascular cerebral e relatar os resultados com base em tal comparação.

5 Um tal aspecto diz respeito a meios legíveis por computador. Em termos gerais, tal meio tem capacidades de armazenar (i) informação identificadora quanto a pelo menos um marcador polimórfico ou um
10 haplótipo; (ii) um indicador da frequência de pelo menos um alelo do dito pelo menos um marcador ou a frequência de um haplótipo, em indivíduos com a arritmia cardíaca (por exemplo, fibrilação atrial e flúter atrial) e/ou
acidente vascular cerebral; e um indicador da frequência de pelo menos um
alelo do dito pelo menos um marcador ou a frequência de um haplótipo, em
uma população de referência. A população de referência pode ser uma
população livre de doença de indivíduos. Alternativamente, a população de
15 referência é uma amostra aleatória da população geral e é assim representativa da população como um todo. O indicador de frequência pode ser uma frequência calculada, uma contagem de cópias de alelos e/ou haplótipo ou valores normalizados ou de outro modo manipulados das frequências reais
que são adequadas para o meio particular.

20 Informação adicional a cerca do indivíduo pode ser armazenada no meio, tal com informação de ancestralidade, informação a cerca do sexo, atributos ou características físicos (incluindo altura e peso),
medições bioquímicas (tais com pressão sanguínea, níveis de lipídeos do
sangue, níveis de glicose no jejum, medições de resposta à insulina),
25 resultados de biomarcador ou outra informação útil que seja desejável para armazenar ou manipular no contexto da situação genotípica de um indivíduo particular.

A invenção além disso diz respeito a um aparelho que é adequado para a determinação ou manipulação de dados genéticos úteis para

determinar uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial e flúter atrial) e acidente vascular cerebral em um indivíduo humano. Um tal aparelho pode incluir uma memória legível por computador, uma rotina para manipular os dados armazenados na memória legível por computador e uma rotina para gerar um produto que inclui uma medida dos dados genéticos. Tal medida pode incluir valores tais com frequências alélicas ou haplótipo, contagens genotípicas, sexo, idade, informação de fenótipo, valores para Razão de Chances (OR) ou Risco Relativo (RR), risco atribuível à população (PAR) ou outra informação útil que seja uma estatística direta dos dados genotípicos originais ou com base nos cálculos com base nos dados genéticos.

As aplicações descritas acima podem ser todas praticadas com os marcadores e haplótipos da invenção que foram descritas em mais detalhes com respeito aos métodos de avaliar a suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial e flúter atrial) e acidente vascular cerebral. Assim, estas aplicações em uma maneira geral podem ser reduzidas para praticar o uso dos marcadores listados nas Tabelas 5, Tabela 4, Tabela 9 e Tabela 19 e marcadores em desequilíbrio de ligação com estes. Em uma forma de realização, os marcadores ou haplótipos estão presentes dentro do segmento genômico cujas sequências são apresentadas na SEQ ID NO: 50. Em uma outra forma de realização, os marcadores ou haplótipos compreendem pelo menos um marcador selecionado dos marcadores apresentados na Tabela 19. Em uma outra forma de realização, os marcadores e haplótipos compreendem pelo menos um marcador selecionado de D4S406 (SEQ ID NO: 45), rs2634073 (SEQ ID NO: 33), rs2200733 (SEQ ID NO: 28), rs2220427 (SEQ ID NO: 1), rs10033464 (SEQ ID NO: 41) e rs13143308 (SEQ ID NO: 51), opcionalmente incluindo os marcadores em desequilíbrio de ligação com estes. Em uma tal forma de realização, o desequilíbrio de ligação é definido pelos valores numéricos para r^2 maiores do que 0,1. Em

uma outra tal forma de realização, o desequilíbrio de ligação é definido pelos valores numéricos para r^2 maiores do que 0,2. Em uma outra forma de realização, o marcador ou haplótipo compreende pelo menos um alelo selecionado dos alelos 2, 4 e/ou 8 no marcador D4S406, alelo A do marcador rs2634073, alelo T do marcador rs2200733, alelo T do marcador rs2220427, alelo T do marcador rs10033464, e/ou alelo G do marcador rs13143308

Ácidos nucleicos e polipeptídeos

Os ácidos nucleicos e polipeptídeos aqui descritos podem ser usados em métodos e kits da presente invenção, como descrito acima.

Uma molécula de ácido nucleico “isolada”, como aqui usada, é uma que é separada de ácidos nucleicos que normalmente flanqueiam o gene ou sequência de nucleotídeo (como em sequências genômicas) e/ou foi completa ou parcialmente purificado de outras sequências transcritas (por exemplo, como em uma biblioteca de RNA). Por exemplo, um ácido nucleico isolado da invenção pode ser substancialmente isolado com respeito ao meio celular complexo em que o mesmo naturalmente ocorre ou meio de cultura quando produzidos pelas técnicas recombinantes ou precursores químicos ou outros produtos químicos quando quimicamente sintetizados. em alguns casos, o material isolado formará parte de uma composição (por exemplo, um extrato bruto contendo outras substâncias), sistema de tampão ou mistura de reagente. Em outras circunstâncias, o material pode ser purificado essencialmente até a homogeneidade, por exemplo como determinado pela eletroforese em gel de poliacrilamida (PAGE) ou cromatografia de coluna (por exemplo, HPLC). Uma molécula de ácido nucleico isolada da invenção pode compreender pelo menos cerca de 50%, pelo menos cerca de 80% ou pelo menos cerca de 90% (em uma base molar) de todas as espécies macromoleculares presentes. Com respeito ao DNA genômico, o termo “isolado” também pode referir-se às moléculas de ácido nucleico que são separadas do cromossoma com que o DNA genômico está naturalmente

associado. Por exemplo, a molécula de ácido nucleico isolada pode conter menos do que cerca de 250 kb, 200 kb, 150 kb, 100 kb, 75 kb, 50 kb, 25 kb, 10 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0,5 kb ou 0,1 kb dos nucleotídeos que flanqueiam a molécula de ácido nucleico no DNA genômico da célula a partir da qual a molécula de ácido nucleico é derivada.

A molécula de ácido nucleico pode ser fundida a outras sequências codificadoras ou reguladoras e ainda ser considerada isolada. Assim, o DNA recombinante contido em um vetor é incluído na definição de “isolado” como aqui usada. Também, moléculas de ácido nucleico isoladas incluem moléculas de DNA recombinante em células hospedeiras heterólogas ou organismos heterólogos, assim como moléculas de DNA parcial ou substancialmente purificadas em solução. As moléculas de ácido nucleico “isoladas” também abrangem transcritos de RNA *in vivo* e *in vitro* das moléculas de DNA da presente invenção. Uma molécula de ácido nucleico ou sequência de nucleotídeo isoladas podem incluir uma molécula de ácido nucleico ou sequência de nucleotídeo que é quimicamente sintetizada ou por meios recombinantes. Tais sequências de nucleotídeo isoladas são úteis, por exemplo, na fabricação do polipeptídeo codificado, como sondas para isolar sequências homólogas (por exemplo, de outra espécie de mamífero), para o mapeamento de gene (por exemplo, pela hibridização *in situ* com cromossomas) ou para detectar a expressão do gene no tecido (por exemplo, tecido humano), tal com pela análise de Northern blot ou outras técnicas de hibridização.

A invenção também diz respeito às moléculas de ácido nucleico que hibridizam sob condições de hibridização de severidade alta, tais com para hibridização seletiva, a uma sequência de nucleotídeo aqui descrita (por exemplo, moléculas de ácido nucleico que especificamente hibridizam a uma sequência de nucleotídeo contendo um sítio polimórfico associado com um marcador ou haplótipo aqui descritos). Em uma forma de realização, a

invenção inclui variantes que hibridizam sob condições de hibridização e lavagem de severidade alta (por exemplo, para a hibridização seletiva) a uma sequência de nucleotídeo que compreende a sequência de nucleotídeo do bloco LD C04 (SEQ ID NO: 50). Tais moléculas de ácido nucleico podem ser detectadas e/ou isoladas pela hibridização específica de alelo ou sequência (por exemplo, sob condições de severidade alta). As condições de severidade e os métodos para as hibridizações de ácido nucleico são bem conhecidos pela pessoa habilitada (ver, por exemplo, *Current Protocols in Molecular Biology*, Ausubel, F. *et al*, John Wiley & Sons, (1998) e Kraus, M. e Aaronson, S., *Methods Enzymol.*, 200: 546-556 (1991), as divulgações inteiras da qual são aqui incorporadas por referência.

A identidade percentual de dois nucleotídeo ou sequências de aminoácido pode ser determinada alinhando-se as sequências com propósitos de comparação ótima (por exemplo, intervalos podem ser introduzidos na sequência de uma primeira sequência). Os nucleotídeos ou aminoácidos nas posições correspondentes são depois comparados e a identidade percentual entre as duas sequências é uma função do número de posições idênticas compartilhadas pelas sequências (isto é, % de identidade = # de posições idênticas / # total de posições x 100). Em certas formas de realização, o comprimento de uma sequência alinhada com propósitos de comparação é de pelo menos 30%, pelo menos 40%, pelo menos 50%, pelo menos 60%, pelo menos 70%, pelo menos 80%, pelo menos 90% ou pelo menos 95%, do comprimento da sequência de referência. A comparação real das duas sequências pode ser realizada por métodos bem conhecidos, por exemplo, usando um algoritmo matemático. Um exemplo não limitante de um tal algoritmo matemático é descrito em Karlin, S. e Altschul, S., *Proc. Natl. Acad. Sci. USA*, 90: 5873-5877 (1993). Um tal algoritmo é incorporado nos programas NBLAST e XBLAST (versão 2.0), como descrito em Altschul, S. *et al.*, *Nucleic Acids Res.*, 25: 3389-3402 (1997). Quando da utilização de

programas BLAST e Gapped BLAST, os parâmetros de default dos respectivos programas (por exemplo, NBLAST) podem ser usados. Ver o sítio da web na world wide web em ncbi.nlm.nih.gov. Em uma forma de realização, os parâmetros para comparação de sequência podem ser ajustados na contagem = 100, comprimento de palavra = 12 ou podem ser variados (por exemplo, $W = 5$ ou $W = 20$).

Outros exemplos incluem o algoritmo de Myers e Miller, CABIOS (1989), ADVANCE e ADAM como descrito em Torellis, A. e Robotti, C., *Comput. Appl. Biosci.* 10: 3-5 (1994); e FASTA descrito em Pearson, W. e Lipman, D., *Proc. Natl. Acad. Sci. USA*, 85: 2444-48 (1988). Em uma outra forma de realização, a identidade percentual entre duas sequências de aminoácido pode ser realizada usando o programa GAP no pacote de software GCG (Accelrys, Cambridge, UK).

A presente invenção também fornece moléculas de ácido nucleico isoladas que contém um fragmento ou porção que hibridiza sob condições altamente severas a um ácido nucleico que compreende ou consiste da sequência de nucleotídeo de bloco LD C04 (SEQ ID NO: 50) ou uma sequência de nucleotídeo que compreende ou que consiste do complemento da sequência de nucleotídeo de bloco LD C04 (SEQ ID NO: 50), em que a sequência de nucleotídeo compreende pelo menos um alelo polimórfico contido nos marcadores e haplótipos aqui descritos. Os fragmentos de ácido nucleico da invenção têm pelo menos cerca de 15, pelo menos cerca de 18, 20, 23 ou 25 nucleotídeos e podem ter 30, 40, 50, 100, 200, 500, 1000, 10,000 ou mais nucleotídeos no comprimento.

Os fragmentos de ácido nucleico da invenção são usados como sondas ou iniciadores em ensaios tais com aqueles aqui descritos. “Sondas” ou “iniciadores” são oligonucleotídeos que hibridizam em uma maneira específica de base a um filamento complementar de uma molécula de ácido nucleico. Além do DNA e RNA, tais sondas e iniciadores incluem ácidos

nucleicos polipeptídicos (PNA), como descrito em Nielsen, P. *et al.*, Science 254: 1497-1500 (1991). Uma sonda ou iniciador compreende uma região de sequência de nucleotídeo que hibridiza a pelo menos cerca de 15, tipicamente cerca de 20 a 25 e em certas formas de realização cerca de 40, 50 ou 75, nucleotídeos consecutivos de uma molécula de ácido nucleico. Em uma forma de realização, a sonda ou iniciador compreende pelo menos um alelo de pelo menos um marcador polimórfico ou pelo menos um haplótipo aqui descrito ou o complemento destes. Em formas de realização particulares, uma sonda ou iniciador pode compreender 100 ou menos nucleotídeos; por exemplo, em certas formas de realização de 6 a 50 nucleotídeos, ou, por exemplo, de 12 a 30 nucleotídeos. Em outras formas de realização, a sonda ou iniciador é pelo menos 70% idêntico, pelo menos 80% idêntico, pelo menos 85% idêntico, pelo menos 90% idêntico ou pelo menos 95% idêntico, à sequência de nucleotídeo contígua ou ao complemento da sequência de nucleotídeo contígua. Em uma outra forma de realização, a sonda ou iniciador são capazes de hibridizar seletivamente à sequência de nucleotídeo contígua ou ao complemento da sequência de nucleotídeo contígua. Frequentemente, a sonda ou iniciador compreende ainda um rótulo, por exemplo, um radioisótopo, um rótulo fluorescente, um rótulo de enzima, um rótulo de co-fator de enzima, um rótulo magnético, um rótulo de spin, um rótulo de epítopo.

As moléculas de ácido nucleico da invenção, tais com aquelas descritas acima, podem ser identificadas e isoladas usando técnicas da biologia molecular padrão bem conhecidas pela pessoa habilitada. O DNA amplificado pode ser rotulado (por exemplo, radiorrotulado) e usado como uma sonda para triar uma biblioteca de cDNA derivada de células humanas. O cDNA pode ser derivado de mRNA e contido em um vetor adequado. Os clones correspondentes podem ser isolados, o DNA pode obtido a seguir da excisão *in vivo* e o inserto clonado pode ser sequenciado em cada uma ou ambas as orientações pelos métodos reconhecidos na técnica para identificar a

matriz de leitura correta que codifica um polipeptídeo do peso molecular apropriado. Usando estes ou métodos similares, o polipeptídeo e o DNA que codifica o polipeptídeo podem ser isolados, sequenciados e ainda caracterizados.

5 Em uma maneira geral, as sequências de ácido nucleico isoladas da invenção podem ser usadas como marcadores de peso molecular em géis de Southern e como marcadores de cromossoma que são rotulados para mapear posições de gene relacionadas. As sequências de ácido nucleico também podem ser usadas para comparar com sequências de DNA endógenas
10 em pacientes para identificar arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral ou uma suscetibilidade para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral e como sondas, tal com para hibridizar e descobrir sequências de DNA relacionadas ou para subtrair sequências conhecidas de
15 uma amostra (por exemplo, hibridização subtrativa). As sequências de ácido nucleico podem ser usadas ainda para derivar iniciadores para a impressão digital genética, para incitar anticorpos anti-polipeptídeo usando técnicas de imunização, e/ou como um antígeno para incitar anticorpos anti-DNA ou evocar respostas imunes.

20 *Anticorpos*

Os anticorpos policlonais e/ou anticorpos monoclonais que especificamente se ligam a uma forma do produto de gene mas não à outra forma do produto de gene também são fornecidos. Os anticorpos também são fornecidos que se ligam a uma porção da variante ou do produto de gene de
25 referência que contém o sítio ou sítios polimórficos. O termo “anticorpo” como aqui usado refere-se às moléculas de imunoglobulina e porções imunologicamente ativas de moléculas de imunoglobulina, isto é, moléculas que contêm sítios de ligação de antígeno que especificamente se ligam a um antígeno. Uma molécula que especificamente se liga a um polipeptídeo da

invenção é uma molécula que se liga àquele polipeptídeo ou um fragmento deste, mas não se ligam substancialmente a outras moléculas em uma amostra, por exemplo, uma amostra biológica, que naturalmente contém o polipeptídeo. Os exemplos de porções imunologicamente ativos de moléculas de imunoglobulina incluem fragmentos de F(ab) e F(ab')₂ que podem ser gerados tratando-se o anticorpo com uma enzima tal com pepsina. A invenção fornece anticorpos policlonais e monoclonais que se ligam a um polipeptídeo da invenção. O termo “anticorpo monoclonal” ou “composição de anticorpo monoclonal”, como aqui usado, refere-se a uma população de moléculas de anticorpo que contêm apenas uma espécie de um sítio de ligação de antígeno capaz de imunorreagir com um epítipo particular de um polipeptídeo da invenção. Uma composição de anticorpo monoclonal assim tipicamente demonstra uma única afinidade de ligação para um polipeptídeo particular da invenção com que o mesmo imunorreage.

Os anticorpos policlonais podem ser preparados como descrito acima pela imunização de um paciente adequado com um imunógeno desejado, por exemplo, polipeptídeo da invenção ou um fragmento deste. O título de anticorpo no paciente imunizado pode ser monitorado com o tempo pelas técnicas padrão, tal como com um ensaio imunossorvente ligado à enzima (ELISA) usando polipeptídeo imobilizado. Se desejado, as moléculas de anticorpo direcionadas contra o polipeptídeo podem ser isoladas do mamífero (por exemplo, a partir do sangue) e purificado ainda pelas técnicas bem conhecidas, tais como cromatografia em proteína A para obter a fração IgG. Em um tempo apropriado depois da imunização, por exemplo, quando os títulos de anticorpo são mais altos, as células que produzem anticorpo podem ser obtidas do paciente e usadas para preparar anticorpos monoclonais pelas técnicas padrão, tais com a técnica de hibridoma originalmente descrita por Kohler e Milstein, *Nature* 256: 495-497 (1975), a técnica de hibridoma de célula B humana (Kozbor *et al.*, *Immunol. Today* 4: 72 (1983)), a técnica de

EBV-hibridoma (Cole *et al.*, Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, 1985, Inc., pp. 77-96) ou técnicas de trioma. A tecnologia para produzir hibridomas é bem conhecida (ver em uma maneira geral Current Protocols in Immunology (1994) Coligan *et al.*, (eds.) John Wiley & Sons, Inc., Nova Iorque, NY). Em resumo, uma linhagem de célula imortal (tipicamente um mieloma) é fundido aos linfócitos (tipicamente esplenócitos) de um mamífero imunizado com um imunógeno como descrito acima e os sobrenadantes da cultura da qual as células de hibridoma resultantes são triados para identificar um hibridoma que produz um anticorpo monoclonal que liga um polipeptídeo da invenção.

Qualquer um dos muitos protocolos bem conhecidos usados para fundir linfócitos e linhagens de célula imortalizadas podem ser aplicados para o propósito de gerar um anticorpo monoclonal para um polipeptídeo da invenção (ver, por exemplo, Current Protocols in Immunology, supra; Gaifre *et al.*, Nature 266: 55052 (1977); R. H. Kenneth, em Monoclonal Antibodies: A New Dimension In Biological Analyses, Plenum Publishing Corp., Nova Iorque, Nova Iorque (1980); e Lerner, Yale J. Biol. Med. 54: 387-402 (1981)). Além disso, o técnico ordinariamente habilitado avaliará que existem muitas variações de tais métodos que também podem ser úteis.

Alternativa para preparar hibridomas que secretam anticorpo monoclonal, um anticorpo monoclonal para um polipeptídeo da invenção pode ser identificado e isolado pela triagem de uma biblioteca de imunoglobulina combinatória recombinante (por exemplo, uma biblioteca de demonstração de fago de anticorpo) com o polipeptídeo para isolar deste modo os membros da biblioteca de imunoglobulina que se ligam ao polipeptídeo. Kits para gerar e triar bibliotecas de demonstração de fago são comercialmente disponíveis (por exemplo, o Pharmacia Recombinant Phage Antibody System, Catálogo N^o 27-940001; e o Kit de Demonstração de Fago SurfZAP[®] da Stratagene, Catálogo N^o 240612). Adicionalmente, os exemplos

de métodos e reagentes particularmente acessíveis para o uso na geração e triagem de biblioteca de demonstração de anticorpo podem ser encontrados, por exemplo, na Patente U.S. Nº 5.223.409; Publicação PCT Nº WO 92/18619; Publicação PCT Nº WO 91/17271; Publicação PCT Nº WO 92/20791; Publicação PCT Nº WO 92/15679; Publicação PCT Nº WO 93/01288; Publicação PCT Nº WO 92/01047; Publicação PCT Nº WO 92/09690; Publicação PCT Nº WO 90/02809; Fuchs *et al.*, Bio/Technology 9: 1370-1372 (1991); Hay *et al.*, Hum. Antibod. Hybridomas 3: 81-85 (1992); Huse *et al.*, Science 246: 1275-1281 (1989); e Griffiths *et al.*, EMBO J. 12: 725-734 (1993).

Adicionalmente, anticorpos recombinantes, tais com anticorpos monoclonais quiméricos ou humanizados, que compreendem tanto porções humanas quanto não humanas, que podem ser fabricados usando as técnicas de DNA recombinante padrão, estão dentro do escopo da invenção. Tais anticorpos monoclonais quiméricos e humanizados podem ser produzidos pelas técnicas de DNA recombinantes conhecidas no ramo.

Em uma maneira geral, os anticorpos da invenção (por exemplo, um anticorpo monoclonal) podem ser usados para isolar um polipeptídeo da invenção pelas técnicas padrão, tais com cromatografia de afinidade ou imunoprecipitação. Um anticorpo específico de polipeptídeo pode facilitar a purificação de polipeptídeo natural das células e de polipeptídeo recombinantemente produzido expressado nas células hospedeiras. Além disso, um anticorpo específico para um polipeptídeo da invenção pode ser usado para detectar o polipeptídeo (por exemplo, em um lisado celular, sobrenadante de célula ou amostra de tecido) de modo a avaliar a abundância e padrão de expressão do polipeptídeo. Anticorpos podem ser usados em diagnóstico para monitorar os níveis de proteína em tecido como parte de um procedimento de teste clínico, por exemplo, para determinar, por exemplo, a eficácia de um dado regime de tratamento. O anticorpo pode ser

ligado a uma substância detectável para facilitar a sua detecção. Os exemplos de substâncias detectáveis incluem várias enzimas, grupos protéticos, materiais fluorescentes, materiais luminescentes, materiais bioluminescentes e materiais radioativos. Os exemplos de enzimas adequadas incluem peroxidase de rábano, fosfatase alcalina, betagalactosidase ou acetilcolinesterase; os exemplos de complexos de grupo protético adequados incluem estreptavidina/biotina e avidina/biotina; os exemplos de materiais fluorescentes adequados incluem umbeliferona, fluoresceína, isotiocianato de fluoresceína, rodamina, diclorotriazinil-amina fluoresceína, cloreto de dansila ou ficoeritrina; um exemplo de um material luminescente inclui luminol; os exemplos de materiais bioluminescentes incluem luciferase, luciferina e aequorina e os exemplos de material radioativo adequados incluem ^{125}I , ^{131}I , ^{35}S ou ^3H .

Anticorpos também podem ser úteis na análise farmacogenômica. Em tais formas de realização, os anticorpos contra proteínas variantes codificadas pelos ácidos nucleicos de acordo com a invenção, tais como proteínas variantes que são codificadas pelos ácidos nucleicos que contém pelo menos um marcador polimórfico da invenção, podem ser usados para identificar indivíduos que requerem modalidades de tratamento modificadas.

Anticorpos além disso podem ser úteis para avaliar a expressão de proteínas variantes em estados de doença, tais como em estágios ativos de uma doença ou em um indivíduo com uma predisposição para uma doença relacionada com a função da proteína, em particular arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral. Anticorpos específicos para uma proteína variante da presente invenção que é codificada por um ácido nucleico que compreende pelo menos um marcador polimórfico ou haplótipo como aqui descritos podem ser usados para triar quanto a presença da proteína variante, por exemplo para triar quanto uma pré

disposição para a arritmia cardíaca (por exemplo, fibrilação atrial ou flúter atrial) e/ou acidente vascular cerebral como indicado pela presença da proteína variante.

Anticorpos podem ser usados em outros métodos. Assim, os anticorpos são úteis como ferramentas de diagnóstico para a avaliação de proteínas, tais com proteínas variantes da invenção, em conjunção com a análise pela mobilidade eletroforética, ponto isoelétrico, digestão triptica ou outra protease ou para o uso em outros ensaios físicos conhecidos por aqueles habilitados na técnica. Anticorpos também podem ser usados na tipagem de tecido. Em uma tal forma de realização, uma proteína variante específica foi correlacionada com a expressão em um tipo de tecido específico e anticorpos específicos para a proteína variante podem ser depois usados para identificar o tipo de tecido específico.

A localização subcelular de proteínas, incluindo proteínas variantes, também pode ser determinada usando anticorpos e podem ser aplicados para avaliar a localização subcelular aberrante da proteína em células em vários tecidos. Tal uso pode ser aplicado em teste genético, mas também na monitoração de uma modalidade de tratamento particular. No caso onde o tratamento é almejado na correção do nível de expressão ou a presença da proteína variante ou distribuição em tecido aberrante ou expressão desenvolvimental da proteína variante, anticorpos específicos para a proteína variante ou fragmentos destes podem ser usados para monitorar a eficácia terapêutica.

Anticorpos são ainda úteis para inibir a função da proteína variante, por exemplo pelo bloqueio da ligação de uma proteína variante a uma molécula ou parceiro de ligação. Tais usos também podem ser aplicados em um contexto terapêutico em que o tratamento envolve inibir uma função da proteína variante. Um anticorpo pode ser por exemplo usado para bloquear ou competitivamente inibir a ligação, modulando deste modo (isto é,

agonizando ou antagonizando) a atividade da proteína. Anticorpos podem ser preparados contra fragmentos de proteína específicos contendo sítios requeridos para a função específica ou contra uma proteína intacta que está associada com uma célula ou membrana de célula. Para a administração *in vivo*, um anticorpo pode ser ligado com uma carga adicional de produto terapêutico, tal com radionuclídeo, uma enzima, um epítipo imunogênico ou um agente citotóxico, incluindo toxinas bacterianas (toxinas da difteria ou vegetais, tais com ricina). A meia vida *in vivo* de um anticorpo ou um fragmento deste pode ser aumentada pela peguilação através da conjugação ao polietileno glicol.

A presente invenção diz respeito ainda a kits para usar anticorpos nos métodos aqui descritos. Isto inclui, mas não é limitado a, kits para detectar a presença de uma proteína variante em uma amostra de teste. Uma forma de realização preferida compreende anticorpos tais com um anticorpo rotulado ou rotulável e um composto ou agente para detectar proteínas variantes em uma amostra biológica, meios para determinar a quantidade ou a presença e/ou ausência de proteína variante na amostra e meios para comparar a quantidade de proteína variante na amostra com um padrão, assim como instruções para o uso do kit.

A presente invenção será agora exemplificada pelo seguinte exemplo não limitante.

EXEMPLIFICAÇÃO

Exemplo 1. Identificação de variantes de risco para a fibrilação atrial no cromossoma 4.

O que segue contém a descrição da identificação de fatores de suscetibilidade descobertos estar associados com a fibrilação atrial e acidente vascular cerebral através da análise de ponto único de marcadores de SNP.

Métodos. O estudo foi aprovado pelo Data Protection Commission of Iceland e pelo National Bioethics Committee.

Grupo AF islandês. Os pacientes foram todos diagnosticados com AF no Landspítali University Hospital em Reykjavik, Islândia, de 1987 a 2003. Os diagnósticos foram confirmados por um eletrocardiograma de 12 fios condutores elétricos não demonstrando nenhuma onda P e intervalos R-R irregularmente irregulares. Todos os ECGs foram manualmente lidos por uma cardiologista.

Grupo de Acidente Vascular Cerebral Islandês. O grupo do acidente vascular cerebral foi derivado de dois hospitais principais na Islândia e da Icelandic Heart Association. Pacientes com acidente vascular cerebral hemorrágico representaram 6% de todos os pacientes (pacientes com o tipo islandês de hemorragia cerebral hereditária com amiloidose e pacientes com hemorragia subaracnóide foram excluídos). O acidente vascular cerebral isquêmico é responsável por 67% dos pacientes totais e TIAs 27%. A distribuição de subtipos de acidente vascular cerebral neste estudo é similar àquele relatado em outras populações Caucasianas (Mohr, J. P., *et al.*, Neurology, 28: 754-762 (1978); L. R. Caplan, Em Stroke, A Clinical Approach (Butterworth-Heinemann, Stoneham, MA, ed 3, (1993)).

Genotipagem. Uma varredura ampla de genoma de 437 indivíduos islandeses diagnosticados com Fibrilação Atrial (AF) e 7406 controles da população foi realizada usando chips de SNP Infinium HumanHap300 da Illumina para ensaiar aproximadamente 317.000 polimorfismos de nucleotídeo único (SNPs) em um único chip (Illumina, San Diego, CA, USA). A genotipagem de SNP para replicação em outros grupos de controle de caso foi realizada usando a plataforma Centaurus (Nanogen). Um total de 347 indivíduos diagnosticados com Acidente Vascular Cerebral e 7497 controles também foi realizado para SNPs dentro do bloco LD descobertos estar associados com a Fibrilação Atrial.

Métodos Estatísticos para a Análise de Associação. Para a associação de marcador único para a fibrilação atrial ou acidente vascular

cerebral, nós usamos um teste de razão de probabilidade para calcular um valor p de lado duplo para cada alelo. Nós calculamos o Risco Relativo (RR) e o risco atribuível à população (PAR) assumindo um modelo multiplicativo (C. T. Falk, P. Rubinstein, *Ann Hum Genet* 51 (Pt 3), 227 (1987); J. D. Terwilliger, J. Ott, *Hum Hered* 42, 337 (1992)). Para os dados de CEPH Caucasian HapMap, nós calculamos LD entre pares de SNPs usando a definição padrão de D' (R. C. Lewontin, *Genetics* 50, 757 (1964)) e R^2 W. G. Hill, A. Robertson, *Genetics* 60, 615 (Nov, 1968). Quando da plotagem de todas as combinações de SNP para elucidar a estrutura LD em uma região particular, nós plotamos D' no canto esquerdo superior e os valores p no canto direito inferior. Nas plotagens de LD que nós apresentamos, os marcadores são plotados equidistantemente ao invés de acordo com as suas posições físicas.

RESULTADOS

Estudo de associação ampla de genoma

Nós genotipamos com êxito 437 pacientes islandeses com fibrilação atrial e 7406 indivíduos de controle da população usando o chip Illumina 330K. A análise de associação foi realizada para SNPs únicos. A associação mais significativa foi descoberta para os marcadores rs2220427 e rs2220733, ambos os quais dão valores p próximos a 10^{-9} . O valor para rs2220427 é significativo depois de corrigir quanto ao número de testes realizados, isto é, a associação é significativa ao nível amplo de genoma.

Existe um excesso evidente de homozigotos em indivíduos afetados. Nós rejeitamos tanto o modelo multiplicativo ($P = 0,002$) quanto o modelo recessivo ($P = 0,001$). O modelo de melhor ajuste dá risco de 1,46 para portadores heterozigotos e 5,17 para portadores homozigotos. O valor P (não corrigido) comparando este modelo completo com o modelo nulo de nenhuma associação é de $5,43^{-11}$. Estes dados mostram que indivíduos com duas cópias do alelo em risco estão em risco maior do que o esperado com

base em um modelo multiplicativo simples.

O ajuste de uma idade no modelo inicial para todos os genótipos dá um valor P de $4,84^{-5}$. Heterozigotos são estimados ter início 1,4090 anos mais cedo do que os não portadores e portadores homozigotos são estimados ter início 9,6126 anos mais cedo do que os não portadores. Isto mostra um efeito significativo de idade no início - indivíduos que carregam a variante de risco estão em risco significativo de desenvolver AF em uma idade mais jovem do que os indivíduos que não são portadores do alelo em risco.

Investigando marcadores na vizinidade de rs2220427, nós constatamos que o marcador de microssatélite D4S406 pode ser usado como um marcador substituto para rs2220427. Em particular, os alelos 2, 4 e 8 (com respeito à referência CEPH) foram descobertos ser suficientes para rotular o SNP com base nas frequências de haplótipo:

Tabel

a 1. Relações entre rs2220427 e D4S406

Frequência	Haplótipo			
	Alelo MS		Alelo SNP	
7,55E-05	-2	D4S406	2	rs2220427
0,000109727	16	D4S406	4	rs2220427
0,000148065	-6	D4S406	4	rs2220427
0,000149685	20	D4S406	2	rs2220427
0,000149756	-4	D4S406	2	rs2220427
0,000210154	8	D4S406	4	rs2220427
0,000225802	-8	D4S406	2	rs2220427
0,000227036	4	D4S406	4	rs2220427
0,000299371	18	D4S406	2	rs2220427
0,000899281	0	D4S406	4	rs2220427
0,00203518	-4	D4S406	4	rs2220427
0,00673851	-6	D4S406	2	rs2220427
0,0245484	2	D4S406	2	rs2220427
0,0394983	-2	D4S406	4	rs2220427
0,0422112	14	D4S406	2	rs2220427
0,0594303	0	D4S406	2	rs2220427
0,0762831	-8	D4S406	4	rs2220427
0,0855451	6	D4S406	2	rs2220427
0,0949753	12	D4S406	2	rs2220427
0,100105	16	D4S406	2	rs2220427
0,145942	4	D4S406	2	rs2220427
0,155838	8	D4S406	2	rs2220427
0,164354	10	D4S406	2	rs2220427

Assim, para indivíduos tipados quanto ao marcador D4S406 mas não rs222047, que fundem os alelos 2, 4 e 8 levam a uma estimativa

5 muito boa da frequência do alelo 4 do SNP. Nós analisamos um grupo de replicação islandês para AF, compreendido de 1269 casos e 69.070 controles, nesta maneira. Os resultados são bastante dramáticos em que a associação é acompanhada por um valor p de $2,94^{-14}$ e um Risco Relativo (modelo multiplicativo) de 1,53. Assim, nossa descoberta inicial foi replicada em um grupo islandês independente.

Tabela 2. Associação de pacientes AF ao Cromossoma 4 (bloco LD C04).

São mostrados os valores para RR sob o modelo multiplicativo.

Valor P	Risco relativo	Freq. Aff	Freq. Con	Alelo	Marcador
0,16839039	0,8473	0,902746	0,916352	3	rs10033464
0,16839039	1,1802	0,097254	0,083648	4	rs10033464
0,24275346	0,8746	0,098398	0,110939	1	rs13105878
0,24275346	1,1433	0,901602	0,889061	2	rs13105878
4,89E-06	1,3816	0,441648	0,364078	1	rs13141190
4,89E-06	0,7238	0,558352	0,635922	3	rs13141190
1,25E-06	0,6905	0,677346	0,752498	1	rs1448817
1,25E-06	1,4483	0,322654	0,247502	3	rs1448817
0,00995996	0,7903	0,811213	0,844653	2	rs16997168
0,00995996	1,2654	0,188787	0,155347	4	rs16997168
1,07E-09	0,5601	0,811213	0,88688	2	rs2200733
1,07E-09	1,7855	0,188787	0,115312	4	rs2200733
7,78E-10	0,557	0,810345	0,884673	2	rs2220427
7,78E-10	1,7953	0,189655	0,115327	4	rs2220427
9,75E-08	1,5803	0,236239	0,163692	1	rs2634073
9,75E-08	0,6328	0,763761	0,836308	3	rs2634073
0,96927011	0,9968	0,768879	0,769444	1	rs2723296
0,96927011	1,0032	0,231121	0,230556	3	rs2723296
0,03281713	0,8529	0,665904	0,700311	2	rs2723316
0,03281713	1,1724	0,334096	0,299689	4	rs2723316
0,01803327	1,1855	0,635632	0,595393	1	rs3853444
0,01803327	0,8435	0,364368	0,404607	3	rs3853444
0,40105752	0,9214	0,146789	0,15734	2	rs4576077
0,40105752	1,0853	0,853211	0,84266	4	rs4576077
0,93269621	1,0084	0,145309	0,144275	1	rs6419178
0,93269621	0,9917	0,854691	0,855725	3	rs6419178

Tabela 3. Associação de Acidente Vascular Cerebral com marcadores dentro do bloco LD C04 (SEQ ID NO: 50)

Valor P	Risco relativo	Freq. Aff	Freq. Con	Alelo	Marcador
0,37701852	0,8872	0,90634	0,916011	3	rs10033464
0,37701852	1,1272	0,09366	0,083978	4	rs10033464
0,2838194	0,8717	0,097983	0,110807	1	rs13105878
0,2838194	1,1472	0,902017	0,889193	2	rs13105878
0,01534596	1,2123	0,412104	0,366378	1	rs13141190
0,01534596	0,8249	0,587896	0,633622	3	rs13141190
0,048556224	0,8418	0,716138	0,7498	1	rs1448817
0,048556224	1,1879	0,283862	0,2502	3	rs1448817
0,14450185	0,8599	0,822767	0,843717	2	rs16997168
0,14450185	1,1629	0,177233	0,156283	4	rs16997168
0,00374992	0,724	0,84438	0,882271	2	rs2200733
0,00374992	1,3812	0,15562	0,117729	4	rs2200733
0,0025713	0,7141	0,842566	0,882274	2	rs2220427
0,0025713	1,4003	0,157434	0,117726	4	rs2220427
0,01664881	1,2682	0,201729	0,166154	1	rs2634073
0,01664881	0,7885	0,798271	0,833846	3	rs2634073
0,58350156	0,9511	0,760807	0,769811	1	rs2723296
0,58350156	1,0514	0,239193	0,230189	3	rs2723296
0,03150475	0,8367	0,661383	0,700107	2	rs2723316
0,03150475	1,1952	0,338617	0,299893	4	rs2723316
0,16517516	1,1172	0,622832	0,596462	1	rs3853444
0,16517516	0,8951	0,377168	0,403538	3	rs3853444
0,24301926	0,8797	0,14121	0,157473	2	rs4576077
0,24301926	1,1367	0,85879	0,842527	4	rs4576077
0,19773377	1,1482	0,161383	0,143543	1	rs6419178
0,19773377	0,8709	0,83867	0,856457	3	rs6419178

Tabela 4. Marcadores em perfeito desequilíbrio de ligação ($r^2 = 1,0$) com rs2220427 na população CEU no conjunto de dados International HapMap (Indivíduos de descendência europeia). Também são mostradas a correlação com amostras de Yuroba (Nigéria) e Ásia (China e Japão) - a descrição do grupo está documentado ainda no <http://www.hapmap.org>.

SNP	Alelo	CEU_R2	CEU_frq	YRI_R2	YRI_frq	ASIA_R2	ASIA_frq
rs17042059	1	1	0,117647	0,500382	0,117647	0,473183	0,30814
rs4529121	1	1	0,116667	0,604601	0,1	0,539766	0,337079
rs4543199	2	1	0,116667	0,502036	0,116667	0,539766	0,337079
rs10019689	1	1	0,116667	0,128175	0,341667	0,664071	0,388889
rs4626276	2	1	0,116667	0,603474	0,10084	0,537439	0,333333
rs17042076	2	1	0,117647	0,128175	0,341667	0,664071	0,388889
rs11098089	2	1	0,117647	0,549368	0,108333	0,539766	0,337079
rs11939528	4	1	0,11017	0,120773	0,321429	0,662926	0,377907
rs17042098	1	1	0,116667	0,669219	0,092437	0,64297	0,355556
rs17042102	1	1	0,091743	NA	NA	0,580822	0,302632
rs17042121	3	1	0,116667	0,736119	0,141667	0,639142	0,353933
rs10516563	3	1	0,109244	0,846743	0,128205	0,636329	0,364706
rs4605724	1	1	0,116667	0,748252	0,083333	0,645257	0,359551
rs2350269	4	1	0,11017	0,495806	0,098214	0,628257	0,346154
rs6533527	1	1	0,116667	0,525151	0,133333	0,804123	0,421348
rs17042144	2	1	0,119658	0,727891	0,077586	NA	NA
rs1906618	2	1	0,115044	NA	NA	NA	NA
rs1906617	2	1	0,116667	0,541206	0,183333	0,977348	0,454545
rs12646447	2	1	0,119658	1	0,108333	1	0,444444
rs12646754	4	1	0,119658	0,681842	0,11017	1	0,425
rs2129981	1	1	0,116667	1	0,108333	1	0,444444
rs12639654	4	1	0,116667	0,139505	0,016667	1	0,438202
rs6817105	2	1	0,117647	0,27862	0,299145	1	0,440476
rs17042171	1	1	0,109244	0,281063	0,302521	1	0,425287
rs1906591	1	1	0,116667	1	0,108333	1	0,444444
rs1906592	3	1	0,109244	0,283489	0,3	1	0,446429
rs2200732	2	1	0,112069	0,272544	0,308334	1	0,449438
rs2200733	4	1	0,116667	0,276161	0,301724	1	0,445783
rs4611994	2	1	0,116667	0,272544	0,308334	1	0,449438
rs4540107	1	1	0,116667	0,27862	0,305085	1	0,44382
rs1906593	4	1	0,117647	0,285134	0,301724	1	0,438202
rs1906596	2	1	0,121739	0,255864	0,330275	0,97478	0,448718

Tabela 5.

A. Marcadores SNP dentro do bloco LD C04 (Entre 111.954.811 e 112.104.250 em C04; Construção NCBI 35; SEQ ID NO: 50).

Marcador ID	Pós-construção 35	Pós na SEQ ID NO:50	Tipo	Filamento
rs1448824	111954811	1	A/G	-
rs1947189	111955221	411	A/G	-
rs1947188	111955479	669	C/G	-
rs1992927	111956353	1543	C/T	-
rs1470619	111957122	2312	A/G	-
rs1448823	111958486	3676	A/G	-
rs4834327	111958676	3866	A/T	+
rs1448822	111958702	3892	C/T	-
rs2044674	111959075	4265	A/G	-
rs28445748	111959470	4660	A/T	+
rs2595116	111959591	4781	C/T	-
rs2595115	111959725	4915	A/C	-
rs13120244	111961948	7138	A/G	+
rs2723296	111962087	7277	A/G	+
rs2723297	111962201	7391	A/T	+
rs10021211	111962246	7436	C/T	+
rs17042011	111962331	7521	C/T	+
rs2595114	111962791	7981	C/G	-
rs2595113	111962792	7982	C/G	-
rs2595112	111963368	8558	C/G	-
rs6831623	111964677	9867	C/T	+
rs6854883	111964919	10109	C/T	+
rs2255793	111965457	10647	A/G	+
rs2723298	111966089	11279	C/T	+
rs12505886	111966218	11408	A/T	+
rs28718263	111966220	11410	A/T	+
rs12501913	111966355	11545	A/C	+
rs13126974	111966385	11575	A/T	+
rs36194761	111966385	11575	A/T	+
rs28473341	111966486	11676	C/T	+
rs36160675	111967780	12970	G/T	+
rs13147139	111968764	13954	A/G	+
rs13147489	111968795	13985	C/T	+
rs13147299	111968812	14002	A/C	+
rs13147726	111968923	14113	C/T	+
rs13147730	111968926	14116	C/T	+
rs13147552	111968949	14139	A/G	+
rs13123918	111968996	14186	A/T	+
rs35610510	111970561	15751	C/T	+
rs36162200	111971480	16670	G/T	+
rs11098086	111971997	17187	C/T	+
rs4034950	111972120	17310	A/G	-
rs11724067	111972144	17334	A/G	+
rs2723299	111972436	17626	A/G	+
rs2723300	111972512	17702	A/G	+
rs13138211	111972606	17796	A/G	+
rs2595075	111973312	18502	C/T	-
rs2723301	111973731	18921	C/G	+
rs2595074	111974709	19899	A/T	-
rs2723302	111974736	19926	C/G	+
rs2723303	111974741	19931	A/G	+
rs2218698	111975356	20546	G/T	-
rs2218697	111975357	20547	C/T	-
rs2595073	111975436	20626	A/G	-
rs2723307	111975800	20990	A/T	+
rs1584430	111976043	21233	C/G	-

Marcador ID	Pós-construção 35	Pós na SEQ ID NO:50	Tipo	Filamento
rs1584429	111976151	21341	C/G	-
rs1900828	111976526	21716	C/T	-
rs7672226	111976785	21975	C/T	+
rs1839189	111976971	22161	C/T	-
rs1579946	111977724	22914	A/G	-
rs12509115	111977892	23082	A/G	+
rs1579945	111978096	23286	A/T	-
rs7661383	111979181	24371	A/C	+
rs2122078	111979201	24391	A/G	-
rs2122077	111979254	24444	C/T	-
rs2723311	111979626	24816	A/G	+
rs7667461	111979738	24928	A/G	+
rs1448799	111980386	25576	C/T	-
rs1448798	111980789	25979	C/T	-
rs12650829	111980880	26070	A/G	+
rs2723312	111980956	26146	A/G	+
rs1900827	111981343	26533	A/G	-
rs6815628	111981980	27170	C/T	+
rs6838131	111981993	27183	A/G	+
rs6838139	111982000	27190	A/C	+
rs6838295	111982012	27202	C/T	+
rs4582211	111982043	27233	A/G	+
rs4353966	111982088	27278	A/T	+
rs6838536	111982144	27334	C/T	+
rs1375302	111983068	28258	C/T	+
rs1375303	111983069	28259	A/G	+
rs7699114	111983094	28284	C/T	+
rs2197814	111983098	28288	A/C	+
rs2218700	111983340	28530	A/G	+
rs969642	111983529	28719	C/T	+
rs17042020	111984067	29257	A/C	+
rs2595099	111984371	29561	A/C	+
rs4371683	111984371	29561	A/C	+
rs2595093	111984960	30150	C/T	-
rs17625509	111984998	30188	A/G	+
rs2723313	111985093	30283	A/G	+
rs2723314	111985111	30301	G/T	+
rs2595092	111985112	30302	A/G	-
rs2595091	111985223	30413	C/G	-
rs1375301	111985458	30648	A/G	-
rs2245595	111985715	30905	C/T	+
rs2595088	111985958	31148	C/T	-
rs981150	111986232	31422	C/T	-
rs16997168	111986643	31833	C/T	+
rs6812840	111986654	31844	A/T	+
rs16997169	111986685	31875	C/T	+
rs4527540	111986742	31932	C/T	+
rs2595078	111987397	32587	A/G	+
rs11098087	111987538	32728	C/T	+
rs6843456	111988165	33355	C/T	+
rs998101	111988219	33409	A/G	-
rs13120535	111989691	34881	A/G	+
rs17042026	111989978	35168	A/G	+
rs6840960	111991045	36235	C/T	+
rs2122079	111991108	36298	C/T	+
rs2166961	111991365	36555	C/T	+
rs2723316	111991891	37081	C/T	+
rs2595079	111992019	37209	A/G	+
rs7665126	111992019	37209	A/G	+
rs2595080	111992042	37232	A/G	+

Marcador ID	Pós-construção 50	Pós na SEQ ID NO:50	Tipo	Filamento
rs12646859	111992237	37427	G/T	+
rs10222783	111992430	37620	C/T	+
rs12498380	111992563	37753	C/T	+
rs2595081	111992761	37951	C/T	+
rs2595082	111992896	38086	G/T	+
rs2723317	111993104	38294	A/G	+
rs6419178	111993104	38294	A/G	+
rs13110876	111993625	38815	A/G	+
rs2595083	111993625	38815	A/G	+
rs7690164	111994069	39259	C/T	+
rs2595084	111994163	39353	A/G	+
rs2595085	111994377	39567	C/G	+
rs2595086	111994385	39575	C/T	+
rs2723318	111994576	39766	G/T	+
rs17042050	111994805	39995	C/T	+
rs9998222	111995088	40278	A/G	+
rs2723319	111995233	40423	A/T	+
rs2595087	111995380	40570	C/T	+
rs17042052	111995521	40711	A/T	+
rs28558677	111995664	40854	G/T	+
rs6812731	111995691	40881	A/C	+
rs2723320	111997050	42240	C/T	+
rs12644107	111997588	42778	C/T	+
rs28482179	111998237	43427	C/T	+
rs28759131	111998559	43749	C/T	+
rs1448817	111998657	43847	A/G	+
rs28526075	111998725	43915	A/G	+
rs17042059	111998790	43980	A/G	+
rs10014075	112000023	45213	G/T	+
rs10026140	112000455	45645	G/T	+
rs13351232	112000455	45645	G/T	+
rs7666806	112000477	45667	G/T	+
rs10028327	112000489	45679	G/T	+
rs12650941	112002415	47605	A/T	+
rs28650220	112002617	47807	C/T	+
rs13113361	112002671	47861	G/T	+
rs13113522	112002686	47876	G/T	+
rs4529121	112003159	48349	A/G	+
rs6831284	112003582	48772	G/T	+
rs10009621	112003846	49036	C/T	+
rs10021534	112003945	49135	C/T	+
rs10032150	112004222	49412	A/G	+
rs10024267	112004571	49761	C/T	+
rs10012705	112004726	49916	C/T	+
rs11943627	112005073	50263	C/T	+
rs4543199	112005744	50934	C/T	+
rs28410055	112006340	51530	A/G	+
rs7693227	112006532	51722	C/T	+
rs6852197	112006679	51869	A/G	+
rs12647316	112006855	52045	C/T	+
rs12647393	112006886	52076	G/T	+
rs10019645	112007248	52438	G/T	+
rs10019689	112007473	52663	A/C	+
rs4626276	112007593	52783	A/C	+
rs10022067	112007672	52862	C/T	+
rs4469143	112007678	52868	C/G	+
rs6836206	112007902	53092	C/T	+
rs13150693	112008086	53276	G/T	+
rs11737632	112008416	53606	C/T	+
rs5011975	112008427	53617	A/G	+

Marcador ID	Pós-construção 35	Pós na SEQ ID NO:50	Tipo	Filamento
rs6811511	112008429	53619	A/C	+
rs4383676	112008437	53627	A/G	+
rs28392642	112009161	54351	C/T	+
rs17631468	112009386	54576	A/G	+
rs17042076	112009942	55132	C/T	+
rs4434326	112010480	55670	C/T	+
rs17042081	112010815	56005	G/T	+
rs4833436	112011350	56540	C/T	+
rs7679623	112011519	56709	A/C	+
rs11098088	112011728	56918	C/T	+
rs4530699	112011761	56951	A/T	+
rs11098089	112011830	57020	A/C	+
rs17042088	112012418	57608	C/T	+
rs12648785	112013496	58686	A/G	+
rs12639820	112013644	58834	C/T	+
rs10001807	112013708	58898	A/G	+
rs10024486	112013722	58912	G/T	+
rs12648889	112013890	59080	C/G	+
rs28376747	112013925	59115	A/G	+
rs11098090	112014012	59202	C/T	+
rs11944778	112014571	59761	A/G	+
rs7436333	112014951	60141	A/C	+
rs4307025	112015107	60297	A/T	+
rs4447925	112015252	60442	C/T	+
rs28523292	112015772	60962	C/T	+
rs28635581	112015858	61048	C/T	+
rs28508237	112016004	61194	C/T	+
rs28521134	112016167	61357	C/T	+
rs17042093	112017716	62906	C/G	+
rs11930438	112017749	62939	C/T	+
rs28542185	112017795	62985	C/T	+
rs11930528	112017798	62988	G/T	+
rs13121382	112020177	65367	G/T	+
rs7439625	112021082	66272	A/T	+
rs28501998	112021318	66508	A/T	+
rs10016838	112021718	66908	C/T	+
rs17042098	112021762	66952	A/G	+
rs10005076	112021953	67143	C/T	+
rs10027473	112022056	67246	A/G	+
rs2634073	112023387	68577	A/G	-
rs1906611	112023520	68710	A/G	-
rs1906610	112023521	68711	C/T	-
rs28446238	112024025	69215	A/C	+
rs1906609	112024055	69245	A/C	-
rs34916665	112025294	70484	G/T	+
rs17042102	112026230	71420	A/G	+
rs17042104	112026555	71745	C/T	+
rs10015819	112026628	71818	C/T	+
rs2634071	112026824	72014	A/G	-
rs10007386	112028021	73211	C/T	+
rs10007547	112028050	73240	A/G	+
rs12647522	112028465	73655	C/T	+
rs1906614	112028900	74090	A/G	-
rs2723335	112029230	74420	A/G	+
rs17042112	112029281	74471	C/T	+
rs17042115	112029414	74604	A/G	+
rs10013510	112029527	74717	C/G	+
rs11939057	112029755	74945	C/T	+
rs2634076	112029877	75067	A/G	-
rs2723293	112031377	76567	A/G	+

Marcador ID	Pós-construção 35	Pós na SEQ ID NO:50	Tipo	Filamento
rs28494131	112032777	77967	C/T	+
rs2634075	112033582	78772	A/G	-
rs13121715	112034239	79429	G/T	+
rs2634074	112034645	79835	A/T	-
rs17042121	112034705	79895	A/G	+
rs10516563	112035326	80516	G/T	+
rs17042125	112035883	81073	A/G	+
rs13136439	112036254	81444	G/T	+
rs13114686	112036503	81693	C/T	+
rs36166388	112037782	82972	A/G	+
rs2450934	112038532	83722	A/C	-
rs36138049	112040474	85664	A/C	+
rs36168695	112040495	85685	A/G	+
rs36129850	112040548	85738	A/G	+
rs12513264	112041966	87156	C/T	+
rs2882365	112041975	87165	A/G	+
rs36139649	112042072	87262	C/T	+
rs4033107	112042072	87262	C/T	+
rs36176419	112042092	87282	A/G	+
rs4033108	112042092	87282	A/G	+
rs4450997	112042160	87350	A/C	+
rs2350268	112042213	87403	A/G	+
rs4613627	112042225	87415	A/C	+
rs4033109	112042227	87417	C/T	+
rs4833443	112042247	87437	C/T	+
rs2723336	112042258	87448	C/T	-
rs4033111	112042302	87492	A/G	+
rs1807360	112042333	87523	C/T	-
rs4605724	112042685	87875	A/C	+
rs6856879	112043066	88256	A/T	+
rs6834418	112043172	88362	C/T	+
rs2466455	112043219	88409	A/G	-
rs6857810	112043220	88410	A/G	+
rs2634079	112043541	88731	C/G	-
rs28366840	112043710	88900	A/T	+
rs2350269	112044728	89918	C/T	+
rs7665409	112045070	90260	C/T	+
rs6533527	112045118	90308	A/C	+
rs12649717	112045283	90473	A/C	+
rs6822831	112045374	90564	A/G	+
rs35916701	112046074	91264	C/T	+
rs6829419	112046178	91368	C/T	+
rs2723334	112046356	91546	A/G	-
rs2634078	112046528	91718	C/T	-
rs12512819	112046597	91787	C/T	+
rs17042144	112047270	92460	C/T	+
rs2171594	112047908	93098	A/G	-
rs6842887	112048170	93360	A/G	+
rs2171593	112048375	93565	G/T	-
rs7690874	112048681	93871	A/G	+
rs17042145	112049101	94291	G/T	+
rs17042146	112049106	94296	C/T	+
rs9998815	112049904	95094	C/G	+
rs7683336	112051207	96397	C/T	+
rs17042150	112051452	96642	A/T	+
rs10016842	112051810	97000	C/T	+
rs10005432	112052219	97409	A/G	+
rs1906620	112052624	97814	C/T	-
rs1906619	112052670	97860	C/T	-
rs1906618	112053026	98216	C/T	-

Marcador ID	Pós-construção 35	Pós na SEQ ID NO:50	Tipo	Filamento
rs1906617	112053418	98608	C/T	-
rs6847935	112054255	99445	A/T	+
rs6831873	112055138	100328	C/T	+
rs1906616	112055172	100362	C/T	-
rs6837901	112055712	100902	C/T	+
rs2723333	112056695	101885	C/T	-
rs12646447	112056930	102120	C/T	+
rs6820568	112057435	102625	C/T	+
rs1906615	112059402	104592	A/C	-
rs2634077	112061112	106302	A/G	-
rs7689774	112061114	106304	G/T	+
rs12646754	112061176	106366	C/T	+
rs35807830	112061497	106687	G/T	+
rs2129983	112061684	106874	C/T	-
rs2129982	112061747	106937	C/T	-
rs2129981	112061803	106993	A/C	-
rs6854111	112062140	107330	A/T	+
rs12639654	112062899	108089	C/T	+
rs4515229	112062985	108175	A/G	+
rs2129984	112063010	108200	C/T	+
rs6817105	112063372	108562	C/T	+
rs12503217	112063765	108955	C/T	+
rs2634070	112064016	109206	A/C	+
rs17042171	112065891	111081	A/C	+
rs7434417	112066042	111232	A/G	+
rs1906591	112066493	111683	A/G	+
rs1906592	112066608	111798	G/T	+
rs12510087	112066632	111822	A/G	+
rs7661554	112067221	112411	A/G	+
rs34796144	112067333	112523	A/C	+
rs2200732	112067646	112836	C/T	+
rs2200733	112067773	112963	C/T	+
rs17042175	112068571	113761	A/T	+
rs4611994	112068645	113835	C/T	+
rs4540107	112068706	113896	A/C	+
rs1906593	112069526	114716	C/T	+
rs4371684	112069651	114841	A/G	+
rs1906594	112069739	114929	A/G	+
rs1906595	112069788	114978	G/T	+
rs1906596	112069840	115030	C/T	+
rs6838775	112069908	115098	G/T	+
rs2129977	112070036	115226	A/G	+
rs2129978	112070158	115348	A/C	+
rs1906597	112070190	115380	G/T	+
rs1906598	112070229	115419	C/T	+
rs1906599	112070290	115480	C/T	+
rs1906600	112070480	115670	C/T	+
rs1906601	112070883	116073	C/T	+
rs1906602	112070927	116117	C/T	+
rs1906603	112071040	116230	C/T	+
rs28645285	112071426	116616	A/G	+
rs2171590	112071435	116625	C/T	+
rs6852357	112071939	117129	C/T	+
rs13143308	112072023	117213	G/T	+
rs2220427	112072493	117683	C/T	+
rs17632693	112072538	117728	C/T	+
rs11935917	112072850	118040	A/G	+
rs4833456	112073911	119101	C/T	+
rs12644625	112074117	119307	C/T	+
rs4400058	112074277	119467	A/G	+

Marcador ID	Pós-construção 35	Pós na SEQ ID NO:50	Tipo	Filamento
rs1906604	112074452	119642	A/G	+
rs1906605	112074796	119986	C/T	+
rs13126975	112075129	120319	A/T	+
rs6837490	112075447	120637	C/T	+
rs6843082	112075671	120861	A/G	+
rs13105878	112075751	120941	A/C	+
rs6533528	112076843	122033	A/G	+
rs7692272	112076857	122047	G/T	+
rs2171591	112077012	122202	A/G	+
rs17042195	112077142	122332	C/G	+
rs11931959	112077289	122479	A/G	+
rs17042198	112077582	122772	G/T	+
rs10033464	112078365	123555	G/T	+
rs2171592	112078392	123582	C/T	+
rs13121924	112078423	123613	A/G	+
rs2129979	112078601	123791	G/T	+
rs2350539	112078814	124004	G/T	+
rs1906606	112080996	126186	A/C	+
rs7672570	112081189	126379	C/T	+
rs4834418	112081408	126598	A/G	+
rs723364	112082075	127265	C/G	-
rs723363	112082105	127295	A/G	-
rs7697491	112083422	128612	A/T	+
rs13125644	112083505	128695	A/G	+
rs2350294	112084449	129639	A/G	-
rs2350293	112084451	129641	A/G	-
rs3855819	112084767	129957	C/G	-
rs2220428	112085064	130254	A/G	+
rs2220429	112085089	130279	A/C	+
rs11727566	112085934	131124	A/T	+
rs13141190	112086218	131408	A/G	+
rs4032976	112086371	131561	A/G	-
rs3866829	112086379	131569	A/C	-
rs7671348	112086408	131598	A/G	+
rs3866830	112086598	131788	C/G	-
rs6811267	112086942	132132	C/T	+
rs3853440	112087213	132403	C/T	-
rs3853441	112087344	132534	A/G	-
rs3853442	112087632	132822	C/T	-
rs3853443	112087733	132923	A/G	-
rs4124158	112087798	132988	C/T	+
rs4124159	112087847	133037	A/G	+
rs12506083	112088016	133206	A/C	+
rs34809282	112088051	133241	A/G	+
rs7683219	112088051	133241	A/G	+
rs7683618	112088259	133449	A/C	+
rs7683625	112088269	133459	A/G	+
rs7662050	112088325	133515	C/T	+
rs36183416	112088804	133994	C/T	+
rs4447926	112088804	133994	C/T	+
rs4594787	112088813	134003	A/G	+
rs10390275	112089009	134199	A/T	+
rs36179422	112089009	134199	A/T	+
rs10006659	112089030	134220	G/T	+
rs36181695	112089078	134268	A/G	+
rs7440730	112089078	134268	A/G	+
rs10006881	112089277	134467	C/T	+
rs36149087	112089277	134467	C/T	+
rs6533530	112089540	134730	C/T	+
rs6533531	112089569	134759	G/T	+

Marcador ID	Pós-construção 35	Pós na SEQ ID NO:50	Tipo	Filamento
rs3866831	112089718	134908	C/T	-
rs4269241	112089833	135023	A/G	+
rs4032975	112089842	135032	A/C	-
rs4032974	112090140	135330	A/G	-
rs4124160	112090452	135642	A/G	+
rs3866832	112091304	136494	C/G	-
rs3853444	112091740	136930	A/G	-
rs7662345	112091766	136956	A/G	+
rs2350545	112092258	137448	C/T	-
rs17042215	112092562	137752	C/T	+
rs2003121	112093023	138213	C/T	+
rs880309	112093143	138333	A/G	+
rs9991046	112093346	138536	G/T	+
rs17042216	112094463	139653	C/T	+
rs17570669	112094486	139676	A/T	+
rs17042218	112094520	139710	A/G	+
rs17042223	112094922	140112	C/T	+
rs3866833	112095138	140328	C/T	-
rs17042224	112096509	141699	G/T	+
rs13130446	112096760	141950	C/T	+
rs10516564	112096896	142086	A/G	+
rs7686320	112097215	142405	A/T	+
rs7686499	112097282	142472	C/T	+
rs17042230	112097319	142509	C/T	+
rs4124161	112097459	142649	C/T	+
rs4576077	112098061	143251	C/T	+
rs4260600	112098098	143288	C/T	+
rs12644093	112098445	143635	A/G	+
rs4124162	112098593	143783	A/G	+
rs7674295	112099042	144232	A/G	+
rs11938968	112100356	145546	A/G	+
rs28601812	112101457	146647	A/C	+
rs4032983	112101551	146741	G/T	+
rs3866834	112101617	146807	A/G	+
rs6852021	112101716	146906	A/G	+
rs28580491	112102583	147773	C/T	+
rs13110989	112102671	147861	G/T	+
rs3866835	112102983	148173	C/T	-
rs4124163	112103203	148393	A/G	+
rs3866836	112103244	148434	A/G	+
rs17042238	112103458	148648	A/G	+
rs4124164	112104250	149440	C/T	+

B. Marcadores de microssatélite dentro do bloco LD C04 (Entre 111.954.811 e 112.104.250 em C04; Construção NCBI 35; SEQ ID NO: 50).

Marcador	Posição de partida	Posição final	Iniciador dianteiro	Iniciador reverso
D4S193	112062811	112062911	ACAACCCCATTTGTGAAGAC	TTTATAGAAAATTTAGCATGGA
D4S2940	112070055	112070267	CTAAGTTGTGCAGCCATGAA	TGGAACCACTTTTGCAGTAA
D4S406	112076047	112076292	CTGGTTTTAAGGCATGTTG	TCCTCAGGGAGGTCTAATCA

Tabela 6. Chave para as sequências apresentadas na Listagem de Sequência

SEQ ID NO	Marcador ID
1	rs2220427
2	rs17042059
3	rs4529121
4	rs4543199
5	rs10019689
6	rs4626276
7	rs17042076
8	rs11098089
9	rs11930528
10	rs17042098
11	rs17042102
12	rs17042121
13	rs10516563
14	rs4605724
15	rs2350269
16	rs6533527
17	rs17042144
18	rs1906618
19	rs1906617
20	rs12646447
21	rs12646754
22	rs2129981
23	rs12639654
24	rs6817105
25	rs17042171
26	rs1906591
27	rs2200732
28	rs2200733
29	rs4611994
30	rs4540107
31	rs1906593
32	rs1906596
33	rs2634073
34	rs1906592
35	rs2723296
36	rs16997168
37	rs2723316
38	rs6419178
39	rs1448817
40	rs13105878
41	rs10033464
42	rs13141190
43	rs3853444
44	rs4576077
45	D4S406
46	rs7668322
47	rs2197815
48	rs6831623
49	rs2595110
50	LD Block C04
51	rs13143308

Exemplo 2. Caracterização de variantes de risco de AF

O que segue contém descrição adicional da identificação de variantes que conferem risco para a fibrilação atrial no cromossoma 4q25

5

A fibrilação atrial (AF) é a arritmia cardíaca prolongada mais

comum no ser humano e é caracterizada pela atividade elétrica caótica do átrio. Ela afeta um em dez indivíduos acima dos oitenta, causa morbidez significativa e é um prognosticador independente de mortalidade². Estudos recentes têm fornecido evidência de uma contribuição genética para AF³⁻⁵.

5 Mutações nos genes do canal de potássio foram associados com a AF familiar⁶⁻¹⁰ mas responsável apenas por uma fração pequena de todos os casos de AF^{11,12}. Nós constatamos uma varredura de associação ampla de genoma, seguida pelos estudos de replicação em três populações de descendência europeia e uma população chinesa de Hong Kong e descobrimos uma forte
10 associação entre as duas variantes de sequência no cromossoma 4q25 para AF. Aproximadamente 35% dos indivíduos de descendência europeia têm pelo menos uma das variantes e o risco de AF aumenta em 1,72 e 1,39 por cópia. A associação para a variante mais forte foi replicada na população chinesa, onde a mesma é carregada por 75% dos indivíduos e o risco de AF é
15 aumentado em 1,42 por cópia. Uma associação mais forte foi observada em indivíduos com flúter atrial típico (AFI). Ambas as variantes são adjacentes ao PITX2, que é conhecido desempenhar um papel crítico na assimetria esquerda-direita do coração¹³⁻¹⁵. Nós conduzimos um estudo de associação ampla de genoma usando o Illumina Hap300 BeadChip em uma população
20 islandesa com AF e/ou AFI. 316.515 SNPs que satisfizeram nossos critérios de qualidade foram testados individualmente quanto a associação para AF/AFI em uma amostra de 550 pacientes e 4.476 controles da Islândia. Três SNPs fortemente correlacionados, todos localizados dentro de um único bloco de desequilíbrio de ligação (LD) no cromossoma 4q25, foram os únicos SNPs
25 descobertos ser significantes de amplo genoma depois de considerar os 316.515 SNPs testados ($P < 0,05/316,515 = 1,58 \times 10^{-7}$): rs2200733 (OR = 1,75; $P = 1,6 \times 10^{-10}$), rs2220427 (OR = 1,75; $P = 1,9 \times 10^{-10}$) e rs2634073 (OR = 1,60; $P = 2,1 \times 10^{-9}$). Estes resultados e todos os outros resultados com base na população islandesa foram ajustados quanto a relacionabilidade de

indivíduos. Os dois SNPs mais significantes, rs2200733 e rs2220427, são substitutos perfeitos um do outro no conjunto de dados de CEPH CEU HapMap¹⁶ e são próximos a serem substitutos perfeitos um do outro no conjunto de dados islandês ($D' = 1$, $r^2 = 0,999$), portanto, apenas rs2200733

5 será aludido no debate seguinte. A correlação de rs2634073 para rs2200733 é mais fraca no conjunto de dados islandês ($D' = 0,95$, $r^2 = 0,605$). No outro estudo dos SNPs Illumina Hap300 na vicinidade dos primeiros três SNPs e condicionando a associação para rs2200733, uma associação a um novo SNP, rs10033464, foi identificada ($OR = 1,42$; $P = 0,0024$). Depois de considerar a

10 associação para rs2200733 e rs10033464, a associação para rs2634073 não foi mais significativa ($P = 0,30$). Daqui em diante, todos os resultados de associação para rs2200733 T e rs10033464 T, incluindo aqueles apresentados na Tabela 7, são com base na comparação com o haplótipo do tipo selvagem que não carrega nenhum dos dois alelos em risco, ao invés da comparação

15 com os alelos principais de cada SNP separadamente. Especificamente, razões ímpares para rs2200733 T e rs10033464 T são cada uma computadas condicionalmente e seriam interpretadas como o Risco Relativo estimado de cada variante comparado com o tipo selvagem. Nos alelos de risco T de rs2200733 e T de rs10033464 têm frequências alélicas de população estimada

20 de 12,05% e 8,53% na Islândia, respectivamente e nunca são observados juntos no mesmo cromossoma, no conjunto de dados islandês ou no conjunto de dados de CEU HapMap. Um terceiro SNP, rs13143308, que tem um alelo menor que corresponde completamente aos cromossomas que carregam o alelo T de rs2200733 ou o alelo T de rs10033464, foi identificado através do

25 conjunto de dado de CEU HapMap. A Figura 2 demonstra a estrutura de haplótipo em relação aos SNPs chave da região associada. Os conjuntos de SNPs, que são substitutos perfeitos (isto é, substitutos perfeitos, $r^2 = 1,0$ para o SNP de rotulação) de cada um destes três SNPs chave nas amostras de CEU HapMap, são fornecidos na Tabela 9 e em locais relacionados demonstrados

na Figura 3. Nós enfatizamos que os SNPs indicados devem ser considerados representativos dos haplótipos definidos pelos SNPs que eles são equivalentes e são primariamente escolhidos por questão de conveniência.

Um marcador de microssatélite, D4S406, localizado no mesmo bloco LD como os dois SNPs foi identificado. Na Islândia, três dos quatro alelos mais curtos de D4S406 (8, 4 e 2) combinam para formar um substituto quase perfeito para o alelo T de rs2200733 ($D' = 0,995$, $r^2 = 0,98$) e os dois alelos mais curtos remanescentes (6 e 0) formam um bom substituto do alelo T de rs10033464 ($D' = 0,98$, $r^2 = 0,75$) (Tabela 10). Nenhum dos alelos remanescentes (mais longos) de D4S406 estão associados ao AF/AFI depois de levar em conta o efeito dos alelos curtos. Para a replicação da observação original na Islândia os genótipos D4S406 foram usados para fornecer informação quando genótipos SNP não foram disponíveis.

Em uma tentativa para duplicar a nossa descoberta original nós analisamos uma amostra islandesa adicional que consiste de 2.251 pacientes AF/AFI e 13.238 controles (Tabela 7). A associação de ambos os SNPs para AF/AFI foi replicada nestas amostras ($OR = 1,64$, $P = 2,7 \times 10^{-23}$ para rs2200733, $OR = 1,40$, $P = 8,2 \times 10^{-8}$ para rs10033464) e ambas alcançam significância ampla de genoma nas amostras islandesa combinadas ($OR = 1,68$, $P = 1,9 \times 10^{-30}$ para rs2200733, $OR = 1,38$, $P = 9,4 \times 10^{-9}$ para rs10033464). Nós também tipamos todos os 18 SNPs Hap300 da Illumina na região em torno do nosso sinal em 404 dos casos de AF adicionais e 2.036 dos controles adicionais. Nenhum destes SNPs permaneceram significantes depois de levar em consideração a associação ao rs2200733 e rs10033464 (Tabela 11).

Em outras tentativas para duplicar nossos resultados, nós testamos estas variantes quanto a uma associação para AF em duas populações de ancestralidade européia, uma da Suécia, que consiste de 143 casos e 738 controles e a outra dos Estados Unidos (U.S.), que consiste de

636 casos e 804 controles (Tabela 7). A associação para rs2200733 foi fortemente duplicada em ambas as populações (OR = 2,01, P = 0,00027 na Suécia ou = 1,84, P = $9,8 \times 10^{10}$ nos U.S.). A associação ao rs10033464 é mais fraca, mas não obstante foi duplicada na população Sueca (OR = 1,65, P = 0,0087) e foi quase significativa na população U.S. (OR = 1,30, P = 0,052). Quando combinado com as amostras islandesas, a associação a rs2200733 foi inequívoca (OR = 1,72, P = $3,3 \times 10^{-41}$) e a significância de rs10033464 foi bem além do limiar de significância ampla de genoma (OR = 1,39, P = $6,9 \times 10^{-11}$). Assumindo o modelo multiplicativo, o risco atribuível à população (PAR) das duas variantes combinadas é de aproximadamente 20% em populações de ancestralidade europeia.

Finalmente, nós tentamos replicar estes sinais em uma população Han Chinesa de Hong Kong que consiste de 333 casos e 2.836 controles. A associação ao rs2200733 T foi significativamente duplicada (OR = 1,42, P = 0,00064), mas a associação ao rs10033464 T não foi significativa, embora na direção certa (OR = 1,08, P = 0,55) (Tabela 7). De maneira interessante, o alelo T de rs2200733 é muito mais frequente nos chineses (frequência alélica nos controles: 0,528) do que naqueles de descendência europeia (frequência alélica nos controles: 0,098 a 0,139) (Figura 2) que é refletida em um PAR de união maior de aproximadamente 35%, embora o risco estimado seja menor. O bloco LD contendo as duas variantes é mais fragmentado nas amostras HapMap de CHB Chinês e JPT Japonês do que nas amostras HapMap de CEU (Figura 3). Nós portanto analisamos diversos marcadores na população de Hong Kong que foram em LD perfeito com rs2200733 nas amostra de CEU, mas em LD imperfeito nas amostras CHB e JPT (Tabela 12). Estes marcadores tiveram associação evidente mais fraca para AF do que rs2200733, sugerindo que as variantes funcionais que dirigem a associação estão localizadas na região de aproximadamente 20 kb em torno da variante rs2200733 original e definido pelo SNPs que permanece

equivalente a rs2200733 nas amostras CHB e JPT (de cor vermelha na figura 3).

Para as amostras da descoberta islandesa inicial, rs2200733 teve um OR significativamente mais alto do que rs10033464 ($P = 0,041$). Isto confirmou-se nas amostras de duplicação e de ponta a ponta existe uma diferença significativa nos riscos associados com as duas variantes ($P = 0,00019$ nas amostras europeias combinadas e $P = 0,0099$ nas de Hong Kong). Quando as Razões de Chance específicas de genótipo foram estudadas, algum desvio fora do modelo multiplicativo é detectável no conjunto de dados combinados ($P = 0,018$ para as amostras europeias, ver a Tabela 13). Os riscos estimados de carregadores heterozigotos em relação aos não carregadores foram similares, mas os carregadores homozigotos de rs2200733 T e rs10033464 T têm riscos estimados que foram, respectivamente, mais altos e mais baixos do que aquele prognosticado por um modelo multiplicativo. Uma tendência similar foi observada nas amostras de Hong Kong; embora o tamanho da amostra fosse muito pequeno para ter força para detectar tais desvios com significância. Nas populações combinadas de descendência europeia o OR observado para indivíduos homozigotos para rs2200733 T foi 3,64 quando comparado aos indivíduos homozigotos para o haplótipo do tipo selvagem e 1,77 para a população chinesa demonstrando que estas variantes são componentes importantes em qualquer modelagem preditiva de AF.

A idade no diagnóstico de AF/AFI para as amostras islandesas correlaciona-se com os dois SNPs (o diagnóstico ocorre 2,28 anos mais cedo para o alelo T de rs2200733 e 1,10 anos mais cedo para o alelo T de rs10033464, P da união = $1,29 \times 10^{-6}$). O efeito da idade no diagnóstico também foi avaliado medindo-se a força de associação enquanto estratificando pela idade no diagnóstico. A associação das duas variantes é mais forte naqueles diagnosticados em uma idade mais jovem, embora o risco

permaneça significativa mesmo naqueles diagnosticados depois de atingir 80 anos de idade (Tabela 8). A informação sobre a idade no diagnóstico de AF não foi disponível para as amostras suecas. As amostras U.S. foram compreendidas de dois grupos principais, pacientes mais jovens com AF sozinho ou AF e hipertensão (HTN) e casos de AF mais velhos que são a maioria das vezes hemorrágicos e pacientes com acidente vascular cerebral isquêmico. Em ambas as populações existe uma tendência clara para uma associação mais forte em casos de AF mais novos do que em casos mais velhos. Nossa análise dos dados não sugere qualquer associação diferencial pelo sexo (Tabela 8).

AF1 frequentemente acompanha AF, mas pode ocorrer em isolamento¹⁷. De maneira interessante nós observamos uma forte associação entre as variantes e o subconjunto pequeno (N = 116) dos pacientes AF1 islandeses (OR = 2,60, 95% de intervalo de confiança (CI) = 1,83 a 3,68, P = $7,5 \times 10^{-8}$ para rs2200733 ou = 1,94, 95% de CI = 1,26 a 3,00, P = 0,0028 para rs10033464). De fato, para rs2200733, a OR para estes casos AFI definidos é significativamente mais alta do que aquela para os casos com um fenótipo AF (P = 0,0026) e próximo a significativamente mais alta para rs10033464 (P = 0,084). Nossos resultados sugerem que embora estes traços compartilhem fatores de risco genético, AFI é menos influenciado pelas fenocópias do que AF.

Nenhuma variante mostrou uma associação com a obesidade, hipertensão ou infartação do miocárdio nas amostras islandesas, todos fatores de risco conhecidos para AF (observado OR < 1,1 em todos os casos, Tabela 14). Embora estes resultados negativos não excluam a possibilidade de que as novas variantes associem-se com estes fenótipos, eles sugerem, junto com o alto risco nos U.S. AF sozinho e idade mais cedo no início em carregadores, que as novas variantes não estão afetando o risco de AF através destes fatores de risco conhecidos.

Não existe nenhum gene conhecido presente no bloco LD contendo rs2200733 e rs10033464 (Figura 3). O bloco LD contém um EST unido (DA725631) e dois ESTs de exon único (DB324364 e AF017091). A RT-PCR de bibliotecas de cDNA de vários tecidos não detectaram a expressão destes ESTs (Tabela 16). O gene PITX2 localizado no bloco LD adjacente a montante é o gene mais próximo às variantes de risco. Diversos marcadores dentro do bloco LD contendo o gene PITX2 são correlacionados com os marcadores que mostram a associação ao AF e Af1, como mostrado na Tabela 18. A proteína codificada por este gene, o fator de transcrição 2 de homodomínio emparelhado, é um candidato interessante para AF/AFI visto que é conhecido desempenhar um papel importante no desenvolvimento cardíaco pela direção da morfogênese assimétrica do coração¹³. Em um modelo de camundongo silenciado Pitx2 foi mostrado suprimir um caminho de default para a formação do nodo sinoatrial no átrio esquerdo^{14,15}. Existe muito pouca expressão de mRNA de PITX2 em todos os tecidos facilmente acessíveis, tais com sangue e tecido adiposo, dificultando o estudo de correlação entre genótipos e níveis de expressão. O gene seguinte a montante de PITX2 é ENPEP, uma aminopeptidase responsável pela ruptura da angiotensina II no endotélio vascular¹⁸. Este gene é expressado mais amplamente, mas as variantes associadas com AF não mostraram nenhuma correlação para a sua expressão no sangue ou tecido adiposo. Nenhum outro dos genes anotados está localizado dentro de uma região de 400 kb a montante e regiões de 1,5 Mb a jusante das variantes associadas.

Em resumo, nós identificamos duas variantes no cromossoma 4q25 que estão fortemente associados com AF em três populações distintas de descendência européia. A variante mais forte também duplica bem em uma população chinesa onde a mesma é muito mais comum e tem PAR mais alta do que em populações de descendência européia. Esta associação é particularmente obrigatória em pacientes mais jovens e naqueles com AF

sozinha, mas também está presente em pacientes mais velhos com formas mais habitualmente encontradas de AF. Embora o mecanismo para esta associação seja desconhecido, nossos resultados fornecem um alicerce para mais estudos nos suportes molecular de AF.

5 MÉTODOS

Pacientes

Os casos islandeses consistiram de todos os pacientes diagnosticados com AF e/ou AFI nos dois maiores hospitais no país de 1987 a 2005. Os casos suecos foram recrutados de 1996 a 2002 como uma parte de um estudo de epidemiologia genética em andamento, o Estudo do Acidente Vascular Cerebral Isquêmico do Sul de Stockholm. Os casos U.S. foram uma mistura de pacientes com acidente vascular cerebral com um diagnóstico de AF e pacientes consecutivos mais jovens com AF sozinho ou AF com um diagnóstico co-existente de hipertensão. Os casos de Hong Kong foram uma coleção de pacientes com acidente vascular cerebral e diabete com um diagnóstico AF. O diagnóstico AF foi confirmado por um eletrocardiograma de doze fios condutores elétricos em todas as populações de estudo.

Os controles islandeses foram escolhidos aleatoriamente de indivíduos que têm participado em outros estudos genéticos no deCODE, excluindo primeiro grau em relação de pacientes e controles (Tabela 15). Os controles suecos foram recrutados da mesma região como pacientes de doadores de sangue (em 2001) e voluntários saudáveis (1990-1994). Os controles U.S. foram recrutados de uma prática de primeiros socorros grande e de pacientes que participam em um estudo de acidente vascular cerebral hemorrágico. Os controles de Hong Kong foram indivíduos sem um diagnóstico de AF.

População de estudo islandesa

Este estudo inicialmente incluiu o consentimento de todos os pacientes para participação, que foram diagnosticados com AF e/ou AFI (ICD

10 diagnóstico 148 e ICD 9 diagnóstico 427,3) no Landspítali University Hospital em Reykjavik, o único centro de referência terciária na Islândia e no Akureyri Regional Hospital, o segundo maior hospital no país, de 1987 a 2005. Todos os diagnósticos foram confirmados por um eletrocardiograma de
 5 doze fios condutores elétricos (EKG) que foi manualmente lido por um cardiologista. Todos os casos foram incluídos, independente se os pacientes tiveram sintomas clínicos ou não, exceto aqueles diagnosticados apenas imediatamente depois de cirurgia cardíaca aberta.

Um conjunto de 550 casos foram genotipados com êxito de
 10 acordo com nossos critérios de controle de qualidade em um esforço de genotipar SNP de genoma amplo, usando o método de ensaio Infinium II e o Sentrix HumanHap300 BeadChip (Illumina, San Diego, CA, USA). A idade média no diagnóstico para este grupo inicial de 550 pacientes (370 homens e 180 mulheres) foi de 72,5 (SD = 11,0) anos e a faixa foi de 34,7 a 96,2 anos.
 15 O grupo de validação de 2.273 pacientes (1.359 homens e 913 mulheres) teve uma idade média no diagnóstico de 70,5 (SD = 13,0) e a faixa foi de 16,8 a 100,6. Os controles livres de AF/AFI (2.201 homens e 2.275 mulheres na triagem de genoma amplo inicial com idade média de 61,5 (SD = 15,8) e 5.654 homens e 7.597 mulheres no estágio de validação com idade média de
 20 61,9 (SD = 18,4)) usados neste estudo consistiram de controles aleatoriamente selecionados da base de dados genealógica islandesa e indivíduos de outros estudos genéticos relacionados em andamento no deCODE. Os controles tendo primeiro grau em relação (irmãos, pais ou progênie) com AF/AFI ou um controle de primeiro grau em relação, foram excluídos da análise.

25 ***Populações MI islandesa, obesidade e hipertensão***

Indivíduos que sofreram um MI foram identificados de um registro de mais de 10.000 indivíduos que: a) tiveram um MI antes da idade de 75 na Islândia nos anos de 1981 a 2002 e satisfazem o critério MONICA 9 (REF II) ou tiveram diagnóstico de descarga de MI dos maiores hospitais em

Reykjavik nos anos 2003 e 2005. Os diagnósticos de MI de todos os indivíduos no registro seguem critérios de diagnóstico estritos com base nos sinais, sintomas, eletrocardiogramas, enzimas cardíacas e observações de necrópsia. A informação de genótipo foi disponível para 2.462 homens e 1.114 mulheres, idade média 72,6 (SD = 11,7). O índice de massa corporal (BMI) foi medido para indivíduos que participam nos programas genéticos de fibrilação atrial cardiovascular e/ou acidente vascular cerebral (CVD) no deCODE (pacientes com CVD, seu primeiro grau em relação ou esposos). Para o propósito deste estudo pacientes com BMI > 35 foram definidos como obeso. Informação de genótipo foi disponível para 555 homens e 1.046 mulheres, idade média 53,2 (SD = 16,1). Pacientes hipertensivos incluíram aqueles que foram atendidos na clínica de hipertensão ambulatorial no Landspítali, University Hospital na Islândia e/ou foram diagnosticados na descarga do hospital. O diagnóstico foi verificado pela confirmação de que eles estavam tomando medicações anti-hipertensivas como um tratamento para a hipertensão. A informação de genótipo foi disponível para 1.293 homens e 1.327 mulheres, idade média 71,5 (SD = 12,5). O estudo foi aprovado pela Data Protection Commission of Iceland e pelo National Bioethics Committee of Iceland. Consentimento informado por escrito foi obtido de todos os pacientes, em parentes e controles.

População de estudo sueca

Pacientes com acidente vascular cerebral isquêmico ou TIA tratando na unidade de acidente vascular cerebral ou a clínica de paciente externo com acidente vascular cerebral no Karolinska University Hospital, unidade de Huddinge em Stockholm, Suécia foram recrutados de 1996 a 2002 como parte de um estudo de epidemiologia genética em andamento, o Estudo do Acidente Vascular Cerebral Isquêmico no Sul de Stockholm (SSISS). O estudo foi aprovado pelo Bioethics Committee of Karolinska Institutet (Dnr 286/96 e 08/02). O diagnóstico de AF nas amostras suecas foi com base em

um EKG de doze fios condutores elétricos. A fração de homens nos casos de AF suecos foi de 46,2% e a idade média no diagnóstico de acidente vascular cerebral para os casos de AF suecos foi de 74,4 (SD = 8,7).

Os controles suecos usados neste estudo são controles com base na população recrutados da mesma região na Suécia central como os pacientes, representando a população geral nesta área. Os indivíduos foram doadores de sangue (recrutados em 2001) ou voluntários saudáveis (coletados em 1990-1994) recrutados pelo departamento químico clínico no Karolinska University Hospital para representar uma população de referência normal. A fração de controles de homens na Suécia foi de 59,7% e a idade média no recrutamento para os controles suecos foi de 43,1 (SD = 12,3).

População de estudo dos U.S.

Os pacientes U.S. foram alistados em estudos de controle e grupos de casos em andamento no Massachusetts General Hospital (MGH) entre Janeiro de 1998 e Julho de 2006. Todos os aspectos destes estudos foram aprovados pelo Institutional Review Board local. Pacientes alistados no estudo de controle de caso consistiu de pacientes hospitalizados com acidente vascular cerebral isquêmico ou hemorrágico agudo confirmado pela CT ou MRI, admitidos em um único hospital de cuidado agudo. Dos 328 pacientes de acidente vascular cerebral hemorrágico recrutados 78 foram diagnosticados com AF e foram usados como casos para o estudo corrente, os 250 remanescentes foram usados como controles. 170 pacientes com acidente vascular cerebral isquêmico tiveram um diagnóstico de AF e foram tratados como casos mas nenhum paciente com acidente vascular cerebral isquêmico foi tratado como controle. Pacientes foram excluídos para hemorragia subaracnóide primária e para hemorragia intracerebral secundária para trauma de cabeça, tumor, má formação vascular ou vasculite. 624 controles livres de acidente vascular cerebral foram recrutados de uma prática de primeiros socorros grande (> 18.000 pacientes) servindo a área de captação do hospital

assim como o Serviço de Controle Anticoagulação do hospital. 70 dos 624 indivíduos coletados como controles foram diagnosticados com AF e casos tratados com o propósito do estudo corrente. 50,9% de todos os indivíduos usados como controles foram homens e a sua idade média foi de 67,4 (SD = 12,3). Todos os pacientes ou um informante acompanhante forneceram consentimento informado para a participação em estudos genéticos e foram entrevistadas de forma previdente com respeito à história médica, medicações, história social e familiar. A presença ou ausência de fibrilação atrial foram documentadas de forma previdente através de entrevista e de revisões de registros médicos.

A segunda parte dos pacientes U.S. consistiu de pacientes consecutivos só com AF ou AF com diagnóstico coexistente de hipertensão encaminhados ao serviço de arritmia que forneceram consentimento informado por escrito para a participação em estudos genéticos. Os critérios de inclusão foram AF documentado pelo EKG e uma idade menor do que ou igual a 65 anos. Os critérios de exclusão foram fibrilação atrial cardíaca estrutural e/ou acidente vascular cerebral como avaliado pela ecocardiografia, fibrilação atrial cardíaca reumática e/ou acidente vascular cerebral, hipertireoidismo, infartação do miocárdio ou insuficiência cardíaca congestiva. Cada paciente passou por uma exame física e uma entrevista padronizada para identificar condições médicas passadas, medicações, sintomas e deflagradores possíveis para o início da AF. Todos os pacientes foram avaliados pelo EKG de doze fios condutores elétricos, ecocardiograma e estudos de laboratório. EKGs e ecocardiogramas foram interpretados usando critérios padrão.

População de estudo de Hong Kong

Todos os pacientes na população de estudo de Hong Kong foram de ancestralidade chinesa Han do sul residindo em Hong Kong. Os casos consistiram de 217 indivíduos (49,1% homens, idade média 68,1 (SD =

9,6)) selecionados do Prince of Wales Hospital Diabetes Registry²³ e 116 pacientes (30,2% homens, idade média 76,1 (SD = 10,9)) do Stroke Registry²⁴. Todos os pacientes foram diagnosticados ter fibrilação atrial pelo EKG. Os controles consistiram de 2.836 pacientes sem evidência de AF. O consentimento informado foi obtido para cada paciente participante. Este estudo foi aprovado pelo Clinical Research Ethics Committee of the Chinese University of Hong Kong.

Genotipagem de genoma amplo Illumina

Todas as amostras de caso e controle islandesas foram ensaiadas com os chips de SNP Infinium HumanHap300 (Illumina, SanDiego, CA, USA), contendo 317 503 SNPs que rotulam haplótipo derivados da fase I do projeto International HapMap. Dos SNPs ensaiados no chip, 162 SNPs gerados nos genótipos e um adicional de 178 SNPs produziram mais baixo do que 90%. Quarenta e oito SNPs foram monomórficos e 107 outros quase monomórficos (isto é, a menor frequência de alelo no grupo combinado de pacientes e controles foi menor do que 0,001). Um adicional de 475 SNPs mostraram distorção muito significativa do equilíbrio de Hardy-Weinberg nos controles ($p < 1 \times 10^{-10}$). Por último, uns poucos marcadores ($n = 18$) foram determinados ter problemas de genotipagem depois da investigação de regiões particulares e sinais possíveis em diversos estudos de associação de genoma amplo em andamento diferentes internos. Assim, as análises finais apresentadas no texto utiliza 316.515 SNPs. Quaisquer amostras com uma taxa chamada abaixo de 98% foram excluídas da análise.

Genotipagem de SNP único e Microsatélite.

A genotipagem de SNP foi realizada pela plataforma Centaurus (Nanogen)²⁵. A qualidade de cada ensaio de SNP da Centaurus foi avaliada pela genotipagem de cada ensaio nas amostras CEU e/ou YRI HapMap e comparando os resultados com os dados de HapMap. Os ensaios com >1,5% de taxa de desemparelhamento não foram usados e um teste de

desequilíbrio de ligação (LD) foi usado para marcadores conhecidos estar em LD.

Análise de Associação

Uma tentativa foi feita para genotipar todos os indivíduos participantes quanto a rs2200733, rs4611994 (um substituto perfeito para rs2200733), rs13143308 e rs6843082 (um substituto perfeito para rs13143308). Para cada um dos SNPs, o rendimento foi mais alto do que 90% em cada grupo. Além dos genótipos para o microssatélite D4S406 foram disponíveis para todos os pacientes islandeses e suecos. Por causa da redundância na genotipagem, genótipos observados reduziram a quantidade de informação perdida devido aos genótipos perdidos através de um método de probabilidade que nós usamos antes²⁶. Isto garantiu que os resultados apresentados nas tabelas foram sempre com base no mesmo número de indivíduos, permitindo comparações significativas de resultados. Como os dados em rs10033464 foram apenas diretamente disponíveis nas amostras de descoberta islandesa iniciais e no projeto HapMap os haplótipos rs2200733 C rs13143308 T foram usados para rotular este SNP. Esta rotulação foi perfeita tanto nas amostras da descoberta inicial quanto nas amostras CEPH CEU HapMap.

Um procedimento de probabilidade descrito em um anterior e implementado no software NEMO, foi usado para a análise de associação. Nós testamos a associação de um alelo para cada fenótipo usando uma estatística de razão de probabilidade padrão, que, se os pacientes não fossem relacionados, teriam assintoticamente uma distribuição de qui quadrado, com um grau de liberdade, sob a hipótese nula. A OR específica de alelo foi calculada assumindo um modelo multiplicativo para os dois cromossomas de um indivíduo⁴. Resultados de grupos de controle de caso múltiplos foram combinados usando um modelo de Mantel-Haenszel⁵ em que os grupos foram deixados ter frequências de população alélica diferentes, haplótipos e

genótipos mas foram assumidos ter Riscos Relativos comuns. Não houve nenhum desvio significativo do equilíbrio de Hardy-Weinberg (HWE) em nenhum grupo de controle.

Nas Tabelas 7 e 8, os valores P tanto para rs2200733 quanto rs10033464 foram computados com base nas comparações com haplótipos do tipo selvagem rs2200733 C, rs13143308 G, rs10033464 G que não carregam nenhum dos alelos em risco. A Razão de Chances condicionais correspondentes para rs2200733 T é definida como $[f(rs2200733 T)/f(WT)]/[p(rs2200733 T)/p(WT)]$ onde WT denota o haplótipo tipo selvagem e $f(.)$ e $p(.)$ denotam frequências em casos e controles respectivamente. Sob o modelo multiplicativo e quando os controles puderam ser considerados como controles de população, esta Razão de Chances condicional é a estimativa apropriada do Risco Relativo de rs2200733 T versus o tipo selvagem. A razão impar condicional para rs13143308 T é similarmente definido e tem uma interpretação similar.

Correção para a relacionabilidade e Controle Genômico.

Alguns dos indivíduos nos grupos de caso-controle islandeses foram relacionados entre si, fazendo com que a estatística do teste de qui quadrado anteriormente mencionado tivesse uma média >1 e intermediário $> 0,675^{26}$. Nós estimamos o fator de inflação pelo uso de um procedimento anteriormente descrito onde nós simulamos genótipos através da genealogia de 731.175 islandeses³⁰. Para as amostras da descoberta inicial, onde genótipos para os 316.515 SNPs de varredura de genoma amplo foram disponíveis, nós também estimamos o fator de inflação pelo uso de controles genômicos e calculando a média das 316.515 estatísticas qui quadrado e pela computação da média das estatísticas de qui quadrado 316.515 e dividindo-a por $0,675^{26}$ como descrito anteriormente^{31,32}. Para estas amostras iniciais os fatores de inflação, estimados pelo método da genealogia ORr e os dois métodos de controle genômico deram estimativas de fator de inflação

similares; 1,047, 1,058 e 1,054 respectivamente. Os valores P e intervalo de confidências apresentadas são fundamentados no ajuste pelo fator de inflação estimado pelo método da genealogia.

Triagem por PCR de bibliotecas de cDNA.

5 Para confirmar a expressão das ESTs unidas (DA725631, DB324364 e AF017091) dentro do bloco LD nós triamos bibliotecas de cDNA comercialmente disponíveis e bibliotecas geradas no deCODE. As bibliotecas comerciais triadas foram bibliotecas de cDNA Marathon Ready de coração (Clontech-639304), aorta (Clontech-639325) medula óssea (Clontech 10 7416-1), testículo (Clontech 7414-1) e cérebro integral (BD 50598). Além disso as bibliotecas de cDNA foram construídas para sangue integral e células de linfoblastóide humano transformadas com EBV. O RNA total foi isolado das linhagens de célula de linfoblastóide e sangue integral, usando o kit de isolamento de RNA RNeasy da Qiagen (Cat. 75144) e a isolamento de RNA 15 RNeasy do kit sangue integral (Cat. 52304), respectivamente. As bibliotecas de cDNA foram preparadas na deCODE usando o Kit High Capacity cDNA Archive com iniciadores aleatórios (Applied Biosystems PN 4322171).

 A triagem pela PCR foi realizada usando o Advantage® 2 PCR Enzyme RT _PCR System (Clontech) de acordo com as instruções dos 20 fabricantes e usando iniciadores de PCR da Operon Biotechnologies. As reações de PCR foram feitas em volume de 10 µl a uma concentração final de 3,5 µM de iniciadores avançados e reversos (Tabela 16), 2 mM de dNTP, tampão de PCR 1x Advantage 2 e 0,5 µl de biblioteca de cDNA.

Análise de Northern Blot.

25 Northern blots poli-A de tecido múltiplo comercial foram obtidos da Clontech (Sistema Cardiovascular Humano, Cat. 636825).

Sondas usadas:

i) o clone de cDNA PITX2 (HU3_p983E0327D), obtido da RZPD Deutsches Resfontenzentrum fur Genomforschung GmbH, Alemanha

<http://www.rzpd.de/products/genomecube.shtml>) (sequência verificada, dados não mostrados);

ii) Clone de cDNA que correspondeu aos exons 1 a 12 dos transcritos ENPEP obtidos de experimentos de RT-PCR. O clone ENPEP foi verificado na sequência:

```
TCCTGCTCCAGCTTGTGGATATTTTGCAAAAAGCTCTCCATCTGCCACAGTTGCAGTTCAGTGTTG
AATGGCTCTGCTATTGTGACAATTCGGCCAAGGTTTCTGTTATTGAGTGTATATCTGTTGACTAGAT
AGTCCCAGTTGAGTTGTATCCAATTCAGGCCATGTTCTTCCATAGCTGTTATATGAGATATATCG
AATGACTGTAAACACATCCTGAGTTTTAATAAGGTTCTGTCTCTTGAGCAAATCCAAATACCTTGAC
AAAAGAGTAACGTTCTTCACTGATGCTAATCCATACAGCAGTTTTTCTTTTCTTGAGCTAATGAAGT
TTTCTGGTATTGCTCAAGAGTGTAGTTCCATGAAATCTCATTGCCAGAGTTCTGCATCCCATACCGA
TACACCAGAAGCCTGAGATTTACGGGAAGGCTTACAGTCCCATTTAGCCACTGCTCAAATAACGAG
GAAGCATTGTTCAAGGCTTCTCTGTCTCCCATCTTGCACGCAAACCCTAACACGGAGGAACGGAGT
AACTTTGTGACATGGTCTCCAGCATCATTCATCCCAGAGAATCTGCAATAGGCTTCACTTGACCTT
GGAAGTATTCCTCAATCATAGGATATAGCTCTTTATCATCTTCAAACATGCTAATGATGTAGGTTACA
GCTGAAATTACTCTCTGCCATGGTAAAAAATTCTCTTCCCTTTTGAGATACTTGGTCAAGTTCAAAG
CCACCTTATAATCTAGAAGTTGAGCTCTTGCCAAGGCAAAAGCATCATCAATAAGACTTGCACGATC
TGCTGAAGAAAATGTCTTGTGGTTCAAGGAGAGCGCTGTAGCTATCGAGTCCCAAGTTGCTACTTC
ATAATTTACACGATAAAACCCAATATGATCTGGGTTTATTTTGAGAAAAGCATTTCCACTAGGATTA
GAGGAGTTCAAAGTGATTCTTCTTTTCTGACCTATTAAATAACACACTGCTTGTTATATTATCTTC
AGTCCATTTAACTGGGATATTCCATGTATAACCAAGATCTGAAGGGGGCTGAGAAGGGTTAGCTCT
TGGGTCCAACAAAAGCGTTTCTGTGTGATGTTCTTGACACCGTTCACGTTAAGCACAGGATAACC
CATCTgTCTGGTCCAGGTGTCCATTCTTCTTCACTGGTAGCCTACTTGCTCTTCCAGTGCTGC
CCAAAAAT
```

Os fragmentos de cDNA foram radiorrotulados com [α - 32 P]dCTP (atividade específica 6000 Ci/mmol), usando o kit de rotulação Megaprime (GE Healthcare Cat. RPN 1607) e os nucleotídeos não incorporados removidos da reação usando microcolunas ProbeQuant G-50 (GE Healthcare Cat. 275335-01). As membranas foram pré-hibridizadas em tampão Rapid-hyb (GE Healthcare Cat. RPN 1635) por pelo menos 30 minutos e subsequentemente hibridizados com 100 a 300 ng da sonda de cDNA rotulada. As hibridizações foram realizadas em tampão Rapid-hyb a 65°C durante a noite. As sondas rotuladas foram aquecidas por 5 minutos a 95°C antes da adição aos filtros na solução de pré hibridização. Depois da hibridização, as membranas foram lavadas em baixa severidade em 2x SSC, 0,05% de SDS na temperatura ambiente por 30 a 40 minutos seguido por duas lavagens de alta severidade em 0,1x SSC, 0,1% de SDS a 50°C por 40 minutos. As manchas foram imediatamente seladas e expostas à película de raio X BioMax MR da Kodak (Cat. 8715187).

Avaliação para variantes reguladoras candidatas na região AF

O browser UCSC foi usado para posições extraídas de SNPs e sítios de ligação de fator de transcrição conservados (TFBS) para uma região de 172,5 kb em torno dos SNPs associados com AF (hg libera 17, cromossoma 4, 111.942.401-112.114.901). As duas tabelas foram de referência cruzada e SNPs que acabaram nos sítios de ligação foram ainda interrogados quanto a LD com rs2220427 ou rs6843082 nos dados HapMap. Isto foi feito para as liberações 16, 17 e 18 do genoma humano, mas os resultados são relatados nas coordenadas hg 17. Isto produziu 3 SNPs que acabam nos sítios de ligação conservados para fatores de transcrição conhecidos (Tabela 17). Nota, esta análise detecta apenas uma amostra limitada de candidatos funcionais como i) os haplótipos AF não foram totalmente sequenciados, ii) diversos SNPs candidatos não são tipados em Hapmap e é desconhecido se eles estão nos haplótipos AF, iii) os polimorfismos em regiões menos conservadas podem ser funcionais.

Conservação evolucionária de três TFBS

A utilização do alinhamento Multiz no genoma UCSC permitiu uma avaliação da conservação evolucionária das regiões afetadas por estes SNPs. Em todos os três casos é a parte de núcleo dos sítios de ligação TF intactos, mas as posições afetadas são preservadas a um grau diferente. O SOX5 afetado pelo rs12510087 é o menos conservado em mamíferos mas o segundo (afetado pelo rs2220427) é estritamente preservado (com a exceção de Opossum é mantido em todas as espécies para a galinha). A mutação rs17042171 está na última posição no motivo GGAAAA de núcleo do sítio de ligação NFAT. A conservação indica que um G é preferido nesta localização, que resulta em um motivo GGAAAG.

Correlação entre o genótipo e expressão de ENPEP

Os sangue foi coletado pela manhã, entre 8 e 10 am, depois durante a noite jejum (a partir das 9 pm) e o RNA extraído dentro de 2 horas

da flebotomia de 1.002 indivíduos. A isolamento de RNA foi realizada usando o Kit RNeasy Midi (QIAGEN GmbH, Hilden, Alemanha). Amostras de gordura subcutâneas (5 a 10 cm³) foram removidas através de uma incisão de 3 cm na linha do biquini (sempre do mesmo sítio para evitar a variação específica de sítio) depois da anestesia local usando 10 ml de lidocaína-adrenalina (1%) de 673 indivíduos. Purificação do RNA total foi realizada com o Kit RNeasy Mini (QIAGEN GmbH, Hilden, Alemanha).

A integridade do RNA total foi avaliado através da análise no Bioanalisador Agilent 2100 (Agilent Technologies, Palo Alto, U.S., CA). Cada amostra de RNA rotulado incluiu grupos de referência, 1.765 amostras no total, foi hibridizada a um arranjo Human 25K fabricado pela Agilent Technologies. As imagens de arranjo foram processadas como anteriormente descrito para obter o ruído de fundo, intensidade de canal único e estimativas de erro associadas com a medição¹¹. As mudanças na expressão entre duas amostras foram quantificadas como logaritmo médio ($\log_{10}$) da razão de expressão (MLR), isto é, as razões de expressão comparadas com os valores de intensidade corrigidos com o fundo para os dois canais para cada ponto no arranjo¹². As hibridizações foram através do processo QC padrão, isto é, razão de sinal para ruído, reprodutibilidade e precisão nos compostos injetados, comparando as intensidades Cy3 a Cy5.

Nenhum SNP associado foi correlacionado com a expressão de ENPEP ajustada para a idade e sexo no sangue ($P = 0,90$ e $P = 0,82$ para rs2200733 e rs10033464, respectivamente) ou tecido adiposo ($P = 0,23$ e $P = 0,37$ para rs2200733 e rs10033464, respectivamente)

Tabela 7. Análise da associação de rs2200733 e rs10033464 no cromossoma 4q25 para AF/AFI.

Amostra (N casos/ N controles)	Rs2200733 T ^a			rs10033464 T ^{a,b}			Comparação P ^d	União PAR
	Freq. ^c	OR (95% CI)	P	Freq. ^c	OR (95% CI)	P		
Descoberta								
Islândia ^e (550/4.476)	0,191 0,114	1,84 (1,54-2,21)	2,0x10 ⁻¹¹	0,110 0,080	1,42 (1,13-1,77)	0,0024	0,041	0,216
Replicação (2.251/13.238)	0,166 0,108	1,64 (1,49-1,81)	2,7x10 ⁻²³	0,108 0,080	1,40 (1,24-1,58)	8,2x10 ⁻⁵	0,028	0,176
Combinado (2.801/17.714)	0,171 0,110	1,68 (1,53-1,83)	1,9x10 ⁻³⁰	0,108 0,080	1,40 (1,25-1,55)	9,4x10 ⁻⁹	0,0025	0,180
Outras ancestralidades europeias								
Suécia (143/738)	0,179 0,098	2,01 (1,38-2,93)	0,00027	0,172 0,111	1,65 (1,14-2,41)	0,0087	0,41	0,272
U.S. (636/804)	0,229 0,139	1,84 (1,151-2,23)	9,8x10 ⁻¹⁰	0,105 0,083	1,30 (1,00-1,69)	0,052	0,026	0,232
Combinado ^f	- -	1,88 (1,58-2,23)	1,2x10 ⁻¹²	- -	1,41 (1,13-1,75)	0,0019	0,027	0,237
Todas as ancestralidades europeias								
Combinado ^f	- -	1,72 (1,59-1,86)	3,3x10 ⁻⁴¹	- -	1,39 (1,26-1,53)	6,9x10 ⁻¹¹	0,00019	0,206
Hong Hong								
Hong Hong (333/2.836)	0,605 0,528	1,42 (1,16-1,73)	0,00064	0,190 0,218	1,08 (0,84-1,39)	0,55	0,0099	0,346

Cada linha contém os resultados de uma análise conjunta de duas variantes, rs2200733 T e rs10033464 T^b. Os números de casos e controles (N) são mostrados para cada estudo de caso-controle e para cada variante as frequências alélicas da variante em casos e controles, a OR com um 95% CI e valores P de lado duplo, são mostrados. Além disso, um Valor p para comparar o efeito das duas variantes e seu risco atribuível à população (PAR) conjunto são relatados. Por exemplo, a primeira linha indica que, para as amostras da descoberta islandesa inicial, rs2200733 T tem uma Razão de Chances (OR) estimada de 1,84 (95% CI (1,54-2,21), P = 4,1 x 10⁻¹¹) vs o tipo selvagem (rs2200733 C, rs13143308 G, rs10033464 G haplótipo) e rs10033464 T tem uma OR estimada de 1,42 (95% CI (1,13-1,77), P = 0,0024) vs o tipo selvagem.

^a Resultados de comparar rs2200733 T e rs10033464 T aos haplótipos rs2200733 C, rs13143308 G, rs10033464 G do tipo selvagem.

^b Nas amostras suecas e nas U.S. rs10033464 T foi rotulado pelos haplótipos rs2200733 C, rs13143308 T

5 ^c A frequência nos casos (acima) e controles (abaixo)

^d Valor p para comparar as ORs de rs2200733 T e rs10033464 T.

^e A análise de associação foi ajustada quanto a relacionabilidade de alguns dos indivíduos.

10 ^f Para as populações de estudo combinadas de descendência europeia, o PAR foi calculado pelo uso da frequência de controle média, não ponderada das populações, enquanto que a OR e o Valor p foram estimados usando o modelo de Mantel-Haenszel.

Tabela 8. Associação pela idade no diagnóstico na Islândia e pelo sub-fenótipo de AF no U.S.

Amostra		Rs220073 3 ^a	Rs10033464 ^{a,b}		P	Sexo P
(N casos) N controles	Sexo masculino %	Idade ±SD	OR (95% CI)	OR (95% CI)		
Islândia ^c						
Diagn. ≤60 (510/17.714)	77,8	50,7 ±8,4	2,12 (1,77-2,54)	1,69 (1,34-2,12)	6,3x10 ⁻¹⁸	0,82
Diagn. 60-70 (654/17.714)	66,2	65,6 ±2,9	1,88 (1,60-2,21)	1,44 (1,18-1,77)	6,7x10 ⁻¹⁵	0,58
Diagn. 70-80 (958/17.714)	58,9	75,0 ±2,8	1,60 (1,39-1,84)	1,23 (1,03-1,47)	7,5x10 ⁻¹¹	0,96
Diagn. >80 (679/17.714)	47,4	85,6 ±4,2	1,20 (1,01-1,43)	1,31 (1,08-1,60)	0,0044	0,36
U.S.						
AF sozinho (251/804)	81,7	46,1 ±11,5	2,32 (1,80-2,99)	1,68 (1,19-2,37)	1,2x10 ⁻¹⁰	0,46
AF/HTN (67/804)	74,6	54,5 ±10,2	2,23 (1,43-3,48)	1,66 (0,90-3,04)	0,0010	0,54
Outro AF (318/804)	52,8	75,2 ±1,3	1,44 (1,12-1,64)	0,97 (0,69-1,37)	0,015	0,85

15 Cada linha contém os resultados de uma análise conjunta de duas variantes, rs2200733 T e rs10033464 T^a. Os números de casos e controles (N), a

porcentagem de casos masculinos e a idade média ($\pm$ SD) para os casos, são mostrados para cada estudo de caso-controle. A OR, com um 95% CI e valores P são mostrados para cada variante. Além disso, um Valor p conjunto para o efeito combinado das duas variantes e um Valor p conjunto para testar se existe uma diferença da frequência alélica das variantes entre os sexos dentro de cada subgrupo de pacientes.

^a Resultados de comparação de rs2200733 T e rs10033464 T com os haplótipos rs2200733 C, rs13143308 G, rs10033464 G do tipo selvagem.

Tabela 9. Equivalente de SNPs para rs10033464, rs13143308 e rs2200733 nos dados CEU HapMap

SNP	Rotulação SNP	Local constr. 35	Local SEQ ID No. 50
rs12503217	rs10033464	112063765	108955
rs12510087	rs10033464	112068632	111822
rs6852357	rs10033464	112071939	117129
rs4400058	rs10033464	112074277	119467
rs10033464	rs10033464	112078365	123555
rs2171592	rs10033464	112078392	123582
rs2350539	rs10033464	112078814	124004
rs1906606	rs10033464	112080996	126186
rs723364	rs10033464	112082075	127265
rs2220429	rs10033464	112085089	130279
rs4032976	rs10033464	112086371	131561
rs3853440	rs10033464	112087213	132403
rs3853441	rs10033464	112087344	132534
rs3853442	rs10033464	112087632	132822
rs3853443	rs10033464	112087733	132923
rs4124158	rs10033464	112087798	132988
rs4124159	rs10033464	112087847	133037
rs12506083	rs10033464	112088016	133208
rs4032975	rs10033464	112089842	135032
rs4032974	rs10033464	112090140	135330
rs2634074	rs13143308	112034645	79835
rs2466455	rs13143308	112043219	88409
rs2723334	rs13143308	112046356	91546
rs1906616	rs13143308	112055172	100362
rs1906615	rs13143308	112059402	104592
rs2129983	rs13143308	112061684	106874
rs2129982	rs13143308	112061747	106937
rs1906599	rs13143308	112070290	115480
rs13143308	rs13143308	112072023	117213
rs6843082	rs13143308	112075671	120861
rs17042059	rs2200733	111998790	43980
rs4529121	rs2200733	112003159	48349
rs4543199	rs2200733	112005744	50934
rs12647316	rs2200733	112006855	52045
rs10019689	rs2200733	112007473	52663
rs4626276	rs2200733	112007593	52783
rs17042076	rs2200733	112009942	55132
rs11098089	rs2200733	112011830	57020
rs17042088	rs2200733	112012418	57608
rs11930528	rs2200733	112017798	62988

SNP	Rotulação SNP	Local constr. 35	Local SEQ. ID. No. 50
rs17042098	rs2200733	112021762	66952
rs17042102	rs2200733	112026230	71420
rs17042121	rs2200733	112034705	79895
rs10516563	rs2200733	112035326	80516
rs4605724	rs2200733	112042685	87875
rs2350269	rs2200733	112044728	89918
rs6533527	rs2200733	112045118	90308
rs17042144	rs2200733	112047270	92460
rs1906618	rs2200733	112053026	98216
rs1906617	rs2200733	112053418	98608
rs12646447	rs2200733	112056930	102120
rs12646754	rs2200733	112061176	106366
rs2129981	rs2200733	112061803	106993
rs12639654	rs2200733	112062899	108089
rs6817105	rs2200733	112063372	108562
rs17042171	rs2200733	112065891	111081
rs1906591	rs2200733	112066493	111683
rs1906592	rs2200733	112066608	111798
rs2200732	rs2200733	112067646	112836
rs2200733	rs2200733	112067773	112963
rs4611994	rs2200733	112068645	113835
rs4540107	rs2200733	112068706	113896
rs1906593	rs2200733	112069526	114716
rs1906596	rs2200733	112069840	115030
rs2220427	rs2200733	112072493	117683

Tabela 10. Estrutura de haplótipo (haplótipos com frequência estimada > 0,1%) em relação aos SNPs chave e o microssatélite D4S406 na Islândia

Frequência	D4S406	Rs2200733	rs13143308	rs10033464
0,0800	-8	T	T	G
0,00647	-6	C	T	T
0,00225	-4	T	T	G
0,0415	-2	T	T	G
0,00108	0	T	T	G
0,0592	0	C	T	T
0,00679	2	C	T	T
0,0169	2	C	G	G
0,00923	4	C	T	T
0,135	4	C	G	G
0,0853	6	C	G	G
0,1587	8	C	G	G
0,163	10	C	G	G
0,0928	12	C	G	G
0,0398	14	C	G	G
0,101	16	C	G	G

Tabela 11. Associação a todos os SNPs Illumina Hap300 em uma região de 200 kb em torno de rs2200733 e rs10033464 em um conjunto prolongado de casos AF/AFI islândicos e controles. Os resultados não foram ajustados quanto a relacionabilidade de indivíduos.

SNP	Local	All.	Freq.	OR	Valor P	Ajuste para rs220427		Ajuste também para rs10033464	
						OR	Valor P	OR	Valor P
rs4834295	111892810	G	0,817	1,0	0,27	1,0	0,39	1,03	0,63
rs2278782	111899758	C	0,883	1,0	0,79	0,9	0,70	0,99	0,93
rs2595110	111902927	T	0,637	1,0	0,13	1,0	1,0	1,01	0,89
rs976568	111908325	A	0,743	1,0	0,83	0,62	0,62	0,97	0,58
rs2197815	111924481	T	0,030	1,1	0,34	0,34	0,34	0,97	0,84
rs2723286	111940938	A	0,231	1,0	0,26	0,50	0,50	1,03	0,59
rs2723296	111962087	G	0,229	1,0	0,38	0,60	0,60	1,03	0,67
rs1699716	111986643	T	0,153	1,3	4,7x10 ⁻⁵	0,59	0,59	0,95	0,53
rs2723316	111991891	T	0,297	1,2	1,9x10 ⁻⁵	0,59	0,59	0,95	0,40
rs6419178	111993104	A	0,143	1,1	0,17	0,25	0,25	0,98	0,77
rs1448817	111998657	G	0,252	1,4	4,2x10 ⁻⁷	0,035	0,035	1,06	0,46
rs2634073	112023387	A	0,167	1,6	2,4x10 ⁻⁷	0,039	0,039	0,90	0,48
rs2200733	112067773	T	0,119	1,7	7,6x10 ⁻⁷	-	-	-	-
rs2220427	112072493	T	0,120	1,7	5,6x10 ⁻⁷	-	-	-	-
rs1310587	112075751	C	0,888	1,0	0,33	0,89	0,89	0,95	0,56
rs1003346	112078365	T	0,082	1,2	0,013	5,1x10 ⁻⁴	5,1x10 ⁻⁴	-	-
rs1314119	112086218	A	0,368	1,3	2,0x10 ⁻⁷	0,0067	0,0067	1,08	0,29
rs3853444	112091740	A	0,604	1,1	0,053	0,45	0,45	1,06	0,24

5 **Tabela 12.** Estudo de associação de SNPs que são equivalentes a rs2200733 em amostras CEU HapMap na amostras Chinesas de Hong Kong.

SNP	Local	All.	Freq.	OR	Valor P	HapMap D'	HapMap R ²
rs11930528	112017798	T	0,472	1,27	0,011	0,91	0,66
rs17042121	112034705	G	0,418	1,32	0,0029	0,97	0,64
rs6533527	112045118	A	0,518	1,37	0,0014	0,95	0,79
rs1906617	112053418	C	0,524	1,35	0,0026	1,00	0,98
rs12639654	112062899	T	0,519	1,39	0,0012	1,00	1,00
rs2200733	112067773	T	0,516	1,42	6,4x10 ⁻⁴	-	-
rs4611994	112068645	C	0,518	1,39	0,0012	1,00	1,00

Os valores LD relatados são para rs2200733 nas amostras de HapMap CHB e JPT combinadas

Tabela 13. Associação para AF/AFI pelo genótipo

	RR alélico		Genótipo RR						Valor P
	1	2	00	01	02	11	12	22	
Islândia	1,68	1,38	1	1,55	1,36	3,42	2,47	1,58	0,12
Suécia	2,01	1,65	1	1,66	1,72	5,86	3,10	2,04	0,68
U.S.	1,84	1,30	1	1,63	1,40	4,86	2,31	0,90	0,25
Combinado	1,71	1,38	1	1,56	1,37	3,64	2,44	1,43	0,018
Hong Kong	1,42	1,07	1	1,15	0,95	1,77	1,34	0,97	0,87

10 Os três haplótipos possíveis são codificados como

0 = rs2200733 C, rs13143308 G, rs10033464 G

1 = rs2200733 T, rs13143308 T, rs10033464 G

2 = rs2200733 C, rs13143308 T, rs10033464 T

Tabela 14. Associação de vários fenótipos, considerados fatores de risco para AF para variantes de risco.

Fenótipo (N casos/N controles)	T rs2200733		T rs10033464	
	OR	Valor P	OR	Valor P
Hipertensão (2.620/19.862)	1,08	0,11	1,05	0,37
Infarto do miocárdio (3.576/19.542)	1,05	0,26	1,04	0,49
Obesidade – BMI >35 (1.601/21.593)	0,96	0,51	1,00	1,00

Tabela 15. Um resumo da fonte dos controles islândicos. Observe que os indivíduos podem vir de múltiplos projetos e que alguns indivíduos podem ter sido coletados como parentes de *probands*.

Projeto Fonte	Contagem	Frequência de T rs2200733	Frequência de T rs10033464
Controles de Descoberta			
Dependência	376	0,096	0,082
Ansiedade	337	0,110	0,088
Câncer de bexiga	876	0,116	0,085
Câncer de Cólon	370	0,119	0,070
Doença Infecciosa	297	0,109	0,096
MI	454	0,104	0,076
Controles de População	389	0,099	0,077
Câncer da Próstata	713	0,123	0,081
Esquizofrenia	291	0,110	0,091
Diabete Tipo II	551	0,102	0,078
Controles de Replicação			
Câncer de Mama	228	0,122	0,074
Diabete Tipo II	340	0,097	0,082
Alzheimer	459	0,107	0,061
Osteoartrite	1,175	0,107	0,081
PAD	479	0,096	0,083
Doença pulmonar obstrutiva crônica	326	0,125	0,082
Acidente vascular cerebral	414	0,092	0,069
Osteoporose	1,155	0,109	0,072
MI	390	0,112	0,075
Hipertensão	210	0,118	0,101
Depressão	152	0,128	0,061
Asma	538	0,106	0,076
Parkinson	173	0,102	0,058
Controles de População	305	0,105	0,097
Espondilite Ancilosante	155	0,095	0,077
Apnéia do Sono	422	0,118	0,074
AMD	442	0,101	0,067
Artrite Reumatóide	430	0,100	0,094
Câncer Pulmonar	237	0,106	0,084
FCH	265	0,112	0,057
Longevidade	392	0,09	0,077
Hiperplasia Prostática Benigna	245	0,101	0,058
Pré-eclampsia	262	0,129	0,083

Enurese	249	0,104	0,087
Enxaqueca	590	0,112	0,085
Miopia	353	0,123	0,085
Câncer da Tireóide	104	0,121	0,097
ADHD	123	0,119	0,089
Câncer da Próstata	580	0,117	0,073
Ansiedade	546	0,121	0,096
Obesidade	162	0,081	0,092
Endometriose	258	0,106	0,084
Câncer Renal	174	0,099	0,100
Melanoma	283	0,088	0,089
Dependência	201	0,138	0,098
Psoríase	392	0,136	0,079
IBD	356	0,093	0,102

Tabela 16. Iniciadores usados para triar ESTs de bibliotecas de cDNA

ESTs*	Iniciador avançado	Iniciador reverso
DA725631	AGTGGAGGCTGCCAGACTTC	TGCACCACTCATCACCAACA
DB324364	CCGAGGATGTCTTTAGTCTGCAA	ATCATACAGCAGGAATGCAAACA
AF017091	TGAGATTCCACATCCAACATCTTT	TGGCAAACCTTGATATTGTTCTTG

*Os nomes EST são da NCBI BUILD 35

Tabela 17. SNPs que acabam em TFBS conservados na região associada com AF.

SNP	Local	Filamento	Ancestral	Polím.	TFBS	Início de TF	Fim de TF
rs17042171	112065890	+	C	A/C	NFAT	112065889	112065900
rs12510087	112066631	+	A	A/G	SOX5	112066632	112066641
rs2220427	112072492	+	C	C/T	SOX5	112072483	112072493

5 Filamento indica o filamento no alinhamento de genoma que a mutação aporta.

Polím. é os dois alelos do polimorfismo neste sítio.

Tabela 18. Marcadores no ou próximo ao gene PITX2 em LD com marcadores no bloco LD C04. São mostrados marcadores em ou próximos ao PITX2 (marcador 1) e sua correlação com os marcadores no bloco LD C04 (marcador 2).

Marcador 1	Marcador 2	D'	r ²	Valor P
rs7668322	rs10033464	0.46291	0.133423	0.000953
rs2197815	rs10033464	0.660377	0.300172	2.55E-06
rs6831623	rs2200733	1	0.02834	0.025473
rs2595110	rs2200733	0.699643	0.02896	0.067245

Tabela 19. Marcadores em desequilíbrio de ligação com marcador rs2220427 e marcador rs10033464 pelos valores para r^2 maiores do que 0,1. LD foi calculado com base na amostra da população HapMap CEU.

Marcador 1	âncora	D'	r^2	Valor P	Pós em B35	Pós na SEQ ID NO:50
rs9994891	rs2220427	1	0.128329	0.002914	111149057	
rs11568995	rs2220427	1	0.128329	0.002914	111255189	
rs4698804	rs2220427	1	0.128329	0.002914	111297649	
rs721413	rs2220427	1	0.128329	0.002914	111305212	
rs10488883	rs2220427	1	0.128329	0.002914	111305486	
rs6854883	rs2220427	0.788889	0.510189	4.17E-09	111964919	10109
rs2255793	rs2220427	1	0.245283	9.27E-09	111965457	10647
rs2723298	rs2220427	1	0.274924	8.39E-09	111966089	11279
rs2723300	rs2220427	1	0.236507	1.40E-08	111972512	17702
rs2723307	rs2220427	1	0.176558	2.98E-07	111975800	20990
rs1584429	rs2220427	1	0.245283	9.27E-09	111976151	21341
rs1448799	rs2220427	1	0.245283	9.27E-09	111980386	25576
rs1448798	rs2220427	1	0.245283	9.27E-09	111980789	25979
rs1900827	rs2220427	1	0.246741	8.60E-08	111981343	26533
rs2197814	rs2220427	1	0.240506	1.20E-08	111983098	28288
rs969642	rs2220427	1	0.245283	9.27E-09	111983529	28719
rs2595093	rs2220427	0.830131	0.513828	1.59E-10	111984960	30150
rs2245595	rs2220427	1	0.252078	6.90E-09	111985715	30905
rs2595088	rs2220427	1	0.254302	1.99E-08	111985958	31148
rs981150	rs2220427	1	0.245283	9.27E-09	111986232	31422
rs16997168	rs2220427	0.819277	0.507451	2.11E-10	111986643	31833
rs16997169	rs2220427	1	0.245283	9.27E-09	111986685	31875
rs4527540	rs2220427	1	0.245283	9.27E-09	111986742	31932
rs17042026	rs2220427	0.833488	0.554106	1.23E-11	111989978	35168
rs2723316	rs2220427	1	0.245283	9.27E-09	111991891	37081
rs2595081	rs2220427	0.832621	0.549261	3.07E-11	111992761	37951
rs2595085	rs2220427	1	0.242283	1.16E-08	111994377	39567
rs2723318	rs2220427	1	0.236507	1.40E-08	111994576	39766
rs1448817	rs2220427	1	0.296277	9.75E-10	111998657	43847
rs17042059	rs2220427	1	1	1.62E-20	111998790	43980
rs4529121	rs2220427	1	1	1.43E-20	112003159	48349
rs10032150	rs2220427	1	0.296277	9.75E-10	112004222	49412
rs4543199	rs2220427	1	1	1.43E-20	112005744	50934
rs12647316	rs2220427	1	1	1.43E-20	112006855	52045
rs12647393	rs2220427	1	0.917379	1.57E-15	112006886	52076
rs10019689	rs2220427	1	1	1.43E-20	112007473	52663
rs4626276	rs2220427	1	1	1.43E-20	112007593	52783
rs17042076	rs2220427	1	1	1.62E-20	112009942	55132
rs11098089	rs2220427	1	1	1.62E-20	112011830	57020
rs17042088	rs2220427	1	1	1.62E-20	112012418	57608
rs11944778	rs2220427	0.91509	0.811642	4.20E-12	112014571	59761
rs4307025	rs2220427	1	0.296277	9.75E-10	112015107	60297
rs11930528	rs2220427	1	1	1.42E-19	112017798	62988
rs17042098	rs2220427	1	1	1.43E-20	112021762	66952
rs2634073	rs2220427	1	0.523052	1.42E-12	112023387	68577
rs17042102	rs2220427	1	1	2.07E-16	112026230	71420
rs2634071	rs2220427	1	0.528302	2.16E-13	112026824	72014
rs2634074	rs2220427	1	0.433962	5.12E-12	112034645	79835
rs17042121	rs2220427	1	1	1.43E-20	112034705	79895
rs10516563	rs2220427	1	1	1.43E-20	112035326	80516
rs4605724	rs2220427	1	1	1.43E-20	112042685	87875
rs2466455	rs2220427	1	0.491956	5.72E-12	112043219	88409
rs2350269	rs2220427	1	1	1.42E-19	112044728	89918
rs6533527	rs2220427	1	1	1.43E-20	112045118	90308

Marcador 1	âncora	D'	r2	Valor P	Pós em B35	Pós na SEQ ID NO:50
rs2723334	rs2220427	1	0.433962	5.12E-12	112046356	91546
rs17042144	rs2220427	1	1	1.43E-20	112047270	92460
rs1906618	rs2220427	1	1	2.67E-19	112053026	98216
rs1906617	rs2220427	1	1	1.43E-20	112053418	98608
rs6847935	rs2220427	1	0.921053	1.10E-18	112054255	99445
rs1906616	rs2220427	1	0.433962	5.12E-12	112055172	100362
rs12646447	rs2220427	1	1	1.84E-20	112056930	102120
rs1906615	rs2220427	1	0.433962	5.12E-12	112059402	104592
rs12646754	rs2220427	1	1	2.08E-20	112061176	106366
rs2129983	rs2220427	1	0.428571	6.61E-12	112061684	106874
rs2129982	rs2220427	1	0.433962	5.12E-12	112061747	106937
rs2129981	rs2220427	1	1	1.43E-20	112061803	106993
rs12639654	rs2220427	1	1	1.43E-20	112062899	108089
rs6817105	rs2220427	1	1	1.62E-20	112063372	108562
rs17042171	rs2220427	1	1	1.43E-20	112065891	111081
rs1906591	rs2220427	1	1	1.43E-20	112066493	111683
rs1906592	rs2220427	1	1	1.26E-19	112066608	111798
rs2200732	rs2220427	1	1	1.85E-19	112067646	112836
rs2200733	rs2220427	1	1	1.43E-20	112067773	112963
rs4611994	rs2220427	1	1	1.43E-20	112068645	113835
rs4540107	rs2220427	1	1	1.43E-20	112068706	113896
rs1906593	rs2220427	1	1	1.62E-20	112069526	114716
rs1906596	rs2220427	1	1	2.68E-20	112069840	115030
rs1906599	rs2220427	1	0.433962	5.12E-12	112070290	115480
rs13143308	rs2220427	1	0.438445	5.36E-12	112072023	117213
rs6843082	rs2220427	1	0.433962	5.12E-12	112075671	120861
rs11931959	rs2220427	1	0.249653	7.85E-09	112077289	122479
rs13121924	rs2220427	1	0.156089	9.36E-07	112078423	123613
rs2129979	rs2220427	1	0.256789	5.78E-09	112078601	123791
rs723363	rs2220427	1	0.156089	9.36E-07	112082105	127295
rs7697491	rs2220427	1	0.154058	1.07E-06	112083422	128612
rs13141190	rs2220427	1	0.156089	9.36E-07	112086218	131408
rs6533530	rs2220427	1	0.156089	9.36E-07	112089540	134730
rs6533531	rs2220427	1	0.156089	9.36E-07	112089569	134759
rs3866831	rs2220427	1	0.156089	9.36E-07	112089718	134908
rs3866832	rs2220427	0.857992	0.109603	0.000186	112091304	136494
rs11098083	rs10033464	0.407276	0.12964	0.00205	111855920	
rs11721423	rs10033464	0.365905	0.11129	0.003321	111858873	
rs10005945	rs10033464	0.433962	0.101848	0.004763	111860013	
rs7668322	rs10033464	0.46291	0.133423	0.000953	111906200	
rs2197815	rs10033464	0.660377	0.300172	2.55E-06	111924481	
rs6831623	rs10033464	1	0.511236	3.33E-10	111964677	9867
rs7661383	rs10033464	0.637611	0.21478	0.000011	111979181	24371
rs7667461	rs10033464	0.637611	0.21478	0.000011	111979738	24928
rs1900827	rs10033464	0.735008	0.134278	0.000368	111981343	26533
rs998101	rs10033464	0.635887	0.20914	0.000014	111988219	33409
rs12646859	rs10033464	0.719793	0.271646	0.000191	111992237	37427
rs12498380	rs10033464	0.60232	0.178165	0.000097	111992563	37753
rs7690164	rs10033464	0.496308	0.148782	0.005455	111994069	39259
rs11098090	rs10033464	0.551083	0.169161	0.000104	112014012	59202
rs2634073	rs10033464	0.640189	0.223694	8.35E-06	112023387	68577
rs2634071	rs10033464	0.637611	0.21478	0.000011	112026824	72014
rs2634074	rs10033464	1	0.433962	5.12E-12	112034645	79835
rs2466455	rs10033464	1	0.428256	1.77E-09	112043219	88409
rs2723334	rs10033464	1	0.433962	5.12E-12	112046356	91546
rs1906616	rs10033464	1	0.433962	5.12E-12	112055172	100362
rs1906615	rs10033464	1	0.433962	5.12E-12	112059402	104592

Marcador 1	âncora	D'	r2	Valor P	Pós em. B35	Pós na SEQ ID NO:50
rs2129983	rs10033464	1	0.428571	6.61E-12	112061684	106874
rs2129982	rs10033464	1	0.433962	5.12E-12	112061747	106937
rs12503217	rs10033464	1	1	1.43E-20	112063765	108955
rs12510087	rs10033464	1	1	1.43E-20	112066632	111822
rs1906599	rs10033464	1	0.433962	5.12E-12	112070290	115480
rs6852357	rs10033464	1	1	1.43E-20	112071939	117129
rs13143308	rs10033464	1	0.421583	2.22E-11	112072023	117213
rs4833456	rs10033464	1	0.923858	6.67E-19	112073911	119101
rs4400058	rs10033464	1	1	1.43E-20	112074277	119467
rs6843082	rs10033464	1	0.433962	5.12E-12	112075671	120861
rs2171592	rs10033464	1	1	1.43E-20	112078392	123582
rs13121924	rs10033464	1	0.156089	9.36E-07	112078423	123613
rs2350539	rs10033464	1	1	1.43E-20	112078814	124004
rs1906606	rs10033464	1	1	1.43E-20	112080996	126186
rs723364	rs10033464	1	1	1.43E-20	112082075	127265
rs723363	rs10033464	1	0.156089	9.36E-07	112082105	127295
rs7697491	rs10033464	1	0.154058	1.07E-06	112083422	128612
rs2220429	rs10033464	1	1	1.43E-20	112085089	130279
rs13141190	rs10033464	1	0.156089	9.36E-07	112086218	131408
rs4032976	rs10033464	1	1	1.62E-20	112086371	131561
rs3853440	rs10033464	1	1	1.62E-20	112087213	132403
rs3853441	rs10033464	1	1	1.43E-20	112087344	132534
rs3853442	rs10033464	1	1	1.43E-20	112087632	132822
rs3853443	rs10033464	1	1	1.43E-20	112087733	132923
rs4124158	rs10033464	1	1	1.10E-17	112087798	132988
rs4124159	rs10033464	1	1	1.43E-20	112087847	133037
rs12506083	rs10033464	1	1	1.43E-20	112088016	133206
rs6533530	rs10033464	1	0.156089	9.36E-07	112089540	134730
rs6533531	rs10033464	1	0.156089	9.36E-07	112089569	134759
rs3866831	rs10033464	1	0.156089	9.36E-07	112089718	134908
rs4032975	rs10033464	1	1	7.06E-20	112089842	135032
rs4032974	rs10033464	1	1	1.43E-20	112090140	135330
rs3866832	rs10033464	1	0.148936	1.44E-06	112091304	136494
rs7654080	rs10033464	1	0.151515	0.002495	112585323	

REFERÊNCIAS

1. Go, A. S. *et al.* Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. *Jama* 285, 2370-5 (2001).
2. Miyasaka, Y. *et al.* Secular trends in incidence of atrial fibrillation in Olmsted County, Minnesota, 1980 to 2000, and implications on the projections for future prevalence. *Circulation* 114, 119-25 (2006).
3. Arnar, D. O. *et al.* Familial aggregation of atrial fibrillation in Iceland. *Eur Heart J* 27, 708-12 (2006).
4. Fox, C. S. *et al.* Parental atrial fibrillation as a risk factor for atrial fibrillation in offspring. *Jama* 291, 2851-5 (2004).
5. Ellinor, P. T., Yoerger, D. M., Ruskin, J. N. & MacRae, C. A. Familial aggregation in lone atrial fibrillation. *Hum Genet* 118, 179-84 (2005).

6. Chen, Y. H. *et al.* KCNQ1 gain-of-function mutation in familial atrial fibrillation. *Science* 299, 251-4 (2003).
7. Yang, Y. *et al.* Identification of a KCNE2 gain-of-function mutation in patients with familial atrial fibrillation. *Am J Hum Genet* 75, 899-905 (2004).
- 5 8. Xia, M. *et al.* A Kir2.1 gain-of-function mutation underlies familial atrial fibrillation. *Biochem Biophys Res Commun* 332, 1012-9 (2005).
9. Olson, T. M. *et al.* Kv1.5 channelopathy due to KCNA5 loss-of-function mutation causes human atrial fibrillation. *Hum Mol Genet* 15, 2185-91 (2006).
- 10 10. Hong, K., Bjerregaard, P., Gussak, I. & Brugada, R. Short QT syndrome and atrial fibrillation caused by mutation in KCNH2. *3 Cardiovasc Electrophysiol* 16, 394-6 (2005).
11. Ellinor, P. T. *et al.* Mutations in the long QT gene, KCNQ1, are an uncommon cause of atrial fibrillation. *Heart* 90, 1487-8 (2004).
- 15 12. Ellinor, P. T., Petrov-Kondratov, V. I., Zakharova, E., Nam, E. G. & MacRae, C. A. Potassium channel gene mutations rarely cause atrial fibrillation. *BMC Med Genet* 7, 70 (2006).
13. Franco, D. & Campione, M. The role of Pitx2 during cardiac development. Linking left-right signaling and congenital heart atrial fibrillation and /or strokes. *Trends Cardiovasc Med* 13, 157-63 (2003).
- 20 14. Faucourt, M., Houliston, E., Besnardeau, L., Kimelman, D. & Lepage, T. The pitx2 homeobox protein is required early for endoderm formation and nodal signaling. *Dev Biol* 229, 287-306 (2001).
15. Mommersteeg, M. T. *et al.* Molecular Pathway for the Localized Formation of the Sinoatrial Node. *Circ Res* (2007).
- 25 16. A haplotype map of the human genome. *Nature* 437, 1299-320 (2005).
17. Waldo, A. L. The interrelationship between atrial fibrillation and atrial flutter. *Prog Cardiovasc Dis* 48, 41-56 (2005).
18. Zini, S. *et al.* Identification of metabolic pathways of brain angiotensin II

and III using specific aminopeptidase inhibitors: predominant role of angiotensin III in the control of vasopressin release. *Proc Natl Acad Sci U S A* 93, 11968-73 (1996).

19. Gretarsdottir, S. *et al.* The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat Genet* 35, 131-8 (2003).

20. Falk, C. T. & Rubinstein, P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann Hum Genet* 51 (Pt 3), 227-33 (1987).

21. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of atrial fibrillation and /or stroke. *J Natl Cancer Inst.* 22, 719-48 (1959).

22. Grant, S. F. *et al.* Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 38, 320-3 (2006).

23. Yang, X. *et al.* Development and validation of stroke risk equation for Hong Kong Chinese patients with type 2 diabetes: the Hong Kong Diabetes Registry. *Diabetes Care* 30, 65-70 (2007).

24. Baum, L. *et al.* Methylenetetrahydrofolate reductase gene A222V polymorphism and risk of ischemic stroke. *Clin Chem Lab Med* 42, 1370-6 (2004).

25. Kutuyavin, I. V. *et al.* A novel endonuclease IV post-PCR genotyping system. *Nucleic Acids Research* 34, e128 (2006).

26. Amundadottir, L. T. *et al.* A common variant associated with prostate cancer in European and African populations. *Nat Genet* 38, 652-8 (2006).

27. Gretarsdottir, S. *et al.* The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat Genet* 35, 131-8 (2003).

28. Falk, C. T. & Rubinstein, P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann Hum Genet* 51 (Pt 3), 227-33 (1987).

29. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from

retrospective studies of atrial fibrillation and /or stroke. J Natl Cancer Inst. 22, 719-48 (1959).

30. Grant, S. F. *et al.* Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet 38, 320-3 (2006).

5 31. Devlin, B. & Roeder, K. Genomic Control for association studies. Biometrics 55, 997-1004 (1999).

32. Devlin, B., Bacanu, S.-A. & Roeder, K. Genomic control to the extreme. Nature Genetics 36, 1129-1130 (2004).

33.. Nomenclature and criteria for diagnosis of ischemic heart atrial
10 fibrillation and /or stroke. Report of the Joint International Society and Federation of Cardiology/World Health Organization task force on standardization of clinical nomenclature. Circulation 59, 607-9 (1979).

34. Alpert, J. S., Thygesen, K., Antman, E. & Bassand, J. P. Myocardial infarction redefined--a consensus document of The Joint European Society of
15 Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction. J Am Coll Cardiol 36, 959-69 (2000).

35. Monks, S. A. *et al.* Genetic inheritance of gene expression in human cell lines. Am J Hum Genet 75, 1094-105 (2004).

36. Schadt, E. E. *et al.* Genetics of gene expression surveyed in maize, mouse
20 and man. Nature 422, 297-302 (2003).

Exemplo 3. Associação de variantes do Cromossoma 4 ao Acidente Vascular Cerebral Isquêmico

O acidente vascular cerebral é uma causa comum de morte e a causa condutora da incapacidade do adulto nas sociedades Ocidentais. Está
25 agora também começando um problema de saúde maior em países de baixa renda e renda intermediária devido ao envelhecimento da população e mudanças nos fatores de risco modificáveis para as doenças cardiovasculares¹. O acidente vascular cerebral não é uma doença isolada mas uma síndrome altamente complexa que consiste de um grupo de distúrbios

heterogêneos com muitos fatores de risco genéticos e ambientais²³. Estudos de gêmeos, história familiar e modelos de animal⁴⁻⁸ fornecem evidência para a contribuição genética para as formas comuns de acidente vascular cerebral mas nenhuma variante de risco principal foi ainda identificada mostrando resultados compatíveis através de populações.

Os acidentes vasculares cerebrais isquêmicos (IS), responsáveis pela maioria dos insultos cerebrais (>80%), resultam de trombose ou embolismo que levam à obstrução das artérias cerebrais. Vários mecanismos fisiopatológicos podem causar IS mas aqueles mais comuns são aterosclerose de artéria grande (LAA), acidente vascular cerebral cardioembólico (CES) e doença de vaso pequeno (SVD)⁹.

Métodos

Populações de estudo.

Islândia: Os pacientes com acidente vascular cerebral islandeses foram recrutados de um registro de mais de 4.000 indivíduos diagnosticados com acidente vascular cerebral isquêmico ou TIA no único hospital universitário em Reykjavik, the Landspítali University Hospital, durante os anos de 1993 a 2006. Os pacientes com acidente vascular cerebral foram alistados durante os últimos nove anos através do programa genético de doença cardiovascular (CVD) no deCODE. O diagnóstico de acidente vascular cerebral foi clinicamente confirmado pelos neurologistas (ver abaixo). O grupo de descoberta incluiu 1.661 pacientes e quando da análise dos SNPs em 4q25 nós usamos um conjunto adicional de 282 pacientes (idade média $\pm$ SD: 77,2 $\pm$ 11,3 anos, 45% mulheres). Nós usamos 25.708 controles (idade média $\pm$ SD: 59,2 $\pm$ 21,1 anos, 59% mulheres) de vários programas genéticos sob estudo no Decode, incluindo: aneurisma abdominal (250), fibrilação atrial (1.150), adicção (750), Alzheimer (350), ansiedade (200), asma (1300), COPD (850), câncer de cólon (200), trombose de veia profunda (550), dislexia (200), doenças infecciosas (250), longevidade (400), câncer

pulmonar (750), infartação do miocárdio (2.400), enxaqueca (1.100), doença da artéria periférica (1.200), síndrome do ovário policístico (1.200), pré-eclampsia (700), câncer prostático (400), psoríase (750), artrite reumática (550), síndrome das pernas inquietas (350) e diabetes tipo 2 (400).

5 O estudo foi aprovado pelo Data Protection Commission of Iceland (DPC) e pelo National Bioethics Committee of Iceland. Todos os participantes deram consentimento informado.

Suécia: Os pacientes suecos com acidente vascular cerebral isquêmico atendidos na unidade de acidente vascular cerebral ou na clínica de
10 paciente externo de acidente vascular cerebral no Karolinska University Hospital, unidade Huddinge em Stockholm, Suécia, foram recrutados de 1996 a 2002 como parte de um estudo de epidemiologia genética em andamento, o Estudo de Acidente Vascular Cerebral Isquêmico de Stockholm do Sul (SSISS) (idade média $\pm$ SD: 67,3 $\pm$ 11,8 anos, 44% mulheres). Os controles
15 suecos usados neste estudo são controles com base na população recrutados da mesma região na Suécia central como os pacientes, representando a população geral nesta área. Os indivíduos foram doadores de sangue recrutados no Huddinge ou Karolinska University Hospitals ou voluntários saudáveis (recrutados em 1990-1994) recrutados pelo Clinical Chemistry
20 Department no Karolinska University Hospital para representar uma população de referência normal (idade média $\pm$ SD: 46,8 $\pm$ 15,9 anos para controles do hospital Huddinge, 41% mulheres, informação de idade não disponível para os doadores de sangue recrutados no hospital Karolinska). O estudo foi aprovado pelo Bioethics Committee of Karolinska Institutet.

25 Sul da Alemanha: A população germânica, aqui contida aludida como Alemanha-S, consistiu de pacientes IS consecutivamente recrutados durante o período de 2001 a 2006 na unidade de acidente vascular cerebral do Department of Neurology, Klinikum Grosshadern, University of Munich, Alemanha (idade média $\pm$ SD: 65,3 $\pm$ 13,7 anos, 38% mulheres). O

grupo de controle consistiu de indivíduos igualados em idade e gênero sem uma história de doença cardiovascular (idade média $\pm$ SD: 62,7 $\pm$ 10,9 anos, 38% mulheres). Estes foram selecionados do estudo KORA S4, um projeto epidemiológico com base na comunidade próximo a Munich²³. O estudo foi
5 aprovado pelo comitê de ética local e o consentimento informado foi obtido de todos os indivíduos (ou parentes ou guardiões legais).

Região de Westphalia, Alemanha: A segunda população alemã, aludida como Alemanha-W, recrutou pacientes de acidente vascular cerebral isquêmico através de hospitais que participam no Registro de
10 Acidente Vascular Cerebral Regional de Westphalian, localizado no oeste do país, durante o período de 2000-2003 (idade média $\pm$ SD: 70,4 $\pm$ 12,6 anos, 53% mulheres). Os controles de população sem uma história auto-relatada de acidente vascular cerebral foram tirados da população de seção cruzada, prospectiva, com base no Dortmund Health Study²⁴, conduzido na mesma
15 região e a frequência subsequentemente emparelhada para os casos (idade média $\pm$ SD: 52,3 $\pm$ 13,7 anos, 53% mulheres). Ambos os estudos foram aprovados pelos comitês de ética da University of Munster. Todos os participantes deram seu consentimento informado.

SE-Inglatera, Reino Unido. Pacientes de acidente vascular
20 cerebral isquêmico de descendência europeia que compareceram a um serviço cerebrovascular foram recrutados de 1995 a 2002. Todos os casos foram fenotipado por um neurologista experimentado em acidente vascular cerebral com revisão de formação de imagem original (idade média $\pm$ SD: 64,6 $\pm$ 12,7 anos, 41% mulheres). Os controles da comunidade livres de doença
25 cerebrovascular sintomática foram também recrutados pela amostragem na lista do médico familiar da mesma região geográfica como os pacientes. A amostragem foi estratificada para fornecer uma distribuição similar de idade e gênero como no grupo de paciente (idade média $\pm$ SD: 64,8 $\pm$ 8,6 anos, 41% mulheres). O estudo foi aprovado pelos comitês de ética de pesquisa locais e

o consentimento informado foi obtido de todos os participantes.

Fenotipagem. Apenas pacientes com acidentes vasculares cerebrais isquêmicos mas não com hemorrágicos foram incluídos no estudo. Todos os pacientes tiveram trabalho de diagnóstico clinicamente relevante realizado, incluindo formação de imagem do cérebro com tomografia computadorizada (CT) e/ou formação de imagem pela ressonância magnética (MRI) assim como investigações de diagnóstico subordinadas incluindo ultrassonografia duplex das artérias carótidas e vertebrais, ecocardiografia, monitoramento Holter, MR-angiografia, CT-angiografia e testes de sangue.

10 Pacientes com Ataque Isquêmico Transitório (TIA) clinicamente confirmado foram incluídos no grupo de Acidente Vascular Cerebral Isquêmico da Islândia, Alemanha-S e Suécia. Os pacientes foram classificados em subtipos etiológicos de acordo com o Teste de Org 10172 no Tratamento de Acidente Vascular Cerebral Agudo (TOAST)²⁵. Esta classificação inclui seis categorias:

15 (1) doença oclusiva de artéria grande (doença de vaso grande), (2) cardioembolismo (acidente vascular cerebral cardiogênico), (3) doença de vaso pequeno (acidente vascular cerebral lacunar), (4) outra etiologia determinada, (5) etiologia desconhecidas a despeito dos esforços de diagnóstico ou (6) mais do que uma etiologia. Pacientes classificados nas

20 categorias TOAST 4 a 6 foram excluídos da população de acidente vascular cerebral da Alemanha-W. Na Islândia, os pacientes foram classificados como tendo doença oclusiva de artéria grande se estenose foi 70% que é um critério mais rigoroso do que habitualmente usado isto é, 50%. A classificação de pacientes com acidente vascular cerebral em subtipos de acordo com o Teste

25 de Org 10172 no sistema de classificação do Tratamento de Acidente Vascular Cerebral Agudo (TOAST)²⁵ na descoberta islandesa e os quatro conjuntos de amostra de replicação está listada na Tabela 1.

Genotipagem de genoma amplo Illumina. Todos os casos islandeses e as amostras de controle foram ensaiados com os chips de SNP

Infinium HumanHap300 (Illumina), contendo 317.503 SNPs de rotulação derivados da fase 1 do projeto International HapMap. Dos SNPs ensaiados no chip, 6.622 SNPs foram excluídos porque apresentaram (i) uma taxa de chamada mais baixa do que 95% em casos ou controles, (ii) frequência de alelo menor do que 1% na população ou (iii) distorção significativa do equilíbrio de Hardy-Weinberg nos controles ($P < 1 \times 10^{-10}$). Qualquer amostra com rendimento <98% foi excluída da análise. Na análise final 310.881 SNPs foram usados.

Genotipagem de SNP Único. A genotipagem de SNP único para todos os 121 SNP foi realizada na deCODE genetics em Reykjavik, Islândia usando a plataforma Centaurus (Nanogen)²⁶. A qualidade de cada ensaio de SNP foi avaliada pela comparação da genotipagem das amostras CEU HapMap com os dados de HapMap publicamente disponíveis. Todos os SNPs passaram no teste de desemparelhamento, testes de desequilíbrio de ligação (LD) e foram em equilíbrio de Hardy-Weinberg.

Análise de Associação. Para a análise de associação uma estatística de razão de probabilidade padrão foi usada, como implementado no software NEMO criado na deCODE²⁷, para calcular valores P de lado duplo e Razão de Chances (OR) para cada alelo individual, assumindo um modelo multiplicativo para o risco, isto é, qual o risco dos dois alelos uma pessoa carrega multiplica-se. As frequências alélicas, ao invés das frequências de portador são apresentadas para os marcadores.

No local no cromossoma 4q25, nós analisamos 3 SNPs, rs2200733, rs10033464 e rs13143308. O terceiro SNP, rs13143308, está em LD alto tanto com rs2200733 quanto com rs10033464 ($D' = 0,99$ para ambos) e tem um alelo menor que corresponde completamente aos cromossomas que carregam rs2200733 alelo T ou rs10033464 alelo T. Este foi genotipado em todas as populações usando um ensaio Centaurus e foi usado para inferir genótipos para aqueles indivíduos que tiveram dados perdidos para rs2200733

ou rs10033464 na plataforma Illumina Infinium. Na Tabela 21 e Tabela Suplementar 22, os valores P e OR para ambos os alelos de risco rs2200733-T e rs10033464-T foram computados com base na comparação com os haplótipos rs2200733 alelo C, rs13143308 alelo G, rs10033464 alelo G do tipo selvagem, que não contém nenhum dos alelos em risco.

Para os grupos de estudo islandeses, os valores P são dados depois do ajuste quanto a relacionabilidade dos pacientes e outra estratificação da população possível usando o método de controle genômico¹⁰. Os fatores de inflação para as estatísticas de qui quadrado são estimados ser 1,07, 1,04, 1,06 e 1,02 para a análise de associação de genoma amplo dos grupos de paciente IS, CES, LAA ou SVD respectivamente. Com os casos adicionais e controles tipados quanto ao local 4q, nós estimamos os fatores de inflação usando simulações como anteriormente descrito²⁸. Os fatores de inflação resultantes são 1,09, 1,03, 1,06, 1,05, 1,01, 1,00, 1,01 e 1,00, para os grupos IS, CES, IS excl CES, LVD, SVD, outros, desconhecidos e mais do que uma causa, respectivamente.

Devido ao número grande de controles usados, o tamanho de amostras eficaz depois de ajustar quanto a relacionabilidade dos casos e controles corresponde a testar 2.690 pacientes IS e 2.690 controles. Os tamanhos de amostra eficazes correspondentes para os pacientes CES, LAA e SVD são 710, 417 e 467, respectivamente.

Os resultados de grupos de caso-controle múltiplos foram combinados usando um modelo de Mantel-Haenszel em que os grupos foram deixados ter frequências de população diferentes para os alelos, haplótipos e genótipos mas foram assumidos ter um Risco Relativo comum²⁹

Resultados

A associação de variantes dentro da região do bloco LD C04 para o Acidente vascular cerebral isquêmico foi investigada. De modo a investigar ainda mais a contribuição das duas variantes de risco de AF no

4q25, rs2200733 e rs10033464, para o risco de desenvolver Acidente vascular cerebral isquêmico e seus subtipos, aterosclerose de artéria grande (LAA), acidente vascular cerebral cardioembólico (CES) e doença de vaso pequeno (SVD), nós genotipamos o marcador rs2200733 e o marcador rs10033464 em amostras islandesas e para propósitos de replicação nós também analisamos conjuntos de dados de replicação em grupos da Alemanha do Sul (1.181 casos e 1.189 controles, Alemanha-S), Suécia (1.032 casos e 1.387 controles), região de Westphalia na Alemanha (1.388 casos e 1.106 controles, Alemanha-W) e Reino Unido (654 casos/676 controles, UK). A classificação de fenótipo dos grupos de estudo é mostrada na Tabela 20.

Tabela 20. TOAST subclassificação de casos de acidente vascular cerebral genotipados, n (%)

	Grupo de descoberta	Grupos de replicação			
	Islândia	Alemanha-S	Suécia	Alemanha-W	Reino Unido
Derrame isquêmico	1943	1183	1066	1391	654
Subtipagem TOAST	1443	1183	1061	1389	654
Cardioembolismo	385 (45)	297 (38)	185 (37)	554 (40)	78 (18)
Aterosclerose de artéria grande	229 (27)	372 (47)	230 (46)	560 (40)	232 (55)
Doença de vasos pequenos	246 (29)	118 (15)	82 (16)	275 (20)	114 (27)
Outra causa	42	67	56	não recrutado	3
Mia do que uma causa	34		42	não recrutado	40
Causa desconhecida	507	329	466	não recrutado	187

TOAST = Teste de Org 10172 no Tratamento do Acidente vascular cerebral Agudo.

Pacientes (282) e controles (14.893) adicionais da Islândia também foram genotipados quanto a estes SNPs particulares. O teste de associação foi feito pela comparação de cada SNP com o haplótipo do tipo selvagem (ver Métodos). Como mostrado na Tabela 21, rs2200733 conferiu um risco aumentado de Acidente vascular cerebral isquêmico em todos os conjuntos de amostra e a associação com Acidente vascular cerebral isquêmico foi altamente significativa com um OR combinado = 1,26 ($P = 8,8 \times 10^{-11}$). Para rs10033464, a associação com Acidente vascular cerebral

isquêmico não foi significativa (OR = 1,03, P = 0,45). Ambos SNPs entretanto, associaram significativamente com o Acidente vascular cerebral cardioembólico e este risco foi significativamente maior do que no grupo do acidente vascular cerebral isquêmico como um todo (rs2200733: OR = 1,53, P = $1,5 \times 10^{-12}$; rs10033464: OR = 1,27, P = $5,9 \times 10^{-4}$). Isto é como esperado dada a contribuição conhecida da Fibrilação atrial para este subfenótipo. Removendo-se os pacientes com acidente vascular cerebral cardioembólico do grupo de acidente vascular cerebral isquêmico, o efeito observado para ambos os SNPs foi mais fraco nos pacientes com acidente vascular cerebral isquêmico remanescentes, mas permaneceu significativa para a variante mais forte (rs2200733: OR = 1,18, P = $1,5 \times 10^{-5}$, rs10033464: OR = 0,96, P = 0,39). À parte do acidente vascular cerebral cardioembólico, a aterosclerose de artéria grande e o acidente vascular cerebral de causa não determinada foram os únicos subfenótipos que mostraram associação significativa com rs2200733 (OR = 1,22, P = $1,5 \times 10^{-3}$, Tabela 2 e OR = 1,18, P = 0,01). Estes resultados sugerem que uma porção significativa de acidentes vasculares cerebrais classificados como acidente vascular cerebral criptogênico ou aterosclerose de artéria grande pode ser devido a AF não diagnosticada, intermitente.

Acidente vascular cerebral isquémico excl									
Cardioembolismo									
Islândia (25708/1558)	0,119	0,136	1,17 (1,05-1,31)	0,01	0,082	0,081	1,00 (0,87-1,14)	0,95	
Alemanha-S (1186/886)	0,118	0,125	1,06 (0,87-1,28)	0,57	0,093	0,078	0,83 (0,67-1,04)	0,11	
Alemanha-W (1107/837)	0,114	0,136	1,22 (1,01-1,48)	0,04	0,092	0,091	1,02 (0,82-1,28)	0,84	
Suécia (740/881)	0,098	0,115	1,19 (0,95-1,50)	0,13	0,113	0,106	0,95 (0,76-1,19)	0,66	
Reino Unido (676/576)	0,087	0,123	1,48 (1,14-1,91)	0,003	0,090	0,083	0,96 (0,73-1,28)	0,80	
Todos os grupos (2941/4738)	0,107	0,127	1,18 (1,10-1,28)	1,5x10 ⁻⁵	0,094	0,088	0,96 (0,88-1,05)	0,39	
Aterosclerose de artéria grande									
Islândia (25708 (229)	0,119	0,157	1,41 (1,08-1,86)	0,112	0,082	0,096	1,25 (0,89-1,74)	0,19	
Alemanha-S (1186/372)	0,118	0,117	0,96 (0,75-1,25)	0,78	0,093	0,071	0,74 (0,54-1,00)	0,05	
Alemanha-W (1107/560)	0,114	0,140	1,28 (1,03-1,59)	0,03	0,092	0,100	1,14 (0,89-1,46)	0,30	
Suécia (740/230)	0,098	0,094	0,94 (0,65-1,34)	0,72	0,113	0,096	0,82 (0,58-1,17)	0,27	
Reino Unido (676/232)	0,087	0,138	1,66 (1,19-2,31)	3,0x10 ⁻³	0,090	0,071	0,82 (0,55-1,23)	0,34	
Todos os grupos (2941/1623)	0,107	0,129	1,22 (1,08-1,38)	1,5x10 ⁻³	0,094	0,087	0,96 (0,83-1,11)	0,57	
Doença de vaso pequeno									
Islândia (25708/246)	0,119	0,112	0,94 (0,71-1,24)	0,64	0,082	0,085	1,03 (0,75-1,42)	0,86	
Alemanha-S (1186/118)	0,118	0,145	1,23 (0,83-1,83)	0,30	0,093	0,063	0,68 (0,40-1,14)	0,14	
Alemanha-W (1107/275)	0,114	0,126	1,10 (0,83-1,47)	0,51	0,092	0,075	0,81 (0,57-1,14)	0,22	
Suécia (740/82)	0,098	0,110	1,11 (0,66-1,88)	0,70	0,113	0,091	0,80 (0,46-1,37)	0,42	
Reino Unido (676/114)	0,087	0,101	1,18 (0,73-1,91)	0,50	0,090	0,087	0,99 (0,60-1,63)	0,97	
Todos os grupos (2941/835)	0,107	0,119	1,07 (0,91-1,26)	0,39	0,094	0,080	0,88 (0,73-1,05)	0,16	

Tabela 22. Resultados de Associação para rs2200733 alelo T e rs10033464 alelo T para os subfenótipos TOAST; outras causas, mais do que uma causa e causa desconhecida em três ou quatro populações de estudo. Os resultados para cada fenótipo também estão incluídos depois de combinar as populações de estudo usando um modelo de Mantel-Haenszel (Todos os grupos). Número de controles (m) e casos (n) é mostrado em parênteses, as frequências alélicas em cada grupo, a OR com um 95% CI e Valor p de lado duplo para comparação com o haplótipo do tipo selvagem (ver Métodos Suplementares). Os resultados para a população islandesa são ajustados quanto a relacionabilidade dos indivíduos.

Fenótipo	rs2200733-T					rs10033464-T				
	População de estudo (m/n)		frequência		P	frequência		OR (95% CI)	P	OR (95% CI)
	Controles	Casos	Controles	Casos		Controles	Casos			
Outra causa										
Islândia (25708/42)	0,119	0,155			0,37	0,082	0,060	0,73 (0,31-1,75)	0,48	
Alemanha-S (1186/67)	0,118	0,119			0,91	0,093	0,105	1,14 (0,64-2,04)	0,66	
Suécia (740/56)	0,098	0,125			0,32	0,113	0,134	1,26 (0,70-2,26)	0,44	
Todos os grupos (27634/168)	0,111	0,133			0,32	0,096	0,099	1,06 (0,74-1,54)	0,74	
Mais do que uma causa										
Islândia (25708/34)	0,119	0,088			0,35	0,082	0,044	0,49 (0,17-1,39)	0,18	
Suécia (740/42)	0,098	0,112			0,52	0,113	0,187	1,84 (0,99-3,41)	0,05	
Reino Unido (676/40)	0,087	0,213			8,9x10 ⁻⁴	0,090	0,088	1,15 (0,51-2,61)	0,74	
Todos os grupos (27124/116)	0,101	0,138			0,06	0,095	0,106	1,21 (0,78-1,88)	0,1	
Causa desconhecida										
Islândia (25708/507)	0,119	0,135			0,15	0,082	0,073	0,89 (0,70-1,13)	0,35	
Alemanha-S (1186/329)	0,098	0,129			0,46	0,093	0,087	0,94 (0,69-1,28)	0,70	
Suécia (740/466)	0,087	0,126			0,04	0,113	0,104	0,94 (0,72-1,23)	0,65	
Reino Unido (760/187)		0,102			0,35	0,090	0,096	1,10 (0,74-1,64)	0,63	
Todos os grupos (28310/1489)	0,101	0,123			0,01	0,095	0,090	0,94 (0,82-1,08)	0,41	

Como debatido no acima (Exemplo 2), os alelos de risco de rs2200733 e rs10033464 correlacionam-se significativamente com a idade de diagnóstico de Fibrilação atrial. Uma tendência não significativa na mesma direção foi observada em nosso estudo para a idade no diagnóstico de acidente vascular cerebral cardioembólico (0,62 anos por cópia de T rs2200733, $P = 0,33$ e 0,29 anos por cópia de T rs10033464, $P = 0,71$, Tabela 23), sugerindo que o efeito da idade observada sobre a AF pode se aplicar também ao acidente vascular cerebral cardioembólico, embora seja um efeito mais fraco.

Tabela 23. Regressão linear da idade no diagnóstico no número de alelos de risco de rs2200733 alelo T e rs10033464 alelo T. São mostrados os coeficientes de regressão e os valores P de lado duplo correspondentes obtidos usando a idade nos diagnósticos como uma resposta (em anos) e o número de alelos em risco como variáveis prognosticadoras. O sexo foi incluído como um fator covariado em todos os testes e também a população no teste para todos os grupos combinados. Os números de casos usados na análise são mostrados em parênteses (n).

	rs2200733-t	rs2200733-t	rs10033464-t	rs10033464-t
	reg. coef.	P	reg.coef.	P
Isquêmico				
Islândia (1830)	0,11	0,85	-0,35	0,62
Alemanha-S (114)	0,29	0,73	0,85	0,43
Suécia (780)	0,56	0,56	1,11	0,23
Alemanha-W (1352)	0,17	0,80	1,21	0,14
Reino Unido (654)	0,24	0,83	0,47	0,71
Todos os grupos (5790)	0,40	0,25	0,68	0,10
Cardioembólico	reg. coef.	P	reg.coef.	P
Islândia (356)	-1,52	0,16	-0,29	0,85
Alemanha-S (296)	-1,72	0,21	-2,53	0,18
Suécia (173)	-2,09	0,20	0,81	0,62
Alemanha-W (1352)	-0,04	1,00	0,82	0,53
Reino Unido (78)	7,11	0,084	-5,29	0,16
Todos os grupos (1441)	-0,62	0,33	-0,29	0,71

Debate

Através deste estudo nos 1.661 pacientes IS islandeses e 10.815 controles e o seguimento da replicação nos conjuntos de caso/amostra de controle de acidente vascular cerebral isquêmico europeu grande e bem caracterizado nós identificamos e validamos uma variante de risco no

cromossoma 4q25, rotulado por rs2200733, que associa com acidente vascular cerebral isquêmico. Em nosso estudo, como esperado, estas variantes associaram-se mais fortemente com o subfenótipo de acidente vascular cerebral cardioembólico, que é uma complicação principal da Fibrilação atrial. O risco que é observado em pacientes de acidente vascular cerebral isquêmico sem acidente vascular cerebral cardioembólico é possivelmente devido a um sub-diagnóstico de Fibrilação atrial e deste modo acidente vascular cerebral cardioembólico, visto que a fibrilação atrial é frequentemente assintomática ou intermitente e pode ser consequentemente difícil para detectar em pacientes com acidente vascular cerebral.

Até 30% do acidente vascular cerebral isquêmico são causados pelo cardioembolismo (5, 6) dos quais uma proporção grande ocorre na presença de fibrilação atrial (7, 8). A fibrilação atrial é a arritmia cardíaca prolongada mais comum do ser humano e a sua prevalência aumenta com a idade, afetando aproximadamente 10% daqueles acima dos 80 anos de idade (3, 9). Como tal, AF é um dos fatores de risco independentes mais poderosos para o acidente vascular cerebral e em um nível de população, a AF está associada com um aumento de quatro a cinco vezes no risco de acidente vascular cerebral (3, 7, 8, 10). Além disso, o acidente vascular cerebral cardioembólico é em uma maneira geral grave, refletido pela incapacidade maior, taxas mais altas de recorrência de acidente vascular cerebral e mortalidade mais alta do que em outros subtipos de acidentes vasculares cerebrais (6, 11). A detecção precoce daqueles em risco para AF é importante de modo a reduzir o risco de sofrer um acidente vascular cerebral futuro. Os testes clínicos sobre a prevenção do acidente vascular cerebral em pacientes com AF têm mostrado que as medicações anticoagulantes (por exemplo varfarina) reduzem o risco de acidente vascular cerebral substancialmente (7, 12) e é muito mais eficaz do que os agentes anti-plaqueta tais com aspirina e clopidogrel. Nossos resultados fortemente sugerem que uma porção

significante de pacientes com acidente vascular cerebral tem fibrilação atrial não diagnosticada e são classificados como acidente vascular cerebral criptogênico ou como acidente vascular cerebral de vaso grande. Tais pacientes podem ter AF assintomática, intermitente que não é detectada durante o trabalho de rotina de 24 a 48 horas de monitoramento cardíaco. Isto é sustentado por dois estudos de pacientes pós-acidente vascular cerebral que passaram por mais 4 a 7 dias de monitoramento cardíaco em ambulatório; as taxas de AF intermitente anteriormente não diagnosticadas foram de 5,6 e 14,3% (13, 14). Os pacientes de acidente vascular cerebral com AF assintomática ou intermitente seriam inadequadamente tratados se mal diagnosticados ao invés como por exemplo de acidente vascular cerebral criptogênico ou acidente vascular cerebral de vaso grande visto que tais pacientes são colocados em um agente anti-plaqueta ao invés de varfarina. Portanto, estes marcadores para AF pode ajudar a determinar qual paciente beneficiar-se-ia de monitoramento cardíaco prolongado como um paciente externo para documentar a presença ou ausência de AF. Estudos prospectivos são necessários para determinar se estas descobertas podem ser traduzidas em melhor prevenção ou tratamento para o acidente vascular cerebral.

Referências

1. Strong, K., Mathers, C. & Bonita, R. Preventing stroke: saving lives around the world. *Lancet Neurol* 6, 182-7 (2007).
2. Hassan, A. & Markus, H. S. Genetics and ischaemic stroke. *Brain* 123 (Pt 9), 1784-812 (2000).
3. Markus, H. Genes for stroke. *J Neurol Neurosurg Psychiatry* 75, 1229-31 (2004).
4. Flossmann, E., Schulz, U. G. & Rothwell, P. M. Systematic review of methods and results of studies of the genetic epidemiology of ischemic stroke. *Stroke* 35, 212-27 (2004).
5. Brass, L. M., Isaacsohn, J. L., Merikangas, K. R. & Robinette, C. D. A

study of twins and stroke. *Stroke* 23, 221-3 (1992).

6. Jerrard-Dunne, P., Cloud, G., Hassan, A. & Markus, H. S. Evaluating the genetic component of ischemic stroke subtypes: a family history study. *Stroke* 34, 1364-9 (2003).

5 7. Jousilahti, P., Rastenyte, D., Tuomilehto, J., Sarti, C. & Vartiainen, E. Parental history of cardiovascular disease and risk of stroke. A prospective follow-up of 14371 middle-aged men and women in Finland. *Stroke* 28, 1361-6 (1997).

8. Rubattu, S. *et al.* Chromosomal mapping of quantitative trait loci contributing to stroke in a rat model of complex human disease. *Nat Genet* 13, 429-34 (1996).

9. Dichgans, M. Genetics of ischaemic stroke. *Lancet Neurol* 6, 149-61 (2007).

10. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* 55, 997-1004 (1999).

11. Gudbjartsson, D. F. *et al.* Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature* 448, 353-7 (2007).

12. Ferro, J. M. Cardioembolic stroke: an update. *Lancet Neurol* 2, 177-88 (2003).

20 13. Murtagh, B. & Smalling, R. W. Cardioembolic stroke. *Curr. Atheroscler Rep* 8, 310-6 (2006).

14. Lip, G. Y. & Lim, H. S. Atrial fibrillation and stroke prevention. *Lancet Neurol* 6, 981-93 (2007).

25 15. Wolf, P. A. & Singer, D. E. Preventing stroke in atrial fibrillation. *Am Fam Physician* 56, 2242-50 (1997).

16. Feinberg, W. M., Blackshear, J. L., Laupacis, A., Kronmal, R. & Hart, R. G. Prevalence, age distribution, and gender of patients with atrial fibrillation. Analysis and implications. *Arch Intern Med* 155, 469-73 (1995).

17. Go, A. S. *et al.* Prevalence of diagnosed atrial fibrillation in adults:

- national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. *Jama* 285, 2370-5 (2001).
18. Wolf, P. A., Abbott, R. D. & Kannel, W. B. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke* 22, 983-8 (1991).
19. Lip, G. Y. & Boos, C. J. Antithrombotic treatment in atrial fibrillation. *Heart* 92, 155-61 (2006).
20. Hart, R. G., Pearce, L. A. & Aguilar, M. I. Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. *Ann Intern Med* 146, 857-67 (2007).
21. Barthelemy, J. C. *et al.* Automatic cardiac event recorders reveal paroxysmal atrial fibrillation after unexplained strokes or transient ischemic attacks. *Ann Noninvasive Electrocardiol* 8, 194-9 (2003).
22. Jabaudon, D., Sztajzel, J., Sievert, K., Landis, T. & Sztajzel, R. Usefulness of ambulatory 7-day ECG monitoring for the detection of atrial fibrillation and flutter after acute stroke and transient ischemic attack. *Stroke* 35, 1647-51 (2004).
23. Wichmann, H. E., Gieger, C. & Illig, T. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* 67 Supl 1, S26-30 (2005).
24. Berger, K. *et al.* The glu298asp polymorphism in the nitric oxide synthase 3 gene is associated with the risk of ischemic stroke in two large independent case-control studies. *Hum Genet* 121, 169-78 (2007).
25. Adams, H. P., Jr. *et al.* Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke* 24, 35-41 (1993).
26. Kutyavin, I. V. *et al.* A novel endonuclease IV post-PCR genotyping system. *Nucleic Acids Res* 34, e128 (2006).

27. Gretarsdottir, S. *et al.* The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat Genet* 35, 131-8 (2003).
28. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nat Genet* 37, 129-37 (2005).
- 5 29. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 22, 719-48 (1959).

REIVINDICAÇÕES

1. Método para determinar uma suscetibilidade a arritmia cardíaca ou acidente vascular cerebral em um indivíduo humano, caracterizado pelo fato de que compreende determinar a presença ou ausência de pelo menos um alelo de pelo menos um marcador polimórfico em uma amostra de ácido nucleico do indivíduo, em que o pelo menos um marcador polimórfico é selecionado dentre rs12503217 (posição 108955 na SEQ ID NO:50), rs12510087 (posição 111822 na SEQ ID NO:50), rs6852357 (posição 117129 na SEQ ID NO:50), rs4400058 (posição 119467 na SEQ ID NO:50), rs10033464 (posição 123555 na SEQ ID NO:50), rs2171592 (posição 123582 na SEQ ID NO:50), rs2350539 (posição 124004 na SEQ ID NO:50), rs1906606 (posição 126186 na SEQ ID NO:50), rs723364 (posição 127265 na SEQ ID NO:50), rs2220429 (posição 130279 na SEQ ID NO:50), rs4032976 (posição 131561 na SEQ ID NO:50), rs3853440 (posição 132403 na SEQ ID NO:50), rs3853441 (posição 132534 na SEQ ID NO:50), rs3853442 (posição 132822 na SEQ ID NO:50), rs3853443 (posição 132923 na SEQ ID NO:50), rs4124158 (posição 132988 na SEQ ID NO:50), rs4124159 (posição 133037 na SEQ ID NO:50), rs12506083 (posição 133206 na SEQ ID NO:50), rs4032975 (posição 135032 na SEQ ID NO:50), rs4032974 (posição 135330 na SEQ ID NO:50), rs17042059 (posição 43980 na SEQ ID NO:50), rs4529121 (posição 48349 na SEQ ID NO:50), rs4543199 (posição 50934 na SEQ ID NO:50), rs12647316 (posição 52045 na SEQ ID NO:50), rs10019689 (posição 52663 na SEQ ID NO:50), rs4626276 (posição 52783 na SEQ ID NO:50), rs17042076 (posição 55132 na SEQ ID NO:50), rs11098089 (posição 57020 na SEQ ID NO:50), rs17042088 (posição 57608 na SEQ ID NO:50), rs11930528 (posição 62988 na SEQ ID NO:50), rs17042098 (posição 66952 na SEQ ID NO:50), rs17042102 (posição 71420 na SEQ ID NO:50), rs17042121 (posição 79895 na SEQ ID NO:50), rs10516563 (posição 80516 na SEQ ID NO:50), rs4605724 (posição

87875 na SEQ ID NO:50), rs2350269 (posição 89918 na SEQ ID NO:50), rs6533527 (posição 90308 na SEQ ID NO:50), rs17042144 (posição 92460 na SEQ ID NO:50), rs1906618 (posição 98216 na SEQ ID NO:50), rs1906617 (posição 98608 na SEQ ID NO:50), rs12646447 (posição 102120 na SEQ ID NO:50), rs12646754 (posição 106366 na SEQ ID NO:50), rs2129981 (posição 106993 na SEQ ID NO:50), rs12639654 (posição 108089 na SEQ ID NO:50), rs6817105 (posição 108562 na SEQ ID NO:50), rs17042171 (posição 111081 na SEQ ID NO:50), rs1906591 (posição 111683 na SEQ ID NO:50), rs1906592 (posição 111798 na SEQ ID NO:50), rs2200732 (posição 112836 na SEQ ID NO:50), rs2200733 (posição 112836 na SEQ ID NO:50), rs4611994 (posição 113835 na SEQ ID NO:50), rs4540107 (posição 113896 na SEQ ID NO:50), rs1906593 (posição 114716 na SEQ ID NO:50), rs1906596 (posição 115030 na SEQ ID NO:50) e rs2220427 (posição 117683 na SEQ ID NO:50),

e em que a determinação da presença ou ausência do pelo menos um alelo é indicativa de uma suscetibilidade à arritmia cardíaca ou acidente vascular cerebral no indivíduo.

2. Método para determinar uma suscetibilidade a arritmia cardíaca ou acidente vascular cerebral em um indivíduo humano, caracterizado pelo fato de que compreende determinar se pelo menos um alelo em risco em pelo menos um marcador polimórfico está presente em um conjunto de dados de genótipo derivado do indivíduo, em que o pelo menos um marcador polimórfico é selecionado dentre rs12503217 (posição 108955 na SEQ ID NO:50), rs12510087 (posição 111822 na SEQ ID NO:50), rs6852357 (posição 117129 na SEQ ID NO:50), rs4400058 (posição 119467 na SEQ ID NO:50), rs10033464 (posição 123555 na SEQ ID NO:50), rs2171592 (posição 123582 na SEQ ID NO:50), rs2350539 (posição 124004 na SEQ ID NO:50), rs1906606 (posição 126186 na SEQ ID NO:50), rs723364 (posição 127265 na SEQ ID NO:50), rs2220429 (posição 130279 na

SEQ ID NO:50), rs4032976 (posição 131561 na SEQ ID NO:50), rs3853440 (posição 132403 na SEQ ID NO:50), rs3853441 (posição 132534 na SEQ ID NO:50), rs3853442 (posição 132822 na SEQ ID NO:50), rs3853443 (posição 132923 na SEQ ID NO:50), rs4124158 (posição 132988 na SEQ ID NO:50),
5 rs4124159 (posição 133037 na SEQ ID NO:50), rs12506083 (posição 133206 na SEQ ID NO:50), rs4032975 (posição 135032 na SEQ ID NO:50), rs4032974 (posição 135330 na SEQ ID NO:50), rs17042059 (posição 43980 na SEQ ID NO:50), rs4529121 (posição 48349 na SEQ ID NO:50), rs4543199 (posição 50934 na SEQ ID NO:50), rs12647316 (posição 52045 na
10 SEQ ID NO:50), rs10019689 (posição 52663 na SEQ ID NO:50), rs4626276 (posição 52783 na SEQ ID NO:50), rs17042076 (posição 55132 na SEQ ID NO:50), rs11098089 (posição 57020 na SEQ ID NO:50), rs17042088 (posição 57608 na SEQ ID NO:50), rs11930528 (posição 62988 na SEQ ID NO:50), rs17042098 (posição 66952 na SEQ ID NO:50), rs17042102
15 (posição 71420 na SEQ ID NO:50), rs17042121 (posição 79895 na SEQ ID NO:50), rs10516563 (posição 80516 na SEQ ID NO:50), rs4605724 (posição 87875 na SEQ ID NO:50), rs2350269 (posição 89918 na SEQ ID NO:50), rs6533527 (posição 90308 na SEQ ID NO:50), rs17042144 (posição 92460 na SEQ ID NO:50), rs1906618 (posição 98216 na SEQ ID NO:50), rs1906617
20 (posição 98608 na SEQ ID NO:50), rs12646447 (posição 102120 na SEQ ID NO:50), rs12646754 (posição 106366 na SEQ ID NO:50), rs2129981 (posição 106993 na SEQ ID NO:50), rs12639654 (posição 108089 na SEQ ID NO:50), rs6817105 (posição 108562 na SEQ ID NO:50), rs17042171 (posição 111081 na SEQ ID NO:50), rs1906591 (posição 111683 na SEQ ID NO:50), rs1906592 (posição 111798 na SEQ ID NO:50), rs2200732 (posição 112836 na SEQ ID NO:50), rs2200733 (posição 112836 na SEQ ID NO:50),
25 rs4611994 (posição 113835 na SEQ ID NO:50), rs4540107 (posição 113896 na SEQ ID NO:50), rs1906593 (posição 114716 na SEQ ID NO:50), rs1906596 (posição 115030 na SEQ ID NO:50) e rs2220427 (posição 117683

na SEQ ID NO:50),

e em que a determinação da presença do pelo menos um alelo em risco é indicativa de suscetibilidade aumentada a arritmia cardíaca ou acidente vascular cerebral no indivíduo.

5 3. Método de acordo com a reivindicação 1 ou 2, caracterizado pelo fato de que o pelo menos um marcador polimórfico é selecionado dentre rs2200733, rs2220427 e rs10033464.

10 4. Método de acordo com qualquer uma das reivindicações de 1 a 3, caracterizado pelo fato de que a presença do alelo T do marcador rs2200733, alelo T do marcador rs2220427 e/ou alelo T do marcador rs10033464 é indicativa de suscetibilidade aumentada a arritmia cardíaca ou acidente vascular cerebral no indivíduo.

15 5. Método de acordo com qualquer uma das reivindicações 1 a 4, caracterizado pelo fato de que compreende adicionalmente avaliar pelo menos um biomarcador para fibrilação atrial, flúter atrial e/ou acidente vascular cerebral em uma amostra de um indivíduo, em que o pelo menos um biomarcador é selecionado do grupo que consiste em dímero D da fibrina, fragmento de ativação da protrombina 1,2 (F1,2), complexos de trombina-antitrombina III (TAT), fibrinopeptídeo A (FPA), fosfolipase associada à
20 lipoproteína A2 (Ip-PLA2), beta-tromboglobulina, fator de plaqueta 4, P-selectina, Fator de von Willebrand, peptídeo pró-natriurético (BNP), metaloproteinase de matriz-9 (MMP-9), PARK7, difosfato cinase de nucleosídeo (NDKA), tau, enolase específica de neurônio, fator de crescimento neurotrófico tipo B, proteína astrogial S-100b, proteína ácida
25 fibrilar glial, proteína reativa C, amiloide A sérico, metaloproteinase de matriz-9, moléculas de adesão celular vascular e intracelular, fator alfa de necrose de tumor e interleucinas.

6. Aparelho para determinar um indicador genético para a arritmia cardíaca e/ou acidente vascular cerebral em um indivíduo humano,

caracterizado pelo fato de que compreende:

uma memória legível por computador; e

uma rotina armazenada na memória legível por computador;

em que a rotina é adaptada para ser executada em um

5 processador para analisar dados de genótipo e/ou haplótipo quanto a pelo menos um indivíduo humano com respeito a pelo menos um marcador polimórfico selecionado dos marcadores rs12503217 (posição 108955 na SEQ ID NO:50), rs12510087 (posição 111822 na SEQ ID NO:50), rs6852357 (posição 117129 na SEQ ID NO:50), rs4400058 (posição 119467 na SEQ ID NO:50), rs10033464 (posição 123555 na SEQ ID NO:50), rs2171592 (posição 123582 na SEQ ID NO:50), rs2350539 (posição 124004 na SEQ ID NO:50), rs1906606 (posição 126186 na SEQ ID NO:50), rs723364 (posição 127265 na SEQ ID NO:50), rs2220429 (posição 130279 na SEQ ID NO:50), rs4032976 (posição 131561 na SEQ ID NO:50), rs3853440 (posição 132403 na SEQ ID NO:50), rs3853441 (posição 132534 na SEQ ID NO:50), rs3853442 (posição 132822 na SEQ ID NO:50), rs3853443 (posição 132923 na SEQ ID NO:50), rs4124158 (posição 132988 na SEQ ID NO:50), rs4124159 (posição 133037 na SEQ ID NO:50), rs12506083 (posição 133206 na SEQ ID NO:50), rs4032975 (posição 135032 na SEQ ID NO:50), rs4032974 (posição 135330 na SEQ ID NO:50), rs17042059 (posição 43980 na SEQ ID NO:50), rs4529121 (posição 48349 na SEQ ID NO:50), rs4543199 (posição 50934 na SEQ ID NO:50), rs12647316 (posição 52045 na SEQ ID NO:50), rs10019689 (posição 52663 na SEQ ID NO:50), rs4626276 (posição 52783 na SEQ ID NO:50), rs17042076 (posição 55132 na SEQ ID NO:50), rs11098089 (posição 57020 na SEQ ID NO:50), rs17042088 (posição 57608 na SEQ ID NO:50), rs11930528 (posição 62988 na SEQ ID NO:50), rs17042098 (posição 66952 na SEQ ID NO:50), rs17042102 (posição 71420 na SEQ ID NO:50), rs17042121 (posição 79895 na SEQ ID NO:50), rs10516563 (posição 80516 na SEQ ID NO:50), rs4605724 (posição

87875 na SEQ ID NO:50), rs2350269 (posição 89918 na SEQ ID NO:50), rs6533527 (posição 90308 na SEQ ID NO:50), rs17042144 (posição 92460 na SEQ ID NO:50), rs1906618 (posição 98216 na SEQ ID NO:50), rs1906617 (posição 98608 na SEQ ID NO:50), rs12646447 (posição 102120 na SEQ ID NO:50), rs12646754 (posição 106366 na SEQ ID NO:50), rs2129981 (posição 106993 na SEQ ID NO:50), rs12639654 (posição 108089 na SEQ ID NO:50), rs6817105 (posição 108562 na SEQ ID NO:50), rs17042171 (posição 111081 na SEQ ID NO:50), rs1906591 (posição 111683 na SEQ ID NO:50), rs1906592 (posição 111798 na SEQ ID NO:50), rs2200732 (posição 112836 na SEQ ID NO:50), rs2200733 (posição 112836 na SEQ ID NO:50), rs4611994 (posição 113835 na SEQ ID NO:50), rs4540107 (posição 113896 na SEQ ID NO:50), rs1906593 (posição 114716 na SEQ ID NO:50), rs1906596 (posição 115030 na SEQ ID NO:50) e rs2220427 (posição 117683 na SEQ ID NO:50),

e gerar um produto com base nos dados de marcador ou haplótipo, em que o produto compreende uma medição de risco do pelo menos um marcador ou haplótipo como um indicador genético de arritmia cardíaca e/ou acidente vascular cerebral para o indivíduo humano.

7. Método ou aparelho de acordo com qualquer uma das reivindicações 1 a 6, caracterizado pelo fato de que a arritmia cardíaca é fibrilação atrial ou flúter atrial.

8. Método ou aparelho de acordo com a reivindicação 7, caracterizado pelo fato de que a fibrilação atrial ou flúter atrial é definida pela idade no início de menos do que 60 anos.

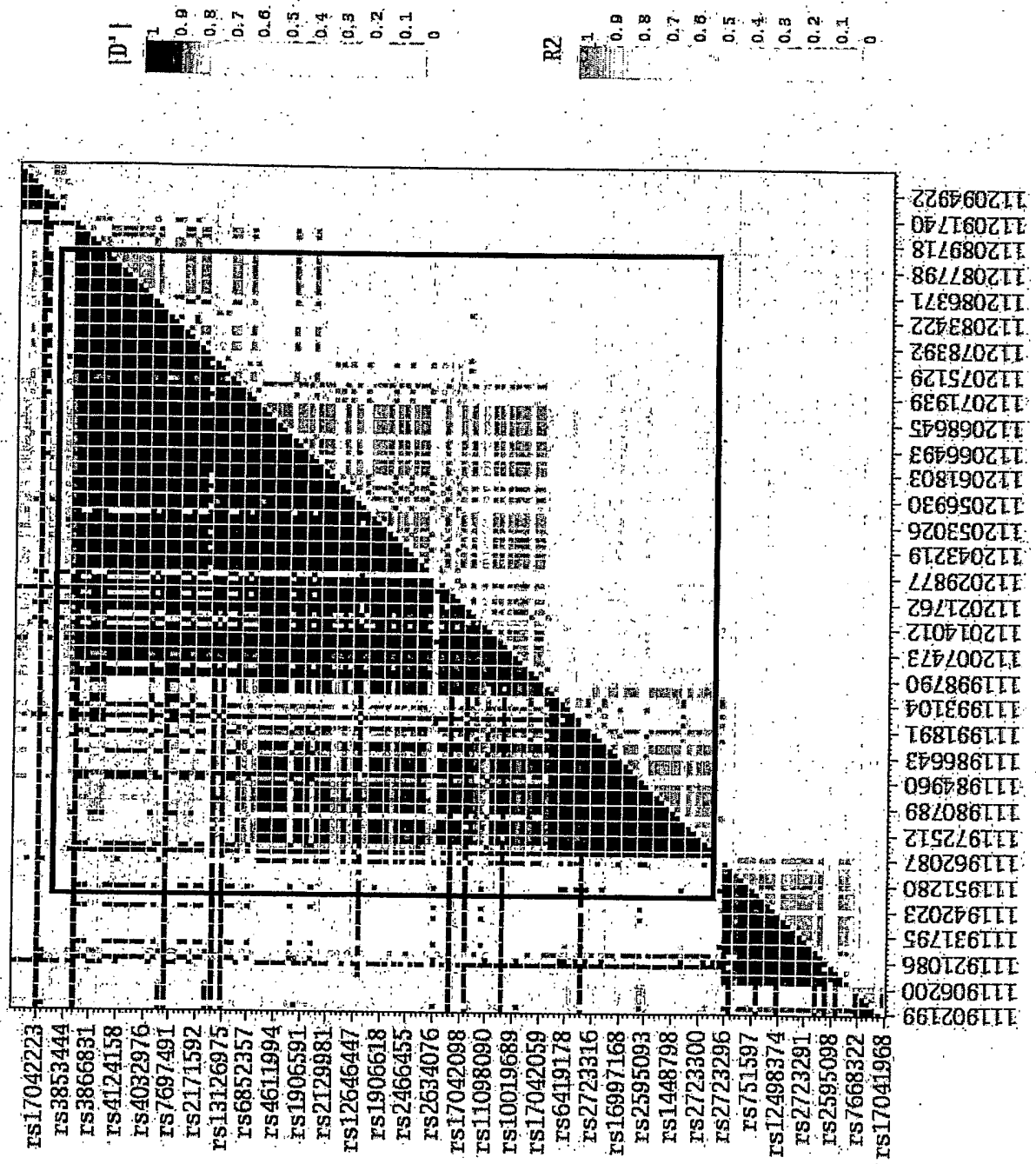


FIG. 1

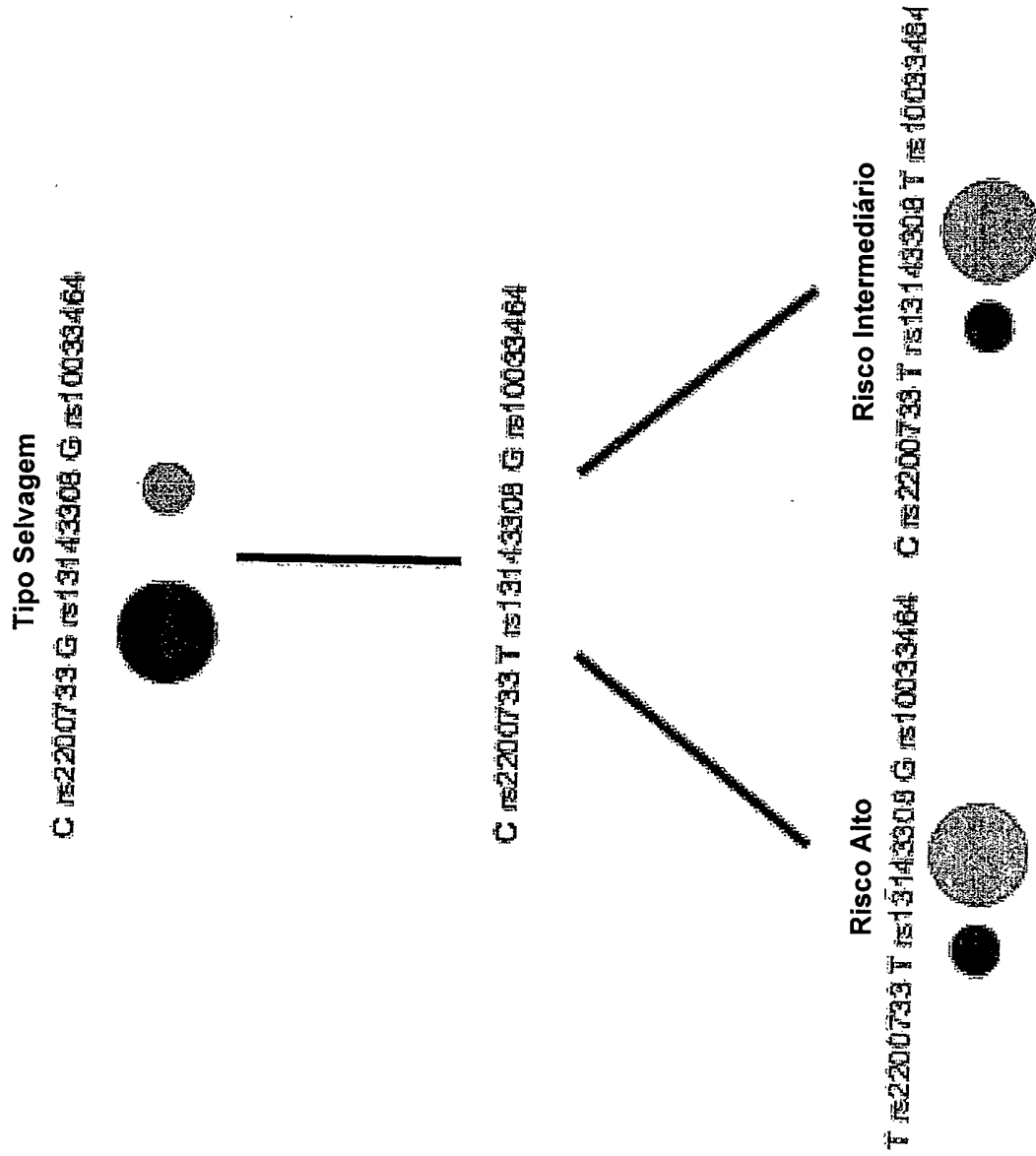


FIG. 2

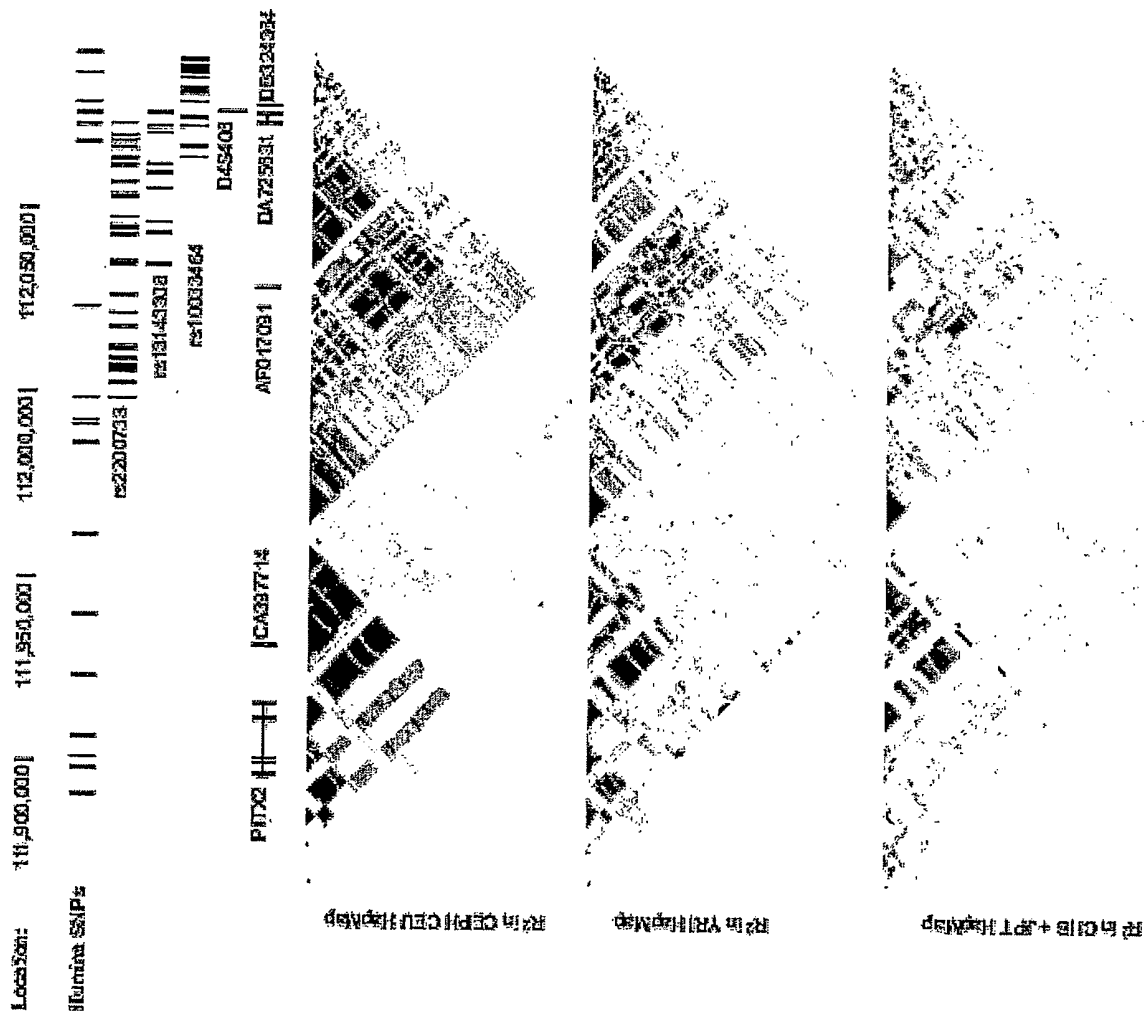


FIG. 3

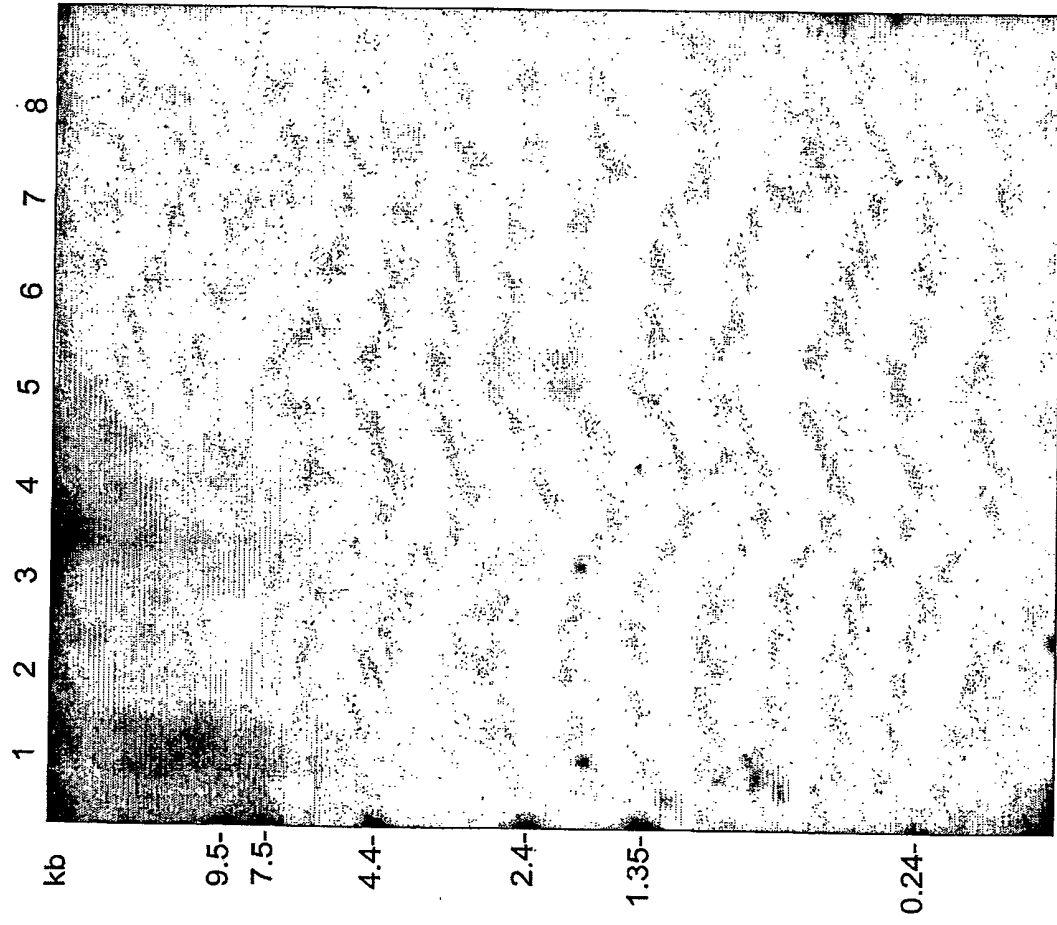


FIG. 4