República Federativa do Brasil

Ministério do Desenvolvimento, Indústria, Comércio e Serviços

Instituto Nacional da Propriedade Industrial

(11) BR 112017005730-1 B1

(22) Data do Depósito: 03/11/2014

(45) Data de Concessão: 12/12/2023

(54) Título: MÉTODO PARA TRIAGEM DE UM INDIVÍDUO QUANTO AO RISCO DE UM EVENTO CARDIOVASCULAR OU PARA PREDIZER A PROBABILIDADE QUE UM INDIVÍDUO TENHA TAL EVENTO

(51) Int.Cl.: C12Q 1/68; G01N 33/68.

(30) Prioridade Unionista: 26/09/2014 US 62/055,984.

(73) Titular(es): SOMALOGIC OPERATING CO., INC..

(72) Inventor(es): DAVID STERLING; SHINTARO KATO; EDWARD N. BRODY; STEPHEN A. WILLIAMS.

(86) Pedido PCT: PCT US2014063714 de 03/11/2014

(87) Publicação PCT: WO 2016/048388 de 31/03/2016

(85) Data do Início da Fase Nacional: 21/03/2017

(57) Resumo: MÉTODO PARA TRIAGEM DE UM INDIVÍDUO QUANTO AO RISCO DE UM EVENTO CARDIOVASCULAR OU PARA PREDIZER A PROBABILIDADE QUE UM INDIVÍDUO TENHA TAL EVENTO. Métodos e métodos de computador usados para avaliar um indivíduo para a predição de risco de desenvolver um Evento Cardiovascular ao longo de um período de 1 a 5 anos são fornecidos. Os métodos empregam pelo menos dois biomarcadores selecionados de MMP12, agiopoietina-2, complemento C7, troponina cardíaca I, proteína 4 relativa a anfioproteína, CCL18/PARC, complexo alfa-1-antiquimotripsina, GDF11 e alfa-2-antiplasmina ou GDF11 em combinação com FSTL3. Os métodos são particularmente úteis na predição de eventos CV em pacientes que sofrem de doença cardíaca coronariana (CHD).

Relatório Descritivo da Patente de Invenção para **"MÉTODO PARA TRIAGEM DE UM INDIVÍDUO QUANTO AO RISCO DE UM EVENTO CARDIOVASCULAR OU PARA PREDIZER A PROBABILIDADE QUE UM INDIVÍDUO TENHA TAL EVENTO".**

CAMPO DA INVENÇÃO

[0001] O presente pedido se refere genericamente à detecção de biomarcadores e um método para avaliar o risco de um futuro evento cardiovascular em um indivíduo e, mais especificamente, a um ou mais biomarcadores, métodos, aparelhos, reagentes, sistemas e kits utilizados para avaliar uma individual para a predição de risco de desenvolvimento de um evento cardiovascular (CV) ao longo de um período de 1 a 5 anos. Tais eventos incluem, mas não estão limitados a enfarte do miocárdio, acidente vascular cerebral, insuficiência cardíaca congestiva ou morte.

FUNDAMENTOS

[0002] A descrição a seguir proporciona um resumo de informação relevante ao presente pedido e não é uma admissão de que qualquer uma das informações proporcionadas ou publicações mencionadas neste documento são o estado da técnica para o presente pedido.

[0003] A doença cardiovascular é a principal causa de morte nos EUA. Há uma série de indicadores existentes e importantes de risco de eventos primários (D'Agostino, R et al., "General Cardiovascular Risk Profile for Use in Primary Care: The Framingham Heart Study" Circulation 117: 743-53 (2008); and Ridker, P. et al., "Development and Validation of Improved Algorithms fo rthe Assessment of Global Cardiovascular Risk in Women" JAMA 297(6): 611-619 (2007)) and secondary events (Shlipak, M. et al. "Biomarkers to Predict Recurrent Cardiovascular Disease: The Heart & Soul Study" Am. J. Med. 121: 50-57 (2008)), que são amplamente utilizados na prática clínica e ensaios terapêuticos. Infelizmente, as curvas características de

receptor-operação, as taxas de perigo, e a concordância mostram que o desempenho de fatores e biomarcadores de risco existentes é modesto (AUC de ~ 0,75 significa que esses fatores estão apenas a meio caminho entre uma virada de moeda e perfeição). Além da necessidade de melhorar o desempenho de diagnóstico, existe uma necessidade para um produto de risco, que é tanto de curto prazo e pessoalmente responsivo em indivíduos para intervenções benéficas (e destrutivas) e alterações do estilo de vida. A equação de Framingham comumente utilizada tem três problemas principais. Em primeiro lugar, é de prazo longo demais: ela dá cálculos de risco de 10 anos, mas os seres humanos desconsideram riscos futuros e se mostram relutantes em fazer modificações de comportamento e estilo de vida com base nela. Em segundo lugar, não é muito responsiva a intervenções: é fortemente dependente da idade cronológica, o que não pode declinar; e sexo, que não pode mudar. Em terceiro lugar, dentro da população de alto risco imaginada aqui, os fatores de Framingham não discriminam bem entre alto e baixo risco: a taxa de perigo entre os quartis de alta e baixa é de apenas 2, e quando se tenta usar pontuações de Framingham para personalizar risco através da estratificação de pacientes em camadas mais finas (decis por exemplo) as taxas de eventos observadas são semelhantes para muitos dos decis.

[0004] Os fatores de risco para as doenças cardiovasculares são amplamente utilizados para dirigir a intensidade e a natureza do tratamento médico, e seu uso tem, sem dúvida, contribuído para a redução da morbidade e mortalidade cardiovascular que tem sido observada ao longo das últimas duas décadas. Esses fatores têm sido rotineiramente combinados em algoritmos, mas infelizmente eles não capturam todos os riscos (a apresentação inicial mais comum de doença cardíaca ainda é a morte). Na verdade, eles provavelmente só capturam metade do risco. Uma área sob a curva ROC de ~ 0,76 é

típica para tais fatores de risco na prevenção primária, com desempenho muito pior na prevenção secundária (0,62 é típica), números de apenas cerca de um quarto a um meio do desempenho entre uma virada de moeda em 0. 5 e perfeição em 1,0.

[0005] A adição de novos biomarcadores para graus de risco clínicos tem sido decepcionante. Por exemplo, no estudo de Framingham (Wang et al., "Multiple Biomarkers for the Prediction of First Major Cardiovascular Events and Death" N. Eng. J. Med. 355: 2631-2637 (2006)) em 3209 pessoas, a adição de 10 biomarcadores (CRP, BNP, NT-proBNP, aldosterona, renina, fibrinogênio, D-dimer, inibidor de ativador de plasminogênio do tipo 1, homocisteína e a albumina urinária creatinina), não melhorou significativamente a AUC, quando adicionados aos fatores de risco existentes: a AUC para eventos de 0-5 anos foi de 0,76 com a idade, sexo e fatores de risco convencionais e 0,77 com a melhor combinação de biomarcadores adicionados à mistura, e para a prevenção secundária, a situação é pior.

[0006] A identificação precoce de pacientes com maior risco de um evento cardiovascular em uma janela de 1-5 anos é importante porque um tratamento mais agressivo de indivíduos com risco elevado pode melhorar o resultado. Assim, a gestão ótima requer intervenção agressiva para reduzir o risco de um evento cardiovascular nos pacientes que são considerados de maior risco, enquanto os pacientes com um menor risco de um evento cardiovascular podem ser poupados de tratamentos caros e potencialmente invasores, que são suscetíveis de não ter qualquer efeito benéfico para o paciente.

[0007] A seleção de biomarcador para predizer o risco de ter estado de doença específica ou condição dentro de um período de tempo definido envolve primeiro a identificação de marcadores que têm uma relação mensurável e estatisticamente significativa com a probabilidade e/ou o momento de um evento para uma aplicação médica específica.

Os biomarcadores podem incluir moléculas secretadas ou derramadas que estão, tanto na via causal para a condição de interesse, ou que estão a jusante ou em paralelo com a doença ou estado de desenvolvimento ou progressão, ou ambos. Elas são liberadas para a corrente sanguínea a partir de tecido cardiovascular ou de outros órgãos e tecidos circundantes e células circulantes em resposta aos processos biológicos que predispõem a um evento cardiovascular ou podem ser refletores de efeitos a jusante da fisiopatologia, tais como um declínio na função renal. Os biomarcadores podem incluir pequenas moléculas, peptídeos, proteínas e ácidos nucleicos. Algumas das questões-chave que afetam a identificação de biomarcadores incluem excesso de encaixe dos dados disponíveis e predisposição nos dados.

[0008] Uma variedade de métodos tem sido utilizada em uma tentativa de identificar biomarcadores e diagnosticar ou prever o risco de ter uma doença ou uma condição. Para os marcadores com uma base proteica, estes incluem a eletroforese bidimensional, espectrometria de massa e métodos de imunoensaio. Para marcadores de ácidos nucleicos, que incluem perfis de expressão de mRNA, perfis de microRNA, FISH, análise em série da expressão gênica (SAGE), matrizes de expressão gênica em larga escala, sequenciamento de genes e genotipagem (SNP ou análise variante pequena).

[0009] A utilidade de electroforese bidimensional é limitada pela baixa sensibilidade de detecção; problemas com a solubilidade da proteína, carga e hidrofobicidade; reprodutibilidade de gel; e a possibilidade de uma única mancha representando múltiplas proteínas. Por espectrometria de massa, dependendo do formato utilizado, limitações giram em torno do processamento de amostras e a separação, a sensibilidade às proteínas de baixa abundância, para considerações de sinal de ruído, e a incapacidade de identificar imediatamente a proteína detectada. Limitações nas abordagens de

imunoensaio para a descoberta de biomarcadores são centradas sobre a incapacidade dos ensaios multiplex à base de anticorpos para medir um grande número de analitos. Alguém poderia simplesmente imprimir uma matriz de anticorpos de alta qualidade e, sem sanduíches, medir os analitos ligados a esses anticorpos. (Isto seria o equivalente formal da utilização de um genoma completo de sequências de ácidos nucleicos para medir por hibridização de todas as sequências de DNA ou de RNA de um organismo ou de uma célula. O experimento de hibridização funciona porque a hibridização pode ser um teste rigoroso de identidade.)No entanto, até anticorpos muito bons não são normalmente suficientemente rigorosos na escolha dos seus parceiros de ligação para trabalhar no contexto de extratos de sangue ou mesmo de células, porque o conjunto de proteínas nessas matrizes têm abundâncias amplamente variadas, que podem conduzir a um sinal pobre para razões de ruído. Assim, deve-se usar uma abordagem diferente com abordagens baseadas em imunoensaios para a descoberta de biomarcadores - seria preciso usar ensaios ELISA multiplexados (isto é, sanduíches) para obter rigor suficiente para medir muitos analitos simultaneamente para decidir quais analitos são realmente biomarcadores. Imunoensaios de sanduíche não escalam para alto conteúdo, e assim, a descoberta de biomarcador utilizando imunoensaios de sanduíche rigorosos não é possível utilizando formatos de matriz padrão. Por último, reagentes de anticorpos estão sujeitos a variabilidade muito substancial e instabilidade de reagente. A plataforma instantânea para a descoberta de biomarcadores de proteína supera esse problema.

[00010] Muitos destes métodos baseiam-se em ou requerem algum tipo de fracionamento da amostra antes da análise. Assim, a preparação da amostra necessária para executar um estudo suficientemente mecânico destinado a identificar e descobrir biomarcadores

estatisticamente relevantes, de uma série de populações de amostras bem definidas é extremamente difícil, dispendiosa e demorada. Durante o fracionamento, uma vasta variedade de variabilidade pode ser introduzida nas várias amostras. Por exemplo, um marcador em potencial pode ser instável para o processo, a concentração do marcador pode ser alterada, agregação inadequada ou desagregação pode ocorrer, e contaminação inadvertida da amostra pode ocorrer e, consequentemente, obscurecer as alterações sutis antecipadas no início da doença.

[00011] É amplamente aceito que a descoberta de biomarcadores e métodos de detecção que utilizam estas tecnologias têm sérias limitações para a identificação de biomarcadores de diagnóstico ou de predição. Estas limitações incluem a incapacidade para detectar biomarcadores de baixa abundância, uma incapacidade para cobrir consistentemente toda a variedade dinâmica do proteoma, irreprodutibilidade no processamento e fracionamento da amostra, e irreprodutibilidade geral e falta de robustez do método. Além disso, estes estudos introduziram polarizações para os dados e não trataram adequadamente a complexidade das populações de amostras, incluindo os controles adequados, em termos de distribuição e randomização necessária para identificar e validar os biomarcadores dentro de uma população alvo da doença.

[00012] Embora os esforços que visam a descoberta de biomarcadores novos e eficazes tenham passado por várias décadas, os esforços têm sido em grande parte sem êxito. Biomarcadores para várias doenças têm sido tipicamente identificados em laboratórios acadêmicos, geralmente através de uma descoberta acidental enquanto se faz a pesquisa básica em algum processo da doença. Com base na descoberta e com pequenas quantidades de dados clínicos, documentos foram publicados que sugeriram a identificação de um

novo biomarcador. A maioria destes biomarcadores propostos, no entanto, não foi confirmada como biomarcadores reais ou úteis, principalmente por que o pequeno número de amostras clínicas testadas proveu apenas estatística fraca de que um biomarcador eficaz foi de fato encontrado. Isto é, a identificação inicial não foi rigorosa no que diz respeito aos elementos básicos de estatística. Em cada um dos anos de 1994 a 2003, uma pesquisa da literatura científica mostra que milhares de referências voltadas para biomarcadores foram publicadas. Durante esse mesmo período de tempo, no entanto, a FDA aprovou para uso em diagnóstico, no máximo, três novas biomarcadores de proteína em um ano, e em vários anos foram aprovados novos biomarcadores de proteína.

[00013] Com base na história dos esforços falhos de descoberta de biomarcador, teorias têm sido propostas que promovem ainda mais o entendimento geral de que os biomarcadores para o diagnóstico, prognóstico ou predição de risco de desenvolvimento de doenças e condições são raros e difíceis de encontrar. Pesquisa com biomarcadores com base em géis 2D ou espectrometria de massa suporta essas noções. Muito poucos biomarcadores úteis têm sido identificados através destas abordagens. No entanto, geralmente é ignorado que o gel 2D e proteínas espectrometria de massa medem proteínas que estão presentes no sangue em concentrações de cerca de 1 nM e mais elevadas, e que este conjunto de proteínas pode muito bem ser o menos provável a alterar com a doença ou o desenvolvimento de uma condição particular. À exceção da plataforma de descoberta de biomarcador instantânea, plataformas de descoberta de biomarcadores proteômicos que são capazes de medir com precisão os níveis de expressão de proteína em concentrações muito mais baixas não existem.

[00014] Muito se sabe sobre vias bioquímicas para a biologia

humana complexa. Muitas vias bioquímicas culminam com ou são iniciados por proteínas secretadas que trabalham localmente dentro da patologia; por exemplo, fatores de crescimento são secretados para estimular a replicação de outras células na patologia, e outros fatores são secretados para afastar o sistema imunológico, e assim por diante. Enquanto muitas destas proteínas secretadas funcionam de uma forma paracrina, algumas operam de modo distal no corpo. Um versado na técnica com uma compreensão básica de vias bioquímicas iria entender que muitas proteínas específicas da patologia deveriam existir no sangue em concentrações abaixo (ainda muito abaixo) dos limites de detecção de géis 2D e espectrometria de massa. O que deve preceder a identificação deste número relativamente abundante de biomarcadores de doença é uma plataforma proteômica que pode analisar proteínas a concentrações inferiores às detectáveis por géis 2D ou espectrometria de massa.

[00015] Como é discutido acima, eventos cardiovasculares podem ser prevenidos por tratamento agressivo se a propensão para tais eventos puder ser determinada com precisão, e alvejando tais intervenções para as pessoas que deles necessitam mais e/ou longe das pessoas que deles necessitam o mínimo, a eficiência de recursos médicos pode ser melhorada e os custos podem ser reduzidos ao mesmo tempo. Além disso, quando o paciente tem o conhecimento de informações precisas e de curto prazo sobre a sua probabilidade pessoal de eventos cardiovasculares, isto pode ser menos negável do que a informação de base populacional a longo prazo e vai levar a melhores escolhas de estilo de vida e melhor cumprimento da medicação que irá adicionar para os benefícios. Testes de multimarcador existentes tanto requerem a coleta de múltiplas amostras de um indivíduo ou exigem que uma amostra seja dividida entre vários ensaios. Idealmente, um teste melhorado pode requerer apenas uma

única amostra de sangue, urina, ou outra amostra, e um único ensaio. Por conseguinte, existe uma necessidade de biomarcadores, métodos, aparelhos, reagentes, sistemas e kits que permitam a predição de eventos cardiovasculares dentro de um período de 5 anos.

SUMÁRIO DA INVENÇÃO

[00016] O presente pedido inclui os biomarcadores, métodos, reagentes, equipamentos, sistemas e kits para a predição de risco de ter um evento cardiovascular (CV) dentro de um período de 1 ano, um período de 2 anos, um período de 3 anos, ou um período de 4 anos. Os biomarcadores do presente pedido de patente foram identificados utilizando um ensaio à base de aptâmero de taxa de desaceleração multiplex (SOMAmer) que é descrito aqui em detalhe. Ao utilizar o método de identificação biomarcador baseado em SOMAmer aqui descrito, este pedido descreve um conjunto de biomarcadores que são úteis para predizer a probabilidade de um evento CV dentro de 1 ano, 2 anos, 3 anos, ou 4 anos.

[00017] Os eventos cardiovasculares podem ser evitados por tratamento agressivo se a propensão para tais eventos puder ser determinada com precisão. Testes de multimarcador existentes tanto requerem a coleta de múltiplas amostras de um indivíduo, ou requerem que uma amostra seja dividida entre vários ensaios. Seria preferível prover um ensaio de prognóstico que exigiria apenas uma única amostra biológica, medido em um ensaio, em vez de várias amostras de diferentes tipos de analitos (lipídios, proteínas, metabolitos) ou de painéis de analitos. O benefício central para um único teste da amostra é a simplicidade no ponto de uso, uma vez que um teste com vários conjuntos de amostras e/ou vários tipos de tecnologia (como a integração de resultados de sangue com uma ou mais fontes de informação de cortesia, como a demografia, a ecocardiografia, imagiologia, testes de urina, pressão arterial ou complacência vascular)

é mais complexo de administrar e isto constitui uma barreira para a adoção. Uma vantagem adicional deriva da execução daquela única amostra em um único ensaio para várias proteínas. Um único ensaio deve mitigar a variação indesejada devido à calibração de vários resultados de ensaio ou formatos de tecnologia juntos. O teste que constitui a base do presente pedido é um teste de "única amostra, ensaio único". Esta combinação de única amostra e ensaio único é uma nova característica deste teste de risco de evento cardiovascular, que aborda a complexidade logística de coleta de amostras múltiplas e usando várias modalidades de medição e os problemas e perigos biológicos envolvidos na divisão de amostras em várias alíquotas para múltiplos procedimentos analíticos independentes.

[00018] A doença cardiovascular é conhecida por envolver vários processos biológicos e tecidos. Exemplos bem conhecidos de sistemas biológicos e processos associados com as doenças cardiovasculares são a inflamação, trombose, angiogese associada à doença, ativação de plaquetas, ativação de macrófagos, a resposta aguda do fígado, remodelação da matriz extracelular, e a função renal. Estes processos podem ser observados como uma função do sexo, estado de menopausa, e a idade, e de acordo com o estado de coagulação e função vascular. Uma vez que estes sistemas comunicam parcialmente através de sistemas de sinalização baseados em proteínas, e proteínas múltiplas podem ser medidas em uma única amostra de sangue, a invenção provê um teste baseado em proteínas múltiplas de única amostra, ensaio único focado em proteínas dos sistemas biológicos específicos e processos envolvidos na doença cardiovascular.

[00019] Como é aqui discutido, uma das funções centrais de medir o risco de um evento cardiovascular é permitir a avaliação dos progressos em resposta ao tratamento e mudanças comportamentais, tais como dieta e exercício. Métodos de predição de riscos atuais, como a

equação de Framingham, incluem claramente as informações covariáveis clínicas sem resposta, principais fatores são a idade e o sexo do indivíduo. Isso faz com que a equação de Framingham seja menos útil para monitorizar a alteração no risco de um indivíduo, embora possa ser precisa para uma população. Uma nova característica do presente teste de risco de evento CV é que ele não necessita de idade como uma parte do modelo prognóstico. A presente invenção baseia-se na premissa de que, no âmbito da biologia do envelhecimento, existem fatores biológicos subjacentes que são mais diretamente relacionados com o risco, mas que são variáveis entre os indivíduos e, portanto, melhor usados para avaliar o risco do que a idade cronológica. A invenção tem como premissa a crença de que a idade em si não é um fator causal da doença, e que a idade está agindo como um substituto ou representante para a biologia subjacente. Enquanto a idade de fato é prognóstico de eventos CV, a mesma não pode ser usada para avaliar a melhoria individual, e presumivelmente o efeito da idade é mediado através da função biológica. Este efeito pode ser melhor determinado através da medição da biologia relevante. Na presente invenção, as proteínas que são visadas estão envolvidas na biologia da doença. Assim, o invento captura a informação biológica que se reflete na correlação entre idade e o risco de um evento CV.

[00020] A estratégia para identificar proteínas de vários processos envolvidos na doença cardiovascular exigiu a escolha de parâmetros que proveram uma vasta variedade/ diversidade de pacientes com doença cardiovascular que apresentam uma variedade de eventos ou sintomas. Eventos devido à doença cardiovascular são heterogêneos, envolvendo a morte súbita de causa desconhecida, e duas classes principais de evento conhecido: trombótica (acidente vascular cerebral, ataques isquêmicos transitórios, infarto do miocárdio) e eventos relacionados a CHF. Alguns eventos que se apresentam podem não ter

informações específicas de diagnóstico (por exemplo, a morte em casa). Tendo em conta estas características de doença cardiovascular, o teste da invenção foi desenvolvido medindo proteínas envolvidas dos processos biológicos associados com a doença cardiovascular, em amostras de sangue a partir de uma vasta variedade de eventos. Esta estratégia resulta na inclusão de informação de múltiplos processos envolvidos na doença (por exemplo, a angiogênese, a ativação de plaquetas, ativação de macrófagos, a resposta aguda do fígado, outras inflamações dos linfócitos, remodelação da matriz extracelular, e a função renal). A fim de desenvolver um teste de amostra única de prognóstico com base em múltiplas proteínas para a doença cardiovascular, a população de estudo escolhida foi um estudo de coorte de grupo de alto risco de indivíduos com doença cardíaca coronária aparentemente estável: The Heart & Soul Study" Am. Ao escolher este conjunto de indivíduos com uma taxa elevada de eventos CV, foi possível determinar o risco associado às medições de proteínas com mais precisão do que seria possível na população em geral (em que os eventos são mais raros). O desenvolvimento do teste em indivíduo neste grupo de alto risco, permitiu a identificação de combinações de biomarcadores de proteínas que podem ser generalizadas devido à biologia comum. Como resultado, o teste inventivo em indivíduo e os biomarcadores são suscetíveis de serem eficazes além da predição de evento em uma população maior do que aqueles indivíduos que correspondem aos critérios de entrada do estudo "Heart & Soul".

[00021] Em algumas modalidades, métodos para a triagem de um indivíduo para o risco de um evento cardiovascular (CV) são proporcionados evento. Em algumas modalidades, o método compreende:

(a) formar um painel de biomarcador compreendendo N

biomarcadores selecionados de MMP12, angiopoietina-2, complemento C7, troponina I cardíaca, proteína relacionada com angiopoietina 4, CCL18/PARC, complexo de alfa-1-antiquimotripsina, GDF11 e alfa-2-antiplasmina, em que N é um número inteiro dentre 2 e 9; e

(b) detectar o nível de cada um dos N biomarcadores do painel em uma amostra do indivíduo.

[00022] Em algumas modalidades, os métodos para predizer a probabilidade de um indivíduo ter um evento CV são proporcionados. Em algumas modalidades, o método compreende:

(a) formar um painel de biomarcador compreendendo N biomarcadores selecionados de MMP12, angiopoietina-2, complemento C7, troponina I cardíaca, proteína relacionada com angiopoietina 4, CCL18/PARC, complexo de alfa-1-antiquimotripsina, GDF11 e alfa-2-antiplasmina, em que N é um número inteiro dentre 2 e 9; e

(b) detectar o nível de cada um dos N biomarcadores do painel em uma amostra do indivíduo.

[00023] Em algumas modalidades, métodos para a triagem de um indivíduo para o risco de um evento cardiovascular (CV) são providos, compreendendo a detecção do nível de, pelo menos cinco, pelo menos seis, pelo menos sete, pelo menos oito, ou todos os nove biomarcadores selecionados a partir de MMP12, angiopoietina-2, complemento C7, troponina I cardíaca, proteína relacionada com angiopoietina 4, CCL18/PARC, complexo de alfa-1-antiquimotripsina, GDF11 e alfa-2-antiplasmina em uma amostra do indivíduo.

[00024] Em algumas modalidades, métodos para predizer a probabilidade de que um indivíduo vai ter um evento CV são providos, compreendendo a detecção do nível de, pelo menos cinco, pelo menos seis, pelo menos sete, pelo menos oito, ou todos os nove

biomarcadores selecionados a partir de MMP12, angiopoietina-2, complemento C7, troponina I cardíaca, proteína relacionada com angiopoietina 4, CCL18/PARC, complexo de alfa-1-antiquimotripsina, GDF11 e alfa-2-antiplasmina em uma amostra do indivíduo.

[00025] Em algumas modalidades, a probabilidade do indivíduo ter um evento CV dentro de 4 anos é elevada se o nível de, pelo menos cinco, pelo menos seis, ou todos os sete biomarcadores selecionados a partir do nível de MMP12, angiopoetina-2, complemento C7, troponina I, proteína relacionada com angiopoietina 4, CCL18/PARC e complexo de alfa-1 antiquimotripsina forem superiores a um nível da respectiva proteína de controle, e se o nível de pelo menos um biomarcador ou ambos os biomarcadores selecionados a partir de GDF11 e a2-antiplasmina for menor do que um nível de controle da respectiva proteína.

[00026] Em algumas modalidades, métodos para a triagem de um indivíduo para o risco de um evento cardiovascular (CV) são proporcionados, compreendendo a detecção do nível de GDF11 e FSTL3 em uma amostra do indivíduo.

[00027] Em algumas modalidades, métodos para predizer a probabilidade de que um indivíduo terá um evento CV são proporcionados, compreendendo a detecção do nível de GDF11 e FSTL3 em uma amostra do indivíduo. Em algumas modalidades, métodos para predizer a probabilidade de que um indivíduo terá um evento trombótico são proporcionados, compreendendo a detecção do nível de GDF11 e FSTL3 em uma amostra do indivíduo. Em algumas modalidades, o evento trombótico é selecionado entre enfarte do miocárdio, acidente vascular cerebral e ataque isquêmico transitório.

[00028] Em algumas modalidades, a probabilidade do indivíduo ter um evento CV (tais como um evento trombótico) dentro de 4 anos é elevada se o nível de GDF11 for inferior a um nível de GDF11 e/ou o

nível de FSTL3 de controle for maior do que um nível de controle de FSTL3.

[00029] Em algumas modalidades, o método compreende a detecção do nível de MMP12. Em algumas modalidades, o método compreende a detecção do nível de angiopoietina-2. Em algumas modalidades, o método compreende a detecção do nível de complemento C7. Em algumas modalidades, o método compreende a detecção do nível de troponina cardíaca I. Em algumas modalidades, o método compreende a detecção do nível de proteína relacionada a angiopoietina 4. Em algumas modalidades, o método compreende a detecção do nível de CCL18/ PARC. Em algumas modalidades, o método compreende a detecção do nível de complexo de alfa-1-antiquimotripsina. Em algumas modalidades, o método compreende a detecção do nível de GDF11. Em algumas modalidades, o método compreende a detecção do nível de alfa-2-antiplasmina. Em algumas modalidades, o método compreende a detecção do nível de MMP12, angiopoietina-2, complemento C7, troponina I cardíaca, proteína relacionada com angiopoietina 4, CCL18/PARC, complexo de alfa-1-antiquimotripsina, GDF11 e alfa-2-antiplasmina.

[00030] Em algumas modalidades, o indivíduo tem doença da artéria coronária. Em algumas modalidades, o indivíduo não tem um histórico de eventos cardiovasculares. Em algumas modalidades, o indivíduo tem um grau de risco do American College of Cardiology (ACC) elevado. Em algumas modalidades, o indivíduo tem um grau de risco do ACC intermediário. Em algumas modalidades, o indivíduo tem um grau de risco do ACC baixo. Em algumas modalidades, o indivíduo teve pelo menos um evento CV. Em algumas modalidades, o evento CV é selecionado dentre enfarte do miocárdio, acidente vascular cerebral, insuficiência cardíaca congestiva, ataque isquêmico transgênico, e morte.

[00031] Em algumas modalidades, a amostra é selecionada dentre uma amostra de sangue, uma amostra de soro, uma amostra de plasma, e uma amostra de urina. Em algumas modalidades, a amostra é uma amostra de plasma. Em algumas modalidades, o método é realizado *in vitro*.

[00032] Em algumas modalidades, cada biomarcador é um biomarcador de proteína. Em algumas modalidades, o método compreende o contato de biomarcadores da amostra do indivíduo com um conjunto de reagentes de captura de biomarcadores, em que cada reagente de captura biomarcador do conjunto de reagentes de captura de biomarcadores se liga especificamente a um biomarcador diferente a ser detectado. Em algumas modalidades, cada reagente de captura biomarcador é um anticorpo ou um aptâmero. Em algumas modalidades, cada reagente de captura biomarcador é um aptâmero. Em algumas modalidades, pelo menos um aptâmero é um aptâmero de taxa de desaceleração lenta. Em algumas modalidades, pelo menos um aptâmero taxa de desaceleração lenta compreende, pelo menos um, pelo menos dois, pelo menos três, pelo menos quatro, pelo menos cinco, pelo menos seis, pelo menos sete, pelo menos oito, pelo menos nove, ou pelo menos 10 nucleotídeos com modificações. Em algumas modalidades, cada aptâmero de taxa de desaceleração lenta se liga à sua proteína alvo com uma taxa de desaceleração ($t_{½}$) ≥ 30 minutos, ≥ 60 minutos, ≥ 90 minutos e ≥ 120 minutos, ≥ 150 minutos, ≥ 180 minutos, ≥ 210 minutos, ou ≥ 240 minutos.

[00033] Em algumas modalidades, a probabilidade de um evento CV baseia-se nos níveis de biomarcadores e, pelo menos, um item de informação biomédica adicional selecionada de

a) informação correspondente à presença de fatores de risco cardiovascular selecionado a partir do grupo consistindo em enfarte do miocárdio prévio, evidência angiográfica de maior do que

50% de estenose em um ou mais vasos coronários, isquemia induzida por exercício em esteira ou ensaios nucleares ou revascularização coronária prévia,

b) informação correspondente ao descritores físicos do referido indivíduo,

c) informação correspondente a uma mudança no peso do referido indivíduo,

d) informação correspondente à etnia do referido indivíduo,

e) informação correspondente ao sexo do referido indivíduo,

f) informação correspondente ao histórico de fumo do referido indivíduo,

g) informação correspondente ao histórico de uso de álcool do referido indivíduo,

h) informação correspondente ao histórico ocupacional do referido indivíduo,

i) informação correspondente ao referido histórico da família do indivíduo de doenças cardiovasculares ou outras condições do sistema circulatório,

j) informação correspondente à presença ou ausência no referido indivíduo de pelo menos um marcador genético correlacionando-se com um maior risco de doença cardiovascular no referido indivíduo ou um membro da família do referido indivíduo,

k) informação correspondente aos sintomas clínicos do referido indivíduo,

l) informação correspondente a outros testes de laboratório,

m) informação correspondente a valores de expressão de genes do referido indivíduo, e

n) informação correspondente ao consumo do referido indivíduo de fatores de risco cardiovasculares conhecidos tais como a dieta rica em gorduras saturadas, sal elevado, colesterol elevado,

o) informação correspondente a resultados de imagiologia do indivíduo obtidos por técnicas selecionadas do grupo consistindo de eletrocardiograma, ecocardiograma, ultrassom da carótida para a espessura íntima-média, dilatação mediada por fluxo, velocidade de onda de pulso, índice tornozelo-braquial, ecocardiograma de estresse, cintilografia de perfusão miocárdica, cálcio coronário por tomografia, angiografia de tomografia de alta resolução, imagiologia por ressonância magnética, e outras modalidades de imagiologia,

p) informação sobre medicamentos do indivíduo, e

q) informação sobre a função renal do indivíduo.

[00034] Em algumas modalidades, o método compreende a determinação da probabilidade de um evento CV para a finalidade de determinar um seguro médico ou seguro de vida. Em algumas modalidades, o método compreende ainda a determinação de cobertura ou prêmio para o seguro médico ou seguro de vida. Em algumas modalidades, o método compreende, ainda, utilizar a informação resultante do método para predizer e/ou controlar a utilização dos recursos médicos. Em algumas modalidades, o método compreende ainda, utilizar a informação resultante do método para permitir a decisão de adquirir ou comprar uma prática médica, hospital ou empresa.

[00035] Em algumas modalidades, um método implementado por computador para avaliar o risco de um evento cardiovascular é provido. Em algumas modalidades, o método compreende a recuperação em um computador de uma informação do biomarcador para um indivíduo, em que a informação de biomarcador compreende os níveis de, pelo menos cinco, pelo menos seis, pelo menos sete, pelo menos oito, ou todos os nove biomarcadores selecionados de MMP12, angiopoietina-2, complemento C7, troponina I cardíaca, proteína relacionada com angiopoietina 4, CCL18/ PARC, complexo de alfa-1-antiquimotripsina, GDF11 e alfa-2-antiplasmina em uma amostra do indivíduo; que

executa com o computador uma classificação de cada um dos referidos valores dos biomarcadores; indicando um resultado da avaliação do risco para um evento CV para o referido indivíduo com base em uma pluralidade de classificações. Em algumas modalidades, indicar o resultado da avaliação do risco de um evento CV para o indivíduo compreende exibir o resultado em um monitor de computador.

BREVE DESCRIÇÃO DAS FIGURAS

[00036] **A FIG. 1** mostra diagramas de caixas de distribuição escala de normalização para proteínas medidas em conjuntos de descoberta e validação em cada diluição da amostra. Nos diagramas de caixa, a linha vermelha indica o valor médio, a extensão da caixa exibe o intervalo inter-quartil contendo 50% dos dados e as aparas se estendem 1,5 vezes a faixa inter-quartil para fora da caixa. As amostras com fatores de escala extrema de normalização são marcadas com o sinal "+" vermelho. Aumentos de normalização (diminuições) de níveis de sinal medianos na descoberta (validação) estabelecidos para compensar pelo viés de intensidade sistemática evidente no sinal de proteína medido nas amostras de validação.

[00037] **A FIG. 2** mostra gráficos de vulcão das razões de risco univariadas do modelo de Cox por desvio padrão de RFU (topo) ou entre quartis externos de RFU (parte inferior). A linha pontilhada horizontal indica nível de importância Bonferroni de p = 0. 05. Os nomes de genes do National Center for Biotechnology Information (NCBI) são usados como rótulos sucintos para proteínas com taxas de risco extremas. As proteínas marcadas em vermelho estão incluídas no modelo CVD9: ANGPT2 = "Angiopoietina-2"; C7 = "Complemento C7"; SERPINF2 = "F2 serina-protease inibidor" ou "α2-Antiplasmina"; CCL18 = "quimiocina (motivo C-C) ligando 18" também conhecido como "pulmonar e quimiocina regulada por ativação (PARC)"; ANGL4 = "proteína relacionada com angiopoietina 4"; KLK3. SERPINA3 =

"complexo antiquimotripsina α1-"; e TNNI3 = "troponina-I, cardíaca".

[00038] **A FIG. 3** mostra os níveis médios de sinal das proteínas CVD9 em conjuntos de descoberta e de validação e modelo de regressão linear robusto usado para estimar a intensidade de polarização residual resultante do processo de normalização.

[00039] **A FIG. 4** mostra uma comparação de risco previsto e real gerada pelo modelo de Framingham no conjunto de descoberta antes (esquerda) e depois (direita) da recalibragem com o modelo de calibração de Cox.

[00040] **A FIG. 5** mostra uma comparação de risco previsto e real gerada pelo modelo de Framingham na validação HUNT3 definida antes (esquerda) e depois (direita) da recalibração com o modelo de calibração de Cox.

[00041] **A FIG. 6** mostra um fluxograma do processo estatístico e de amostra, tal como aplicado aos conjuntos de amostra de descoberta (esquerda, cinza) e de validação (direita, rosa).

[00042] **A FIG. 7** mostra a razão de quarto univariado para razões de perigo de primeiro quartil (com intervalos de confiança de 95%) para um grupo complementar de 16 proteínas selecionadas pelo procedimento de LASSO multivariado no conjunto de descoberta (símbolos pretos, linha superior de cada par de linhas) e as mesmas proteínas a partir do conjunto de validação (símbolos vermelhos, a linha inferior de cada par de linhas). As proteínas marcadas com asterisco são incluídas no modelo paramétrico final (CVD9) depois da eliminação de trás para frente passo a passo das proteínas menos importantes. Para propriedades biológicas relevantes destas 16 proteínas, consulte exemplos. Legenda: MMP-7 = matriz metaloproteinase 7; MMP12 = metaloproteinase de matriz 12; TIM3 = imunoglobulina de células T e mucina proteína contendo o domínio 3; CCL18 = ligando de quimiocina 18 (motivo CC), anteriormente conhecido como PARC = quimiocina de

ativação regulada e pulmonar; GDF11=fator de diferenciação de crescimento 11; CDO = adesão celular associada com oncogene regulamentado; EGF = fator de crescimento epidérmico.

[00043] **A FIG. 8** mostra o desempenho de calibragem por decil de risco previsto na validação de HUNT-3 definida para CVD9 (esquerda) e Framingham (direita).

[00044] **A FIG. 9** mostra risco previsto para CVD9 (rosa) e Framingham (cinza) versus percentual de risco CVD9. Pontos sólidos indicam a frequência de evento observada para pacientes em cada decil de risco previsto gerado pelos modelos de CVD9 (rosa) e de Framingham (cinza). A linha horizontal indica a incidência de eventos em 4 anos.

[00045] **A FIG. 10** mostra as curvas ROC para o modelo aplicado ao conjunto de descoberta (preto, indicado pela seta) e conjunto de validação independente (vermelho, indicado pela seta) no ano 1 e no ano 4, o tempo máximo válido para o grau de Framingham nesta população. Também estão incluídas as curvas ROC para o grau de Framingham no coorte de descoberta (verde) e de validação (azul).

[00046] **A FIG. 11** mostra curvas de sobrevivência de Kaplan-Meier para cada quartil de risco previsto de CVD9 nos coortes de descoberta (à esquerda) e de validação (direita). As marcas mostram o tempo de censura (última observação) para indivíduos individuais e os intervalos sombreados indicam intervalos de confiança de 95%.

[00047] **A FIG. 12** ilustra um sistema de computador exemplificativo não limitativo para utilização com vários métodos implementados em computador aqui descritos.

[00048] **A FIG. 13** ilustra um ensaio de aptâmero exemplificativo não limitativo que pode ser usado para detectar um ou mais biomarcadores em uma amostra biológica.

[00049] **A FIG. 14** mostra certas pirimidinas modificadas

exemplificativas que podem ser incorporadas em aptâmeros, tais como aptâmeros de taxa de desaceleração lenta.

[00050] **A FIG. 15** mostra a correlação entre GDF11 e FSTL3.

[00051] **A FIG. 16** mostra as curvas de sobrevivência para cada quartil para cada modelo. O 1º ao 4º quartis são descritos com preto (linha de topo), vermelho (segunda linha embaixo), verde (terceira linha embaixo) e azul (linha de fundo). O sombreamento mostra os intervalos de confiança de 95%. Caractere "+" significa amostras censuradas.

[00052] **A FIG. 17** mostra uma comparação das curvas de sobrevivência entre GDF11 GDF11.FSTL3 e para o grupo de risco baixo e o grupo alto risco. No painel à esquerda, a linha superior representa o modelo GDF11.FSTL3 e a linha de fundo representa o modelo GDF11. No painel da direita, a linha superior representa o modelo GDF11 e a linha de fundo representa o modelo GDF11.FSTL3.

[00053] **A FIG. 18** mostra uma comparação da probabilidade de 4 anos entre GDF11 e GDF11.FSTL3 (esquerda) e entre FSTL3 e GDF11.FSTL3 (à direita).

[00054] **A FIG. 19** mostra a curva ROC no ano 4 para os três modelos.

[00055] **A FIG. 20** mostra as curvas de sobrevivência para cada quartil de preditor linear de cada grupo (tudo, CHF-Morte, e evento trombótico) do modelo GDF11.FSTL3. O 1º ao 4º quartis são descritos com preto (linha de topo), vermelho (segunda linha embaixo), verde (terceira linha embaixo) e azul (linha de fundo). O sombreamento mostra os intervalos de confiança de 95%.

[00056] **A FIG. 21** mostra as curvas de sobrevivência para cada quartil para os modelos GDF11, GDF11.WFIKKN1, GDF11.WFIKKN2 e GDF11.WFIKKN1.WFIKKN2.

[00057] **A FIG. 22** mostra a probabilidade de risco entre o modelo GDF11 e GDF11.WFIKKN1, GDF11.WFIKKN2 e os modelos

GDF11.WFIKKN1.WFIKKN2.

[00058] **A FIG. 23** mostra as curvas ROC para cada modelo: GDF11, GDF11.WFIKKN1, GDF11.WFIKKN2, e GDF11.WFIKKN1.WFIKKN2.

DESCRIÇÃO DETALHADA

[00059] Enquanto a invenção será descrita em conjunto com determinadas modalidades representativas, será entendido que a invenção é definida pelas reivindicações, e não é limitada a essas modalidades.

[00060] Uma pessoa versada na técnica vai reconhecer que muitos métodos e materiais semelhantes ou equivalentes descritos aqui podem ser utilizados na prática da presente invenção. A presente invenção não é de forma alguma limitada aos métodos e materiais descritos.

[00061] Salvo definição em contrário, os termos técnicos e científicos usados aqui têm o mesmo significado que é comumente compreendido por um especialista na técnica à qual pertence esta invenção. Embora quaisquer métodos, dispositivos e materiais semelhantes ou equivalentes àqueles descritos aqui possam ser usados na prática da presente invenção, certos métodos, dispositivos e materiais são descritos abaixo.

[00062] Todas as publicações, documentos de patentes publicados e pedidos de patente citados aqui estão incorporados neste documento por referência na mesma medida como se cada publicação, documento de patente publicado ou pedido de patente individual fosse indicado especificamente e individualmente para ser incorporado por referência.

[00063] Tal como utilizado neste pedido, incluindo as reivindicações anexas, as formas singulares "um", "uma" e "o" incluem o plural, a menos que o contexto dite claramente de outra forma, e podem ser utilizadas alternadamente com "pelo menos um" e "um ou mais". Assim, a referência a "um aptâmero" inclui misturas de aptâmeros, a referência a "uma sonda" inclui misturas de sondas, e semelhantes.

[00064] Como usados aqui, os termos "compreende", "compreendendo", "inclui," "incluindo", "contém," "contendo", e quaisquer variações dos mesmos, se destinam a cobrir uma inclusão não exclusiva, tal que um processo, método, processo de subproduto ou composição da matéria que compreende, inclui ou contém um elemento ou uma lista de elementos pode incluir outros elementos não expressamente enunciados.

[00065] O presente pedido inclui os biomarcadores, métodos, dispositivos, reagentes, sistemas e kits para a predição de risco de eventos cardiovasculares em curto prazo dentro de um período de tempo definido, como dentro de 1 ano, dentro de 2 anos, dentro de 3 anos, ou dentro 4 anos.

[00066] "Eventos Cardiovasculares" significam uma falha ou mau funcionamento de qualquer parte do sistema circulatório. Em uma modalidade, "Eventos Cardiovasculares" significam acidente vascular cerebral, ataque isquêmico transitório (TIA), enfarte do miocárdio (MI), morte súbita atribuída a um mau funcionamento do sistema circulatório, e/ou insuficiência cardíaca, ou morte súbita de causa desconhecida em uma população onde a causa mais provável é cardiovascular. Em uma outra modalidade, "Eventos Cardiovasculares" significam qualquer um dos mau funcionamentos antecedentes e/ou angina instável, necessidade de uma prótese endovascular ou de angioplastia, ou semelhantes.

[00067] Eventos cardiovasculares incluem "insuficiência cardíaca congestiva" ou "CHF" e "eventos trombóticos. " Eventos trombóticos incluem infartos, acidentes isquêmicos transitórios (TIA), acidente vascular cerebral, síndrome coronariana aguda e necessidade de revascularização coronária.

[00068] Em certas modalidades, os biomarcadores são proporcionados para uso por si só ou em várias combinações para

avaliar o risco de morte súbita ou um evento CV futuro dentro de um período de tempo de 4 anos com eventos CV definidos como enfarte do miocárdio, acidente vascular cerebral, morte e insuficiência cardíaca congestiva. Eventos trombóticos consistem de infarto do miocárdio e acidente vascular cerebral combinados. Tal como descrito em detalhe abaixo, modalidades exemplificativas incluem os biomarcadores proporcionados na Tabela 3, que foram identificados utilizando um ensaio baseado em SOMAmer multiplex que está descrito genericamente nos Exemplos.

[00069] Enquanto alguns dos biomarcadores de eventos CV descritos podem ser úteis por si só para avaliar o risco de um evento CV, os métodos são também descritos aqui para o agrupamento dos vários subconjuntos dos marcadores de eventos CV, onde cada agrupamento ou seleção de subconjunto é útil como um painel de três ou mais biomarcadores, indiferentemente aqui referido como um "painel de biomarcador" e um painel. Assim, várias modalidades do presente pedido de patente proveem combinações que compreendem, pelo menos cinco, pelo menos seis, pelo menos sete, pelo menos oito, ou todos os nove biomarcadores na Tabela 3.

[00070] Em uma modalidade, o número de biomarcadores úteis para um subconjunto ou painel de biomarcador baseia-se na sensibilidade e valor de especificidade para a combinação particular de valores dos biomarcadores. Os termos "sensibilidade" e "especificidade" são aqui utilizados no que diz respeito à capacidade de classificar corretamente um indivíduo, com base em um ou mais valores dos biomarcadores detectados na sua amostra biológica, como tendo um risco aumentado de ter um evento CV dentro de 4 anos ou não ter risco aumentado de ter um evento CV dentro do mesmo período de tempo. "Sensibilidade" indica o desempenho do(s) biomarcador (es) com respeito à classificação correta dos indivíduos que têm risco aumentado de um

evento CV. "Especificidade" indica o desempenho do(s) biomarcador(es) com respeito à classificação correta dos indivíduos que não têm maior risco de um evento CV. Por exemplo, 85% de especificidade e 90% de sensibilidade para um painel de marcadores utilizados para testar um conjunto de amostras de evento negativo e as amostras de evento positivo indicam que 85% das amostras de controle foram corretamente classificadas como amostras de eventos negativos pelo painel, e 90% das amostras de evento positivo foram corretamente classificados como amostras de evento positivo pelo painel.

[00071] Em um método alternativo, os resultados podem ser relatados em uma faixa contínua, com um limite de risco alto, intermediário ou baixo de um evento CV dentro de uma unidade de tempo definido, com limites determinados com base em descobertas clínicas; uma expressão alternativa dos mesmos dados consiste em fixar o limite de probabilidade (tal como 50%) e prever o momento em que esta proporção de indivíduos teria seu evento (por exemplo, de forma análoga a meia-vida no decaimento radioativo, o tempo em que metade do isótopo teria deteriorado).

[00072] Um fator que pode afetar o número de biomarcadores a ser usado em um subconjunto ou o painel de biomarcadores são os procedimentos utilizados para obter amostras biológicas a partir de indivíduos vão ser avaliados para o risco de um evento CV. Em um ambiente de coleta de amostra cuidadosamente controlado, o número de biomarcadores necessário para alcançar a sensibilidade e a especificidade e/ou valores de limite desejados será menor do que em uma situação em que não pode haver mais variação na coleta, manuseio e armazenagem de amostras. Alternativamente, uma maior sensibilidade e especificidade podem ser obtidas através utilização de mais marcadores que são menos robustos para a aquisição da amostra (por exemplo, que não sobrevive em situação de coleta variável),

juntamente com marcadores de manipulação de amostras que permitem a rejeição de amostras mal coletadas ou a eliminação de marcadores sensíveis do algoritmo de predição de risco.

[00073] "Amostra biológica", "amostra", e "amostra de teste" são aqui utilizados indiferentemente para se referir a qualquer material, fluido biológico, tecido ou célula obtida de outra forma ou derivado de um indivíduo. Isto inclui sangue (incluindo sangue total, leucócitos, células mononucleares do sangue periférico, camada leucocitária, plasma e soro), saliva, lágrimas, muco, lavagens nasais, aspirado nasal, urina, saliva, lavagens peritoneais, ascites, fluido cístico, fluido glandular, fluido linfático, aspirado brônquico, líquido sinovial, aspirado conjunta, secreções de órgãos, células, um extrato celular, e o líquido cefalorraquidiano. Isto também inclui frações experimentalmente separadas de todos os precedentes. Por exemplo, uma amostra de sangue pode ser fracionada em soro, plasma, ou em frações contendo determinados tipos de células do sangue, tais como glóbulos vermelhos ou células sanguíneas brancas (leucócitos). Em algumas modalidades, uma amostra de sangue é uma mancha de sangue seco. Em algumas modalidades, uma amostra de plasma é uma mancha de plasma seco. Em algumas modalidades, uma amostra pode ser uma combinação de amostras de um indivíduo, tal como uma combinação de um tecido e a amostra de fluido. O termo "amostra biológica" inclui também materiais que contêm material sólido homogeneizado, tais como uma amostra de fezes, uma amostra de tecido ou uma biopsia de tecido, por exemplo. O termo "amostra biológica" também inclui materiais derivados de uma cultura de tecido ou uma cultura de células. Quaisquer métodos adequados para a obtenção de uma amostra biológica podem ser empregados; métodos exemplificativos incluem, por exemplo, flebotomia, zaragatoa (por exemplo, cotonete bucal), e um procedimento de biópsia de aspiração de agulha fina. Tecidos

exemplificativos suscetíveis à aspiração com agulha fina incluem linfonodo, pulmão, tireoide, mama, pâncreas e fígado. As amostras também podem ser coletadas, por exemplo, por microdissecação (por exemplo, captura de laser de microdissecação (LCM) ou microlaser de dissecação (LMD)), lavagem de bexiga, esfregaço (por exemplo, um exame de Papanicolaou), ou lavagem ductal. Uma "amostra biológica" obtida ou derivada de um indivíduo inclui qualquer amostra que tenha sido processada em qualquer forma adequada depois de ser obtida do indivíduo. Em algumas modalidades, uma amostra biológica é uma amostra de plasma.

[00074] Além disso, em algumas modalidades, uma amostra biológica pode ser obtido por coleta de amostras biológicas de um número de indivíduos e reunindo-os, ou reunindo-se uma alíquota da amostra biológica de cada indivíduo. A amostra combinada pode ser tratada tal como aqui descrito para uma amostra de um único indivíduo, e, por exemplo, se um mau prognóstico é estabelecido na amostra reunida, em seguida, cada amostra biológica individual pode ser retestada para determinar qual(is) indivíduo(s) têm um risco aumentado ou diminuído de um evento CV.

[00075] Para os fins desta especificação, a frase "dados atribuídos a uma amostra biológica de um indivíduo" pretende significar que os dados de alguma forma derivada, ou foram gerados utilizando, a amostra biológica do indivíduo. Os dados podem ter sido reformatados, revistos, ou matematicamente alterados até certo ponto depois de terem sido gerados, por exemplo, por conversão das unidades de um sistema de medição para as unidades no outro sistema de medição; mas, os dados são entendidos como tendo sido derivados de, ou foram gerados utilizando, a amostra biológica.

[00076] "Alvo", "molécula alvo", e "analito" são aqui utilizados indiferentemente para se referir a qualquer molécula de interesse que

possa estar presente em uma amostra biológica. Uma "molécula de interesse" inclui qualquer alteração menor de uma molécula em particular, tal como, no caso de uma proteína, por exemplo, pequenas variações na sequência de aminoácidos, formação de ligação de dissulfeto, glicosilação, lipidação, acetilação, fosforilação, ou qualquer outra manipulação ou modificação, tais como conjugação com uma componente de marcação, que não altera substancialmente a identidade da molécula. Uma "molécula alvo", "alvo", ou "analito" se refere a um conjunto de cópias de um tipo ou espécie de molécula ou estrutura multimolecular. "Moléculas alvo", "alvos", e "analitos" referem-se a mais do que um tipo ou espécie de molécula ou estrutura multimolecular. Moléculas alvo exemplificativas incluem proteínas, polipeptideos, ácidos nucleicos, hidratos de carbono, lipídeos, polissacarídeos, glicoproteínas, hormônios, receptores, antígenos, anticorpos, aficorpos, imitadores de anticorpos, vírus, agentes patogênicos, substâncias tóxicas, substratos, metabólitos e análogos de estado de transição, cofatores, inibidores, drogas, corantes, nutrientes, fatores de crescimento, células, tecidos, e qualquer fragmento ou parte de qualquer um dos precedentes. Em algumas modalidades, uma molécula alvo é uma proteína, no caso em que a molécula alvo pode ser referida como uma "proteína alvo".

[00077] Tal como aqui utilizado, um "agente de captura" ou "reagente de captura" se refere a uma molécula que é capaz de se ligar especificamente a um biomarcador. Um "reagente de captura de proteína alvo" se refere a uma molécula que é capaz de se ligar especificamente a uma proteína-alvo. Aptâmeros de reagentes de captura exemplificativos não limitativos, anticorpos, adnectinas, anquirinas, outros miméticos de anticorpos e outros suportes proteicos, autoanticorpos, quimeras, pequenas moléculas, ácidos nucleicos, lectinas, receptores de ligação de ligandos, polímeros imprimidos,

avímeros, peptidomiméticos, receptores de hormônios, receptores de citocina, receptores sintéticos, e modificações e fragmentos de qualquer um dos reagentes de captura acima mencionados. Em algumas modalidades, um reagente de captura é selecionado a partir de um aptâmero e um anticorpo.

[00078] O termo "anticorpo" se refere a anticorpos de comprimento total de todas as espécies e fragmentos e derivados de tais anticorpos, incluindo fragmentos Fab, $F(ab')_2$ fragmentos, anticorpos de cadeia simples, fragmentos Fv e fragmentos Fv de cadeia simples. O termo "anticorpo" também se refere a anticorpos sinteticamente derivados, tais como anticorpos de fagos derivados de exibição e fragmentos, aficorpos, nanocorpos, etc.

[00079] Tal como aqui utilizado, "marcador" e "biomarcador" são utilizados indiferentemente para se referir a uma molécula alvo que indica ou é um sinal de um processo normal ou anormal em um indivíduo ou de uma doença ou outra condição de um indivíduo. Mais especificamente, um "marcador" ou "biomarcador" é um parâmetro anatômico, fisiológico, bioquímico, ou molecular associado com a presença de um estado fisiológico específico ou processo, normal ou anormal, e, se anormal, seja crônico ou agudo. Os biomarcadores são detectáveis e mensuráveis por uma variedade de métodos, incluindo ensaios de laboratório e imagiologia médica. Em algumas modalidades, um biomarcador alvo é uma proteína.

[00080] Tal como aqui utilizado, "nível de biomarcador" e "nível" referem-se a uma medida que é feita através de qualquer método analítico para a detecção do marcador biológico em uma amostra biológica e que indica a presença, ausência, quantidade absoluta ou a concentração, a quantidade relativa ou concentração, título, um nível, um nível de expressão, uma proporção de níveis medidos, ou semelhantes, de, por, ou correspondente ao biomarcador na amostra

biológica. A natureza exata do "nível" depende da estrutura e dos componentes do método analítico particular empregado para detectar o marcador biológico específico.

[00081] Quando um biomarcador indica ou é um sinal de um processo anormal ou uma doença ou outra condição de um indivíduo, este biomarcador é geralmente descrito como sendo ou sobre-expressado ou sub-expressado, em comparação com um nível de expressão ou o valor do biomarcador que indica ou é um sinal de um processo normal ou uma ausência de uma doença ou outra condição de um indivíduo. "Suprarregulação", "suprarregulado", "sobre-expressão", "sobre-expressado", e quaisquer das suas variantes, são utilizados indiferentemente para se referir a um valor ou nível de um biomarcador em uma amostra biológica que é maior do que um valor ou nível (ou intervalo de valores ou níveis) do biomarcador que é normalmente detectado em amostras biológicas semelhantes a partir de indivíduos saudáveis ou normais. Os termos podem também referir-se a um valor ou nível de um biomarcador em uma amostra biológica que é maior do que um valor ou nível (ou intervalo de valores ou níveis) do biomarcador que possa ser detectado em uma fase diferente de uma doença particular.

[00082] "Infrarregulação", "infrarregulado", "subexpressão", "subexpressado", e quaisquer das suas variantes, são utilizados indiferentemente para se referir a um valor ou nível de um biomarcador em uma amostra biológica que é inferior a um valor ou nível (ou intervalo de valores ou níveis) do biomarcador que é normalmente detectado em amostras biológicas semelhantes a partir de indivíduos saudáveis ou normais. Os termos podem se referir também a um valor ou nível de um biomarcador em uma amostra biológica que é inferior a um valor ou nível (ou intervalo de valores ou níveis) do biomarcador que possa ser detectado em uma fase diferente de uma doença em particular.

[00083] Além disso, um biomarcador que é ou sobre-expressado ou sub-expressado também pode ser referido como sendo "expressado diferencialmente" ou como tendo um "nível diferencial" ou "valor diferencial", em comparação com um nível de expressão "normal" ou valor do biomarcador que indica ou é um sinal de um processo normal ou uma ausência de uma doença ou outra condição de um indivíduo. Assim, "expressão diferencial" de um biomarcador pode também ser referida como uma variação a partir de um nível "normal" de expressão do biomarcador.

[00084] Um "nível de controle" de uma molécula alvo se refere ao nível da molécula alvo no mesmo tipo de amostra de um indivíduo que não possui a doença ou condição, ou de um indivíduo que não é suspeito ou está em risco de ter a doença ou condição, ou de um indivíduo que teve um primeiro evento cardiovascular ou um evento cardiovascular primário, mas não um evento cardiovascular secundário, ou de um indivíduo que tem a doença cardiovascular estável. Nível de controle pode se referir ao nível médio da molécula alvo em amostras a partir de uma população de indivíduos que não têm a doença ou condição, ou que não se suspeita ou está em risco de ter a doença ou condição, ou que tenha tido um primeiro evento cardiovascular ou um evento cardiovascular primário mas não um evento cardiovascular secundário, ou que tem doença cardiovascular estável ou uma combinação dos mesmos.

[00085] Tal como aqui utilizado, "indivíduo", "indivíduo" e "paciente" são utilizados indiferentemente para se referir a um mamífero. Um indivíduo mamífero pode ser um ser humano ou não humano. Em várias modalidades, o indivíduo é um humano. Um indivíduo saudável ou normal é um indivíduo em que a doença ou condição de interesse (incluindo, por exemplo, eventos cardiovasculares, tais como enfarte do miocárdio, acidente vascular cerebral e insuficiência cardíaca

congestiva) não é detectável por técnicas de diagnóstico convencionais.

[00086] "Diagnosticar", "diagnóstico", "diagnose" e variações referem-se mesmo à detecção, determinação, ou o reconhecimento de um estado de saúde ou condição de um indivíduo com base em um ou mais sinais, sintomas, dados ou outras informações relativas a este indivíduo. O estado de saúde de um indivíduo pode ser diagnosticado como saudável/ normal (isto é, um diagnóstico da ausência de uma doença ou condição) ou diagnosticado como doente/ anormal (isto é, um diagnóstico da presença ou de uma avaliação das características, de uma doença ou condição). Os termos "diagnosticar", "diagnóstico", "diagnose", etc., abrangem, no que diz respeito a uma doença ou condição em particular, a detecção inicial da doença; a caracterização e classificação da doença; a detecção da progressão, a remissão, ou a recorrência da doença; e a detecção da doença de resposta após a administração de um tratamento ou terapia para o indivíduo. A predição de risco de um evento CV inclui a distinção de indivíduos que têm um risco aumentado de um evento CV de pessoas que não têm.

[00087] "Prognosticar", "prognóstico", "prognose", e variações do mesmo referem-se à predição de um curso futuro de uma doença ou condição em um indivíduo que tem a doença ou condição (por exemplo, prevendo a sobrevivência do paciente), e estes termos englobam a avaliação de resposta à doença ou condição após a administração de um tratamento ou terapia para o indivíduo.

[00088] "Avaliar", "avaliando", "avaliação" e suas variações abranger ambos "diagnosticar" e "prognosticar" e também englobam determinações ou previsões sobre o curso futuro de uma doença ou condição nem um indivíduo que não tem a doença, bem como determinações ou previsões sobre o risco de que uma doença ou condição que se repete em um indivíduo que aparentemente foi curado da doença ou que teve a condição resolvida. O termo "avaliação"

também abrange a avaliação da resposta de um indivíduo a uma terapia, tal como, por exemplo, prever se um indivíduo é suscetível de responder favoravelmente a um agente terapêutico ou é pouco provável de responder a um agente terapêutico (ou vai experimentar efeitos colaterais indesejáveis ou tóxicos, por exemplo), a seleção de um agente terapêutico para administração a um indivíduo, ou de controle ou determinação da resposta de um indivíduo a uma terapia que foi administrada ao indivíduo. Assim, "avaliar" o risco de um evento cardiovascular pode incluir, por exemplo, qualquer um dos seguintes: prever o risco futuro de um evento cardiovascular em um indivíduo; prever o risco de um evento cardiovascular em um indivíduo que aparentemente não tem problemas CV; prever um tipo particular de evento CV; prever o tempo para um evento CV; ou determinar ou prever a resposta de um indivíduo a um tratamento CV ou selecionar um tratamento CV para administrar a um indivíduo com base em uma determinação dos valores dos biomarcadores provenientes de amostra biológica do indivíduo. Avaliação de risco de um evento CV pode incluir modalidades, tais como a avaliação do risco de um evento CV em uma escala contínua, ou classificação de risco de um evento CV na classificação em escala. Classificação de risco inclui, por exemplo, a classificação em duas ou mais classificações tais como "sem risco elevado de um evento CV", "risco elevado de um evento CV"; e/ou "risco abaixo da média de evento CV". Em algumas modalidades, a avaliação do risco de um evento CV é para um período de tempo definido. Períodos não limitativos exemplificativos definidos incluem 1 ano, 2 anos, 3 anos, 4 anos, 5 anos e mais de 5 anos.

[00089] Tal como utilizado aqui, "informação biomédica adicional" se refere a uma ou mais avaliações de um indivíduo, exceto utilizando qualquer um dos marcadores aqui descritos, que estão associados com o risco cardiovascular ou, mais especificamente, o risco de eventos CV.

"Informação biomédica adicional" inclui qualquer dos seguintes procedimentos: descritores físicos de um indivíduo, incluindo a altura e/ou peso de um indivíduo; a idade de um indivíduo; o sexo de um indivíduo; alteração no peso; a etnia de um indivíduo; histórico ocupacional; histórico familiar de doença cardiovascular (ou outras desordens do sistema circulatório); a presença de um marcador(es) genético correlacionando-se com um maior risco de doença cardiovascular (ou outras desordens do sistema circulatório) no indivíduo ou de um membro da família; alterações na espessura da carótida íntima; sintomas clínicos, tais como dor no peito, valores de expressão de gene de ganho ou perda de peso; descritores físicos de um indivíduo, incluindo descritores físicos observados por exames radiológicos; condição de fumante; histórico de uso de álcool; histórico ocupacional; hábitos alimentares - sal, gordura saturada e colesterol; o consumo de cafeína; e informações de imagiologia, como eletrocardiograma, ecocardiograma, ultrassom da carótida para a espessura íntima-média, dilatação mediada por fluxo, a velocidade da onda de pulso, índice de tornozelo-braço, ecocardiograma de estresse, cintilografia de perfusão miocárdica, cálcio coronário por tomografia, angiografia de tomografia de alta resolução, imagiologia de ressonância magnética, e outras modalidades de imagiologia; e os medicamentos do indivíduo. Testes de níveis de biomarcadores em combinação com uma avaliação de qualquer informação biomédica adicional, incluindo outros testes de laboratório (por exemplo, HDL, testes de LDL, níveis de PCR, ensaios de NT-proBNP, BNP, a dosagem de troponina de alta sensibilidade, testes de galectina-3, o teste de albumina de soro, ensaios de creatina), podem, por exemplo, melhorar a sensibilidade, a especificidade e/ou a AUC para a predição de eventos cardiovasculares em comparação com testes de biomarcador isoladamente ou avaliar qualquer elemento de informação biomédica complementar por si só

(por exemplo, imagiologia de espessura da carótida íntima sozinha). Informação biomédica adicional pode ser obtida a partir de um indivíduo utilizando técnicas de rotina conhecidas na técnica, tais como a partir dos próprios indivíduos através da utilização de um questionário sobre a rotina do paciente ou questionário do histórico de saúde, etc., ou de um profissional médico, etc. Testes de níveis de biomarcador em combinação com uma avaliação de qualquer informação biomédica adicional podem, por exemplo, melhorar a sensibilidade, especificidade e/ou os limites para predição de eventos cardiovasculares (ou outros usos cardiovasculares relacionadas), em comparação com testes de biomarcador isolados ou avaliação de qualquer item particular de informação biomédica adicional por si só (por exemplo, imagiologia de tomografia sozinha).

[00090] Tal como aqui utilizado, "detectar" ou "determinar" em relação a um valor de biomarcador inclui o uso ambos ao instrumento utilizado para observar e registar um sinal que corresponde a um nível de biomarcadores e ao material/ is necessário(s) para gerar esse sinal. Em várias modalidades, o nível de biomarcador é detectado utilizando qualquer método adequado, incluindo fluorescência, quimioluminescência, ressonância de plasma de superfície, as ondas acústicas, espectrometria de massa, espectroscopia de infravermelho, espectroscopia Raman, microscopia de força atômica, por microscopia de encapsulamento de varredura, métodos de detecção eletroquímica, ressonância magnética nuclear, pontos quânticos, e semelhantes.

[00091] Tal como aqui utilizado, um "grau de risco do American College of Cardiology (ACC) Resultado" é determinado de acordo com Goff et ai., "2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines," published online in *Circulation* on November 12, 2013 (Print

ISSN: 0009-7322, Online ISSN: 1524-4539). Como aqui utilizado, um grau de risco "alto" é um risco previsto de 10 anos de 20,0% ou mais para um evento de doença aterosclerótica/ cardiovascular dura (ASCVD) (definido como primeira ocorrência de infarto do miocárdio não fatal ou morte por doença cardíaca coronária (CHD), ou acidente vascular cerebral fatal ou não); um grau de risco "intermediário" é um risco de 10 anos previsto de 10,0-19,9% para um evento ASCVD dura; e um grau de "baixo" risco é um risco previsto de 10 anos de <10,0% d para um evento ASCVD dura. *Ver* Goff na página 16, tabela 5.

[00092] "Suporte sólido" se refere aqui a qualquer substrato que tem uma superfície à qual podem ser ligadas moléculas, direta ou indiretamente, através de ligações quer covalentes ou não covalentes. Um "suporte sólido" pode ter uma variedade de formatos físicos, que podem incluir, por exemplo, uma membrana; um chip (por exemplo, um chip de proteína); uma lâmina (por exemplo, uma lâmina de vidro ou lamínula); uma coluna; um poço, sólido, semissólido, ou poço de poro contendo partículas, tais como, por exemplo, um grânulo; um gel; uma fibra, incluindo um material de fibra óptica; uma matriz; e um recipiente de amostra. Recipientes de amostra exemplificativos incluem poços de amostra, tubos, capilares, frascos, e qualquer outro recipiente, ranhura ou indentação capaz de manter uma amostra. Um receptáculo de amostra pode ser contido em uma plataforma multiamostra, tal como uma placa de microtitulação, uma lâmina, o dispositivo de microfluidos, e semelhantes. Um suporte pode ser composto por um material natural ou sintético, um material orgânico ou inorgânico. A composição do suporte sólido na qual estão ligados os reagentes de captura geralmente depende do modo de fixação (por exemplo, afixação covalente). Outros recipientes exemplificativos incluem emulsões de microgotículas e microfluidas controladas e/ou óleo em massa/ aquoso dentro dos quais os ensaios e as manipulações relacionadas podem

ocorrer. Os suportes sólidos adequados incluem, por exemplo, plásticos, resinas, polissacarídeos, sílica ou materiais à base de sílica, vidro funcionalizado, silicone modificado, carbono, metais, vidros inorgânicos, membranas de nylon, fibras naturais (como, por exemplo, seda, lã e algodão), polímeros, e semelhantes. O material que compõe o suporte sólido pode incluir grupos reativos tais como, por exemplo, carboxi, amino ou grupos hidroxil, que são utilizados para a afixação dos reagentes de captura. Suportes sólidos poliméricos podem incluir, por exemplo, poliestireno, tetraftalato polietileno glicol, acetato de polivinil, cloreto de polivinil, pirrolidona de polivinil, poliacrilonitril, polimetacrilato de metil, politetrafluoroetileno, borracha de butil, borracha de estireno-butadieno, borracha natural, polietileno, polipropileno, (poli) tetrafluoroetileno, (poli) fluoreto de vinilideno, policarbonato, e polimetilpenteno. Partículas de suporte sólido adequadas que podem ser utilizadas incluem, por exemplo, partículas codificadas, tais como partículas de tipo Luminex® codificadas, partículas magnéticas e as partículas de vidro.

USOS EXEMPLIFICATIVOS DE BIOMARCADORES

[00093] Em várias modalidades exemplificativas, são providos métodos para a avaliação do risco de um evento cardiovascular em um indivíduo através da detecção de um ou mais valores de biomarcadores que corresponde a um ou mais biomarcadores que estão presentes na circulação de um indivíduo, tal como no soro ou no plasma, por algum número de métodos analíticos, incluindo qualquer um dos métodos analíticos descritos aqui. Esses biomarcadores são, por exemplo, expressados diferencialmente em indivíduos com um risco aumentado de um evento CV, em comparação com os indivíduos sem o aumento do risco de um evento CV. A detecção da expressão diferencial de um biomarcador em um indivíduo pode ser utilizada, por exemplo, para permitir a predição de risco de um evento CV dentro de 1 ano, 2 anos, 3

anos, 4 anos, ou período de tempo de 5 anos.

[00094] Além de testar os níveis de biomarcadores como um teste de diagnóstico autônomo, os níveis de biomarcadores também podem ser feito em conjunto com a determinação de polimorfismos individuais (SNPs) ou de outras lesões genéticas ou variabilidade que são indicativos de um risco aumentado de suscetibilidade a uma doença ou condição. (Consulte, por exemplo, Amos et al., Nature Genetics 40, 616-622 (2009)).

[00095] Além de testar os níveis de biomarcadores como um teste de diagnóstico autônomo, os níveis de biomarcadores também podem ser usados em conjunto com a triagem radiológica. Os níveis de biomarcadores também podem ser usados em conjunto com os sintomas relevantes ou os testes genéticos. A detecção de qualquer dos biomarcadores aqui descritos pode ser útil após o risco de eventos CV ter sido avaliado para guiar o cuidado clínico adequado do indivíduo, incluindo o aumento de níveis mais agressivos de tratamento em indivíduos de alto risco, após o risco de eventos CV ter sido determinado. Além dos níveis de teste de biomarcadores em conjunto com os sintomas relevantes ou fatores de risco, informações sobre os biomarcadores podem também ser avaliadas em conjunto com outros tipos de dados, particularmente dados que indicam o risco de eventos cardiovasculares de um indivíduo (por exemplo, o histórico clínico do paciente e os sintomas, histórico familiar de doença cardiovascular, histórico de consumo de álcool ou fumo, fatores de risco, tais como a presença de um marcador(es) genético(s), e/ou o estado de outros biomarcadores, etc.). Estes vários dados podem ser avaliados por métodos automatizados, tais como um programa de computador/ software, que podem ser incorporados em um computador ou outro aparelho/ dispositivo.

[00096] Além dos níveis de teste de biomarcadores em conjunto com

a triagem radiológica em indivíduos de alto risco (por exemplo, avaliação dos níveis de biomarcadores em conjunto com o bloqueio detectado em um angiograma coronário), informações sobre os biomarcadores podem também ser avaliadas em conjunto com outros tipos de dados, particularmente de dados que indicam o risco de um indivíduo de ter um evento CV (por exemplo, o histórico clínico do paciente, sintomas, histórico familiar de doença cardiovascular, fatores de risco, tais como se o indivíduo é ou não um fumante, o usuário pesado de álcool e/ou o estado de outros biomarcadores, etc.). Estes vários dados podem ser avaliados por métodos automatizados, tais como um programa de computador/ software, que podem ser incorporados em um computador ou outro aparelho/ dispositivo.

[00097] Ensaios de biomarcadores também podem ser associados com as diretrizes e os algoritmos de risco cardiovascular atualmente em utilização na prática clínica. Por exemplo, o Grau de Risco de Framingham usa fatores de risco para proporcionar um grau de risco, tais fatores de risco, incluindo colesterol LDL e os níveis de colesterol HDL, os níveis de glicose diminuída, tabagismo, pressão arterial sistólica e diabetes. A frequência de pacientes de alto risco aumenta com a idade, e homens compreendem uma maior proporção de pacientes de alto risco que as mulheres.

[00098] Qualquer um dos biomarcadores descritos também pode ser utilizado em ensaios de imagiologia. Por exemplo, um agente de imagiologia pode ser acoplado a qualquer um dos biomarcadores descritos, que podem ser utilizados para auxiliar na predição de risco de um evento cardiovascular, para monitorar a resposta a intervenções terapêuticas, para selecionar populações alvo em um ensaio clínico entre outras utilizações.

DETECÇÃO E DETERMINAÇÃO DE BIOMARCADORES E NÍVEIS DE BIOMARCADORES

[00099] Um nível de marcador biológico para os biomarcadores aqui descritos pode ser detectado utilizando qualquer um de uma variedade de métodos analíticos conhecidos. Em uma modalidade, um valor de biomarcador é detectado usando um reagente de captura. Em várias modalidades, o reagente de captura pode ser exposto ao biomarcador em solução ou pode ser exposto ao biomarcador enquanto o reagente de captura está imobilizado em um suporte sólido. Em outras modalidades, o reagente de captura contém uma funcionalidade que é reativa com uma característica secundária em um suporte sólido. Nestas modalidades, o reagente de captura pode ser exposto ao biomarcador em solução, e, em seguida, a característica do reagente de captura pode ser usada em conjunto com o elemento secundário sobre o suporte sólido para imobilizar o biomarcador no suporte sólido. O reagente de captura é selecionado com base no tipo de análise a ser conduzida. Reagentes de captura incluem, mas não estão limitados a, anticorpos, aptâmeros adnectinas, anquirinas, outros miméticos de anticorpos e outros suportes proteicos, autoanticorpos, quimeras, moléculas pequenas, F $(ab')_2$ fragmentos, fragmentos de anticorpo de cadeia única, fragmentos Fv, fragmentos Fv de cadeia simples, ácidos nucleicos, lectinas, receptores de ligação ao ligando, aficorpos, nanocorpos, polímeros impressos, avímeros, peptidomiméticos, receptores de hormônios, receptores de citocinas e receptores sintéticos, e modificações e fragmentos destes.

[000100] Em algumas modalidades, um nível de biomarcador é detectado usando um de complexo de reagente de biomarcador/ captura.

[000101] Em algumas modalidades, o nível de biomarcador é derivado a partir do de complexo de reagente de biomarcador/ captura e é detectado indiretamente, tal como, por exemplo, como resultado de uma reação que é posterior à interação reagente de biomarcador/

captura, mas está dependente da formação do complexo de reagente de biomarcador/ captura.

[000102] Em algumas modalidades, o nível de biomarcador é detectado diretamente com o biomarcador em uma amostra biológica.

[000103] Em algumas modalidades, os biomarcadores são detectados utilizando um formato de multiplexagem que permite a detecção simultânea de dois ou mais biomarcadores em uma amostra biológica. Em algumas modalidades do formato multiplexado, reagentes de captura são imobilizados, direta ou indiretamente, covalentemente ou não covalentemente, em localizações discretas sobre um suporte sólido. Em algumas modalidades, um formato de multiplexagem utiliza suportes sólidos discretos, em que cada suporte sólido tem um único reagente de captura associado com esse suporte sólido, tal como, por exemplo, pontos de quantum. Em algumas modalidades, um dispositivo individual é utilizado para a detecção de cada um dos biomarcadores múltiplos a ser detectado em uma amostra biológica. Os dispositivos individuais podem ser configurados para permitir que cada biomarcador na amostra biológica seja processado simultaneamente. Por exemplo, uma placa de microtitulação pode ser utilizada de tal modo que cada poço da placa que é utilizada para analisar unicamente um ou mais biomarcadores a serem detectados em uma amostra biológica.

[000104] Em uma ou mais das modalidades precedentes, um marcador fluorescente pode ser utilizado para marcar um componente do complexo de reagentes de biomarcador/ captura para permitir a detecção do nível de biomarcador. Em várias modalidades, o marcador fluorescente pode ser conjugado com um reagente de captura específico para qualquer um dos marcadores aqui descritos, utilizando técnicas conhecidas, e o marcador fluorescente pode então ser utilizado para detectar o nível do biomarcador correspondente. Marcadores fluorescentes adequados incluem quelatos de terras raras, a

fluoresceína e seus derivados, rodamina e seus derivados, dansilo, aloficocianina, PBXL-3, Qdot 605, Lissamina, ficoeritrina, Vermelho Texas, e outros tais compostos.

[000105] Em algumas modalidades, o marcador fluorescente é uma molécula de corante fluorescente. Em algumas modalidades, a molécula de corante fluorescente inclui, pelo menos, um sistema de anel indol substituído em que o substituinte no carbono 3 do anel de indol contém um grupo quimicamente reativo ou uma substância conjugada. Em algumas modalidades, a molécula de corante inclui uma molécula de AlexFluor, tais como, por exemplo, AlexaFluor 488, AlexaFluor 532, AlexaFluou 647, AlexaFluor 680, ou AlexaFluor 700. Em outras modalidades, a molécula de corante inclui um primeiro tipo e um segundo tipo de molécula de corante, tal como, por exemplo, duas moléculas diferentes AlexaFluor. Em algumas modalidades, a molécula de corante inclui um primeiro tipo e um segundo tipo de molécula de corante, e as duas moléculas de corante têm diferentes espectros de emissão.

[000106] A fluorescência pode ser medida com uma variedade de instrumentação compatível com uma grande variedade de formatos de ensaio. Por exemplo, espectrofluorímetros foram concebidos para analisar placas de microtitulação, lâminas de microscópio, matrizes impressas, cubetas, etc. Ver Principles of Fluorescence Spectroscopy, de J. R. Lakowicz, Springer Science Business Media, Inc., 2004. Ver Bioluminescence & Chemiluminescence: Progress & Current Applications; Philip E. Stanley e Larry J. Kricka editors, World Scientific Publishing Company, January 2002.

[000107] Em uma ou mais modalidades, uma marcação de quimioluminescência pode ser opcionalmente utilizada para marcar um componente do complexo de biomarcador/ captura para permitir a detecção de um nível de biomarcador. Materiais quimioluminescentes

adequados incluem qualquer um dentre cloreto de oxalilo, Rodamin 6G, Ru $(bipi)_3^{2+}$, TMAE (tetraquis (dimetilamino) etileno), Pirogalol (1,2,3-trihydroxibenzene), lucigenina, peroxioxalates, oxalatos de arilo, ésteres de acridinio, dioxetanos, e outros.

[000108] Em algumas modalidades, o método de detecção inclui uma combinação de enzima/ substrato que gera um sinal detectável que corresponde ao nível de biomarcador. De um modo geral, a enzima catalisa uma alteração química do substrato cromogênico que pode ser medida utilizando várias técnicas, incluindo espectrofotometria, fluorescência, e quimioluminescência. As enzimas adequadas incluem, por exemplo, as luciferases, luciferina, malato-desidrogenase, urease, peroxidase de rábano (HRPO), fosfatase alcalina, beta-galactosidase, glucoamilase, lisozima, glucose-oxidase, galactose-oxidase, e desidrogenase de glucose-6-fosfato, uricase, xantina oxidase, lactoperoxidase, microperoxidase, e semelhantes.

[000109] Em algumas modalidades, o método de detecção pode ser uma combinação de fluorescência, quimioluminescência, radionuclídeos ou combinações de enzima/ substrato que geram um sinal mensurável. Em algumas modalidades, a sinalização multimodal pode ter características únicas e vantajosas em formatos de ensaio biomarcador.

[000110] Em algumas modalidades, os níveis de biomarcadores para os biomarcadores aqui descritos podem ser detectados utilizando quaisquer métodos analíticos, incluindo, ensaios de aptâmeros simplex, ensaios multiplexados aptâmeros, imunoensaios simplex ou multiplexados, perfil de expressão mRNA, perfis de expressão miRNA, análise por espectrometria de massa, métodos histológicos/ citológicos, etc., como discutido abaixo.

DETERMINAÇÃO DOS NÍVEIS DE BIOMARCADORES USANDO ENSAIOS COM BASE EM APTÂMERO

[000111] Os ensaios voltados para a detecção e quantificação de moléculas fisiologicamente significativas em amostras biológicas e outras amostras são ferramentas importantes na investigação científica e no campo da saúde. Uma classe de tais ensaios envolve a utilização de um microarranjo que inclui um ou mais aptâmeros imobilizadas em um suporte sólido. Os aptâmeros são, cada um capaz de se ligar a uma molécula alvo de uma forma altamente específica e com afinidade muito elevada. Ver, por exemplo,PatenteUS5. 475. 096 intitulada "Nucleic Acid Ligands"; ver também, por exemplo, Patente US 6.242.246, Patente US6.458.543, e Patente US 6.503.715, cada uma das quais é intitulada "Nucleic Acid Ligand Diagnostic Biochip". Uma vez que a micromatriz é posta em contato com uma amostra, os aptâmeros se ligam às suas respectivas moléculas alvo presentes na amostra e, assim, permitem uma determinação do nível de um biomarcador que corresponde a um biomarcador.

[000112] Tal como aqui utilizado, um "aptâmero" se refere a um ácido nucleico que tem uma afinidade de ligação específica com uma molécula alvo. É reconhecido que as interações de afinidade são uma questão de grau; no entanto, neste contexto, a "afinidade de ligação específica" de um aptâmero para seu alvo significa que o aptâmero se liga ao seu alvo geralmente com muito maior grau de afinidade do que se liga a outros componentes em uma amostra de teste. Um "aptâmero" é um conjunto de cópias de um tipo ou espécie de molécula de ácido nucléico que tem uma sequência particular de nucleotídeos. Um aptâmero pode incluir qualquer número adequado de nucleotídeos, incluindo qualquer número de nucleotídeos modificados quimicamente. "Aptâmeros" se refere a mais de um tal conjunto de moléculas. Aptâmeros diferentes podem ter números iguais ou diferentes de nucleotídeos. Aptâmeros podem ser DNA ou RNA ou ácidos nucleicos quimicamente modificados e podem ser de cadeia simples, cadeia

dupla, ou conter regiões de cadeia dupla, e podem incluir estruturas de ordem maior. Um aptâmero também pode ser um fotoaptâmero, onde um grupo funcional quimicamente reativo ou fotorreativo está incluído no aptâmero para permitir que seja ligado covalentemente ao seu alvo correspondente. Qualquer um dos métodos aqui descritos aqui podem incluir a utilização de dois ou mais aptâmeros que se ligam especificamente à mesma molécula alvo. Tal como adicionalmente descrito abaixo, um aptâmero podem incluir um marcador. Se um aptâmero inclui um marcador, todas as cópias do aptâmero não precisa ter a mesma etiqueta. Além disso, se diferentes aptâmeros cada incluir um marcador, estes diferentes aptâmeros pode ter tanto a mesma etiqueta ou um marcador diferente.

[000113] Um aptâmero pode ser identificado usando qualquer método conhecido, incluindo o processo SELEX. Uma vez identificado, um aptâmero pode ser preparado ou sintetizado em conformidade com qualquer método conhecido, incluindo os métodos de síntese química e os métodos de síntese enzimática.

[000114] Os termos "SELEX" e "processo SELEX" são usados permutavelmente neste documento para referir-se geralmente a uma combinação da (1) seleção dos aptâmeros que interagem com a molécula alvo de forma desejável, por exemplo, por ligação com alta afinidade a uma proteína, com (2) a ampliação desses ácidos nucleicos selecionados. O processo SELEX pode ser usado para identificar aptâmeros com alta afinidade para um alvo ou biomarcador específico.

[000115] SELEX geralmente inclui preparar uma mistura de candidato de ácidos nucleicos, vinculação da mistura candidata para a molécula alvo desejada para formar um complexo de afinidade, separando os complexos de afinidade dos ácidos nucleicos não ligados candidatos, separando e isolando o ácido nucleico do complexo de afinidade. purificando o ácido nucleico e identificando uma sequência específica

de aptâmero. O processo pode incluir várias rodadas para refinar ainda mais a afinidade do aptâmero selecionado. O processo pode incluir etapas de amplificação em um ou mais pontos no processo. Ver, por exemplo,PatenteUS5. 475. 096, intitulada "Nucleic Acid Ligands". O processo SELEX pode ser usado para gerar um aptâmero que liga covalentemente seu alvo, bem como um aptâmero que liga não covalentemente seu alvo. Ver, por exemplo, Patente US5. 705. 337 intitulada "Systematic Evolution of Nucleic Acid Ligands by Exponential Enrichment: Chemi-SELEX."

[000116] O processo SELEX pode ser usado para identificar aptamers de alta afinidade contendo nucleotídeos modificados que conferem características melhoradas ao aptâmero, tais como, por exemplo, estabilidade *in vivo* melhorada ou características de entrega melhoradasExemplos de tais modificações incluem substituições químicas na ribose e/ou fosfato e/ou posições de base. Aptâmeros indentificados no processo SELEX que contêm nucleotídeos modificados estão descritos na PatentePatenteUS 5. 660. 985, intitulada "High Affinity Nucleic Acid Ligands Containing Modified Nucleotides", que descreve oligonucleotídeos contendo derivados de nucleotídeos modificados quimicamente nas posições 5 'e 2' de pirimidinas. PatenteUS5. 580. 737, ver acima, descreve aptâmeros altamente específicos que contêm um ou mais nucleotídeos modificados com 2'-amino (2'-NH 2), 2'-fluoro (2'-F), e/ou 2'-O-metil (2'-OMe). Ver, também, Publicação de Pedido de Patente US 20090098549, intitulada "SELEX and PHOTOSELEX," que descreve bibliotecas de ácido nucleico com propriedades físicas e químicas expandidas e sua utilização em SELEX e fotoSELEX.

[000117] SELEX também pode ser usado para identificar aptâmeros que possuem características desejáveis de dissociação lenta. Ver Publicação US20090004667, intitulada "Method for Generating

Aptamers with Improved Off-Rates", que descreve métodos melhorados SELEX para a geração de aptâmeros que podem se ligar a moléculas alvo. São descritos métodos para a produção de aptâmeros e fotoaptâmeros com taxas mais lentas de dissociação de suas moléculas alvo correspondentes. Os métodos envolvem pôr em contato a mistura candidata com a molécula alvo, permitindo a formação de complexos de ácido nucleico-alvo ocorrer e realização de um processo de enriquecimento de dissociação lenta onde complexos de ácido nucleico alvo com taxas de dissociação rápida dissociam e não se reformam, enquanto complexos com taxas de dissociação vão permanecer intactos. Além disso, os métodos incluem o uso de nucleotídeos modificados na produção de misturas de ácidos nucleicos candidatos para gerar aptâmeros com desempenho melhorado na taxa de desaceleração. Nucleotídeos modificados não limitativos ilustrativos incluem, por exemplo, as pirimidinas modificadas, apresentadas na Figura 14. Em algumas modalidades, um aptâmero compreende pelo menos um nucleotídeo com uma modificação, tal como uma alteração de base. Em algumas modalidades, um aptâmero compreende pelo menos um nucleotídeo com uma modificação hidrofóbica, tal como uma modificação base hidrofóbica, permitindo os contatos hidrofóbicos com uma proteína alvo. Tais contatos hidrofóbicos, em algumas modalidades, contribuem para uma maior afinidade e/ou a taxa de desaceleração mais lenta de ligação pelo aptâmero. Nucleotídeos exemplifiactivos não limitativos, com modificações hidrofóbicas são mostrados na Figura 14. Em algumas modalidade, um aptâmero compreende pelo menos dois, pelo menos três, pelo menos quatro, pelo menos cinco, pelo menos seis, pelo menos sete, pelo menos oito, pelo menos nove, ou pelo menos 10 nucleotídeos, com modificações hidrofóbicas, em que cada modificação hidrofóbica pode ser a mesma ou diferente das outras. Em algumas modalidades, pelo menos uma,

pelo menos duas, pelo menos três, pelo menos quatro, pelo menos cinco, pelo menos seis, pelo menos sete, pelo menos oito, pelo menos nove, ou pelo menos 10 modificações hidrofóbicas de um aptâmero podem ser independentemente selecionadas dentre as modificações hidrofóbicas mostradas na Figura 14.

[000118] Em algumas modalidades, um aptâmero de taxa de desaceleração lenta (incluindo aptâmeros compreendendo pelo menos um nucleotídeo com uma modificação hidrofóbica) tem uma taxa de desaceleração (t½) de ≥ 30 minutos, ≥ 60 minutos, de ≥ 90 minutos, ≥ 120 minutos, ≥ 150 minutos, 180 minutos, ≥ 210 minutos, ou ≥ 240 minutos.

[000119] Em algumas modalidades, um ensaio emprega aptâmeros que incluem grupos funcionais fotorreativos que permitem que os aptâmeros se liguem de forma covalente ou por "fotorreticulação" as suas moléculas alvo. Ver, por exemplo,PatenteUS6. 544. 776 intitulada "Nucleic Acid Ligand Diagnostic Biochip". Estes aptâmeros fotorreativos também são referidos como fotoaptâmeros. Ver, por exemplo,PatenteUS5. 763. 177,Patente US6. 001. 577, ePatente US6. 291. 184, cada uma das quais é intitulada "Systematic Evolution of Nucleic Acid Ligands by Exponential Enrichment: Photoselection of Nucleic Acid Ligands and Solution SELEX"; ver também por exemplo: PatenteUS6. 458. 539, intitulada "Photoselection of Nucleic Acid Ligands". Depois da micromatriz ser colocada em contato com a amostra e os fotoaptâmeros terem tido uma oportunidade de se ligarem às seus moléculas-alvo, os fotoaptâmeros são fotoativados, e o suporte sólido é lavado para remover quaisquer moléculas não ligadas especificamente. Podem ser usadas condições de lavagem duras, uma vez que as moléculas alvo que estão ligadas aos fotoaptâmeros geralmente não são removidas, devido às ligações covalentes criadas pelo(s) grupo(s) funcional(is) fotoativado(s) sobre os fotoaptâmeros.

Desta maneira, o ensaio permite a detecção de um nível correspondente a um biomarcador na amostra de teste.

[000120] Em alguns formatos de ensaio, os aptâmeros são imobilizados no suporte sólido antes de serem postos em contato com a amostra. Sob certas circunstâncias, no entanto, a imobilização dos aptâmeros antes do contato com a amostra pode não proporcionar um ensaio ideal. Por exemplo, pré-imobilização dos aptâmeros podem resultar na mistura ineficiente dos aptâmeros com as moléculas alvo na superfície do suporte sólido, talvez, levando a tempos de reação longos e, portanto, períodos de incubação prolongados para permitir a eficiente ligação dos aptâmeros com as suas moléculas alvo. Além disso, quando fotoaptâmeros são empregados no ensaio e, dependendo do material utilizado como suporte sólido, o suporte sólido pode ter tendência a dispersar ou absorver a luz utilizada para efetuar a formação de ligações covalentes entre os fotoaptâmeros e as suas moléculas alvo. Além disso, dependendo do método utilizado, a detecção de moléculas alvo ligadas aos seus aptâmeros podem estar sujeitas a imprecisão, uma vez que a superfície do suporte sólido pode também ser exposta e influenciada por quaisquer agentes de marcação que são utilizados. Finalmente, a imobilização dos aptâmeros sobre o suporte sólido geralmente envolve uma etapa de preparação de aptâmero (isto é, a imobilização) antes da exposição dos aptâmeros para a amostra, e este passo de preparação pode afetar a atividade ou funcionalidade dos aptâmeros.

[000121] Ensaios de aptâmero que permitem que um aptâmero capture o alvo em solução, e então empregam etapas de separação que são concebidas para remover componentes específicos da mistura de aptâmero alvo antes da detecção também foram descritos (verPublicação US20090042206, intitulada "Multiplexed Analyses of Test Samples"). Os métodos de ensaio de aptâmero descritos permitem

a detecção e quantificação de um alvo de ácido não nucleico (por exemplo, uma proteína alvo) em uma amostra de ensaio por detecção e quantificação de um ácido nucleico (isto é, um aptâmero). Os métodos descritos criam um substituto de ácido nucleico (isto é, o aptâmero) para a detecção e quantificação de um alvo de ácido não nucleico, permitindo, assim, a grande variedade de tecnologias de ácidos nucleicos, incluindo amplificação, ser aplicada a uma ampla variedade de alvos desejados, incluindo proteínas alvo.

[000122] Os aptâmeros podem ser construídos para facilitar a separação dos componentes de ensaio a partir de um complexo de biomarcador de aptâmero (ou complexo covalente de biomarcador de aptâmero) e permitir o isolamento do aptâmero para a detecção e/ou quantificação. Em uma modalidade, essas construções podem incluir um elemento clivável ou liberável dentro da sequência de aptâmero. Em outras modalidades, a funcionalidade adicional pode ser introduzida no aptâmero, por exemplo, um componente marcado ou detectável, um componente espaçador, ou um marcador de ligação específica ou elemento de imobilização. Por exemplo, o aptâmero pode incluir um marcador ligada ao aptâmero por meio de uma porção clivável, um marcador, um componente espaçador que separa o marcador, e a porção clivável. Em uma modalidade, um elemento clivável é um ligante fotoclivável. O ligante fotoclivável pode ser ligado a uma porção de biotina e uma seção de separador, pode incluir um grupo NHS para derivação de aminas, e pode ser utilizada para introduzir um grupo de biotina a um aptâmero, permitindo assim a liberação do aptâmero mais tarde em um método de ensaio.

[000123] Ensaios homogêneos, feitos com todos os componentes do ensaio em solução, não necessitam de separação de amostras e reagentes antes da detecção do sinal. Estes métodos são rápidos e fáceis de usar. Estes métodos geram sinal com base em uma captura

molecular ou reagente ligando que reage com o seu alvo específico. Em algumas modalidades dos métodos aqui descritos, os reagentes de captura molecular compreendem um aptâmero ou um anticorpo ou semelhante, e o alvo pode ser um biomarcador mostrado na Tabela 3.

[000124] Em algumas modalidades, um método para a geração de sinal toma vantagem da mudança do sinal de anisotropia devido à interação de um reagente de captura marcado com fluoróforo com o seu alvo biomarcador específico. Quando a captura marcada reage com o seu alvo, o aumento do peso molecular faz com que o movimento rotacional do fluoróforo afixado ao complexo se torne muito mais lento mudando o valor de anisotropia. Ao monitorar a alteração de anisotropia, eventos de ligação podem ser utilizados para medir quantitativamente os biomarcadores em soluções. Outros métodos incluem ensaios de polarização de fluorescência, os métodos de farol molecular, resolvida no tempo de extinção de fluorescência, quimioluminescência, transferência de energia de ressonância de fluorescência, e semelhantes.

[000125] Um ensaio de aptâmero baseado em solução exemplificativo que pode ser usado para detectar um nível de biomarcadores em uma amostra biológica inclui o seguinte: (A) preparação de uma mistura por contacto da amostra biológica com um aptâmero aquele inclui uma primeira etiqueta e tem uma afinidade específica para o biomarcador, em que um complexo de afinidade do aptâmero é formado quando o biomarcador está presente na amostra; (b) expor a mistura a um primeiro suporte sólido, incluindo um primeiro elemento de captura, e permitindo que a primeira etiqueta se associe com o primeiro elemento de captura; (c) a remoção de quaisquer componentes da mistura não associado com o primeiro suporte sólido; (d) ligar uma segunda etiqueta ao componente biomarcador do complexo de afinidade do aptâmero; (e) libertar o complexo de afinidade do aptâmero a partir do primeiro

suporte sólido; (f) expor o complexo de afinidade do aptâmero dliberado para um segundo suporte sólido que inclui um segundo elemento de captura e permitindo que a segunda etiqueta de se associar ao segundo elemento de captura; (g) remover qualquer aptâmero não complexado a partir da mistura através da divisão do aptâmero não complexado a partir do complexo de afinidade do aptâmero; (h) eluir o aptâmero a partir do suporte sólido; e (i) detectar o marcador biológico através da detecção do componente de aptâmero do complexo de afinidade do aptâmero.

[000126] Quaisquer meios conhecidos na técnica podem ser usados para detectar um valor de biomarcador, detectando o componente de aptâmero de um complexo de afinidade do aptâmero. Um número de diferentes métodos de detecção pode ser utilizado para detectar o componente de aptâmero de um complexo de afinidade, tais como, por exemplo, ensaios de hibridação, espectroscopia de massa, ou QPCR. Em algumas modalidades, métodos de sequenciação de ácidos nucleicos podem ser utilizados para detectar o componente de aptâmero de um complexo de afinidade de aptâmero e, consequentemente, detectar um valor de biomarcador. Resumidamente, uma amostra de teste pode ser submetida a qualquer tipo de método de sequenciação de ácido nucleico para identificar e quantificar a sequência ou sequências de um ou mais aptâmeros presentes na amostra de teste. Em algumas modalidades, a sequência inclui toda a molécula de aptâmero ou qualquer porção da molécula que pode ser usada para identificar de forma exclusiva a molécula. Em outras modalidades, a sequência de identificação é uma sequência específica para o aptâmero adicionado; Tais sequências são muitas vezes referidas como "etiquetas", "códigos de barras", ou "ceps". Em algumas modalidades, o método de sequenciação inclui etapas enzimáticas para amplificar a sequência de aptâmero ou para converter

qualquer tipo de ácido nucleico, incluindo RNA e DNA que contenha modificações químicas para qualquer posição, para qualquer outro tipo de ácido nucleico apropriado para a sequenciação.

[000127] Em algumas modalidades, o método de sequenciação inclui uma ou mais etapas de clonagem. Em outras modalidades, o método de sequenciação inclui um método de sequenciação direta sem clonagem.

[000128] Em algumas modalidades, o método de sequenciação inclui uma abordagem dirigida com oligonucleotídeos específicos que têm como alvo um ou mais aptâmeros na amostra de teste. Em outras modalidades, o método de sequenciação inclui uma abordagem aleatória que atinge a todos os aptâmeros na amostra de teste.

[000129] Em algumas modalidades, o método de sequenciação inclui etapas enzimáticas para amplificar a molécula alvo para a sequenciação. Em outras modalidades, o método de sequenciação sequencia diretamente moléculas individuais. Um método baseado na sequenciação de ácidos nucleicos exemplificativos que podem ser utilizados para detectar um valor correspondente a um biomarcador em uma amostra biológica inclui o seguinte: (a) conversão de uma mistura de aptâmeros que contêm nucleotídeos quimicamente modificados a ácidos nucleicos não modificados com uma etapa enzimática; (b) sequenciação aleatória dos ácidos nucleicos não modificados resultantes com uma plataforma de sequenciação massivamente paralela, como, por exemplo, o 454 Sequencing System (454 Life Sciences/Roche), o Illumina Sequencing System (Illumina), o ABI SOLiD Sequencing System (Applied Biosystems), o HeliScope Single Molecule Sequencer (Helicos Biosciences), ou o Pacific Biosciences Real Time Single-Molecule Sequencing System (Pacific BioSciences) ou o Polonator G Sequencing System (Dover Systems); e (c) identificar e quantificar os aptâmeros presentes na mistura pela sequência específica e contagem de sequência.

[000130] Um método exemplificativo não limitativo de detectar biomarcadores em uma amostra biológica usando aptâmeros é descrito no Exemplo 1. *Ver também* Kraemer et al., 2011, *PLoS One* 6 (10): e26332.

A DETERMINAÇÃO DE NÍVEIS DE BIOMARCADOR USANDO IMUNOENSAIOS

[000131] Métodos de imunoensaios baseiam-se na reação de um anticorpo ao seu alvo correspondente ou analito e pode detectar o analito em uma amostra de acordo com o formato de ensaio específico. Para melhorar a especificidade e sensibilidade de um método de ensaio baseado na técnica de imunorreatividade, anticorpos monoclonais e fragmentos dos mesmos são muitas vezes utilizados devido ao seu reconhecimento de epítopo específico. Os anticorpos policlonais também têm sido utilizados com sucesso em vários imunoensaios devido à sua maior afinidade para o alvo, em comparação com anticorpos monoclonais. Os imunoensaios foram concebidos para utilização com uma grande variedade de matrizes de amostras biológicas. Formatos de imunoensaio foram concebidos para proporcionar resultados qualitativos, semiquantitativos, e quantitativos.

[000132] Os resultados quantitativos são gerados através do uso de uma curva padrão criada com concentrações conhecidas do analito específico a ser detectado. A resposta ou sinal a partir de uma amostra desconhecida é traçada para a curva padrão, e uma quantidade ou nível correspondente ao alvo na amostra desconhecida é estabelecida.

[000133] Muitos formatos de imunoensaio foram concebidos. ELISA ou EIA pode ser quantitativo para a detecção de um analito. Este método baseia-se na colocação de um marcador para tanto o analito ou o anticorpo e o componente de marcador incluir, quer direta ou indiretamente, uma enzima. ELISA pode ser formatado para a detecção direta, indireta, competitiva ou em sanduíche do analito. Outros

métodos baseiam-se nos marcadores, tais como, por exemplo, radioisótopos (I^{125}) ou fluorescência. Outras técnicas incluem, por exemplo, aglutinação, nefelometria, turbidimetria, Western blot, imunoprecipitação, imunocitoquímica, imunohistoquímica, citometria de fluxo, ensaio Luminex, e outros (ver Immunoassay: A Practical Guide, edited by Brian Law, published by Taylor & Francis, Ltd., edição de 2005).

[000134] Formatos de ensaio exemplificativos incluem ensaio de imunoabsorção ligado a enzima (ELISA), radioimunoensaio, fluorescente, quimioluminescência, e fluorescência de transferência de energia de ressonância (FRET) ou o tempo de imunoensaios resolvidos FRET (TR-FRET). Exemplos de procedimentos para detectar biomarcadores incluem imunoprecipitação de biomarcador seguida por métodos quantitativos que permitem que o tamanho e a discriminação de nível de peptídeo, tais como eletroforese em gel, eletroforese capilar, eletrocromatografia planar, e semelhantes.

[000135] Os métodos de detecção e/ou quantificação de um material gerador de etiqueta ou sinal detectável dependem da natureza do marcador. Os produtos de reações catalisadas por enzimas adequadas (em que o marcador detectável é uma enzima, ver acima) podem ser, sem limitação, fluorescentes, luminescentes, radioativos ou podem absorver a luz visível ou ultravioleta. Exemplos de detectores apropriados para detectar esses marcadores detectáveis incluem, sem limitação, filme em raio X, contadores de radioatividade, contadores de cintilação, espectrofotômetros, colorímetros, fluorímetros, luminômetros e densitômetros.

[000136] Qualquer um dos métodos de detecção pode ser realizado em qualquer formato que permita qualquer preparação adequada, processamento e análise das reações. Isto pode ocorrer, por exemplo, em placas de ensaio de multipoços (por exemplo,96 poços ou de 386

poços) ou usando qualquer matriz ou micromatrizes adequadas. As soluções de reserva para vários agentes pode ser feita manualmente ou roboticamente, e todas as pipetagem, diluição, mistura, distribuição, lavagem, incubação, leitura de amostra, coleta e análise de dados subsequentes podem ser feitas roboticamente usando software de análise disponível comercialmente, robótica, e instrumentação de detecção capaz de detectar um marcador detectável.

DETERMINAÇÃO DOS NÍVEIS DE BIOMARCADORES USANDO PERFILAÇÃO DE EXPRESSÃO DE GENE

[000137] A medição de mRNA de uma amostra biológica pode, em algumas modalidades, ser utilizada como uma substituta para a detecção do nível de proteína correspondente na amostra biológica. Assim, em algumas modalidades, um biomarcador ou um painel de biomarcadores aqui descrito pode ser detectado através da detecção do RNA apropriado.

[000138] Em algumas modalidades, os níveis de expressão de mRNA são medidos pela transcrição reversa quantitativa reação em cadeia da polimerase (RT-PCR seguida por qPCR). RT-PCR é utilizado para criar um cDNA a partir do mRNA. O cDNA pode ser utilizado em um ensaio de qPCR para produzir fluorescência à medida em que processo de amplificação de DNA progride. Em comparação com uma curva padrão, qPCR pode produzir uma medição absoluta como o número de cópias de mRNA por célula. Northern blots, micromatrizes, ensaios de invasor, e RT-PCR combinados com eletroforese capilar têm sido utilizados para medir os níveis de expressão de mRNA em uma amostra. Ver Gene Expression Profiling: Methods and Protocols, Richard A. Shimkets, editor, Humana Press, 2004.

DETECÇÃO DE BIOMARCADORES UTILIZANDO TECNOLOGIAS DE IMAGIOLOGIA MOLECULAR *IN VIVO*

[000139] Em algumas modalidades, um biomarcador aqui descrito

pode ser utilizado em testes de imagiologia molecular. Por exemplo, um agente de imagiologia pode ser acoplado a um reagente de captura, que pode ser usado para detectar o biomarcador *In Vivo*.

[000140] *Tecnologias de* imagiologia in vivo provide proveem métodos não invasivos para a determinação do estado de uma doença em particular no corpo de um indivíduo. Por exemplo, porções inteiras do corpo, ou mesmo todo o corpo, podem ser vistas como uma imagem tridimensional, dando assim informação valiosa com relação à morfologia e estruturas no corpo. Tais tecnologias podem ser combinadas com a detecção dos biomarcadores aqui descritos para proporcionar informação relativa ao biomarcador *in vivo*.

[000141] O uso de tecnologias *in vivo* de imagiologia molecular está se expandindo devido a vários avanços na tecnologia. Estes avanços incluem o desenvolvimento de novos agentes de contraste ou marcadores, tais como radiomarcadores e/ou marcadores fluorescentes, que pode proporcionar sinais fortes dentro do corpo; e o desenvolvimento da nova poderosa tecnologia de imagiologia, a qual pode detectar e analisar estes sinais de fora do corpo, com sensibilidade e precisão suficiente para prover informação útil. O agente de contraste pode ser visualizado em um sistema de imagiologia apropriado, provendo assim uma imagem da porção ou porções do corpo nas quais o agente de contraste está localizado. O agente de contraste pode estar ligado ou associado com um reagente de captura, tal como um aptâmero ou um anticorpo, por exemplo, e/ou com um peptídeo ou proteína, ou um oligonucleotídeo (por exemplo, para a detecção da expressão de genes), ou um complexo contendo qualquer um destes com uma ou mais macromoléculas e/ou outras formas de partículas.

[000142] O agente de contraste pode também possuir um átomo radioativo que é útil na imagiologia. Átomos radioativos adequados

incluem tecnécio-99m ou iodo-123 para estudos de cintigrafia. Outras porções prontamente detectáveis incluem, por exemplo, marcadores de spin para imagiologia de ressonância magnética (MRI), tal como, por exemplo, iodo-123 novamente, iodo-131, índio-111, flúor-19, carbono-13, nitrogênio-15, oxigênio -17, gadolínio, manganês e ferro. Tais marcadores são bem conhecidos na técnica e podem ser facilmente selecionados por um dos peritos na técnica.

[000143] Técnicas de imagiologia padrão incluem, mas não estão limitadas a imagiologia por ressonância magnética, tomografia computadorizada, tomografia por emissão de pósitrons (PET), tomografia de emissão de fóton único (SPECT), e semelhantes. Para imagiologia de diagnóstico *in vivo*, o tipo de instrumento de detecção disponível é um fator principal na seleção de um dado agente de contraste, tal como um dado radioisótopo e o biomarcador particular, que é utilizado para marcar (proteína, mRNA, e semelhantes). O radioisótopo escolhido tipicamente tem um tipo de declínio que pode ser detectado por um dado tipo de instrumento. Além disso, ao selecionar um radioisótopo para diagnóstico *in vivo*, a meia-vida deve ser suficientemente longa para permitir a detecção, no momento de absorção máxima pelo alvo, mas o tecido suficientemente curto para que a radiação deletéria do hospedeiro seja minimizada.

[000144] Técnicas de imagiologia exemplificativas incluem, mas não estão limitadas a PET e SPECT, as quais são técnicas de formação imagem em que um radioisótopo é sinteticamente ou localmente administrado a um indivíduo. A absorção subsequente do radiomarcador é medida ao longo do tempo e usada para obter informação sobre o tecido alvo e o biomarcador. Por causa das emissões de alta energia (raios gama) dos isótopos específicos empregados e a sensibilidade e a sofisticação dos instrumentos utilizados para detectá-los, a distribuição bidimensional de

radioatividade pode ser inferida a partir do exterior do corpo.

[000145] Nuclídeos emitindo pósitrons comumente usados em PET incluem, por exemplo, carbono-11, nitrogênio-13, oxigênio-15, e flúor-18. Isótopos que declinam por captura de elétron e/ou por emissão de gama são utilizados em SPECT e incluem, por exemplo, iodo-123 e tecnécio-99m. Um método exemplificativo para marcação de aminoácidos com tecnécio-99m é a redução do íon de pertecnetato na presença de um precursor quelante, para formar o complexo de tecnécio-99m-precursor instável, o qual, por sua vez, reage com o grupo de ligação de metal de um peptídeo quimiotático para formar um conjugado de peptídeo de tecnécio-99m quimiotático.

[000146] Anticorpos são frequentemente utilizados para tais métodos de diagnóstico de imagiologia *in vivo*. A preparação e utilização de anticorpos para o diagnóstico in vivo são bem conhecidas na técnica. Da mesma forma, aptâmeros podem ser utilizados para tais métodos de diagnóstico de imagiologia *in vivo*. Por exemplo, um aptâmero que foi utilizado para identificar um biomarcador particular aqui descrito pode ser marcado de forma adequada e injectado em um indivíduo para detectar o biomarcador *in vivo*. O marcador utilizado será selecionado de acordo com a modalidade de imagiologia a ser utilizada, tal como anteriormente descrito. Agentes de imagiologia dirigida por aptâmero podem ter características únicas e vantajosas, com relação à penetração do tecido, distribuição no tecido, cinética, eliminação, potência e seletividade em comparação com outros agentes de imagiologia.

[000147] Tais técnicas podem também opcionalmente, ser realizadas com os oligonucleotídeos marcados, por exemplo, para a detecção da expressão de genes por meio de formação de imagens com oligonucleotídeos de antissentido. Estes métodos são utilizados para hibridação in situ, por exemplo, com moléculas fluorescentes ou

radioisótopos como o marcador. Outros métodos para a detecção da expressão de genes incluem, por exemplo, a detecção da atividade de um gene repórter.

[000148] Outro tipo geral de tecnologia de imagiologia é imagiologia óptica, em que sinais fluorescentes dentro do indivíduo são detectados por um dispositivo óptico que está externo ao indivíduo. Estes sinais podem ser devidos à fluorescência real e/ou à bioluminescência. Melhorias na sensitividade de detecção óptica aumentaram a utilidade de imagiologia óptica para ensaios de diagnostico in vivo.

[000149] Para uma revisão de outras técnicas, ver N. Blow, Nature Methods, 6, 465-469, 2009.

DETERMINAÇÃO DOS NÍVEIS DE BIOMARCADORES USANDO MÉTODOS DE ESPECTROMETRIA DE MASSA.

[000150] Uma variedade de configurações de espectrômetros de massa pode ser utilizada para detectar níveis de biomarcadores. Vários tipos de espectrômetros de massa estão disponíveis ou podem ser produzidos com várias configurações. Em geral, um espectrômetro de massa tem os seguintes componentes principais: uma entrada de amostras, uma fonte de íons, um analisador de massa, um detector, um sistema de vácuo, e sistema de controle do instrumento, e um sistema de dados. Diferença na entrada de amostra, fonte de íons, e analisador de massa geralmente definem o tipo de instrumento e suas capacidades. Por exemplo, uma entrada pode ser uma fonte de cromatografia líquida em coluna capilar, ou pode ser uma sonda direta ou fase tal como utilizado na dessorção a laser assistida por matriz. Fontes de íons comuns são, por exemplo, eletropulverização, incluindo nanopulverização e micropulverização ou dessorção por laser assistida por matriz. Analisadores de massa comuns incluem um filtro de massa quadrúpolo, analisador de massa de barreira iônica e analisador de massa de tempo de voo. Métodos de espectrometria de massa

adicionais são bem conhecidos na técnica. (ver Burlingame et al. Anal. Chem. 70: 647 R-716R (1998); Kinter e Sherman, Nova Iorque (2000)).

[000151] Biomarcadores de proteína e níveis de biomarcadores podem ser detectados e medidos por qualquer um dos seguintes: Espectrometria de massa de ionização de eletropulverização (ESI-MS), ESI-MS/MS, ESI-MS/(MS) n, espectometria de massa de tempo de voo de ionização por dessorção a laser assistida por matriz (MALDI-TOF-MS), espectometria de massa de tempo de voo de ionização/ dessorção a laser intensificada pela superfície (SELDI-TOF-MS), dessorção/ ionização em silício (DIOS), espectrometria de massa de íons secundários (SIMS), tempo de voo quádruplo (Q-TOF), tempo de voo em tandem (TOF/TOF) tecnologia, denominada ultraflex III TOF/TOF, espectrometria de massa de ionização a pressão atmosférica química (APCI-MS), APCI-MS/MS, APCI (MS) N, espectrometria de massa de fotoionização de pressão atmosférica(APPI- MS), APPI-MS/MS, e APPI- (MS) N, espectrometria de massa quádrupla, espectrometria de massa da transformada de Fourier (FTMS), espectrometria de massa quantitativa, e espectrometria de massa de barreira iônica.

[000152] Estratégias de preparação de amostras são utilizadas para marcar e enriquecer amostras antes de caracterização espectroscópica de massa de biomarcadores de proteína e níveis de determinação de biomarcadores. Os métodos de marcação incluem, mas não estão limitados a etiqueta isobárica para quantificação relativa e absoluta (iTRAQ) e marcação de isótopo estável com aminoácidos em cultura de células (SILAC). Reagentes de captura utilizadas para enriquecer seletivamente amostras para proteínas candidatas de biomarcadores antes de análise espectroscópica de massa incluem, mas não estão limitados a, aptâmeros, anticorpos, sondas de ácidos nucleicos, quimeras, moléculas pequenas, um $F(ab')_2$ fragmento, um fragmento

de anticorpo de cadeia única, um fragmento Fv, um fragmento Fv de cadeia única, um ácido nucleico, uma lectina, um receptor de ligação de ligando, afficorpos, nanocorpos, anquirinas, anticorpos de domínio, andaimes de anticorpos alternativos (por exemplo diacorpos, etc) polímeros imprimidos, avímeros, peptidomiméticos, peptoides, ácidos nucleicos de peptídios, ácidos nucleicos de treose, um receptor de hormônio, um receptor de citosina, e receptores sintéticos, e modificações e fragmentos destes.

DETERMINAÇÃO DOS NÍVEIS DE BIOMARCADORES USANDO UM ENSAIO DE LIGAÇÃO DE PROXIMIDADE

[000153] Um ensaio de ligação de proximidade pode ser usado para determinar os valores dos biomarcadores. Resumidamente, uma amostra de teste é posta em contato com um par de sondas de afinidade que pode ser um par de anticorpos ou um par de aptâmeros, com cada membro do par estendido com um oligonucleotídeo. Os alvos para o par de sondas de afinidade podem ser dois determinantes distintos sobre uma proteína ou uma determinada em cada uma das duas proteínas diferentes, que podem existir como complexos homo ou heteromultiméricos. Quando as sondas se ligam às determinantes alvo, as extremidades livres das extensões de oliganonucleotídeos são postos em proximidade suficientemente estreita para hibridar em conjunto. A hibridação das extensões de oligonucleotídeos é facilitada por um conector de oligonucleotídeos comum, que serve para preencher as extensões em conjunto de oligonucleotídeos quando eles estão posicionados em proximidade suficiente. Uma vez que as extensões de oligonucleotídeos das sondas são hibridizadas, as extremidades das extensões são unidas entre si por ligação enzimática de DNA.

[000154] Cada extensão de oligonucleotídeo iniciador compreende um local para a amplificação por PCR. Uma vez que as extensões de

oligonucleotídeo são ligadas umas às outras, os oligonucleotídeos formam uma sequência contínua de DNA que, através de amplificação por PCR, revela informação sobre a identidade e a quantidade da proteína alvo, assim como, a informação relativa às interações de proteína-proteína em que os determinantes de destino estão em duas proteínas diferentes. ligação de proximidade podem proporcionar um ensaio altamente sensível e específico para a concentração de proteína em tempo real e a informação interação através da utilização de PCR em tempo real. As sondas que não se ligam a determinantes de interesse não tem as extensões de oligonucleotídeos correspondentes postos em proximidade e nenhuma ligação ou amplificação por PCR pode prosseguir, resultando em nenhum sinal a ser produzido.

[000155] Os ensaios precedentes permitem a detecção dos valores dos biomarcadores que são úteis em métodos para a predição de risco de eventos CV, onde os métodos compreendem a detecção, em uma amostra biológica de um indivíduo, de pelo menos cinco, pelo menos seis, pelo menos sete, pelo menos oito, ou todos os nove biomarcadores selecionados de MMP12, angiopoietina-2, complemento de C7, troponina I cardíaca, proteína relacionada com angiopoietina 4, CCL18/PARC, complexo de alfa-1-antiquimotripsina, GDF11 e alfa-2-antiplasmina, em que uma classificação, como descrita abaixo, utilizando os valores dos biomarcadores indica se o indivíduo tem risco elevado de um evento CV ocorrendo dentro de um período de tempo de 1 ano, 2 anos, 3 anos, ou 4 anos. Em conformidade com qualquer dos métodos aqui descritos, os valores dos biomarcadores podem ser detectados e classificados individualmente ou podem ser detectados e classificados em conjunto, como por exemplo, em um formato de ensaio multiplex.

CLASSIFICAÇÃO DE BIOMARCADORES E CÁLCULO DO GRAU DA DOENÇA

[000156] Em algumas modalidades, a "assinatura" de um biomarcador para um determinado teste de diagnóstico contém um conjunto de biomarcadores, cada biomarcador tendo níveis característicos nas populações de interesse. Os níveis característicos, em algumas modalidades, pode referir-se a média ou média dos níveis de biomarcadores para os indivíduos de um grupo específico. Em algumas modalidades, um método de diagnóstico aqui descrito pode ser utilizado para atribuir uma amostra desconhecida a partir de um indivíduo em um de dois grupos, quer no aumento do risco de um evento CV ou não.

[000157] A atribuição de uma amostra em um dos dois ou mais grupos é conhecida como classificação, e o procedimento usado para realizar esta tarefa é conhecido como um classificador ou um método de classificação. Métodos de classificação também pode ser referidos como métodos de graduação. Existem muitos métodos de classificação que podem ser utilizados para construir um classificador de diagnóstico a partir de um conjunto de níveis de biomarcadores. Em alguns casos, os métodos de classificação são realizados utilizando técnicas supervisionadas de aprendizagem em que um conjunto de dados é coletado usando amostras obtidas de indivíduos dentro de dois (ou mais, para vários estados de classificação) grupos distintos que se deseja distinguir. Uma vez que a classe (ou grupo de população) a que pertence cada amostra é conhecido antecipadamente para cada amostra, o método de classificação pode ser treinados para dar a resposta desejada de classificação. Também é possível usar técnicas de aprendizagem não supervisionada para produzir um classificador de diagnóstico.

[000158] Abordagens comuns para o desenvolvimento de classificadores de diagnóstico incluem árvores de decisão; ensacamento + impulsionamento + florestas; governar a aprendizagem

baseada na inferência; janelas de Parzen; modelos lineares; logística; métodos de redes neurais; agrupamento não supervisionado; K-means; ascendência/ escendência hierárquica; aprendizado semissupervisionado; métodos de protótipos; vizinho mais próximo; estimativa de densidade de grãos; apoiar máquinas de vetores; modelos ocultos de Markov; Aprendizagem de Boltzmann; e classificadores podem ser combinados quer simplesmente ou de forma a minimizar particulares funções de objetivo. Para uma revisão, ver, por exemplo,Pattern Classification, R. O. Duda, et al., editors, John Wiley & Sons, 2nd edition, 2001; ver também, The Elements of Statistical LeRNAing - Data Mining, Inference, and Prediction, T. Hastie, et al., editors, Springer Science+Business Media, LLC, 2a edição, 2009.

[000159] Para produzir um classificador utilizando técnicas de aprendizagem supervisionada, um conjunto de amostras chamado dados de treinamento são obtidos. No âmbito de testes de diagnóstico, dados de treino incluem amostras dos diferentes grupos (categorias) de amostras desconhecidas que irão depois ser atribuídas. Por exemplo, amostras coletadas de indivíduos em uma população de controle e os indivíduos em uma população de doença determinada podem constituir dados de formação para desenvolver um classificador que pode classificar amostras desconhecidas (ou, mais particularmente, os indivíduos de quem foram obtidas as amostras), tanto como tendo a doença ou como estando livre da doença. O desenvolvimento do classificador a partir dos dados de formação é conhecido como treinamento o classificador. Detalhes específicos sobre a treinamento do classificador dependem da natureza da técnica de aprendizado supervisionado. Treinar um classificador bayesiano natural é um exemplo de uma tal técnica de aprendizagem supervisionada (ver, por exemplo,Pattern Classification, R. O. Duda, et al., editors, John Wiley & Sons, 2nd edition, 2001; ver também, The Elements of Statistical

LeRNAing - Data Mining, Inference, and Prediction, T. Hastie, et al., editors, Springer Science+Business Media, LLC, 2a edição, 2009. A formação de um classificador Bayesiano natural é descrita, por exemplo, naPublicação US2012/0101002 e 2012/0077695.

[000160] Uma vez que, normalmente, há muitos mais níveis de biomarcadores em potencial do que amostras em um conjunto de treinamento, deve-se ter cuidado para evitar o excesso de encaixe. O sobre-encaixe ocorre quando um modelo estatístico descreve erro aleatório ou ruído em vez da relação subjacente. O sobre-encaixe pode ser evitado em uma variedade de formas, incluindo, por exemplo, através da limitação do número de biomarcadores utilizados no desenvolvimento do classificador, assumindo-se que as respostas dos biomarcadores são independentes umas das outras, limitando a complexidade do modelo estatístico subjacente empregado, e assegurando que o modelo estatístico subjacente esteja de acordo com os dados.

[000161] Um exemplo ilustrativo do desenvolvimento de um teste de diagnóstico utilizando um conjunto de biomarcadores inclui a aplicação de um classificador de Bayes natural, um classificador de probabilística simples com base no teorema de Bayes, com tratamento independente rigoroso dos biomarcadores. Cada biomarcador é descrito por uma função de densidade de probabilidade dependente da classe (pdf) para os valores de RFU medidos ou valores de RFU (unidades fluorescência relativa) de log em cada classe. As pdfs conjuntas para o conjunto de biomarcadores em uma classe são presumidas como sendo o produto das pdfs individuais dependentes de classe para cada biomarcador. Treinar um classificador de Bayes natural, neste contexto, equivale à atribuição de parâmetros ("parametrização") para caracterizar as pdfs dependentes de classe. Qualquer modelo subjacente para as pdfs dependentes de classe pode ser usado, mas geralmente o modelo deve

estar em conformidade com os dados observados no conjunto de treino.

[000162] O desempenho do classificador de Bayes natural é dependente do número e da qualidade dos biomarcadores utilizados para construir e treinar o classificador. Um único biomarcador irá realizar de acordo com a sua distância de KS (Kolmogorov-Smirnov). A adição de biomarcadores subsequentes com boas distâncias de KS (> 0,3, por exemplo) vai, em geral, melhorar o desempenho de classificação se os biomarcadores subsequentemente adicionados forem independentes do primeiro biomarcador. Utilizando a sensibilidade mais a especificidade como um grau de classificador, muitos classificadores de alto grau podem ser gerados com uma variação de um algoritmo mais. (Um algoritmo guloso é qualquer algoritmo que segue a resolução de problemas meta-heurísticos de fazer a escolha localmente ideal em cada fase com a esperança de encontrar o ideal global.)

[000163] Outra maneira de descrever o desempenho do classificador é através de uma característica de operação de receptor (ROC), ou simplesmente curva ROC ou gráfico de ROC. O ROC é uma representação gráfica da sensibilidade, ou taxa positiva verdadeira, versus taxa de falsos positivos (1 - especificidade ou 1 - true taxa negativa), para um sistema classificador binário como o seu limiar de discriminação é variada. O ROC também pode ser representado de forma equivalente traçando a fração de positivos verdadeiros fora dos pontos positivos (TPR = taxa positiva verdadeira) vs. a fração de falsos positivos a partir dos negativos (FPR = false taxa positiva). Também conhecida como curva de característica de operação relativa, porque é uma comparação de duas características de funcionamento (TPR & FPR) como as alterações critério. A área sob a curva ROC (AUC) é comumente utilizada como uma medida de síntese de precisão do diagnóstico. Ela pode ter valores de 0,0 a 1,0. A AUC tem uma

propriedade estatística importante: a AUC de um classificador é equivalente à probabilidade de que o classificador vai classificar um exemplo positivo escolhido aleatoriamente como maior do que um exemplo negativo escolhido aleatoriamente (Fawcett T, 2006. An introduction to ROC analysis. Pattern Recognition Letters. 27: 861-874). Isto é equivalente ao teste de Wilcoxon de classificações (Hanley, JA,McNeil, B. J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29-36.). Outra forma de descrever o desempenho de um teste diagnóstico em relação a um padrão de referência conhecido é o índice de reclassificação de rede: a capacidade do novo teste para atualizar ou desatualizar corretamente o risco quando comparado com o teste padrão de referência. Ver, por exemplo,Pencina et al., 2011, *Stat. Med*. 30: 11-21. Enquanto a AUC sob a curva ROC é ideal para avaliar o desempenho de um classificador de 2 classes, medicina estratificada e personalizada depende da inferência de que a população contém mais classes de 2. Para tais comparações a taxa de risco dos quartis superiores versus inferiores (ou outros, tais como estratificações de decis) pode ser utilizada de forma mais adequada.

[000164] As previsões de risco ativadas através desta invenção podem ser aplicadas a indivíduos em cuidados primários ou em centros cardiovasculares especializados, ou mesmo diretamente ao consumidor. Em algumas modalidades, os classificadores utilizados para predizer eventos podem envolver alguma calibração para a população para a qual eles são aplicados - por exemplo, pode haver variações devidas a etnia ou geografia. Tais calibrações, em algumas modalidades, pode ser estabelecida com antecedência a partir de grandes estudos populacionais, por isso, quando aplicadas a um paciente individual estas são incorporadas antes de fazer uma predição de risco. Uma amostra de sangue venoso é feita, processada de forma

adequada e analisada tal como aqui descrito. Uma vez que a análise for concluída, as previsões de risco podem ser feitas matematicamente, com ou sem incorporar outros metadados a partir de registros médicos aqui descritos, tais como genética ou demografia. Várias formas de saída de informação são possíveis, dependendo do nível de conhecimento do consumidor. Para os consumidores procurando o tipo mais simples de saída, as informações podem ser, em algumas modalidades, "esta pessoa está suscetível a ter um evento nos próximos anos x (em que x é 1-4), sim/não", ou, alternativamente, semelhante a uma "sinal de trânsito" vermelho/ laranja/ verde ou seu equivalente verbal ou por escrito, tal como de risco alto/médio/baixo. Para os consumidores que procuram maiores detalhes, em algumas modalidades, o risco pode ser liberado como um número ou um gráfico que ilustra a probabilidade de um evento por unidade de tempo como um grau contínuo, ou um número maior de camadas (tais como decis), e/ou o tempo médio para o evento e/ou o tipo de evento mais provável. Em algumas modalidades, o resultado pode incluir indicações terapêuticas. Acompanhamento longitudinal do mesmo paciente ao longo do tempo permitirá que os gráficos mostrem a resposta a intervenções ou mudanças de estilo de vida. Em algumas modalidades, mais do que um tipo de saída pode ser proporcionado ao mesmo tempo para cumprir as necessidades do paciente e de membros individuais da equipe de cuidados com diferentes níveis de especialização.

[000165] Em algumas modalidades, os nove biomarcadores mostrados na Tabela 3 (os "marcadores CVD9") são detectados em uma amostra de sangue (tal como uma amostra de plasma ou uma amostra de soro) a partir de um indivíduo, por exemplo, utilizando aptâmeros, tais como aptâmeros de taxa de desaceleração lenta. Os valores de log de RFU são utilizados para calcular um índice de prognóstico (PI). Uma fórmula PI exemplificativa não limitativa é

mostrado abaixo:

$$PI = -16.61+1.55\times ANGPT2 - 1.22\times GDF11+2.12\times C7- 2.64\times SERPINF2+0.57\times CCL18+ 1.02\times ANGPTL4 + 1.43\times KLK3.SERPINA3+0.72\times MMP12+0.59\times TNNI3,$$
$$s = 0.85,$$

[000166] onde os níveis de proteína são levados para estarem no log10 RFU. Um versado na técnica irá apreciar que a fórmula PI pode ser recalibrada de acordo com a população da qual o indivíduo é retirado. Tal recalibração pode ser realizada de acordo com os métodos aqui descritos e/ou métodos conhecidos na técnica.

[000167] Dada a PI, a probabilidade de que o indivíduo vai sofrer um evento cardiovascular (CV evento) nos próximos "t" anos é dada pela fórmula:

$$\Pr[T \leq t] = 1 - e^{-e^{\left(\frac{Log(t)-PI}{s}\right)}},$$

[000168] onde PI é o *índice de prognóstico* (ou previsor linear) e s é o parâmetro de escala associado para a distribuição do valor extremo. Em várias modalidades, "t" é 5 anos ou menos, 4 anos ou menos, 3 anos ou menos, ou 2 anos ou menos.

KITS

[000169] Qualquer combinação dos biomarcadores aqui descritos pode ser detectada utilizando um kit adequado, tal como para utilização na realização dos métodos aqui descritos. Além disso, qualquer kit pode conter um ou mais marcadores detectáveis, tal como aqui descrito, tal como um grupo fluorescente, etc.

[000170] Em algumas modalidades, um kit inclui (a) um ou mais reagentes de captura (tal como, por exemplo, pelo menos, um aptâmero ou anticorpo) para a detecção de um ou mais biomarcadores em uma amostra biológica, em que os biomarcadores incluem, pelo menos, cinco, pelo menos seis, pelo menos sete, pelo menos oito, ou todos os nove biomarcadores selecionados de MMP12, angiopoietina-2, complemento C7, troponina I cardíaca, proteína relacionada com

angiopoietina 4, CCL18/PARC, complexo de alfa-1-antiquimotripsina, alfa e GDF11 2-antiplasmina, e opcionalmente (b) um ou mais de produtos de software ou programas de computador para classificar o indivíduo de quem a amostra biológica foi obtida como tendo ou não tendo risco aumentado de um evento CV ou para determinar a probabilidade de que o indivíduo tem risco aumentado de um evento cardiovascular, como descrito aqui adicionalmente. Alternativamente, ao invés de um ou mais produtos de programas de computador, uma ou mais instruções para executar manualmente as etapas acima por um ser humano podem ser providas.

[000171] Em algumas modalidades, um kit compreende um suporte sólido, um reagente de captura, e um material gerador de sinal. O kit pode também incluir instruções para utilizar os dispositivos e reagentes, a manipulação da amostra, e análise de dados. Além disso, o kit pode ser utilizado com um sistema de computador ou software para analisar e registar os resultados da análise da amostra biológica.

[000172] Os kits também podem conter um ou mais reagentes (por exemplo, tampões de solubilização, detergentes, lavagens, ou tampões) para o processamento de uma amostra biológica. Qualquer dos kits aqui descritos podem também incluir, por exemplo, tampões, agentes, materiais de matriz de espectrometria de massa, agentes de captura de anticorpos, amostras de controle positivas e amostras de controle negativas, software e informação, tais como protocolos, orientação, e dados de referência.

[000173] Em algumas modalidades, kits são providos para a análise de estado de risco de evento CV, em que os kits compreendem os iniciadores de PCR para um ou mais aptâmeros específicos para os biomarcadores aqui descritos. Em algumas modalidades, um kit pode ainda incluir instruções para o uso e correlação dos biomarcadores com a predição de risco de um evento CV. Em algumas modalidades, um kit

pode também incluir uma matriz de DNA contendo o complemento de um ou mais dos aptâmeros específicos para os biomarcadores aqui descritos, os reagentes e/ou enzimas para amplificar ou isolar a amostra de DNA. Em algumas modalidades, os kits podem incluir reagentes para PCR em tempo real, por exemplo, sondas e/ou iniciadores TaqMan, e enzimas.

[000174] Por exemplo, um kit pode compreender (a) os reagentes compreendendo pelo menos um reagente de captura para a determinação do nível de um ou mais biomarcadores em uma amostra de teste, e, opcionalmente, (b) um ou mais algoritmos ou programas de computador para executar as etapas de comparar a quantidade de cada biomarcador quantificado na amostra de teste para um ou mais pontos de corte predeterminados. Em algumas modalidades, um algoritmo ou programa de computador atribui um grau para cada biomarcador quantificado com base na referida comparação e, em modalidades, combina os graus atribuídas para cada biomarcador quantificado para obter um grau total. Além disso, em algumas modalidades, um algoritmo ou programa de computador compara o grau total com um resultado predeterminado, e usa a comparação para determinar se um indivíduo tem um maior risco de um evento CV. Alternativamente, ao invés de um ou mais algoritmos ou programas de computador, uma ou mais instruções para executar manualmente as etapas acima por um ser humano podem ser providas.

MÉTODOS DE COMPUTADOR E SOFTWARE

[000175] Uma vez que um painel de biomarcadores ou biomarcador é selecionado, um método para o diagnóstico de um indivíduo pode compreender o seguinte: 1) obtenção de uma amostra biológica; 2) realizar um método analítico para detectar e medir o biomarcador ou biomarcadores no painel na amostra biológica; 3) opcionalmente, realizar qualquer normalização de dados ou padronização; 4)

determinar cada nível de biomarcadores; e 5) relatar os resultados. Em algumas modalidades, os resultados são calibrados para a população/etnia do indivíduo. Em algumas modalidades, os níveis de biomarcadores são de alguma forma combinados e um único valor para os níveis de biomarcador combinado é relatado. Nesta abordagem, em algumas modalidades, o marcador pode ser um único número determinado a partir da integração de todos os biomarcadores que é comparado com um valor limite pré-estabelecido que é uma indicação da presença ou ausência de doença. Ou o diagnóstico ou o grau de predição pode ser uma série de barras, cada uma representando um valor de biomarcador e o padrão das respostas podem ser comparados com um padrão predefinido para a determinação da presença ou ausência de doença, condição ou o risco aumentado (ou não) de um evento.

[000176] Pelo menos algumas modalidades dos métodos aqui descritos podem ser implementadas com a utilização de um computador. Um exemplo de um sistema de computador 100 é mostrado na Figura 12. Com referência à Figura 12, o sistema 100 é mostrado constituído por elementos de hardware que estão eletricamente acoplados através do barramento 108, incluindo um processador 101, dispositivo de entrada 102, o dispositivo de saída 103, o dispositivo de armazenamento 104, leitor meios de armazenamento legíveis por computador 105a, sistema de comunicações 106, aceleração de processamento (por exemplo,DSP ou processadores para fins especiais) 107 e memória 109. O leitor de meios de armazenamento legível por computador 105a é ainda acoplado à mídia de armazenamento legível por computador 105b, a combinação abrangente representando dispositivos de armazenamento remotos, locais, fixos e/ou removíveis, além de mídia de armazenamento, memória, etc., para temporariamente e/ou mais permanentemente

contendo informação legível por computador, que pode incluir dispositivo de armazenamento 104, a memória 109 e/ou qualquer outro tal recurso de sistema acessível 100. O sistema 100 também compreende elementos de software (mostrados como estando atualmente localizados dentro da memória de trabalho 191), incluindo um sistema operacional 192 e outro código 193, tais como programas, dados e semelhantes.

[000177] No que diz respeito à Figura 12, o sistema 100 tem grande flexibilidade e capacidade de configuração. Assim, por exemplo, uma única arquitetura pode ser utilizada para implementar um ou mais servidores que podem ser, ainda, configurados de acordo com os protocolos atualmente desejáveis, variações de protocolo, extensões, etc. No entanto, será evidente para os peritos na técnica que modalidades pode também ser utilizadas de acordo com os requisitos das aplicações mais específicas. Por exemplo, um ou mais elementos do sistema podem ser implementados como sub-elementos dentro de um componente do sistema 100 (por exemplo, dentro do sistema de comunicações 106). Hardware personalizado também pode ser utilizado e/ou elementos particulares podem ser implementados em hardware, software ou ambos. Além disso, enquanto a ligação a outros dispositivos de computação, tais como dispositivos de entrada/ saída de rede (não mostrados) podem ser empregados, deve ser entendido que conexão com fio, sem fio, modem, e/ou outra conexão ou conexões para outros dispositivos de computação também pode ser utilizadas.

[000178] Em um aspecto, o sistema pode compreender um banco de dados que contém aspectos de características de biomarcadores de predição de risco de um evento CV. Os dados de biomarcadores (ou informações de biomarcador) podem ser utilizados como uma entrada para o computador para uso como parte de um método implementado por computador. Os dados dos biomarcadores podem incluir os dados

tal como aqui descritos.

[000179] Em um aspecto, o sistema compreende ainda um ou mais dispositivos de fornecimento de dados de entrada para um ou mais processadores.

[000180] O sistema compreende ainda uma memória para armazenar um conjunto de dados de elementos de dados classificados.

[000181] Em um outro aspecto, o dispositivo para fornecimento de dados de entrada compreende um detector para detectar a característica do elemento de dados, por exemplo, tal como um espectrômetro de massa ou leitor de chip de gene.

[000182] O sistema pode compreender adicionalmente um sistema de gerenciamento de base de dados. As solicitações ou consultas do usuário podem ser formatadas em uma linguagem apropriada compreendida pelo sistema de gerenciamento de banco de dados que processa a consulta para extrair as informações relevantes a partir do banco de dados de conjuntos de treinamento.

[000183] O sistema pode ser conectado a uma rede à qual um servidor de rede e um ou mais clientes estão conectados. A rede pode ser uma rede de área local (LAN) ou uma rede de banda larga (WAN), tal como é conhecido na técnica. De preferência, o servidor inclui o hardware necessário para a execução de produtos de programa de computador (por exemplo, software) para acessar dados do banco de dados para o processamento de solicitações do usuário.

[000184] O sistema pode incluir um sistema operacional (por exemplo,UNIX ou Linux) para a execução de instruções de um sistema de gerenciamento de banco de dados. Em um aspecto, o sistema operacional pode funcionar em uma rede global de comunicações, como a Internet, e utilizar um servidor de rede de comunicações global para se conectar a uma rede deste tipo.

[000185] O sistema pode incluir um ou mais dispositivos que

compõem uma interface de visualização gráfica que compreende elementos de interface, como botões, menus, barras de rolagem, campos para a entrada de texto, e outros como são rotineiramente encontrados em interfaces gráficas de usuário conhecidos na técnica. Os pedidos introduzidos e,m uma interface de usuário podem ser transmitidos para um programa de aplicação no sistema para formatar para pesquisar informações relevantes em uma ou mais das bases de dados do sistema. Os pedidos ou consultas inseridos por um usuário podem ser construídos em qualquer linguagem de banco de dados adequada.

[000186] A interface gráfica de usuário pode ser gerada por um código de interface gráfica de usuário, como parte do sistema operacional e pode ser utilizado para introduzir dados e/ou para exibir os dados introduzidos. O resultado dos dados processados podem ser exibido na interface, impressa numa impressora em comunicação com o sistema, salvo em um dispositivo de memória, e/ou transmitido através da rede, ou pode ser provido na forma do meio de leitura por computador.

[000187] O sistema pode estar em comunicação com um dispositivo de entrada para prover dados em relação a elementos de dados ao sistema (por exemplo, valores de expressão). Em um aspecto, o dispositivo de entrada pode incluir um sistema de perfilação de expressão do gene, incluindo, por exemplo, um espectrômetro de massa, chip de gene ou um leitor de matriz, e outros semelhantes.

[000188] Os métodos e aparelhos para a análise de informação de biomarcador para predição de risco de eventos de acordo com várias modalidades podem ser implementados de qualquer forma adequada, por exemplo, usando um programa de computador operando em um sistema de computador. Pode ser utilizado um sistema de computador convencional, que compreende um processador e uma memória de acesso aleatório, tal como um servidor de aplicação acessível

remotamente, um servidor de rede, uma estação de trabalho ou um computador pessoal. Os componentes adicionais do sistema de computador podem incluir dispositivos de memória ou sistemas de armazenamento de informação, tal como um sistema de armazenamento em massa e uma interface de usuário, por exemplo um monitor convencional, teclado e dispositivo de rastreamento. O sistema de computador pode ser um sistema independente ou parte de uma rede de computadores incluindo um servidor e uma ou mais bases de dados.

[000189] O sistema de análise de biomarcador de predição de risco de eventos CV pode prover funções e operações para completar a análise dos dados, tais como a coleta de dados, processamento, análise, relatórios e/ou diagnóstico. Por exemplo, em uma modalidade, o sistema de computador pode executar um programa de computador que pode receber, armazenar, pesquisar, analisar, e comunicar essa informação com relação aos biomarcadores de predição de risco de evento CV. O programa de computador pode compreender vários módulos de execução de várias funções ou operações, tais como um módulo de processamento para o processamento de dados em bruto e a geração de dados adicionais e um módulo de análise para analisar os dados em bruto e os dados suplementares para gerar um estado de predição de risco de evento CV e/ou diagnóstico ou cálculo de risco. Cálculo do estado de risco para um evento CV pode compreender, opcionalmente, a geração ou coleta de qualquer outra informação, incluindo informação biomédica adicional, em relação à condição do indivíduo relativo à doença, condição ou evento, identificar se testes adicionais podem ser desejáveis, ou de outro modo avaliar o estado de saúde do indivíduo.

[000190] Algumas modalidades aqui descritas podem ser implementadas de forma a incluir um produto de programa de

computador. Um produto de programa de computador pode incluir um meio de leitura por computador com código de programa de leitura por computador incorporado no meio para fazer com que um programa de aplicação seja executado em um computador com uma base de dados.

[000191] Como aqui usado, um "produto de programa de computador" se refere a um conjunto organizado de instruções em forma de declarações de linguagem natural ou de programação que estão contidos em um suporte físico de qualquer natureza (por exemplo, por escrito, eletrônico, magnético, óptico ou de outra forma) e que pode ser usado com um computador ou outro sistema de processamento de dados automatizado. Tais declarações de linguagem de programação, quando executadas por um sistema de computador ou processamento de dados, fazem com que o sistema de computador ou de processamento de dados aja de acordo com o conteúdo específico das declarações. Produtos de programa de computador incluem, sem limitação: programas em código fonte e código de objeto e/ou teste ou bibliotecas de dados incorporadas em um meio de leitura por computador. Além disso, o produto de programa de computador que permite que um dispositivo de sistema de computador ou dados de equipamentos de processamento aja de maneira pré-selecionada pode ser provido em uma série de formas, incluindo, mas não se limitando ao código de fonte original, do código de montagem, o código de objeto, linguagem da máquina, versões criptografadas ou compactadas do que vem a seguir e todos e quaisquer equivalentes.

[000192] Em um aspecto, um produto de programa de computador é provido para a avaliação do risco de um evento CV. O produto de programa de computador inclui um meio legível por computador incorporando código de programa executável por um processador de um dispositivo ou sistema de computação, o código de programa compreendendo: código que recupera dados atribuídos a uma amostra

biológica de um indivíduo, em que os dados compreendem níveis de biomarcadores que correspondem a cada um dos biomarcadores na Tabela 3; e código que executa um método de classificação que indica uma situação de risco de eventos CV do indivíduo como uma função dos valores dos biomarcadores.

[000193] Em outro aspecto ainda, um produto de programa de computador é proporcionado para indicar uma probabilidade de risco de um evento CV. O produto de programa de computador inclui um meio legível por computador incorporando código de programa executável por um processador de um dispositivo ou sistema de computação, o código de programa compreendendo: código que recupera dados atribuídos a uma amostra biológica de um indivíduo, em que os dados compreendem um valor de biomarcador correspondente a um biomarcador na amostra biológica selecionada dos biomarcadores providas no quadro 3; e código que executa um método de classificação que indica um estado de risco de eventos CV do indivíduo como uma função do valor do biomarcador.

[000194] Embora várias modalidades tenham sido descritas como métodos ou aparelhos, deve ser entendido que modalidades podem ser implementadas por meio de código acoplado com um computador, por exemplo, o código residente em um computador ou acessível pelo computador. Por exemplo, software e bancos de dados podem ser utilizados para implementar muitos dos métodos discutidos acima. Portanto, além das modalidades realizadas por hardware, é também notado que estas modalidades podem ser realizadas através da utilização de um artigo de fabricação composto por um meio utilizável por computador tendo um código de programa legível por computador incorporado no mesmo, o que causa a capacitação das funções descritas aqui. Portanto, deseja-se que modalidades também sejam consideradas protegidas por esta patente no seus meios de código de

programa também. Além disso, as modalidades podem ser incorporados como código armazenada em uma memória legível por computador de virtualmente qualquer tipo, incluindo, sem limitação, a RAM, ROM, meios magnéticos, meios ópticos, ou meios magneto-ópticos. Mais geralmente, as modalidades podem ser implementadas em software, ou em hardware, ou qualquer combinação dos mesmos incluindo, mas não limitado a, software executado num processador de uso geral, de microcódigo, matrizes lógicos programáveis (PLAs), ou circuitos integrados de aplicação específica (ASIC).

[000195] Prevê-se também que modalidades podem ser realizadas como sinais de computador incorporados em uma onda portadora, assim como sinais (por exemplo, elétricos e ópticos) propagados através de um meio de transmissão. Assim, os vários tipos de informação discutidos acima podem ser formatados em uma estrutura, tal como uma estrutura de dados, e transmitidos como um sinal elétrico através de um meio de transmissão ou armazenados em um meio legível por computador.

[000196] Note-se também que muitas das estruturas, materiais, atos aqui descritos podem ser descritos como meios para realizar uma função ou uma etapa para a realização de uma função. Portanto, deve ser entendido que tal linguagem tem direito de cobrir todas essas estruturas, materiais, ou atos divulgados nesta especificação e os seus equivalentes, incluindo a matéria incorporada por referência.

[000197] O processo de identificação de biomarcadores, a utilização dos biomarcadores aqui descritos, bem como os vários métodos para determinar valores de biomarcadores são descritos em detalhe acima com respeito à avaliação do risco de um evento CV. No entanto, a aplicação do processo, o uso de marcadores identificados, e os métodos para a determinação dos valores dos biomarcadores são

inteiramente aplicáveis para outros tipos específicos de condições cardiovasculares, para qualquer outra doença ou condição médica, ou para a identificação de indivíduos que podem ou não se beneficiar de um tratamento médico auxiliar.

OUTROS MÉTODOS

[000198] Em algumas modalidades, os biomarcadores e os métodos descritos aqui são usados para determinar um prêmio de seguro médico ou decisão de cobertura e/ou um prêmio de seguro de vida ou decisão de cobertura. Em algumas modalidades, os resultados obtidos pelos métodos aqui descritos são utilizados para determinar um prêmio de seguro médico e/ou um prêmio de seguro de vida. Em alguns desses exemplos, uma organização que proporciona seguro médico ou seguro de vida pede ou, de outra maneira, obtém informações sobre o risco de um evento CV em um indivíduo e usa essa informação para determinar um prêmio apropriado de seguro médico ou seguro de vida para o indivíduo. Em algumas modalidades, o teste é solicitada por, e paga por, a organização que proporciona seguro médico ou seguro de vida. Em algumas modalidades, o teste é utilizado com o comprador em potencial de uma prática ou sistema de saúde ou empresa para predizer riscos no futuro ou custos se a aquisição for adiante.

[000199] Em algumas modalidades, os biomarcadores e os métodos aqui descritos são utilizados para predizer e/ou gerenciar a utilização de recursos médicos. Em algumas de tais modalidades, os métodos não são realizados para o propósito de tal predição, mas a informação obtida do método é utilizada em uma tal predição e/ou gerenciamento da utilização de recursos médicos. Por exemplo, uma instalação de testes ou hospital pode montar informações dos métodos atuais para muitos indivíduos, a fim de prever e/ou gerenciar a utilização de recursos médicos em uma instalação em particular ou em uma área geográfica em particular.

EXEMPLOS

[000200] Os seguintes exemplos são providos apenas para fins ilustrativos e não se destinam a limitar o âmbito do pedido, tal como definido pelas reivindicações anexas. Técnicas de biologia molecular de rotina descritas nos seguintes exemplos podem ser realizadas como descrito nos manuais de laboratório padrão, tais como Sambrook et al., Clonagem molecular: A Laboratory Manual, 3rd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y. 2001). (2001).

EXEMPLO 1: DETECÇÃO DE BIOMARCADORES EXEMPLIFICATIVOS USANDO APTÂMEROS

[000201] Um método exemplificativo de detecção de um ou mais biomarcadores em uma amostra é descrito, por exemplo, em Kraemer et al., *PLoS One* 6 (10): e26332, e é descrito abaixo. Três métodos diferentes de quantificação: hibridização baseada em micromatriz, um método baseado em microesfera Luminex, e qPCR, são descritos.

REAGENTES

[000202] HEPES, NaCl, KCl, EDTA, EGTA, $MgCl_2$ e Tween-20 podem ser adquiridos, por exemplo, na Fisher Biosciences. O sal de sódio de sulfato de dextrano (DxSO4), nominalmente 8000 de peso molecular, pode ser adquirido, por exemplo, na AIC e é dialisado contra água desionizada durante pelo menos 20 horas com uma troca. KOD EX DNA polimerase pode ser adquirida, por exemplo, na VWR. O cloreto de tetrametilamônio e o CAPSO podem ser adquiridos, por exemplo, na Sigma-Aldrich e a estreptavidina-ficoeritrina (SAPE) pode ser adquirida, e. g., na Moss Inc. O cloridrato de 4-(2-aminoetil)-benzenossulfonilfluoreto (AEBSF) pode ser adquirido, por exemplo, na Gold Biotechnology. Placas de 96 poços revestidas com estreptavidina podem ser adquiridas, por exemplo, na Thermo Scientific. (Pierce Streptavidin Coated Plates HBC, claras, de 96 poços, número de produto 15500 ou 15501)NHS-PEO4-biotina pode ser

adquirida, por exemplo, na Thermo Scientific (EZ-Link NHS-PEO4-Biotin, número de produto 21329), dissolvida em DMSO anidro e pode ser armazenada congelada em alíquotas de uso único. IL-8, MIP-4, Lipocalina-2, RANTES, MMP-7 e MMP-9 podem ser adquiridos, por exemplo, na R&D Systems. Resistina e MCP-1 podem ser adquiridos, por exemplo, na PeproTech, e tPA pode ser adquirido, por exemplo, na VWR.

ÁCIDOS NUCLEICOS

[000203] Oligodeoxinucleotídeos convencionais (inclusive substituidos por aminas e biotina) podem ser adquiridos, por exemplo, na Integrated DNA Technologies (IDT). O bloco Z é um oligodeoxinucleotídeo de cadeia única de sequência 5'- (AC-BnBn) 7-AC-3', em que Bn indica um resíduo de deoxiuridina substituida por benzil. O bloco Z pode ser sintetizado utilizando química de fosforamidita convencional. Os reagentes de captura de aptâmero também podem ser sintetizados por química convencional de fosforamidita e podem ser purificados, por exemplo, em uma coluna PRP-3 de 21,5 × 75 mm, operando a 80°C em um sistema Waters Autopurification 2767 (ou sistema semiautomatizado da série Waters 600) por exemplo, usando um aquecedor TL-600 ou TL-150 de madeira timberline e um gradiente de bicarbonato de trietilamônio (TEAB)/ACN para eluir o produto. A detecção é realizada a 260 nm e as frações são recolhidas através do pico principal antes do agrupamento das frações melhores.

TAMPÕES

[000204] O tampão SB18 consiste em 40 mM de HEPES, 101 mM de NaCl, 5 mM de KCl, 5 mM de MgCl2 e 0,05% (v/v) de Tween 20 ajustado a pH 7,5 com NaOH. O tampão SB17 é SB18 suplementado com 1 mM de EDTA trissódico. O tampão PB1 consiste em 10 mM de HEPES, 101 mM de NaCl, 5 mM de KCl, 5 mM de MgCl2, 1 mM de EDTA trissódico e

0,05% (v/v) de Tween-20 ajustado a pH 7,5 com NaOH. O tampão de eluição de CAPSO consiste em 100 mM de CAPSO pH 10,0 e 1 M de NaCl. O tampão de neutralização contém HEPES 500 mM, HCl a 500 mM e 0,05% (v/v) de Tween-20. O tampão de hibridização Agilent é uma formulação proprietária que é proporcionada como parte de um kit (Oligo aCGH/ Kit de hibridação chip-on-chip). O tampão de lavagem Agilent 1 é uma formulação proprietária (Oligo aCGH/Tampão de lavagem ChIP-on-chip 1, Agilent). O tampão de lavagem Agilent 2 é uma formulação proprietária (Oligo aCGH/Tampão de lavagem ChIP-on-chip 2, Agilent). A solução de hibridação TMAC consiste em cloreto de tetrametilamônio 4,5 M, EDTA trissódico 6 mM, Tris-HCl 75 mM (pH 8,0) e Sarkosyl a 0,15% (v/v). O tampão KOD (concentrado 10 vezes) consiste em Tris-HCl 1200 mM, MgSO4 15 mM, KCl 100 mM, (NH4) 2SO4 60 mM, Triton-X 100 a 1% v/v e BSA a 1 mg/mL.

PREPARAÇÃO DA AMOSTRA

[000205] Soro (armazenado ao -80° C em alíquotas de 100 ul) é descongelado em um banho de água ao 25° C durante 10 minutos, e depois armazenado em gelo antes da diluição da amostra. As amostras são misturadas por agitação em vórtex suave por 8 segundos. Prepara-se uma solução de amostra de soro a 6% por diluição em 0,94 x SB17 suplementado com MgCl2 0,6 mM, EGTA trissódico 1 mM, AEBSF a 0,8 mM e 2 µM de bloco Z. Uma porção da solução de carga de soro de 6% é diluída 10 vezes em SB17 para criar uma carga de soro 0,6%. Cargas de 6% e 0,6% são utilizadas, em algumas modalidades, para detectar analitos de alta e baixa abundância, respectivamente.

REAGENTE DE CAPTURA (APTÂMERO) E PREPARAÇÃO DE PLACA ESTREPTAVIDINA

[000206] Os aptâmeros são agrupados em 2 misturas de acordo com a abundância relativa aos seus analitos cognatos (ou biomarcadores). As concentrações de carga são de 4 nM para cada aptâmero, e a

concentração final de cada aptâmero é de 0,5 nM. Misturas de carga de aptâmero são diluídas 4 vezes em tampão SB17, aquecidas ao 95° C por 5 minutos e resfriado até 37° C durante um período de 15 minutos antes da utilização. Este ciclo de desnaturação-renaturação é destinado a normalizar distribuições de confôrmeros de aptâmeros e, assim, garantir a atividade reprodutível de aptâmero, apesar de histórias variáveis. Placas de estreptavidina são lavadas duas vezes com tampão de PB1 de150 µl antes da utilização.

EQUILÍBRIO E CAPTURA DE PLACA

[000207] Misturas de aptâmero resfriadas a quente 2x (55 ul) são combinadas com um volume igual de diluições de soro de 6% ou 0,6%, produzindo misturas de equilíbrio contendo 3% e 0,3% de soro. As placas são vedadas com uma tampa de vedação de silicone (tampa de vedação de silicone Axymat, VWR) e incubadas durante 1,5 h a 37 ° C. Misturas de equilíbrio são então transferidas para os poços de uma placa lavada de 96 poços de estreptavidina e incubadas adicionalmente em um termomisturador Eppendorf fixado em 37 ° C, com agitação a 800 rpm, por duas horas.

ENSAIO MANUAL

[000208] A menos que especificado de outro modo, o líquido é removido por operações de imersão, seguido por duas punções sobre toalhas de papel em camadas. Os volumes de lavagem são de 150 µL e todas as incubações de agitação são feitas em um termomisturador Eppendorf fixado em 25 ° C, 800 rpm. Misturas de equilíbrio são removidas por pipetagem, e as placas são lavadas duas vezes por 1 minuto com tampão PB1 suplementado com sulfato de dextrano a 1 mM e 500 µM de biotina, depois 4 vezes durante 15 segundos com tampão PB1. Uma solução feita recentemente de 1 mM NHS-PEO4-biotina em tampão PB1 (150 µl/poço) é adicionada, e as placas são incubadas por 5 minutos com agitação. A solução de NHS-biotina é removida, e as

placas lavadas 3 vezes com tampão PB1 suplementado com glicina 20 mM, e 3 vezes com tampão PB1. Oitenta e cinco µl de tampão PB1 suplementado com 1 mM de DxSO4 é então adicionado a cada poço, e as placas são irradiadas sob uma lâmpada de UV BlackRay (comprimento de onda nominal de 365 nm) a uma distância de 5 cm por 20 minutos com agitação. As amostras são transferidas para uma placa nova, lavada e revestida com estreptavidina, ou um poço não utilizado da placa lavada de estreptavidina existente, combinando misturas de alta e baixa diluição da amostra para um único poço. As amostras são incubadas à temperatura ambiente com agitação por 10 minutos. O material não adsorvido é removido e as placas lavadas 8 vezes por 15 segundos cada com tampão PB1 suplementado com 30% de glicerol. As placas são então lavadas uma vez com tampão PB1. Os aptâmeros são eluídos por 5 minutos à temperatura ambiente com 100 µL de tampão de eluição de CAPSO. 90 µL do eluato são transferidos para uma placa HybAid de 96 poços e 10 µL de tampão de neutralização são adicionados.

ENSAIO SEMIAUTOMATIZADO

[000209] Placas de estreptavidina carregando misturas de equilíbrio adsorvidas são colocadas na parte traseira de um lavador de placas Biotek EL406, que é programado para executar as seguintes etapas: o material não adsorvido é removido por aspiração, e os poços são lavados 4 vezes com 300 µL de tampão PB1 suplementado com 1 mM de sulfato de dextrano e 500 µL de biotina. Os poços são, então, lavados 3 vezes com 300 µL de tampão PB1. Cento e cinquenta µL de uma solução preparada fresca (a partir de um estoque de 100 mM em DMSO) de 1 mM de NHS-PEO4-biotina em tampão PB1 são adicionados. As placas são incubadas durante 5 minutos com agitação. O líquido é aspirado, e os poços são lavados 8 vezes com 300 µL de tampão PB1 suplementado com 10 mM de glicina. Cem µL de tampão

PB1 suplementado com 1 mM de sulfato de dextrano são adicionados. Após estas etapas automatizadas, as placas são removidas do lavador de placas e colocadas em um termoagitador montado sob uma fonte de luz UV (BlackRay, comprimento de onda nominal de 365 nm) a uma distância de 5 cm por 20 minutos. O termoagitador é ajustado a 800 rpm e 25° C. Após 20 minutos de irradiação, as amostras são transferidas manualmente para uma placa de estreptavidina nova e lavada (ou para um poço não utilizado da placa lavada existente). Misturas de reação de alta abundância (3% de soro+3% de mistura de aptâmero) e de baixa abundância (0,3% de soro+ 0,3% de mistura de aptâmero) são combinadas em um único poço, neste ponto. Esta placa de "captura-2" é colocada sobre a plataforma de lavagem de placas Biotek EL406, que é programada para executar as seguintes etapas: a placa é incubada durante 10 minutos com agitação. O líquido é aspirado, e os poços são lavados 21 vezes com 300 µL de tampão PB1 suplementado com 30% de glicerol. Os poços são lavados 5 vezes com 300 µL de tampão PB1, e a lavagem final é aspirada. Cem µL de tampão de eluição CAPSO são adicionados, e aptâmeros são eluídos durante 5 minutos com agitação. Após estas etapas automatizadas, a placa é então removida da parte traseira do lavador de placas, e alíquotas de 90 µL das amostras são transferidas manualmente para os poços de uma placa de 96 poços HybAid que contém 10 µL de tampão de neutralização.

HIBRIDAÇÃO PARA CUSTOMIZAR MICROMATRIZES DE AGILENT 8×15K

[000210] 24 µL do eluato neutralizado são transferidos para uma nova placa de 96 poços e 6 µL de Agilent Block 10x (Kit de hibridização Oligo aCGH/ChIP-on-chip, grande volume, Agilent 5188-5380), que contém um conjunto de controles de hibridação compostos por aptâmeros de 10 Cy3 são adicionados a cada poço. Trinta µL de tampão de hibridização 2× Agilent são adicionados a cada amostra e misturados. Quarenta µL

da solução de hibridação resultante são pipetados manualmente em cada "poço" da lâmina de gaxeta de hibridização (placa de gaxeta de hibridização, 8 micromatrizes por formato de lâmina, Agilent). Lâminas de micromatriz da Agilent sob encomenda, tendo 10 sondas por matriz complementares a 40 nucleotídeos de região aleatória de cada aptâmero com um ligante de 20 × dT, são colocadas nas lâminas de gaxeta de acordo com o protocolo dos fabricantes. O conjunto (kit de câmara de hibridização - ativado por SureHyb, Agilent) é apertado e incubado durante 19 horas a 60° C durante a rotação a 20 rpm.

LAVAGEM PÓS-HIBRIDIZAÇÃO

[000211] Aproximadamente 400 mL de tampão de lavagem Agilent 1 são colocados em cada uma de dois pratos de coloração de vidro separados. As lâminas (não mais do que duas por vez) são desmontadas e separadas enquanto estão submersas no tampão de lavagem 1 e, em seguida transferidas para um porta-lâminas, em um segundo prato de coloração contendo também o tampão de lavagem 1. As lâminas são incubadas durante um período adicional de 5 minutos no tampão de lavagem 1 com agitação. As lâminas são transferidas para o tampão de lavagem 2 pré-equilibrado a 37° C e incubadas durante 5 minutos com agitação. As lâminas são transferidas para um quarto prato de coloração contendo acetonitrilo, e incubadas durante 5 minutos com agitação.

IMAGIOLOGIA DE MICROMATRIZ

[000212] Lâminas de micromatrizes têm a imagem formada com um sistema de scanner Microarray Agilent G2565CA, usando o canal Cy3 em resolução 5 mm no ajuste PMT 100%, e a opção DRX habilitado em 0,05. As imagens TIFF resultantes são processadas usando extração de característicasde Agilent de software versão 10. 5. 1. 1 com o protocolo GE1_105_Dec08.

PROJETO DE SONDA LUMINEX

[000213] As sondas imobilizadas para as microesferas têm 40 desoxinucleotídeos complementares à extremidade 3' da região aleatória de 40 nucleótidos do aptâmero alvo. A região complementar de aptâmero é acoplada a microesferas Luminex através de um ligante de hexaetilenoglicol (HEG) que suporta um terminal amino 5'. Desoxioligonucleotídeos de detecção biotinilada compreendem 17-21 desoxinucleotídeos complementares à região do iniciador 5'de aptâmeros alvo. Uunidades de biotina são acrescentadas às extremidades 3' dos oligos de detecção.

ACOPLAMENTO DE SONDAS PARA MICROESFERAS LUMINEX

[000214] As sondas são acopladas a Microesferas Luminex Microplex essencialmente de acordo com as instruções do fabricante, mas com as seguintes modificações: Quantidades de oligonucleotídeos amino-terminais são microesferas de 0,08 nmol por $2,5 \times 10^6$, e a segunda adição de EDC é 5 µL a 10 mg/mL. Reações de acoplamento são realizadas em um conjunto de termoagitador Eppendorf a 25° C e 600 rpm.

HIBRIDAÇÃO DE MICROESFERA

[000215] Soluções carga de microesfera (cerca de 40000 microesferas/ µL) são agitadas e sonicadas em um limpador ultrassônico da Health Sonics (Modelo: T1. 9C) durante 60 segundos para suspender as microsferas. Microesferas suspensas são diluídas para 2000 microsferas por reação em 1,5 × soluções de hibridação TMAC e misturadas por agitação em vórtex e sonicação. Trinta e três µL por reação da mistura de microesferas são transferidos para uma placa de 96 poços Hybaid. Sete µL de 15 nM de carga de oligonucleotídeo biotinilada de detecção em um tampão 1x TE são adicionados a cada reação e misturados. Dez µL de amostra de ensaio neutralizada são adicionados e a placa é vedada com uma tampa de vedação de silicone. A placa é incubada primeiro a 96° C durante 5 minutos e incubada a 50°

C sem agitação durante a noite em um forno de hibridação convencional. Uma placa de filtro (Durapore, Millipore número de parte MSBVN1250, 1,2 µm de tamanho de poro) é pré-umedecida com uma solução de hibridização de 75 µL 1 × TMAC, suplementada com 0,5% (w/v) de BSA. O volume da amostra inteira a partir da reação de hibridização é transferido para a placa de filtro. A placa de hibridização é lavada com uma solução de hibridação de 75 µL 1x TMAC contendo 0,5% de BSA e qualquer material remanescente é transferido para a placa de filtro. As amostras são filtradas sob vácuo lento, com 150 µL de tampão evacuado durante cerca de 8 segundos. A placa de filtro é lavada uma vez com a solução de hibridização com 75 µL de 1x TMAC contendo 0,5% de BSA e as microesferas na placa de filtro são ressuspensas em uma solução de hibridização com 75 µL de 1x TMAC contendo 0,5% de BSA placa de filtro é protegida da luz e incubada em um misturador térmico R da Eppendorf durante 5 minutos a 1000 rpm. A placa de filtro é, em seguida, lavada uma vez com uma solução de hibridização com 75 µL de 1x TMAC contendo 0,5% de BSA. 75 µL de 10 µg/mL de ficoeritrina estreptavidina (SAPE-100, Moss, Inc.) em uma solução de hibridação 1x TMAC são adicionados a cada reação e incubados no misturador térmico R Eppendorf a 25° C a 1000 rpm durante 60 minutos. A placa de filtro é lavada duas vezes com a solução de hibridização com 75 µL de 1x TMAC contendo 0,5% de BSA e as microesferas na placa de filtro são ressuspensas em uma solução de hibridização com 75 µL de 1x TMAC contendo 0,5% de BSA. A placa de filtro é então incubada protegida da luz em um misturador térmico R da Eppendorf durante 5 minutos a 1000 rpm. A placa de filtro é, em seguida, lavada uma vez com uma solução de hibridização com 75 µL de 1x TMAC contendo 0,5% de BSA. As microesferas são ressuspensas em uma solução de hibridização com 75 µL de 1x TMAC contendo 0,5% de BSA, e analisadas em um instrumento Luminex 100 executando o

software Xponent 3. 0. Pelo menos 100 microesferas são contadas por tipo de microesfera, sob calibração alta PMT e um cenário discriminador de par de 7. 500-18. 000.

LEITURA DE QPCR

[000216] As curvas padrão para qPCR são preparadas em água variando de 108 a 102 cópias com diluições de 10 vezes e um controle sem modelo. As amostras do ensaio neutralizadas são diluídas 40 vezes em diH2O. A mistura principal qPCR é preparada a uma concentração final 2x (tampão de KOD 2x, mistura de 400 µM de dNTP, mistura para frente e reversa de 400 nM de iniciadores, 2× SYBR Green I e 0,5 U de KOD EX). Dez µL de 2× qPCR da mistura principal são adicionados a 10 µL de amostra de ensaio diluída. qPCR é executado em um BioRad MyIQ iCycler com 2 minutos a 96° C, seguidos de 40 ciclos de 96° C durante 5 segundos e 72° C durante 30 segundos.

EXEMPLO 2. MÉTODOS

PROJETO DE ESTUDO E COLETA DE AMOSTRAS

[000217] Amostras de plasma arquivadas de pacientes com doença coronariana estável foram obtidas a partir de dois estudos bem conhecidos, de coorte independentes. As características da população do estudo são mostradas na Tabela 1. Realizamos a descoberta de biomarcadores de proteína e treinamento do modelo em 938 amostras de plasma a partir do estudo Heart and Soul, com a subsequente acompanhamento de 10 anos. *Ver, por exemplo*, Shlipak et al., *Am J Med.* 2008; 121: 50-57; Whooley et al., *JAMA*. 2008; 300: 2379-2388. Validamos o modelo em 971 amostras de HUNT3, um estudo prospectivo de coorte norueguês com acompanhamento de 5 anos. *Ver* Krokstad et al., *Int J Epidemiol*. 2013;42: 968-977. Foram utilizados os critérios de inclusão e exclusão de Heart and Soul na seleção de todos os participantes com doença coronariana estável do coorte HUNT3 maior para esta análise. A descoberta das amostras de plasma foi

representante de um estudo prospectivo acadêmico bem controlado: Os indivíduos foram mantidos em jejum, as amostras coletadas na mesma hora do dia e centrifugadas e congeladas a -80° C no decurso de uma hora após a coleta. Em contraste, a coleta de amostras no conjunto de validação HUNT3 foi representativa das condições mais prováveis do "mundo real"; os indivíduos não estavam em jejum, foram vistos em horas variadas do dia, e o plasma não foi separado a partir de células até 24 horas enquanto que as amostras permaneceram em 4°C. Avaliar o desempenho do modelo dessa maneira nos permite verificar a robustez do modelo de fatores associados à coleta prática de amostras clínicas, uma consideração importante para validação de biomarcadores. *Ver* McShane et al., *Nature*. 2013;502: 317-320. Ambos os estudos foram aprovados pelos conselhos de revisão institucionais relevantes.

TABELA 1: Características da população de estudo

	Descoberta (Heart and Soul)	**Validação (HUNT3)**
Origem da amostra	Estudo de coorte baseado em UCSF prospectivo em 12 ambulatórios na área da Baía de San Francisco	Coorte aninhado de 1017 pacientes do estudo prospectivo de coorte norueguês em 50. 807 participantes
Critério de entrada	Diagnósticos de doenças cardíacas coronárias estáveis por MI prévia, estenose de > 50% na angiografia, esquemia induzida por exercício, revascularização prévia	Todos os indivíduos tinham doença coronariana estável, selecionados por critérios idênticos para coorte Heart and Soul, exceto que dados do exercício não estavam disponíveis.
Processamento das amostras	Datas de coleta: 2000-2002 Jejum Tempo fixo do dia Plasma EDTA Tempo para a separação das células, tipicamente <1 h Armazenamento a -80°C	Datas de coleta: 2006-2008 Sem Jejum Tempo aleatório do dia Plasma EDTA Tempo para separação de células até 24 horas, a amostra mantida a 4 °C

	Descoberta (Heart and Soul)	Validação (HUNT3)
		Armazenamento a -80ºC
Evento e adjudicação de evento	Ponto final evento compósito definido como o primeiro de: morte por qualquer causa; hospitalização por infarto do miocárdio; acidente vascular cerebral ou ataque isquêmico transitório; hospitalização por sinais e sintomas de insuficiência cardíaca. Cada evento foi julgado por 2 revisores independentes e cegos. Em caso de desacordo, os juízes conferiram, reconsideraram sua classificação, e, se necessário, solicitaram a consulta de um terceiro juiz cego.	Definições de eventos são as mesmas para a descoberta, adjudicadas de revisão de prontuários por um cardiologista experiente
Tempo de acompanhamento	Data do último acompanhamento: 11,09 anos Tempo de acompanhamento mediano: (IQR)7,9 (5,5) anos	Data do último acompanhamento: 5,57 anos Tempo de acompanhamento mediano: (IQR)4,3 (1,0) anos
Interrupção	Técnicos de laboratório cegos para as características clínicas e resultados. Adjudicação de resultados cegos para resultados proteômicos.	Técnicos de laboratório cegos para as características clínicas e resultados. Adjudicação de resultados cegos para resultados proteômicos.
Aplicação do modelo	Biomarcadores identificados, modelos treinados, modelos aplicados	Nenhum biomarcadores identificados, nenhum modelos treinados, apenas modelos treinados Heart and Soul aplicados

ENSAIO PROTEÔMICO SOMAscan

[000218] Os reagentes de afinidade individuais utilizados no ensaio de proteína são aptâmeros de DNA modificado de taxa de desaceleração lenta (SOMAmers) com uma afinidade muito elevada para os seus alvos de proteína. *Ver* Vaught et al., *J Am Chem Soc*. 2010; 132: 4141-4151. Modificações químicas das bases de DNA nos aptâmeros intensificam as suas características de ligação. *Ver* Davies et al., *Proc Natl Acad Sci*

USA. 2012;109: 19971-19976. Utilizamos 1130 destes reagentes no ensaio multiplex SOMAscan™ 8-10. Em resumo, uma amostra de plasma em cada poço de uma placa de 96 poços foi incubada com uma mistura de SOMAmers que se ligam às suas proteínas alvo. Duas etapas de imobilização baseadas em microesferas permitem a eliminação das proteínas não especificamente ligadas ou não ligados e a eliminação de SOMAmers não ligados. Apenas reagentes ligados a proteínas alvo sobrevivem ao ensaio, com o número de cada um quantitativamente proporcional à concentração da proteína na amostra original. O DNA em cada um dos reagentes é quantificado em uma matriz de hibridação Agilent, e as amostras normalizadas e calibradas de modo que o grau de fluorescência no local na matriz se refere à concentração de uma proteína específica. As 1054 proteínas que passaram pelo controle de qualidade média tinham intra-ensaio e inter-ensaio de coeficiente de variação <5%. Ver Gold et al., *PLoS One*. 2010;5: e15004.

Ensaio e dados SOMAscan

[000219] As amostras de plasma foram analisadas ao longo de um período de 3 semanas de trabalho em 32 percursos de ensaio separados. As amostras do estudo foram aleatoriamente designadas para percursos de ensaio juntamente com um conjunto de amostras de calibração e controle. Nenhuma informação de identificação estava disponível para os técnicos de laboratório que operam o ensaio.

[000220] A normalização intrapercurso e a calibração interpercurso foram realizadas de acordo com os procedimentos de controle de qualidade (QC) de dados de ensaio SOMAscan Versão 3, conforme definido no sistema de qualidade de boas práticas de laboratório (GLP) SomaLogic. A calibração interpercurso é projetada para remover "efeitos de mistura" entre os percursos sucessivos do ensaio, enquanto a normalização intra-percurso remove alterações em massa em

concentração de proteína (e, portanto, a intensidade do sinal) entre amostras dentro de cada prazo.

[000221] Resumidamente, a calibração interpercurso dimensiona o nível de sinal para cada proteína de modo que esse nível observado no padrão de calibração de execução corresponda ao nível esperado representado pela referência de calibração externa. As tolerâncias QC são definidas em termos da magnitude do dimensionamento multiplicativo necessário para igualar o nível de sinal médio sobre os padrões de calibração de réplica para os níveis de sinal gerado por referência externa.

[000222] Os controles de normalização intrapercurso para predisposições de intensidade de sinal "em massa" que podem resultar tanto da eficiência de hibridação diferencial ou da diluição da amostra diferencial (ou outros artefatos de protocolo de coleta) que alteram a concentração de proteína total da amostra. O primeiro efeito é capturado por um conjunto de controles utilizados para monitorar a reação de hibridação para cada amostra e a segunda usa a média da razão de níveis de sinal médio de cada amostra para o nível de sinal médio sobre todas as amostras dentro do percurso. Não é incomum que as diferenças no protocolo de coleta de amostra gerem um uma predisposição de intensidade sistemática nos níveis de sinal para um grande número de proteínas. A Figura 1 mostra diagramas de caixa dos fatores de escala *multiplicativos* nas duas coortes quando as proteínas 1130 são agrupadas por diluição da amostra. Proteínas medidas nas diluições de amostras de 40% e 1% tinham sistematicamente *maiores (menores)* níveis de sinal no conjunto de validação (descoberta), resultando em fatores de escala de normalização correspondentes menores (maiores) que um. Após o procedimento de normalização, o nível de sinal médio para as proteínas em cada uma das três diluições é o mesmo nos conjuntos de descoberta e de validação.

[000223] Os níveis de proteína foram apresentados em unidades de fluorescência relativas (RFU) e foram transformados em log antes da análise subsequente.

Amostras e proteínas excluídas da análise

[000224] As proteínas foram excluídas da análise se a tolerância inter-percurso de controle de qualidade de calibragem associada (QC) fosse excedida em, pelo menos, um dos 32 percursos do ensaio independente. Isso ocorreu com 76 proteínas; em muitos casos, a *maioria* dos percursos estava dentro da tolerância necessária, mas para simplificar optou-se por excluir todas as 76 proteínas na análise de descoberta de biomarcadores aqui apresentada.

[000225] As amostras foram excluídas da análise de descoberta de biomarcadores pelos seguintes motivos: 1) não cumprimento da tolerância de normalização QC intra-percurso, 2) um número anormalmente elevado de casos anômalos, ou 3) evidência de hemólise indicada por níveis extremos de hemoglobina ou de técnicos de ensaio notando cor de plasma aberrante (vermelha). Valores aberrantes de proteínas individuais foram definidos como as proteínas com os níveis do sinal fora da variedade dada pelos desvios absolutos de média ± 6 *(MDNA)[1] - amostras de pacientes com valores extremos em mais de 5% das proteínas medidas foram excluídas da análise. A Tabela 2 resume o número de amostras excluídas com base em cada critério.

TABELA 2: Amostras excluídas por critério

	QC de normalização Critérios falharam	**Potencial Hemólise**	**>5% de Proteína Casos Atípicos**	**Total**
Conjunto de descoberta	18	22	15	55
Conjunto de Validação	10	27	10	47

[1] Deixe $\Phi^{-1}(Z)$ denotar o inverso da função de distribuição cumulativa normal.Em seguida, para dados distribuídos normalmente a estimativa robusta MDNA (x) = $\sigma*\Phi^{-1}(3/4)$, de modo 3 * MDNA ≈ 2σ e a faixa indicada é ± 4σ para medições de Gauss.

total	28	49	25	

MÉTODOS ESTATÍSTICOS

[000226] O resultado deste estudo foi definido como o primeiro evento entre morte, infarto do miocárdio (MI), acidente vascular cerebral, ataque isquêmico transitório (TIA), ou hospitalização por insuficiência cardíaca. Usamos modelos de perigo Cox proporcionais para estimar as associações univariadas entre os níveis de proteína e risco de eventos cardiovasculares, como se segue.

Seleção de proteínas previsoras de risco cardiovascular

[000227] Modelos de perigo proporcional de Cox única variável foram utilizados para identificar um conjunto de proteínas individualmente associadas com risco aumentado de eventos cardiovasculares secundários. Em um nível de significância corrigido de 5% Bonferroni, exatamente 200 proteínas foram associadas com risco aumentado de resultados cardiovasculares. Os gráficos de "vulcão" na Figura 2 mostram o logaritmo negativo do valor de p da estatística de Wald como uma função da razão de risco, quer por desvio padrão de unidades relativas de fluorescência (RFU) (em cima) ou entre os níveis extremos do indicador categórico de filiação de quartil RFU (parte inferior). Neste último caso, a taxa de risco relatada dá o aumento de perigo experimentado por um indivíduo no quartil de maior risco (4) em comparação com um indivíduo no quartil de menor risco (1º).

[000228] Algumas destas 200 proteínas são associadas com tamanhos de efeito relativamente pequenos, mas as 117 indicadas na Tabela 14 têm taxas de risco fora do intervalo [0,75-1,25]. Examinando a estrutura de correlação correspondente (dados não mostrados) entre estas 200 proteínas revela vários conjuntos de proteínas com correlações de pares semelhantes. Uma discussão abrangente sobre a função biológica destes aglomerados de proteína está além do escopo deste manuscrito e será discutida em outro lugar.

[000229] O LASSO (Tibshirani, *Stat Med* 1997; 16:385-95) foi utilizado

como um processo de triagem variável para identificar um subconjunto de proteínas associadas conjuntamente com o aumento do risco cardiovascular. Validação cruzada generalizada usando *coxnet* (Simon et ai., *Journal of Statistical Software* 2011; 39: 1-13) no pacote R *glmnet* (Friedman et ai., *Journal of Statistical Software* 2010; 33: 1-22) foi usado para definir o parâmetro de regularização LASSO. Usamos a heurística "um erro padrão" (Hastie et al., Elements of Statistical LeRNAing, Second ed.. 2 ed: Springer; 2009) para estabelecer o nível de regularização. Perturbação a etapa de validação cruzada foi utilizada como uma simples verificação da "estabilidade" do conjunto resultante de proteínas selecionadas. Esta análise nos deu a confiança de que as proteínas incluídas em CVD9 são "estáveis" na medida em que seria selecionada a maioria do tempo que o procedimento "LASSO seguido por eliminação contrária" foi aplicado. Para gerar resultados reproduzíveis para a análise seguinte, a semente do número aleatório foi fixada em 1 antes da validação cruzada LASSO. A inicialização deste valor e definição do parâmetro de regularização LASSO para o erro padrão de valor 1 acima que minimiza os resultados de desvio de probabilidade parcial de validação cruzada em um modelo LASSO contendo as 16 proteínas aqui discutidas.

[000230] Usamos LASSO apenas para seleção de variáveis, preferindo o modelo de sobrevivência totalmente paramétrico (Weibull) como um modelo prognóstico final. Este último tem uma representação simples e uma forma matemática passível (Royston et al., *BMC medical research methodology* 2013;13: 33; van Houwelingen, *Stat Med* 2000;19: 3401-15) para calibração para uso em estudos de validação externos. A eliminação reversa por etapas iniciada a partir do modelo de LASSO completo foi utilizada para remover as proteínas que não foram previsoras significativas na *ausência* da restrição imposta pela penalidade LASSO. Ao utilizar o critério de paragem de critério de

informação Bayesiano (BIC) para equilibrar o desempenho e complexidade do modelo, a eliminação reversa descartou 7 proteínas: CDO catepsina H, receptor de EGF, receptor de hormônio de crescimento, proteína de membrana de células T TIM-3, MMP-7, CDO relacionado a oncogene de adesão de célula e trombospondina-2 resultando no modelo CVD9 de 9 proteínas mostrado na Tabela 3.

TABELA 3: Ccaracterísticas de desempenho de análise dos biomarcadores CVD9

Alvo	**% de diluição da amostra**	**Calibrador % CV**	**Limites de quantificação (pg/ ml)**		**Faixa (Logs)**	**Precisão de Inter-Ensaio (n = 3) (%CV)**		
			Inferior	Superior		Baixa	Méd.	Alta
Angiopoietina-2	0,01	5,66	$8,4x10^{0}$	$1,8x10^{4}$	3,3	1,6	2,0	0,6
Complemento C7	0,00	8,66	$1,7x10^{1}$	$3,0x10^{4}$	3,2	7,0	7,2	6,3
MMP12	0,40	4,89	$2,8x10^{-1}$	$1,4x10^{4}$	4,7	1,1	2,0	2,5
Troponina I, cardíaca	0,40	7,62	-	-	-	-	-	-
Proteína 4 relacionada a angiopoietina	0,40	6. 46	$1,4x10^{1}$	$1,4x10^{4}$	3,0	9,6	4,9	7,6
Complexo de antiquimotripsina α1	0,00	5,97	$1,8x10^{0}$	$1. 9x10^{4}$	4,0	4,5	5,8	5,3
GDF11	0,01	6. 03	$6,5x1^{0}$	$4,2x10^{3}$	2,8	1,5	1,2	3,6
CCL18/PARC	0,00	6,83	$1,6x10^{-1}$	$2,6x10^{3}$	4,2	5,1	1,6	5,6
Antiplasmina α2	0,00	7,46	$9,8x10^{0}$	$1,7x10^{4}$	3,2	3,7	2,7	1,8

Modelo CVD9

[000231] O modelo final (CVD9) contém as 9 proteínas. Enquanto que o ajuste deste modelo para as variáveis clínicas melhorou o encaixe ligeiramente (ver abaixo), estes ajustes não conseguiram produzir uma melhoria *significativa* tanto no desempenho de calibração ou de discriminação com relação ao que foi obtido com o modelo "apenas proteínas" no conjunto de descoberta. Isso nos levou a designar CVD9 como nosso modelo "primário" para avaliar o desempenho de validação e um modelo incluindo idade, sexo, estado de diabetes e taxa de

filtração glomerular estimada (EGFR) como um modelo secundário.

[000232] Para um modelo do tempo de falha acelerado, a probabilidade de um evento ocorrer no intervalo [0, T] é dada por

$$\Pr[T \leq t] = 1 - e^{-e^{\left(\frac{Log(t)-PI}{s}\right)}},$$

[000233] onde PI é o *índice de prognóstico* (ou previsor linear) e s é o parâmetro de escala associado para a distribuição do valor extremo. Ao montar o modelo, trabalhou-se com variáveis padronizadas - aqui absorveu-se o meio de população e o desvio padrão para o termo de intercepto para que se pudesse relatar o índice prognóstico e fator de escala como,

$$PI = -16{,}61+1{,}55\times ANGPT2 - 1{,}22\times GDF11+2{,}12\times C7- 2{,}64\times SERPINF2+0{,}57\times CCL18+ 1{,}02\times ANGPTL4 + 1{,}43\times KLK3, SERPINA3+0{,}72\times MMP12+0{,}59\times TNNI3,$$
$$s= 0{,}85,$$

[000234] onde os níveis de proteína são levados para estarem no log10 RFU.

Incorporando variáveis clínicas

[000235] O estudo HUNT3 não foi projetado especificamente como um estudo de doenças cardiovasculares por isso alguns parâmetros de histórico médico e medições laboratoriais clínicas que estavam disponíveis no conjunto de descoberta não estavam disponíveis no conjunto de validação (por exemplo, fração de ejeção ventricular esquerda ecocardiográfica, hipertrofia ventricular esquerda, função diastólica). Com isto em mente, só se considera o ajuste para variáveis clínicas que estavam disponíveis em ambas as coletas e diferiram entre os pacientes com eventos e aqueles sem.

[000236] Quando adicionadas ao CVD9, as variáveis clínicas sexo (masculino), idade, diabetes (sim), inibidores da ECA (sim), e taxa de filtração glomerular estimada (EGFR) individualmente (e em conjunto) aumentaram o encaixe do modelo combinado resultante ($p < 0{,}001$). Inibidor da ECA ou a utilização ARB não foram incluídos no modelo final, porque a medicação não estava disponível na coorte HUNT3.

[000237] Além das 9 proteínas utilizadas na CVD9, adicionou-se primeiro idade e sexo, e então estado de diabetes e EGFR para dar aos modelos adicionais que combinam proteínas e variáveis clínicas comumente disponíveis que foram previsores dos resultados. As estimativas pontuais para os coeficientes do modelo linear de tempo de falha acelerado (AFT) e o parâmetro de escala estimado para a distribuição de valores extremos estão listados na Tabela 4. Na Tabela 4, são abreviações ANGPT2 = "Angiopoietina-2"; C7 = "Complemento C7"; SERPINF2 = "α2-Antiplasmina"; CCL18 = "ligando 18 de quimiocina (motivo C-C)" também conhecido como "quimiocina regulada por ativação e pulmonar (PARC)", ANGL4 = "proteína relacionada com angiopoietina 4; KLK3. SERPINA3 = "complexo de antiquimotripsina α1-"; TNNI3 = "troponina-I, cardíaca"; e EGFR = "taxa de filtração glomerular".

TABELA 4: Coeficientes estimados para 3 modelos candidatos

Termo de Modelo	CVD9	CVD9+ Idade+ Sexo	CVD9+ Idade+ Sexo+ Diabetes+ eGFR
Escala	0,848	0,849	0,845
Intercepto	-16,612	-18,614	-17,478
Diabetes=Sim	0	0	0,277
eGFR	0	0	-0,005
Idade	0	0,012	0,012
Sexo=Masculino	0	0,358	0,391
ANGPT2	1,547	1,712	1,510
GDF11	-1,224	1,320	-1,347
C7	2,115	2,092	2,025
SERPINF2	-2,643	-2,057	-1,663
CCL18	0,574	0,554	0,375
ANGL4	1. 022	0,902	0,848
KLK3. SERPINA3	1,433	1,409	1. 361
MMP12	0,718	0,522	0,436
TNNI3	0,588	0,587	0,610

[000238] Várias medidas diferentes de desempenho de discriminação

são comumente relatadas - relatamos uma "estatística c", o Índice Integrado de Discriminação (IDI) e o índice de reclassificação de rede livre de categoria (NRI).

[000239] A Tabela 5 lista estas medidas de discriminação, juntamente com a taxa de risco Q4/Q1 e a estatística de Hosmer-Lemeshow para o desempenho de calibração resumido para os 3 modelos. Os intervalos de confiança reportados são empíricos IC 95% gerados usando 100 amostras de bootstrap. A primeira coluna lista o valor p para o teste da razão de verossimilhança comparando os modelos ampliados à linha de base do modelo (proteína apenas). A primeira coluna indica o valor p para o teste da razão de verossimilhança (LR) comparando o modelo ampliado ao modelo de proteína única. As medidas subsequentes de discriminação são a área sob a curva ROC dinâmica/ incidente (C^T), o índice de discriminação integrado (IDI), o índice de reclassificação de rede (NRI) e da quarta à primeira taxa de risco de quartil (Q4/Q1). Desempenho de calibração avaliado com a estatística de Hosmer-Lemeshow.

[000240] A adição de variáveis clínicas cujos valores de base distinguiem os grupos do evento e não evento dá uma ligeira melhoria nas estimativas pontuais das razões de perigo de IDI, NRI (> 0) e Q4/Q1, embora a "estatística c" de AUC integrada permaneça inalterada.

TABELA 5: Medidas de desempenho de discriminação e calibração no conjunto de descoberta para o modelo CVD9

Modelo	Teste LR	C[τ]		IDI	NRI			Razão de perigo	Calibração	
	Valor p	Ano 1	Ano 4		NRI (> 0)	Evento	Nenhum Evento	Q4/Q1	Hosmer-Lemeshow χ^2	Valor p
Apenas Proteína (CVD9)	NA	0,76 (0,72-0,79)	0,74 (0,71-0,77)	0,15 (0,11-0,18)	0,57 (0,42-0,70)	0,16 (0,01-,30)	0,41 (0,35-0,47)	8,2	8,14	0,42
CVD9 + Idade+ Masculino	0,0002	0,76 (0,73-0,79)	0,75 (0,72-0,77)	0,16 (0. 12-0. 20)	0,64 (0,50-0,76)	0,19 (0,08-0,30)	0,45 (0,37-0,50)	10,0	2,69	0,95
CVD9 +Idade +Homem +Diabetes +eGFR[2]	6. 3e-6	0,77 (0,73-0,79)	0,75 (0,j72-0,77)	0,17 (0,13-0,20)	0,68 (0,54-0,80)	0,22 (0,11-0,32)	0,46 (0,39-0,52)	10,3	3,81	0,87

[2] Fórmula CKD-EPI 2009 eGFR foi usada porque ele estava disponível em ambos os conjuntos de descoberta e de validação.

Recalibração de CVD9 para validação

[000241] Antes de comparar o desempenho de CVD9 ao grau de Framingham, ambos os modelos foram recalibrados para o seu uso no conjunto de validação. Como em van Houwelingen (*Stat Med* 2000; 19:3401-15), foi utilizado um modelo de calibração de tempo de falha acelerado de Weibull para recalibrar os coeficientes do modelo para uso na população de validação. Se deixarmos PI ser o índice de prognóstico e H (t|PI) denotar a função de perigo cumulativo, então o modelo de calibração é

$$\log(H(t|PI)) = \gamma_0 + \gamma_1 PI + \gamma_2 e \ ,$$

[000242] onde o termo de erro e, tem uma distribuição de valor extremo. Denotar o perigo cumulativo de linha de base por $H_0(T)$ e usar $H(T|PI) = H_0(T)\, e^{PI}$ dá,

$$\log(H_0(t)) = \gamma_0 + (\gamma_1 - 1)PI + \gamma_2 e. \quad (1)$$

[000243] Uma avaliação de calibração formal (chamada de "validação por calibração" por Van Houwelingen) envolve testar a hipótese de calibração perfeita, $H_0: \gamma_0 = 0, \gamma_1 = 0, \gamma_2 = 1$. Encaixe do modelo (1) usando *survreg* do pacote R de *survival* (Therneau, um pacote para análise de sobrevivência em S. R versão do pacote 237-7 2014) dá os coeficientes de calibração listados na Tabela 6.

TABELA 6: Coeficientes para o modelo de calibração de Weibull aplicado a CVD9 para utilização no conjunto de validação.

CVD9				
	Estimativa	**95% de CI**		**Valor p**
$\widehat{\gamma_0}$	-0,230	-0,4189	-0,0412	0,02
$\widehat{\gamma_1} - 1$	-0. 998	-1,1752	-0,8212	0,98
$\log(\widehat{\gamma_2})$	0,149	0,0384	0,2598	0,008

[000244] A intercepção ($\widehat{\gamma_0}$) o termo de escala ($\widehat{\gamma_2}$) indicam que CVD9 necessita de calibração antes de ser aplicado à coorte de validação, embora, como discutido abaixo, a intensidade de polarização sistemática no conjunto de validação é responsável pela maior parte da

contribuição para o termo de intercepção.

[000245] As amostras de sangue no conjunto de validação HUNT3 foram coletadas por meio de um protocolo de coleta mais brando do que no conjunto de descoberta e, como resultado observou-se uma predisposição de intensidade sistemática na maior parte das 1054 proteínas medidas nas amostras de validação. Tal como aqui discutido, esta predisposição foi em grande parte removida pelas etapas de normalização, embora, como mostrado a seguir uma predisposição *residual* permanece nos níveis de sinal para as 9 proteínas utilizadas no modelo CVD9. Esta predisposição é um artefato do processo de normalização (amostras de validação têm níveis *superiores* de sinal do que amostras de descoberta antes da normalização, mas os níveis de sinal inferiores posteriormente) e como mostrado abaixo é em grande parte responsável pelo valor estimado do coeficiente, $\widehat{\gamma_0}$.

[000246] A intercepção da linha de regressão robusta na Figura 3 dá uma estimativa da predisposição de intensidade comum a todas as 9 proteínas em CVD9. Se deixarmos Δ denotar a predisposição estimada e β_j e σ_j serem o coeficiente de modelo e desvio padrão da população para a j° proteína, em seguida, aplicando CVD9 para os resultados dos dados de validação na adição do fator constante,

$$\Delta \sum_{j=1}^{9} \frac{\beta_j}{\sigma_j} \qquad (2),$$

[000247] sobre o que seria a contribuição do modelo de interceptação. Desta maneira, a intensidade de polarização nos sinais de proteína aparece como uma discrepância na *escala de tempo* da função de sobrevivência da linha de base nos conjuntos de descoberta e de validação, precisamente o termo relacionado com o parâmetro γ_0 no modelo de calibração (1). Usando a estimativa Δ = -0,056 gerada pela intercepção da regressão linear robusta em (2), juntamente com os coeficientes do modelo CVD9 e desvios-padrão da população subtrai 0,23145 do previsor linear, quase exatamente o valor ($\widehat{\gamma_0}$)

estimado para o interceptador no modelo de calibração. Assim, a predisposição de intensidade "residual" remanescente após o procedimento de normalização é a grande responsável pela magnitude ($\widehat{\gamma_0}$) ao invés de uma discrepância real entre as funções de sobrevivência da linha de base nas coortes de descoberta e validação.

[000248] A predisposição de intensidade de sinal na coleta de amostras HUNT3 é um aspecto deste conjunto de validação especial, que não esperamos para generalizar as amostras coletadas no âmbito de protocolos de coleta mais rigorosos. Com isto em mente, avaliou-se o desempenho no conjunto de validação através do modelo CVD9 *recalibrado* descrito abaixo.

[000249] Quando a distribuição do tempo de evento é Weibull com escala *a* e formato *b,* a função de sobrevivência da linha de base correspondente é

$$\Pr_{\text{cal}}[T \geq t] = e^{-\left(\frac{t}{a}\right)^b},$$

[000250] que escreve-se em termos de risco de linha de base cumulativa (H_0) como log (H_0) = Log b (t/a). Substituindo isto pelo lado esquerdo da equação (1), e utilizando os coeficientes de calibração na Tabela 5, a expressão resultante para gerar valores de risco pode ser apresentada sob a forma do modelo do tempo de falha acelerado,

$$\log(t|z) = \beta_0^{cal} + \beta^{cal^T} z + s^{cal} e,$$

[000251] com coeficientes "calibrados" do modelo:

$$\beta_0^{cal} = \log\left(a e^{\frac{\gamma_0}{b}}\right), \beta_j^{cal} = -\frac{(\gamma_1 - 1)}{b}\beta_j^{cox}, s^{cal} = \frac{\gamma_2}{b}.$$

[000252] Usando esses coeficientes do modelo, o grau de risco associado *calibrado* é gerado usando

$$\Pr_{\text{cal}}[T \leq t \mid z] = 1 - e^{-e^{u^{cal}(z)}},$$

[000253] em que:

$$u^{cal}(z) = \frac{\left(\log t - \left[\beta_0^{cal} + \beta^{cal^T} z\right]\right)}{\left(s^{cal}\right)}.$$

[000254] O índice prognóstico resultante (PI) e fator de escala de valor extremo para o modelo CVD9 *recalibrado* utilizado no conjunto de validação são:

$$\begin{cases} PI = -16,39 + 1,55 \times ANGPT2 - 1,22 \times GDF11 + 2,11 \times C7 \\ \quad -2,64 \times SERPINF2 + 0,57 \times CCL18 \\ \quad + 1,02 \times ANGPTL4 + 1,43 \times KLK3, SERPINA3 + 0,72 \times MMP12 + 0,59 \times TNNI3 \\ s = 0,98 \end{cases}$$

[000255] Modelos de calibração semelhantes foram construídos para as variantes de CVD9 que incluem variáveis clínicas. Os coeficientes de modelo de calibração resultantes estão listados na Tabela 7. Como foi o caso para CVD9, os modelos que incluem variáveis clínicas tinham a mesma intensidade de polarização sistemática nas 9 proteínas e juntos esta predisposição gerou uma contribuição de -0,254 e -0,245 às ($\widehat{\gamma_0}$) estimativas nos respectivos modelos de calibração.

TABELA 7: Coeficientes para modelos de calibração Weibull aplicados a variantes de CVD9 que incluem variáveis clínicas.

CVD9 + Idade + Sexo="Masculino"				
	Estimativa	95% de CI		Valor p
$\widehat{\gamma_0}$	-0,275	-0,4633	-0,08623	0,004
$\widehat{\gamma_1} - 1$	-1,042	-1,2388	-0,8460	0,67
$\log(\widehat{\gamma_2})$	0,1577			0,005
CVD9 Idade Sexo = "Masculino" Diabetes = "Sim" eGFR				
	Estimativa	95% de CI		Valor p
$\widehat{\gamma_0}$	0,228	-0,4228	-0,03383	0,03
$\widehat{\gamma_1} - 1$	-1,079	-1,3041	-0,8532	0,49
$\log(\widehat{\gamma_2})$	0,1737			0,002

[000256] Após a identificação de proteínas significativamente associadas a eventos cardiovasculares (após a correção de Bonferroni em um nível de significância de 5%), utilizou-se regressão de Cox de L1 penalizado (LASSO, ver Tibshirani, *Stat Med*. 1997; 16:385-395) para propósitos de seleção variáveis (proteína). Em virtude de selecionar simultaneamente variáveis e encolher coeficientes de atendimento,

LASS dá bons modelos preditivos, como tem sido amplamente demonstrado. *Ver* Hastie et al., *Elements of statistical learning*, second ed. Springer; 2009. Tais abordagens de penalização de L1 são especialmente eficazes em ambientes de predição de alta dimensão exemplificados pelas nossas 1054 proteínas. Para obter um modelo totalmente paramétrico, foi aplicada eliminação reversa em etapas para um tempo de falha acelerada de Weibull usando o conjunto completo de proteínas selecionadas LASSO. Isto removeu os 7 contribuintes menos importantes e resultou no modelo de 9 proteínas parcimonioso (CVD9), um modelo de prognóstico totalmente paramétrico no espírito de Framingham.

[000257] Como padrão de referência comparativa respeitados, as previsões de risco foram geradas a partir do modelo de risco de evento secundário de Framingham (D'Agostino et al., *Am Heart J*. 2000; 139: 272-281) recalibrado para utilização nos conjuntos de dados de identificação e de validação, como se segue.

[000258] D'Agostino apresenta o seguinte modelo de tempo de falha acelerado para predição de eventos cardiovasculares secundários:

$$\Pr[T \leq t] = 1 - e^{-e^{\left(\frac{Log(t) - PI_{FR}}{s}\right)}},$$

[000259] onde o índice de prognóstico e o parâmetro de escala para os homens é

$$Male: \begin{cases} PI_{FR} = 4.995 - 0.0145 \times \text{Age} - 0.6738 \times \text{Ln}\left(\frac{\text{TotalChol}}{\text{HDL}}\right) - 0.3042 \times \text{DiabetesStatus}, \\ s = 0.9994 \end{cases}$$

$$Female: \begin{cases} PI_{FR} = 13.537 - 0.0225 \times \text{Age} - 0.834 \times \text{Ln}\left(\frac{\text{TotalChol}}{\text{HDL}}\right) - 0.7829 \times \text{DiabetesStatus} - 1.3713 \times \text{Ln}(\text{SBP}) - 0.3669 \times \text{Smoker}, \\ s = 1.031 \end{cases}$$

[000260] Antes de comparar o modelo de Framingham para CVD9, recalibrou-se o modelo para utilização nos conjuntos de descoberta e de validação.

Recalibração de Framingham para os conjuntos de descoberta e de validação

[000261] Para recalibrar o grau de risco de Framingham secundário para uso no conjunto de descoberta e no conjunto de validação, foi utilizada uma calibração de perigo proporcional de Cox variável única (van Houwelingen, *Stat Med* 2000; 19:3401-15; Steyerberg. Modelos de predição clínicos: Springer; 2010) modelo. Denotar a função sobrevivente da linha de base por $S_0(T)$, grau de risco de Framingham calibrado de 4 anos é

$$\Pr_{\text{cal}}[T \leq t] = 1 - \widehat{S_0}(4)^{e^{\widehat{\beta^*} \times PI_{FR}}},$$

[000262] onde $\widehat{\beta^*}$ é a estimativa do encaixe de modelo de calibração para os valores do índice de prognóstico de Framingham no conjunto de amostra em particular e $\widehat{S_0}(t)$ é a estimativa de Kaplan-Meir da função de sobrevivência naquela população. A Tabela 8 lista os coeficientes resultantes do modelo de calibração.

TABELA 8: Coeficientes de calibração para o modelo de calibração proporcional de Cox variável único usado para recalibrar o risco secundário de Framingham para o respectivo conjunto de descoberta ou de validação.

	Calibração de Cox estimada Coeficiente ($\widehat{\boldsymbol{\beta}^*}$)	**Erro Padrão**	**Valor p**
Descoberta	0,472	0,066	<0,001
Validação	0,396	0,067	<0,001

[000263] O desempenho de calibração foi avaliado através da avaliação do acordo entre a frequência de eventos observados e previstos. A Figura 4 mostra a frequência de eventos previstos e observados para cada decil de risco para o modelo de Framingham na descoberta e Figura 5 nos conjuntos de validação. Em cada caso, o quadro à esquerda mostra o grau original de Framingham e o quadro à direita mostra a o grau recalibrado utilizando o modelo com coeficientes listados na Tabela 8.

[000264] Embora o desempenho de calibração tenha sido aceitável no

coorte de descoberta, um nível semelhante de acordo entre frequências de eventos observados e previstos não foi alcançado na coorte de validação, como pode ser visto na Figura 5. Relatamos a estatística χ^2 (Hosmer-Lemeshow) para resumir o desempenho de calibração apresentado graficamente na Figura 4 e na Figura 5 - esta estatística e valor p associado foram computados com a função de plotCalibration no pacote R previsível PredictABLE. Kundu et al., PredictABEL: Avaliação dos modelos de predição de risco. pacote R versão 12-1 2012. Os valores de p de 0,70 e 0,02 para o teste Hosmer-Lemeshow são consistentes com boa calibração do modelo de Framingham na descoberta e na calibração pobre na coorte de validação.

[000265] As entradas na Tabela 9 resumem o desempenho de discriminação e de calibração dos graus de recalibração de Framingham em ambos os conjuntos de descoberta e de validação. Como discutido em maior detalhe aqui, relatamos duas medidas de desempenho de discriminação, a razão de risco entre o quarto e o primeiro quartil e a "estatística c". Para o último índice de concordância relatamos a área ponderada sob a curva de incidentes/ dinâmica de ROC, C^{τ} por τ= 4 anos. As estatísticas c são consistentes com discriminação relativamente pobre do modelo de Framingham nas coortes de descoberta e de validação.

TABELA 9: Desempenho de discriminação e calibração de modelos de Framingham recalibrados nos conjuntos de descoberta e de validação.

Conjunto de dados	Discriminação			Calibração	
	C^1 (Ano 1)	C^4 (Ano 4)	HR Q4/Q1	Hosmer-Lemeshow χ^2	Valor p
Descoberta	0,620	0,615	2,8	5,54	0,70
Validação	0,616	0,609	2,3	18,75	0,02

[000266] Como este grau foi validado para as previsões até e incluindo 4 anos, foi utilizado o intervalo de tempo de quatro anos para as

comparações de desempenho com o modelo de proteína CVD9. Também foi calculado o índice de reclassificação de rede livre de categoria (NRI; Pencina et al, *Stat Med*. 2011; 30:11-21) para o modelo de proteína CVD9 vs. Framingham tal como discutido abaixo. O grau de risco de Framingham foi previamente validado apenas para previsões de MI e morte, mas também foram previstos eventos de derrame e de insuficiência cardíaca. Reservamo-nos o grau de risco de evento de Framingham secundário como um comparador porque neste estudo o seu desempenho é semelhante em todos os tipos de eventos e porque é visto como o resultado mais provável de interesse para a comunidade científica para esta população. O processo que gerou o modelo de predição de risco cardiovascular de multiproteína e as métricas que comparam o mesmo com o grau de risco de evento secundário de Framingham (D'Agostino et al., *Am Heart J*. 2000; 139:272-281) encontra-se resumidos na Figura 6 e discutido adiante. O impacto da adição de parâmetros clínicos comumente disponíveis (selecionado de variáveis que estavam disponíveis em ambos os grupos e diferiam entre pacientes com eventos e aqueles sem) para CVD9 também foi avaliado em modelos secundários (ver acima). Toda a computação estatística foi realizada utilizando a linguagem R para Computação Estatística. *Ver* R Core Team RFfSC, Vienna, Austria R: A language and environment for statistical computing. Manual. 2013.

Desempenho de validação

[000267] O gráfico de floresta mostrada na Figura 7 mostra uma comparação das taxas de risco para as 16 proteínas LASSO em ambos os conjuntos de detecção e validação. Com a excepção da proteína 4 relacionada com Angiopoietina 4 e do complemento C7, as razões de risco para as proteínas individuais no modelo CVD9 são semelhantes nos conjuntos de descoberta e de validação. Isto é uma medida do desempenho validação das proteínas individuais - o restante desta seção discute o desempenho de validação da combinação específica de

proteínas destes resultados no modelo CVD9.

[000268] O desempenho de calibração é particularmente importante quando as previsões do modelo são usadas para informar decisões clínicas. Primeiro avaliou-se as estimativas CVD9 de risco *absoluto* na população de validação como em Steyerberg (*Epidemiologia* 2010; 21:128-38) e, em seguida, avaliou-se o desempenho de discriminação em termos de mudança na estatística C e reclassificação de risco em relação ao modelo de Framingham.

Calibração

[000269] O desempenho de calibração foi avaliado através da avaliação do acordo entre a frequência de eventos observados e previstos no intervalo de quatro anos após a amostra de sangue de linha de base. A figura 8 mostra a frequência de eventos previstos e observados para cada decil de risco na validação definida para CVD9 (à esquerda) e o modelo de Framingham (à direita) recalibrados para uso no conjunto de validação.

[000270] Em toda a variedade da frequência de eventos previstos em um decil de risco dado gerado pelo CVD9 está dentro de 8 (e, normalmente, dentro de 3) por cento da frequência de evento observada. Cada barra direita de cada par de barras representa cerca de 100 pacientes e, como as barras de erro indicam, graus de risco para pacientes em cada decil são mais semelhantes entre si do que os dos pacientes nos decis de risco vizinhos. É neste sentido que falamos de avaliação de risco "individualizada" quando se considera a informação proporcionada pelas proteínas em CVD9.

[000271] Em geral, o acordo entre as frequências de eventos previstos e observados é mais fraco no modelo de Framingham (particularmente para os pacientes no 10-20° percentis de risco). A Figura 9 mostra as curvas de predição (Pepe et al., *Stat Med* 2013; 32: 1467-1482) para CVD9 e o grau de Framingham recalibrado para utilização no conjunto de validação.

[000272] Com as classificações de risco dos dois modelos na mesma escala na Figura 9, vemos que o modelo CVD9 gera uma representação mais precisa do risco absoluto de Framingham em ambas as extremidades do espectro de risco por prever corretamente o risco (baixo) dos indivíduos abaixo do 10o percentil e o risco catastroficamente alto (65%) para indivíduos acima do percentil de risco 90. Além disso, a inclinação da curva de predição para CVD9 é mais acentuada na metade superior de percentis de risco indicando que CVD9 proporciona uma estimativa de resolução mais fina do risco absoluto para os pacientes em cada decil de risco do que o modelo tradicional de Framingham.

Discriminação

[000273] As entradas na Tabela 10 resumem o desempenho de discriminação dos modelos CVD9 e de Framingham recalibrados para a coorte de validação. Como um índice de concordância, relatou-se a área ponderada sob a curva de incidentes/ dinâmica de ROC, C^{τ}, associada com o intervalo de seguimento fixo $[0,\tau]$ (Heagerty et al., *Biometrics* 2005; 61: 92-105), que é equivalente à estatística c de Harrell, Pencina e D'Agostonio. Ver, por exemplo, Harrel et al., *Stat Med* 1996; 15: 361-87; e Pencina et al., *Stat Med* 2012; 31:1543-1553. Calculamos C^{τ} por τ= 1 e 4 anos utilizando a função risksetAUC no pacote R riskSetROC. *Ver* Heagerty et al., riskset ROC: Riskset ROC curve estimation from censores survival data. R Package versão 104 2012.

TABELA 10: Resumo de discriminação e calibração de desempenho para CVD e os modelos de Framingham recalibrados para uso em coorte de validação.

Modelo	C^{τ}		NRI			IDI	Razão de perigo	Calibração	
	Ano 1	Ano 4	NRI(>0)	Evento	Nenhum Evento		Q4/Q1	Hosmer-Lemeshow $\chi 2$	Valor p

Modelo	C$^{\tau}$		NRI			IDI	Razão de perigo	Calibração	
	Ano 1	Ano 4	NRI(>0)	Evento	Nenhum Evento		Q4/Q1	Hosmer-Lemeshow $\chi 2$	Valor p
CVD9	0,71 (0,68-0,74)	0,70 (0,67-0,73)	52% (37-67%)	18. (5-31%)	34% (26-41%)	0,10 (0,07-0,13)	6,0	7,90	0,44
CVD9 + Idade+ Masculino	0,69 (0. 66,0. 73)	0,68 (0. 65,0. 71)	41% (27-55%)	16% (3-29%)	26% (19-33%)	0,08 (0,05-0,12)	4,9	1,51	0,99
CVD9 +Idade +Homem +Diabetes +eGFR	0,67 (0,64-0,71)	0,65 (0,63-0,69)	35% (22-51%)	11% (1-25%)	24% (18-30%)	0,07 (0,04-0,09)	4,6	9,99	0,266
Framingham	0,616 (0,58-0,66)	0,609 (0,58, 0,64)	-	-	-	0,02	2,3	18,75	0,02

Curvas ROC

[000274] A Figura 10 mostra as curvas ROC para os modelos CVD9 e Framingham gerados com o pacote riskSetROC, tanto para os conjuntos de descoberta e de validação. Geramos curvas ROC em um ano e quatro para cada modelo.

Reclassificação de risco

[000275] As probabilidades de eventos de quatro anos foram geradas com o modelo CVD9 e o modelo secundário de Framingham com o último recalibrado para uso no conjunto de descoberta. O índice de reclassificação de rede livre de categoria[19] NRI (> 0), foi calculado utilizando o pacote R *nricens*. *Ver* Eisuke. NRI para modelos de predição de risco com tempo para eventos e de dados de resposta binária. Pacote R versão 12 2013.

[000276] A Tabela 11 lista os termos em NRI (> 0) e as probabilidades de reclassificação comparando CVD9 a Framingham, tanto nos conjuntos de descoberta e de validação. Intervalos de confiança reportados são intervalos de 95% empíricos computados com 100 amostras de bootstrap. Como discutido na Seção 2.1, ambos o CVD9 e

o modelo de Framingham foram recalibrados para uso no conjunto de validação anterior para esse cálculo.

TABELA 11: Índices de reclassificação de rede e probabilidades de reclassificação para CVD9 em comparação com o modelo de Framingham no conjunto de descoberta.

	Descoberta			Validação		
	Estimação pontual	CI 95% inferior	CI 95% superior	Estimação pontual	CI 95% inferior	CI 95% superior
NRI(>0)	0,57	0,43	0,71	0,52	0,38	0,67
Evento NRI	0,16	0,04	0,29	0,18	0,06	0,31
Sem Evento NRI	0,41	0,34	0,47	0,34	0,27	0,41
Pr (Risco superior|Evento)	0,58	0,52	0,64	0,59	0,53	0,65
Pr (Risco Inferior|Evento)	0,42	0,36	0,48	0,41	0,35	0,47
Pr (Risco Inferior|sem evento)	0,71	0,67	0,74	0,67	0,63	0,70
Pr (Risk Superior|sem evento)	0,29	0,26	0,33	0,33	0,30	0,37

EXEMPLO 3. RESULTADOS

CARACTERÍSTICAS DE LINHA DE BASE

[000277] As características clínicas das duas populações de estudo no início do estudo estão resumidas na Tabela 12. Como esperado, os fatores de risco conhecidos são significativamente mais prevalentes nos grupos com eventos. Houve menos eventos globais em HUNT3 do que em Heart and Soul, devido ao curto acompanhamento; no entanto, as populações eram geralmente comparáveis nas taxas de eventos por unidade de tempo e a distribuição dos tipos de evento. Na Tabela 12, os valores P estão associados com o teste exato de Fisher para covariáveis categóricas e o teste de Mann-Whitney U para covariáveis contínuas. Valores contínuos resumidos com variedade média e inter-quartil (IQR). O conjunto de validação HUNT3 não foi concebido como um estudo de CHD e, como resultado alguma informação clínica não estava disponível e é marcada N/ A. Legendas: IMC = índice de massa corporal; ACE = enzima de conversão da angiotensina; ARB = bloqueador do receptor da angiotensina; LDL-C = colesterol de

lipoproteína de baixa densidade; HDL-C = colesterol de lipoproteína de alta densidade; TG = triglicerídeos; eGFR = taxa de filtração glomerular.

TABELA 12: Características da linha de base da população de estudo

	Conjunto de descoberta (Heart and Soul)			Conjunto de validação (HUNT3)		
Resumo de Evento	**Nenhum Evento (10 anos)**	**Evento (10 anos)**	**Valor P**	**Nenhum Evento (5 anos)**	**Evento (5 anos)**	**Valor P**
# Indivíduos	473	465		699	272	
Variáveis demográficas: mediana (intervalo inter-quartil)						
Idade (anos)	64,0 (57,0-71,0)	71,0 (63,0-78,0)	<0,001	67,6 (60,0-75,3)	75,9 (67,9-81,0)	<0,001
Masculino (%)	361 (76,3)	412 (88,6)	<0,001	508 (72,7)	192 (70,6)	<0,001
Caucasiano (%)	275 (58,1)	290 (62,5)	0,18	NA	NA	NA
Diabetes (%)	96 (20,3)	151 (32,6)	<0,001	84 (12,0)	49 (18,0)	0,02
Fumante, atual (%)	88 (18,6)	96 (20,7)	0,46	142 (20,3)	56 (20,6)	0,93
IMC (kg/m^2)	28,2 (25,2-31,6)	27,1 (24,4-30,5)	<0,001	27,9 (25,8-30,7)	28,0 (25,3-30,8)	0,89
Medicações Cardiovasculares						
Estatina (%)	275 (58,1)	290 (62,5)	0,17	NA	NA	NA
ACE/ARB (%)	89 (23,5)	89 (39,9)	<0,001	NA	NA	NA
Bloqueador beta (%)	88 (18,6)	96 (20,7)	0,69	NA	NA	NA
Aspirina (%)	28,2 (25. 2-31. 6)	27,1 (24. 4-30. 5)	0,26	NA	NA	NA
Testes laboratoriais						
LDL -C (mg/dL)	98,0 (83,0-121)	100 (81,2-124,0)	0,76	NA	NA	NA
HDL-C (mg/dL)	44 (36-54)	42 (35-52)	0,044	46,4 (38,7-54,1)	42,5 (34,8-50,3)	0,002
Colesterol total (mg/dL)	174 (152-196)	169 (146-201)	0,29	174 (151-197)	178 (147-209)	0,32
TG (mg/dL)	110,0 (74.8-169)	111,0 (74,0-163,0)	0,78	142,0 (106,9-195,0)	142,0 (106. 0, 195. 0)	0,22
eGFR[3] (mL/min)	78,2 (65,3-91,9)	65,8 (52,3-82,7)	<0,001	70,9 (58,5-82,9)	60,3 (46,7-73,6)	<0,001

[3] CKD-EPI 2009, no conjunto de validação e CKD-EPI 2012 no conjunto descoberta, sempre que possível com CKD-EPI 2009, quando os valores em falta impedirem o cálculo da fórmula 2012.

Creatinina (mg/dL)	1,0 (0,8-1,1)	1,1 (0,9-1,3)	<0,001	1,0 (0,9-1,1)	1,1 (0,9-1,3)	<0,001

[1]CKD-EPI 2009, no conjunto de validação e CKD-EPI 2012 no conjunto descoberta, sempre que possível com CKD-EPI 2009, quando os valores em falta impedirem o cálculo da fórmula 2012.

PROTEÍNAS RELACIONADAS AO RISCO CARDIOVASCULAR (CV)

[000278] Em um nível corrigido de significância de 5% Bonferroni, a análise de regressão de Cox univariada revelou que 117 das 1054 proteínas que passaram pelo controle de qualidade foram associadas com risco aumentado de eventos cardiovasculares e também tinha razões de perigo do quarto ao primeiro quartil de risco univariadas de > 1,25 ou <0,75 (estas 117 proteínas estão listadas na Tabela 14 abaixo). Algumas destas proteínas foram correlacionadas, sugerindo a presença de muito menos do que 117 processos biológicos distintos; a biologia destas proteínas vai ser alvo de uma análise mais aprofundada. As razões de risco para as 16 proteínas selecionadas desta lista pelo processo de LASSO e o subconjunto de 9 proteínas escolhidas para o modelo CVD9 final estão apresentados na Figura 7. As propriedades biológicas relevantes das 16 proteínas selecionadas de LASSO estão resumidas abaixo.

[000279] Os biomarcadores identificados nesta análise, não só servem para derivar um modelo de predição de risco cardiovascular poderoso, mas também informar a compreensão da biologia da doença cardiovascular (DCV) e identificar potenciais alvos para fármacos e opções de tratamento. Abaixo, vamos dar uma breve descrição da(s) função(ões) conhecida(s) das 16 proteínas selecionadas por LASSO no modelo de predição de risco CV.

Crescimento e remodelação

[000280] Diferenciação de fator de crescimento 11 (GDF11) é um exemplo de descoberta biológica usando a ferramenta de ensaio proteômica imparcial com as descobertas de potencial de importância

clínica. Usando SOMAscan, Lee e colegas relataram perda relacionada à idade de GDF11 como a causa de hipertrofia cardíaca relacionada à idade em ratos. *Ver* Loffredo et al., *Célula* 2013; 153: 828-39. GDF11 agora está sob investigação ativa para o seu papel na supressão da hipertrofia cardíaca e falha cardíaca diastólica em seres humanos. *Ver, por exemplo*., Olson et al., *Journal of the American College of Cardiology* 2014; 63: A780. Curiosamente, enquanto concentrações de GFD-11 são reduzidas com o aumento do risco de evento cardiovascular em nosso estudo, um inibidor da atividade de GDF11, Follistatin-like 3 está positivamente associado com o aumento do risco cardiovascular (*Ver* Tabela 14).

[000281] Fator de Crescimento Epidérmico (EGF), o receptor (EGFR) é expressado em monócitos e macrófagos em lesões ateroscleróticas. Ativação pelo ligando de ligação estimula a proliferação celular e quimiotaxia. Dreux et al., Atherosclerosis 2006;186: 38-53. Evidências de estudos com animais mostra EGF recebeou protege contra a hipertrofia cardíaca e suporta a arquitetura de parede vascular adequada e reatividade do vaso. Schreier et al, *Hypertension* 2013;61: 333-40.

[000282] As formas solúveis do receptor de hormônio de crescimento (GHR) e os fatores de crescimento epidérmico podem servir como ambos os reservatórios e inibidores dos fatores circulantes envolvidos na mitogênese, na função de células, e têm funções bem conhecidas no câncer. Curiosamente, a sinalização do receptor de hormônio de crescimento, através da estimulação do seu mediador anabolizante fator de crescimento semelhante à insulina I, já foi demonstrada como tendo uma correlação negativa com o risco de desenvolver doença das artérias coronárias. Juul et al., *Circulation* 2002;106:939-44.

[000283] Angiopoietina-2 (ANGPT2), que antagoniza a atividade de Angiopoietina-1 sobre a o receptor de quinase de proteína quinocina

receptor de Tie 2 e age em concertação com Angiopoietina-1 durante a angiogênese, promove o relaxamento dos contatos da célula-matriz e pode induzir a apoptose de células endoteliais e ruptura do vaso durante a angiogênese[26]. Maisonpierre et al., *Science* 1997;277: 55-60. Um membro da mesma família de genes, proteína 4 relacionada com angiopoietina (ANGPTL4) é induzida por hipoxia e não só afeta a função vascular e interação célula-matriz endotelial, mas também o metabolismo dos lipídios como um potente inibidor da lipase lipoproteica[27]. Li et al., *Current opinion in lipidology* 2006;17: 152-6.

[000284] Interações controladas da matriz extracelular e das células são vitais para a fisiologia normal dos órgãos, durante o desenvolvimento normal, em resposta à lesão vascular e do miocárdio, e durante a metástase do câncer. As metaloproteinases da matriz e seus inibidores têm vários alvos na matriz extracelular vascular e têm sido associadas com a estabilidade da placa aterosclerótica, formação de aneurisma e outras doenças cardiovasculares. Dollery et al., *Cardiovascular research* 2006;69: 625-35. Matriz de metaloproteinase (MMP)-7 e MMP12 estão representadas no nosso modelo preditivo, enquanto o TIMP1 também tem associação significativa com risco cardiovascular (*Ver* Tabela 14). A trombospondina-2 (THBS2) medeia interações vasculares e célula-célula cardíaca e célula-matriz e tem sido implicada na regulação da angiogênese, trombose, e inflamação. Concentração aumentada de soro de A trombospondina-2 está associada com o risco de mortalidade cardíaca em homens mais velhos. Golledge et al., *The American journal of cardiology* 2013;111: 1800-4. CDO relacionado com oncogene de adesão celular (CDON) é um membro da proteína da superfície celular da superfamília de Ig/fibronectina envolvida na miogênese muscular e adesão celular. Tenzen et al., *Developmental cell* 2006;10: 647-56. O seu papel na interação célula-célula foi observado em invasividade tumoral, mas

pouco se sabe sobre a sua relação com o sistema cardiovascular.

Inflamação

[000285] Representando os papéis complexos de inflamação e imunidade na doença cardiovascular, nosso modelo incorpora a quimiocina inflamatória ligando de quimiocina (motivo C-C) 18, anteriormente conhecido como quimiocina pulmonar regulada por ativação CCL18/PARC, que é uma quimiocina elaborada por monócitos/ macrófagos que parece estar envolvida no recrutamento de células T. Chenivesse et al., *J Immunol* 2012; 189: 128-37. Os níveis plasmáticos de CCL18/PARC são elevados durante os episódios de angina instável e também foram encontrados para predizer eventos cardiovasculares em pacientes com angina estável. De Sutter et al., *Journal of molecular and cellular cardiology* 2010;49:894-6. A imunoglobulina de células T e proteína contendo domínio de mucina 3 (TIM-3) está envolvido na ativação de macrófagos e outras atividades do sistema imunitário. Anderson, *Expert opinion on therapeutic targets* 2007;11: 1005-9

[000286] Complemento C7 (C7) é um dos 5 componentes que formam o complexo terminal do complemento bioativo (TCC). TCC depositada nas células endoteliais resulta na proliferação de células, a liberação de fatores de crescimento e citocinas inflamatórias e aumento da expressão do fator de tecido. TCC também estimula a proliferação de células musculares lisas em placas ateroscleróticas. Speidl et al., *JTH* 2011; 9: 428-40. Em pacientes com TCC solúvel em soro com falha cardíaca elevada sintomática prevê resultado adverso (morte, transplante cardíaco de urgência, ou hospitalização com agravamento da insuficiência cardíaca). Clark et al., *Am Heart J.* 2001; 141:684-90. Complemento C9, um outro membro da TCC, é também elevado no nosso estudo (Tabela 14).

Proteases

[000287] Complexo de antiquimotripsina α1 (SERPINA3) representa a

forma ligada do inibidor de protease antiquimotripsina α1 que tem vários substratos biológicos. Ele pode modular vários processos de doenças agudas e crônicas, incluindo pressão arterial. Tang et al., *Clin Exp Hypertens* 2008; 30: 648-61. Antiplasmina α2 (SERPINF2) é um inibidor de serina-protease (serpinas), que inativa a plasmina e, assim, reduz a fibrinólise. Matsuno et al., *Journal of Thrombosis and Haemostasis: JTH* 2003; 1: 1734-9; Mutch et al., *JTH* 2007; 5: 812-7. A catepsina H (CTSH) é uma proteinase cisteína lisossômica importante na degradação de proteínas lisossômicas[39]. Cheng et al., *Circulation* 2012;125: 1551-62. No entanto, sua relação com a doença CV até o presente estudo tem sido incerta. Lutgens et al., *FASEB J.* 2007; 21:3029-41.

Marcador de necrose miocárdica

[000288] Ao contrário de muitas das proteínas acima mencionadas que são potencialmente envolvidas nas vias causais de doenças cardiovasculares, troponina I é um marcador bem estabelecido de necrose de cardiomiócitos e de risco cardiovascular.

APLICAÇÃO DO MODELO DE RISCO CVD9

[000289] O risco CVD9 foi calculado para cada indivíduo, divididos em quartis e as curvas de sobrevivência resultantes de 5 anos livre de eventos são mostradas na Figura 11. As razões de perigo Q4/Q1 para CVD9 são 8,2 no conjunto de descoberta e 6,0 no conjunto de validação; para o grau de risco de Framingham secundário (recalibrados para uso nestas populações) o grau de perigo Q1/ Q4 é de 2,8 na coorte de descoberta e de 2,3 para a coorte de validação.

[000290] Também avaliamos o desempenho comparativo da CVD9 vs. modelos de Framingham usando o índice de reclassificação de rede e estatística C em 1 ano (um ponto de tempo recomendado por um National Heart, Lung and Blood Institute Working Group; *Ver* Eagle et al., *Circulation*. 2010; 121: 1447-1454) e, no horizonte de tempo máximo validado de 4 anos para o modelo de Framingham. *Ver* D'Agostino et al.,

Am Heart J. 2000; 139:272-281. Como mostrado na Tabela 13, o modelo de predição de risco CVD9 provê melhorias substanciais na discriminação, evidenciadas pelo aumento da estatística C de 0,14/0,09 e um NRI livre de categoria de 0,57/0,54 em coortes de descoberta/ validação, respectivamente. Há um bom acordo para o modelo CVD9 entre as taxas de eventos observadas e previstas (calibração) na coorte de validação. A adição de parâmetros clínicos e demográficos comumente disponíveis (idade, sexo, diabetes e taxa de filtração glomerular) não fez nenhuma melhoria significativa para o modelo CVD9 (Tabela 13). Os dados comparativos de desempenho para todos os modelos é mostrado na Tabela 13. Na Tabela 2, NRI (> 0) = índice de reclassificação de rede livre de categoria, eGFR = taxa de filtração glomerular estimada.

TABELA 13. Desempenho comparativo de modelo CVD9 e modelo Framingham

	Framingham secundário	**Apenas Proteína (CVD9)**	**CVD9+ Idade+ Sexo**	**CVD9+ Idade+ Sexo+ Diabetes+ eGFR**
Todas as métricas mostradas no formato: descoberta/validação				
1 ano Estatística C	0,62/0,62	0,76/0,71	0,76/0,69	0,77/0,67
4 anos Estatística C	0,62/0,61	0,74/0,70	0,75/0,68	0,75/0,65
4 anos NRI(>0) VersusFramingham		0,57/0,52	0,64/0,41	0,67/0,35
Evento NRI VersusFramingham		0,16/0,18	0,19/0,16	0,20/0,11
Sem Evento NRI VersusFramingham		0,41/0,34	0,45/0,26	0,47/0,24

Razão de perigo Q4/Q1	2,8/2,3	8,2/6,0	10,0/4,9	10,3/4,6

EXEMPLO 4: DISCUSSÃO

[000291] Neste estudo, procurou-se melhorar a predição de desfechos cardiovasculares, particularmente a curto prazo, usando descoberta de biomarcadores na maior análise proteômica realizada até à data. Nós usamos a tecnologia de aptâmero modificado para analisar 1130 proteínas plasmáticas na coorte de descoberta de 938 pacientes com doença coronariana estável e validou os resultados em uma coorte independente de 971 pacientes. Na coorte de descoberta, encontramos 117 prognósticos de proteínas do ponto final cardiovascular compósito com taxas de risco superiores a 25% da unidade (Tabela 14). Destes, nós construímos um modelo multivariável constituído por 9 proteínas (CVD9) cuja performance foi superior à de fatores de risco tradicionais ou biomarcadores de sangue descritos na literatura (ver, por exemplo, Eagle et al., *Circulation*. 2010; 121:1447-1454; D'Agostino et al., *Am Heart J*. 2000;139:272-281; and Pearson et al., *Circulation*. 2003; 107:499-511), mostrando as vantagens potenciais da proteômica de base ampla em comparação com abordagens baseadas em candidatos. As proteínas de biomarcadores individuais e o modelo CVD9 replicaram a coorte de validação, apesar de inferior qualidade da amostra de sangue consistente com a prática clínica típica.

[000292] A aplicação adequada de estratégias preventivas e terapêuticas depende do sistema de classificação de risco que permite que os profissionais de saúde orientem os tratamentos mais intensivos para os indivíduos de maior risco. *Ver* Eagle et al., *Circulation*. 2010; 121:1447-1454. Abordagens comumente utilizadas dependem de avaliações de risco com base em fatores de risco tradicionais e têm limitações. Muitos desses fatores de risco são condições crônicas ou mesmo fixas, não modificáveis, tais como sexo, raça, idade avançada ou histórico familiar. Não surpreendentemente, eles são muito mais

adequados para predizer um risco a longo prazo (10 anos) ou risco de vida do que o risco de curto prazo. Fatores de risco tradicionais preveem eventos secundários particularmente mal em pacientes com doença coronariana prevalente. A identificação de pacientes com alto risco de curto prazo de eventos cardiovasculares representa uma importante necessidade não atendida, de modo que identificaria indivíduos em necessidade mais urgente de prevenção cardiovascular, intervenção e cumprimento com os tratamentos prescritos.

[000293] Várias tecnologias "genômicas" têm sido propostas para complementar os fatores de risco tradicionais na avaliação do risco cardiovascular. *Ver,* por exemplo, McShane et al., *Nature*. 2013;502: 317-320. Entre eles, os graus de risco genômicos foram investigados mais extensivamente. Abordagens genômicas com base em polimorfismos de nucleotídeo único comuns não conseguiram melhorar a discriminação de risco ou reclassificação sobre fatores de risco tradicionais, julgadas pelas mesmas métricas que foram favoravelmente impactadas pelo grau proteômico CVD9 no presente estudo (estatística c e rede de reclassificação). *Ver, por exemplo*, Paynter et al., *JAMA*. 2010; 303:631-637; Ripatti et al., *Lancet*. 2010; 376:1393-1. 400. Mesmo se abordagens genômicas forem finalmente bem sucedidas, serão na predição de risco de longo prazo, em vez de risco a curto prazo, de modo que fatores de risco genético não mudam ao longo do tempo e exercem o seu efeito através de exposições ao longo da vida. Comparada à genômica, proteômica oferece várias vantagens potenciais. Proteômica integra influências ambientais e genéticas, os níveis de proteínas podem mudar ao longo do tempo, refletindo os benefícios ou malefícios dos tratamentos ou mudanças de estilo de vida e as proteínas estão muitas vezes nas vias causais de doenças e, portanto, alvos potenciais de terapias. *Ver, por exemplo*, Nissen et al., *N Engl J Med*. 2005; 352:29-38; Ridker et al., *Lancet*.

2009;373:1175-1182; e Stein et al., *N Engl J Med*. 2012; 366:1108-1118.

[000294] Utilizou-se uma nova plataforma de proteômica que consiste em aptâmeros modificados para medir 1130 proteínas em um pequeno volume (<100 µl) de plasma. Descobrimos 117 biomarcadores de proteína candidata de risco cardiovascular (Tabela 14). Notavelmente, muitas destas proteínas não foram anteriormente relatadas como biomarcadores de risco cardiovascular. A partir dessas proteínas, foi construído um modelo totalmente paramétrico parcimonioso utilizando uma estatística (LASSO em conjunto com eliminação reversa) ao invés de abordagem biológica. Neste processo, algumas proteínas com taxas de risco razoável são deixadas de fora (CRP, por exemplo), de modo que transmitem a informação que é capturada por proteínas já no modelo, enquanto outras proteínas com taxas de risco univariadas inferiores são retidas devido à informação única que elas proporcionam. As funções biológicas das proteínas de LASSO selecionadas são aqui discutidas.

[000295] O grau de risco de proteína CVD9 desempenhou melhor do que o grau de risco de Framingham secundário (D'Agostino et al., *Am Heart J*. 2000; 139: 272-281), que depende de fatores de risco tradicionais. Incluindo variáveis clínicas que foram significantemente diferentes na população de eventos tais como, idade, sexo, diabetes ou taxa de filtração glomerular (EGFR) em modelos secundárias proveu apenas melhorias modestas em CVD9 nas taxas de risco inter-quartil e índices de reclassificação de rede na coorte de descoberta (Tabela 13). É possível que CVD9 já encapsule a biologia subjacente ao risco associado com os fatores de risco tradicionais, embora não estejamos propondo que a avaliação de proteínas com CVD9 ou modelos similares substitua-os, pois este último ainda pode ser um melhor indicador de risco a longo prazo e um alvo específico de tratamentos. Ainda assim,

níveis de proteínas CVD9 proporcionam uma avaliação individualizada de risco cardiovascular no curto prazo superior do que Framingham, particularmente para pacientes nos extremos de risco (Figuras 6 e 7), presumivelmente porque indicam se as vias associadas a complicações cardiovasculares foram ativadas e se danos em órgãos de extremidade ocorreu (por exemplo, troponina; ver Beatty et al, *JAMA Intern Med*. 2013; 173: 763-769).

[000296] Nosso estudo é a primeira análise proteômica em grande escala de risco cardiovascular, usando uma plataforma proteômica de alto rendimento em grande escala. Esta abordagem resultou na descoberta de numerosos biomarcadores individuais de proteínas novos e levou à construção um modelo robusto de predição de risco multi-variável com um desempenho superior na predição de risco de eventos cardiovasculares secundários no curto prazo. O estudo foi realizado em dois grupos grandes e bem caracterizados com excelente decretação de eventos resultado, em dois continentes. O US National Cancer Institute, em colaboração com um painel de cientistas especialistas, desenvolveu uma lista de verificação de critérios que podem ser usados para determinar a disponibilidade de tecnologias à base de omics para orientar a assistência ao paciente em ensaios clínicos. Qualidade de espécime foi anotada como uma razão importante pela qual achados de omics reportados de um laboratório podem não replicar em outros. Assim, realizamos nossa análise proteômica de uma variedade de qualidades de espécime, representativo de instituições acadêmicas (coração e alma) e padrões clínicos de prática (HUNT3) e nossos resultados são robustos em toda esta variedade de qualidade de espécime.

[000297] Concentramos propositalmente nossa investigação inicial em uma população de indivíduos de alto risco com doença coronária estabelecida (CHD). Há necessidade adicional de predição precisa de

risco cardiovascular na população em menor risco ou em indivíduos com risco ainda mais elevados com CHD. Esses estudos estão em andamento com outros grupos. Outra limitação é que há muito mais proteínas no sangue do que os 1130 que foram quantificados. Ainda não se sabe se a sua apreciação melhora a avaliação de risco cardiovascular, já que podem estar nas mesmas vias e, assim, redundante com as proteínas que já foram avaliadas. Estudos que avaliam um número ainda maior de proteínas do que o relatado no presente estudo estão em andamento também.

[000298] Em resumo, temos realizado com sucesso o maior estudo proteômica do risco cardiovascular até hoje, com mais de 2 milhões de medições individuais de proteínas, identificado inúmeros biomarcadores de risco novos e demonstrou um modelo de predição de risco com desempenho superior e robusto.

TABELA 14. Tabela de proteínas individuais associadas com risco cardiovascular. Biomarcadores no painel CVD9 estão em negrito. Se a relação de risco (HR) é maior do que 1, níveis aumentados do biomarcador são associados com risco aumentado; Se a HR for menor que 1, níveis diminuídos do biomarcador são associados com risco aumentado.

Alvo	UniProt ID	Q4/Q1 HR	HR por desvio padrão	Valor P para HR contínua
Angiopoietina-2	O15123	3,13	1,67	<1e-16
MMP12	P39900	3,52	1,65	<1e-16
Proteína de membrana de células T TIM-3	Q8TDQ0	2,98	1,61	<1e-16
Proteína 2 com fator de ligação de crescimento tipo insulina	P18065	2,93	1,58	<1e-16
TNF-R II	P20333	3,33	1,56	<1e-16
Tipo Folistatina 3	O95633	3,52	1,56	<1e-16
Quimiocina 1 de	Q16627	2,63	1,55	<1e-16

Alvo	UniProt ID	Q4/Q1 HR	HR por desvio padrão	Valor P para HR contínua
hemofiltrado CC				
Microglobulina β 2	P61769	3,58	1,54	<1e-16
Trombospondina-2	P35442	3,19	1,54	<1e-16
MMP-7	P09237	3,54	1,53	<1e-16
Endostatina	P39060	2,45	1,52	<1e-16
Catepsina H	P09668	4,06	1,52	<1e-16
Receptor B2 de Eph	P29323	2,20	1,50	<1e-16
Proteína de ligação da interleucina-18	O95998	2,55	1,49	<1e-16
Tipo cordina 1	Q9BU40	2,87	1,49	<1e-16
Cistatina C	P01034	3,54	1,49	<1e-16
Complemento C9	P02748	2,81	1,48	8. 80E-14
CCL18/PARC	P55774	2,55	1,47	1,11E-16
Complemento C7	P10643	3,07	1,47	<1e-16
receptor de fator de necrose de tumor RELT	Q969Z4	3,23	1,46	<1e-16
Jagged-	P78504	2,17	1,45	3. 66E-15
Receptor de netrina UNC5H3	O95185	3,15	1,44	<1e-16
Efrina-A4	P52798	3,37	1,44	<1e-16
Protease de serina 4 específica do cérebro	Q9GZN4	2,54	1,44	<1e-16
Supressor de neuroblastoma de tumorigenicidade 1|DAN	P41271	3,24	1,43	<1e-16
Receptor 2 do tipo A de efrina	P29317	2,86	1,43	<1e-16
Espondina-1	Q9HCB6	2,99	1,42	<1e-16
Periostina	Q15063	2,11	1,40	6,08E-12
Fator A de Crescimento endotelial Vascular	P15692	2,43	1,40	1,02E-12
Membro 1 da classe F de receptor de removedor	Q14162	2,90	1,39	<1e-16
Complexo de antiquimotripsina α1	P07288, P01011	2,53	1,39	1,95E-14

Alvo	UniProt ID	Q4/Q1 HR	HR por desvio padrão	Valor P para HR contínua
Proteína Crk-I de Adaptor	P46108	2,71	1,39	<1e-16
Efrina-A5	P52803	2,35	1,37	2,89E-15
Molécula de adesão celular endotelial seletiva	Q96AP7	2,13	1,37	1,02E-14
Glutationa S-Transferase Pi1	P09211	2,37	1,37	1,27E-12
Receptor de morte 6|DR6	O75509	1,98	1,36	1,35E-10
Proteína de terminação de macrófago	P40121	3,01	1,36	<1e-16
Proteína 80 contendo domínio bobinado	Q76M96	2,07	1,35	1,36E-10
Gene-3 de ativação de linfócito	P18627	2,06	1,35	2,47E-10
Ck-β-8-1|proteína 3 inflamatória de macrófagos variante de processamento	P55773	1,84	1,35	1,26E-08
Elafina	P19957	2,18	1,35	9,09E-12
TIMP-1	P01033	3,94	1,34	<1e-16
HSP 70	P08107	2,08	1,34	1,24E-10
Estaniocalcina-1	P52823	2,29	1,34	2,62E-11
Receptor III-B de região G Fc de imunoglobulina	O75015	1,91	1,34	2,53E-10
Inibidor de protease de leucócito secretório	P03973	2,37	1,34	7,91E-11
TRAIL R4	Q9UBN6	2,02	1,33	1,29E-09
MMP-3	P08254	2,40	1,33	1,36E-09
hormônio pancreáticoPH	P01298	2,19	1,33	1,18E-10
Fator neurotrófico de dopamina conservada	Q49AH0	1,82	1,32	5,46E-10
cistatina D	P28325	2,11	1,32	9,36E-10
GPVI	Q9HCN6	1,79	1,32	3,09E-09
Catepsina Z/X/P	Q9UBR2	2,07	1,32	9,36E-10
Protein 1 tipo delta	O00548	3,28	1,31	<1e-16
MPIF-1	P55773	1,88	1,31	2,73E-09
Calicreína 11	Q9UBX7	1,86	1,31	1,22E-11

Alvo	UniProt ID	Q4/Q1 HR	HR por desvio padrão	Valor P para HR contínua
A interleucina-1 do receptor de tipo 1|ST2	Q01638	2,02	1,30	1,91E-09
Molécula de ativação de sinalização linfocítica 5	Q9UIB8	2,49	1,30	9,99E-16
TFF3	Q07654	3,62	1,30	3,50E-13
subunidade β PAFAH	P68402	2,41	1,29	7,94E-14
Proteína de ligação de fator de crescimento tipo insulina-1	P08833	1,87	1,29	1,11E-08
CD48	P09326	2,07	1,29	4,61E-10
Renina	P00797	1,70	1,29	1,67E-08
Neuroligin 4, ligada ao X	Q8N0W4	2,24	1,29	1,91E-12
B quimioatrativo linfócitos	O43927	2,34	1,29	1,24E-11
Proteína A de plasma associado com gravidez	Q13219	1,69	1,29	6,82E-08
uPAR	Q03405	3,00	1,28	3,77E-13
Resistina	Q9HD89	1,77	1,28	2,26E-08
fucosiltransferase 5	Q11128	1,55	1,28	8,30E-07
fator 1 derivado de célula estromal	P48061	1,79	1,28	2,17E-07
Nidogênio	P14543	1,67	1,28	1,11E-07
ligando 1A tipo TNF	O95150	2,66	1,28	6,61E-13
Alta temperatura requisito de serina peptidase A2	O43464	2,01	1,28	6,35E-09
Proteína 7 de ligação de fato de crescimento tipo insulina	Q16270	1,78	1,28	1,55E-07
Receptor 1 de interleucina-1	P14778	1,83	1,27	1,12E-06
Fosfolipase secretora não pancreática A2	P14555	1,98	1,27	5,23E-09
Proteína 4 relacionada a angiopoietina	Q9BY76	2,84	1,27	4,93E-11
Proteína de ligação de ácido graxo, tipo coração	P05413	2,58	1,27	1,29E-10

Alvo	UniProt ID	Q4/Q1 HR	HR por desvio padrão	Valor P para HR contínua
Proteína de ligação de lipopolissacarídeo	P18428	1,99	1,27	1,80E-06
Receptor 1 fator de crescimento tipo insulina O IGF-I sR	P08069	1,75	1,27	9,87E-07
Tenascin-C	P24821	1,89	1,27	3,14E-07
Receptor ectodisplasina-A2 ligado ao X|XEDAR	Q9HAV5	3,64	1,27	<1e-16
Troponina I, cardíaca	P19429	2,94	1,27	1,01E-12
Sialoproteína óssea 2	P21815	1,79	1,27	7,52E-08
Proteína 6 de ligação de fator de crescimento tipo insulina	P24592	2,29	1,26	3,04E-11
Matrilin-2	O00339	1,89	1,26	4,44E-07
Antígeno de superfície de linfócitos T Ly-9	Q9HBG7	1,56	1,26	1,79E-06
Lailina	Q6UX15	2,50	1,26	1,42E-09
dCTP pirofosfatásica 1	Q9H773	1,59	1,26	7,35E-06
Dímero de cadeia γ de fibrinogênio	P02679	1,98	1,25	3,29E-06
B6 receptor EPH	O15197	1,79	1,25	3,26E-09
Anidrase Carbônica III	P07451	1,88	1,25	5,33E-07
Receptor 1 de lipoproteína de densidade baixa oxidizada	P78380	1,89	1,25	2,37E-07
Cistatina SA	P09228	1,59	1,25	2,82E-06
Fator 7 de crescimento de fibroblastos	P21781	2,08	1,25	2,90E-12
Neurexophilin-1	P58417	0,58	0,75	1,11E-08
Soggy-1	Q9UK85	0,54	0,75	2,57E-08
15-Hidroxiprostaglandina Desidrogenases	P15428	0,51	0,74	1,93E-08
Proteína C	P04070	0,38	0,74	1,29E-13
Proteína de activação de fibroblastos α	Q12884	0,49	0,74	1,47E-06

Alvo	UniProt ID	Q4/Q1 HR	HR por desvio padrão	Valor P para HR contínua
Ajustar	O43508	0,44	0,74	4,06E-06
Receptor de Fatores de Crescimento do Endotélio Vascular 2	P35968	0,47	0,74	2,21E-10
Complementar proteína de ligação C1q	Q07021	0,49	0,74	9,01E-05
Angiostatinas	P00747	0,48	0,73	5,89E-11
ERBB3	P21860	0,39	0,72	2,03E-10
GDF11	O95390	0,41	0,72	8,75E-09
BMP-1	P13497	0,39	0,71	2,54E-13
CDO devadesão celular regulamentada por oncogene	Q4KMG0	0,39	0,70	9,06E-14
Creatina Quinase Forma MM	P06732	0,45	0,70	3,60E-11
CRNAosina dipeptidase 1	Q96KN2	0,36	0,69	<1e-16
cAMP e cGMP fosfodiesterase 11A	Q9HCR9	0,39	0,69	5,18E-12
CK-MB	P12277 P06732	0,41	0,69	5,06E-13
Caderinas - 3	P22223	0,41	0,67	1,11E-16
Ret receptor de quinase de proteína-tirosina de proto-oncogene	P07949	0,42	0,66	4,89E-13
Antiplasmina α2	P08697	0,37	0,64	<1e-16
Receptores da Somatotrofina	P10912	0,29	0,63	<1e-16
Receptor de EGF	P00533	0,29	0,60	<1e-16

EXEMPLO 5: MODELO GDF11 E FSTL3

[000299] Três modelos de riscos proporcionais de Cox foram gerados e comparados.

- GDF11Modelo de proteína univariada com proteína GDF11
- FSTL3: Modelo de proteína univariada com proteína relacionada com Follistatin 3 (FSTL3)

• GDF11.FSTL3: Modelo de proteína de combinações com GDF11 e FSTL3

[000300] Para a comparação entre os modelos, análise de variância, razão de perigo Q4/Q1 de previsores lineares, NRI de probabilidade de risco de 4 anos, e AUC integrada dentro de 4 anos foram calculados. O modelo GDF11.FSTL3 foi o melhor modelo com todos os métodos de avaliação.

[000301] As amostras atípicas foram excluídas da análise. Todos os modelos foram calculados com transformada de log com base 10 e padronizada.

[000302] Antes de combinar duas proteínas no modelo, a correlação de Spearman foi aplicada para verificar a relação entre GDF11 e FSTL3. A correlação entre as duas proteínas é significativa ($p = 3.123 - 12$), mas a correlação de Spearman não é forte ($rho = -0.2251$). A Tabela 15 mostra o resultado do teste de correlação de R.

TABELA 15: Teste de correlação de Spearman entre GDF11 e FSTL3.

```
    Spearman's rank correlation rho

data:  gdf11 and fstl3
S = 1.68e+08, p-value = 3.123e-12
alternative hypothesis: true rho is not equal to 0
sample estimates:
    rho
-0.2251
```

[000303] A correlação entre GDF11 e FSTL3 é mostrada na Figura 15. A RFU foi convertida para espaço de log com base 10. A figura da esquerda mostra a correlação de todas as amostras e a figura da direita mostra a correlação sem uma amostra omitida, que tinha um valor de GDF11 alto. Círculos pretos e vermelhos significam amostras sem evento evento e amostras de eventos, respectivamente.

[000304] Três modelos de perigo proporcionais de Cox foram gerados: GDF11, FSTL3, e GDF11.FSTL3. O GDF11, FSTL3 modelos

são modelos de Cox com uma única proteína e o GDF11.FSTL3 é um modelo combinado com duas proteínas. Antes da montagem do modelo, os valores extremos foram excluídos e os valores de RFU foram transformadas de log e padronizados. Tabelas 16 e 17 mostram a comparação entre os modelos individuais e modelo combinado com a tabela de desvio ANOVA. O modelo combinado é significativamente melhorado a partir dos modelos de proteínas individuais. Os valores p para GDF11 vs GDF11.FSTL3 e FSTL3 vs GDF11.FSTL3 são 2e-16 e 3,5e-06, respectivamente.

TABELA 16: O teste ANOVA entre os modelos GDF11 e GDF11.FSTL3

```
Analysis of Deviance Table
 Cox model: response is  s
```

```
 Model 1: ~ GDF11.2765.4.3
 Model 2: ~ GDF11.2765.4.3 + FSTL3.3438.10.2
  loglik Chisq Df P(>|Chi|)
1  -2936
2  -2896    79  1    <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

TABELA 17: O teste ANOVA entre os modelos GDF11 e GDF11.FSTL3

```
Analysis of Deviance Table
 Cox model: response is  s
 Model 1: ~ FSTL3.3438.10.2
 Model 2: ~ GDF11.2765.4.3 + FSTL3.3438.10.2
  loglik Chisq Df P(>|Chi|)
1  -2907
2  -2896  21.5  1   3.5e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

[000305] Tabela 18 mostra a taxa de risco Q4/Q1 de previsores lineares para cada modelo. O modelo combinado mostra uma razão de risco mais elevada do que os modelos simples. Os quartis são definidos pelos previsores lineares de cada modelo de Cox.

TABELA 18: A razão de perigo Q4/Q1

	Q4/Q1	HR inferior. 95	superior. 95
GDF11	**2,475**	**1,894**	**3,233**
FSTL3	**3,637**	**2,738**	**4,830**
GDF11.FSTL3	**4,080**	**3,068**	**5,426**

[000306] A Figura 16 mostra as curvas de sobrevivência de cada quartil de todos os modelos. O 1º ao 4º quartis são descritos com preto, vermelho, verde e azul (de cima para baixo, as linhas são preta, vermelha, verde, em seguida, azul). O sombreamento representa os intervalos de confiança de 95%. A distância entre o 1º e 4º quartis do modelo GDF11.FSTL3 é maior do que os modelos de proteínas individuais. Além disso, a distância entre o 2º e 3º quartis do modelo GDF11.FSTL3 também é maior do que os modelos de proteínas individuais.

[000307] Uma comparação entre as curvas de sobrevivência para o modelo GDF11 e o modelo GDF11.FSTL3 é mostrada na Figura 3. A figura da esquerda mostra a comparação dos grupos de baixo risco e a figura da direita mostra a comparação dos grupos de alto risco. O preto e vermelho representam o modelo GDF11 e modelo GDF11.FSTL3, respectivamente. Neste modelo, o grupo de baixo risco identificado pelo modelo GDF11.FSTL3 tem menos amostras de eventos do que o grupo de baixo risco identificado pelo modelo GDF11, e o grupo de alto risco identificado pelo modelo GDF11.FSTL3 teve mais amostras de evento do que o grupo de alto risco identificado pelo modelo GDF11. O modelo GDF11.FSTL3 foi, por conseguinte, mais preciso para os dois grupos de amostras. A inclusão de FSTL3 melhorou o modelo para ambos os grupos de alto e baixo risco.

[000308] O NRI foi calculado entre os modelos únicos de proteínas GDF11 e FSTL3 e GDF11.FSTL3. A probabilidade foi calculada dentro de 4 anos e perigo de linha de base foi estimado com Kaplan-Meier.

Inferior e Superior na tabela são o intervalo de confiança de 95% do NRI, que é estimado com bootstrapping. GDF11 melhora NRI de amostras de eventos e FSTL3 melhora NRI de amostras não evento.

TABELA 19: NIR entre o modelo de proteína única e modelo GDF11.FSTL3.

GDF11 vs GDF11.FSTL3				FSTL3 vs GDF11.FSTL3			
	Estimativa	Inferior	Superior		Estimativa	Inferior	Superior
NRI	0,4326	0,28552	0,5772	NRI	0,31532	0,17287	0,4639
NRI+	0,1446	0,01143	0,2655	NRI=	0,22840	0,10174	0,3612
NRI-	0,2880	0,21878	0,3563	NRI-	0,08692	0,01022	0,1653
Pr (Superior| Caso)	0,5721	0,50573	0,6329	Pr (Superior|Caso)	0,61422	0,55081	0,6807
Pr (Baixo|Caso)	0,4276	0,36706	0,4942	Pr (Baixo|Caso)	0,38582	0,31923	0,4491
Pr (Baixo|Ctrl)	0,6441	0,60935	0,6780	Pr (Baixo|Ctrl)	0,54345	0,50504	0,5828
Pr (Superior| Ctrl)	0,3560	0,32190	0,3906	Pr (Superior|Ctrl)	0,45653	0,41738	0,4949

[000309] A Figura 18 mostra uma comparação da probabilidade de 4 anos entre GDF11 e GDF11.FSTL3 (esquerda) e entre FSTL3 e GDF11.FSTL3 (à direita). Os pontos pretos, vermelhos e verdes descrevem controle, caso, e as amostras censuradas no ano 4.

[000310] O AUC integrado (Cindex) dentro de 4 anos é apresentado na Tabela 20. Intervalos de confiança de 95% foram calculados com bootstrapping, semelhante ao NRI.

TABELA 20: O AUC integrado (Cindex) no prazo de 4 anos.

	Cindex	inferior. 95	superior. 95
GDF11	0,5882	0,5555	0,6096
FSTL3	0,6038	0,5786	0,6321
GDF11.FSTL3	0,6286	0,6050	0,6566

[000311] As curvas ROC no ano 4 são mostrados na Figura 19. Os números na legenda referem-se a AUC no ano 4 (AUC não integrado).

[000312] Três modelos de risco proporcional de Cox foram comparados com várias estatísticas de avaliação, para os modelos de marcadores individuais e o modelo de marcador de combinação. O modelo de combinações, que inclui GDF11 e FSTL3, teve o melhor desempenho de acordo com todos os valores de avaliação.

[000313] Os seguintes quadros mostram os três modelos utilizados neste exemplo.

```
Call:
coxph(formula = f, data = x, x = T)

  n= 937, number of events= 465

                  coef exp(coef) se(coef)     z Pr(>|z|)
GDF11.2765.4.3 -0.3452    0.7081   0.0579 -5.96  2.5e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

               exp(coef) exp(-coef) lower .95 upper .95
GDF11.2765.4.3     0.708       1.41     0.632     0.793

Concordance= 0.604  (se = 0.014 )
Rsquare= 0.04   (max possible= 0.998 )
Likelihood ratio test= 38.1  on 1 df,   p=6.66e-10
Wald test            = 35.6  on 1 df,   p=2.46e-09
Score (logrank) test = 28.4  on 1 df,   p=9.95e-08
```

```
Call:
coxph(formula = f, data = x, x = T)

  n= 937, number of events= 465

                 coef exp(coef) se(coef)    z Pr(>|z|)
FSTL3.3438.10.2 0.436     1.547    0.042 10.4   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                exp(coef) exp(-coef) lower .95 upper .95
FSTL3.3438.10.2      1.55      0.646      1.42      1.68

Concordance= 0.634  (se = 0.014 )
Rsquare= 0.097   (max possible= 0.998 )
Likelihood ratio test= 95.6  on 1 df,   p=0
Wald test            = 108  on 1 df,   p=0
Score (logrank) test = 105  on 1 df,   p=0
```

```
Call:
```

```
coxph(formula = f, data = x, x = T)

  n= 937, number of events= 465

                    coef exp(coef) se(coef)     z Pr(>|z|)
GDF11.2765.4.3  -0.2605    0.7706   0.0577 -4.52  6.2e-06 ***
FSTL3.3438.10.2  0.4064    1.5014   0.0434  9.36  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                exp(coef) exp(-coef) lower .95 upper .95
GDF11.2765.4.3      0.771      1.298     0.688     0.863
FSTL3.3438.10.2     1.501      0.666     1.379     1.635

Concordance= 0.652  (se = 0.014 )
Rsquare= 0.117   (max possible= 0.998 )
Likelihood ratio test= 117  on 2 df,   p=0
Wald test            = 124  on 2 df,   p=0
Score (logrank) test = 119  on 2 df,   p=0
```

EXEMPLO 6: MODELOS GDF11 E FSTL3 PARA GRUPOS DE EVENTO ESPECÍFICO

[000314] Os modelos de Cox de GDF11, FSTL3 e GDF11.FSTL3 foram equipados com CHF e amostras da Morte, e as amostras de eventos trombóticos separadamente, para determinar como o modelo executa cada tipo de evento CV. Razão de perido Q4/Q1 de preditores lineares do modelo, AUC integrado (Cindex) dentro de 4 anos, e NRI de probabilidade de risco de 4 anos foram calculados. Para o cálculo da probabilidade do risco, o estimador de Kaplan-Meier foi usado como linha de base de perigo.

[000315] Encaixou-se o modelo de riscos proporcionais de Cox com GDF11, FSTL3 e GDF11.FSTL3 com grupos de eventos específicos: CHF e grupo de Morte, e o grupo de evento trombótico. CHF e grupo de Morte incluem CHF (125), CVDDeath (55), morte (135) e NONE (472). Grupo de eventos trombóticos inclui MI (104), acidente vascular cerebral (30), a TIA (16) e NONE (472). NONE, que são amostras sem evento, foram usadas em ambos os grupos. Para a avaliação dos modelos, a taxa de risco Q4/Q1 de preditores lineares, AUC integrado dentro de 4 anos, e NRI de probabilidade de risco de quatro anos foram calculados. Probabilidade de risco foi calculada com perigo de linha de

base de Kaplan-Meier.

[000316] Tabela 21 mostra a taxa de risco Q4/Q1, razão de perigo inversa e e seus intervalos de confiança de 95%. Razão de perigo Q4/Q1 de GDF11 e FSTL3 não são significativamente diferentes entre DHF. DEATH e amostras de eventos trombóticos, mas a razão de perigo de GDF11.FSTL3 é melhor com grupos de eventos trombóticos que o grupo CHF. DEATH.

TABELA 21: Razão de perigo Q4/Q1 de cada modelo e de grupo.

$GDF11

	Q4/Q1 HR	Q1/Q4 HR CI	Lower CI	Upper
All	2.475	0.4040	1.894	3.233
CHF.DEATH	2.726	0.3668	1.964	3.784
Thrombotic.Event	2.743	0.3645	1.698	4.432

$FSTL3

	Q4/Q1 HR	Q1/Q4 HR CI	Lower CI	Upper
All	3.637	0.2750	2.738	4.830
CHF.DEATH	4.478	0.2233	3.125	6.416
Thrombotic.Event	4.605	0.2171	2.731	7.765

$GDF11.FSTL3

	Q4/Q1 HR	Q1/Q4 HR CI	Lower CI	Upper
All	4.080	0.2451	3.068	5.426
CHF.DEATH	4.394	0.2276	3.069	6.291
Thrombotic.Event	5.493	0.1821	3.185	9.473

[000317] A Figura 1 mostra as curvas de sobrevivência dos quartis de preditor linear de cada grupo de modelo GDF11.FSTL3. O 1º ao 4º quartis são descritos com preto (linha de topo), vermelho (segunda linha embaixo), verde (terceira linha embaixo) e azul (linha de fundo). O sombreamento mostra os intervalos de confiança de 95%. O 1º quartil do evento trombótico (baixo grupo de risco) mostra menos eventos. Isto sugere que o modelo pode ser bastante sensível para o evento trombótico.

[000318] AUC integrado (Cindex) dentro de 4 anos e os intervalos de confiança de 95% estão apresentados na Tabela 22. Com Cindex, não há diferenças significativas entre o grupo CHF. DEATH e grupo de eventos trombóticos, embora descobriu-se que a razão de perigo Q4/Q1 é diferente entre os grupos.

TABELA 22: AUC integrado (Cindex) dentro de 4 anos.

```
$GDF11
                  Cindex Cindex.CI.lower.95 Cindex.CI.upper.95
All               0.5882             0.5614             0.6165
CHF.DEATH         0.5892             0.5582             0.6220
Thrombotic.Event 0.6057              0.5641             0.6503

$FSTL3
                  Cindex Cindex.CI.lower.95 Cindex.CI.upper.95
All               0.6038             0.5808             0.6344
CHF.DEATH         0.6018             0.5754             0.6482
Thrombotic.Event 0.5994              0.5667             0.6600

$GDF11.FSTL3
```

```
                  Cindex Cindex.CI.lower.95 Cindex.CI.upper.95
All               0.6286             0.6047             0.6558
CHF.DEATH         0.6308             0.6020             0.6645
Thrombotic.Event 0.6292              0.5939             0.6777
```

[000319] Em conclusão, os modelos de Cox de GDF11, FSTL3 e GDF11.FSTL3 foram gerados com grupos de amostras específicos. O modelo GDF11.FSTL3 mostra o melhor resultado com razão de perigo Q4/Q1 com o grupo de evento trombótico. Com Cindex, todos os modelos apresentaram resultados semelhantes.

EXEMPLO 7: GDF11 E GASP1/MODELO GASP2

[000320] A combinação de GDF11 com outras duas proteínas, GASP1 (WFIKKN2, SwissProt Q8TEU8) e GASP2 (WFIKKN1, SwissProt Q96D09), também foi testada.

[000321] Os quatro modelos de Cox seguintes foram gerados: (1) GDF11, modelo de Cox com proteína GDF11; (2) modelo GDF11.WFIKKN1, Cox com GDF11 e GASP2; (3) modelo GDF11.WFIKKN2, Cox com GDF11 e GASP1; e (4) GDF11.WFIKKN1.WFIKKN2, modelo de Cox com GDF11, GASP1 e GASP2. Antes de criar os modelos, a medição de proteína foi normalizada para Gaussiana (0,1).

[000322] Razão de perigo Q4/Q1 de preditores lineares foi calculada para os modelos. O grupo Q1 assume-se como grupo de baixo risco e o grupo Q4 assume-se como grupo de alto risco. Descobriu-se que

adicionar GASP2 (WFIKKN1) não melhora o modelo GDF11, mas adicionar WFIKKN2 mostrou alguma melhora (2,432-2,719). Os valores de taxa de risco Q4/Q1 e curvas de sobrevivência dos quartis são mostrados na Tabela 23 e na Figura 21.

TABELA 23. Razão de perigo Q4/Q1 de cada modelo.

	Q4/Q1 HR	Q1/Q4 HR	CI Lower C	Inferior CI	Superior CI
GDF11	2.432	0.4111	1.864	3.175	
GDF11.WFIKKN1	2.392	0.4180	1.830	3.127	
GDF11.WFIKKN2	2.719	0.3678	2.071	3.569	
GDF11.WFIKKN1.WFIKKN2	2.758	0.3626	2.102	3.619	

[000323] Além disso, os modelos foram comparados com tabelas de desvio ANOVA. O resultado R de comparação entre GDF11 e os modelos combinados são mostrados abaixo. GDF11.WFIKKN2 e GDF11.WFIKKN1.WFIKKN2 foram significativos, quando comparados com o modelo GDF11 (p = 3. 1e-05, 0,00015, respectivamente). A adição de WFIKKN1 não mostrou significância (p = 0,38). Os valores de p são destacados a seguir.

• Comparação entre GDF11 e GDF11.WFIKKN1

```
## Analysis of Deviance Table
##  Cox model: response is  s
##  Model 1: ~ GDF11.2765.4.3
##  Model 2: ~ GDF11.2765.4.3 + WFIKKN1.3191.50.2
##    loglik Chisq Df P(>|Chi|)
## 1  -2938
## 2  -2937  0.77  1      0.38
```

• Comparação entre GDF11 e GDF11.WFIKKN2

```
## Analysis of Deviance Table
##  Cox model: response is  s
##  Model 1: ~ GDF11.2765.4.3
##  Model 2: ~ GDF11.2765.4.3 + WFIKKN2.3235.50.2
##    loglik Chisq Df P(>|Chi|)
## 1  -2938
## 2  -2929  17.4  1   3.1e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

• Comparação entre GDF11 e GDF11.WFIKKN1.WFIKKN2

```
## Analysis of Deviance Table
##  Cox model: response is  s
##  Model 1: ~ GDF11.2765.4.3
##  Model 2: ~ GDF11.2765.4.3 + WFIKKN1.3191.50.2 + WFIKKN2.3235.50.2
##    loglik Chisq Df P(>|Chi|)
## 1  -2938
## 2  -2929  17.6  2   0.00015 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

[000324] Para avaliar os modelos, cálculo de NRI foi também realizado. A probabilidade foi calculada dentro de 4 anos. A adição de GASP1 (WFIKKN2) melhorou o NRI (0,16), especialmente com amostras não-evento (0,12). A partir deste resultado, GASP1 podem ser capazes de melhorar a taxa negativa verdadeira. Em contraste, GASP2 não melhorou NRI mais do que 0,1. O resultado R de NRI é mostrado abaixo.

• NRI entre GDF11 e GDF11.WFIKKN1

```
## ----- GDF11 vs GDF11.WFIKKN1 -----
##                Estimate    Lower   Upper

## NRI             0.05855 -0.08109 0.20126
## NRI+            0.09245 -0.02881 0.20405
## NRI-           -0.03391 -0.11142 0.04671
## Pr(Up|Case)     0.54627  0.48586 0.60175
## Pr(Down|Case)   0.45382  0.39769 0.51467
## Pr(Down|Ctrl)   0.48303  0.44431 0.52336
## Pr(Up|Ctrl)     0.51694  0.47665 0.55573
```

• Comparação entre GDF11 e GDF11.WFIKKN2

```
## ----- GDF11 vs GDF11.WFIKKN2 -----
##                Estimate    Lower  Upper
## NRI             0.16315  0.02639 0.3063
## NRI+            0.04422 -0.07236 0.1727
## NRI-            0.11892  0.04351 0.1919
## Pr(Up|Case)     0.52206  0.46394 0.5861
## Pr(Down|Case)   0.47784  0.41338 0.5363
## Pr(Down|Ctrl)   0.55948  0.52176 0.5960
## Pr(Up|Ctrl)     0.44056  0.40405 0.4783
```

• NRI entre GDF11 e GDF11.WFIKKN1.WFIKKN2

```
## ----- GDF11 vs GDF11.WFIKKN1.WFIKKN2 -----
##                Estimate    Lower  Upper
## NRI             0.13460 -0.01276 0.2759
## NRI+            0.02732 -0.09310 0.1643
## NRI-            0.10728  0.02863 0.1758
## Pr(Up|Case)     0.51364  0.45354 0.5820
## Pr(Down|Case)   0.48632  0.41779 0.5467
## Pr(Down|Ctrl)   0.55365  0.51428 0.5879
## Pr(Up|Ctrl)     0.44637  0.41211 0.4857
```

[000325] A probabilidade de quatro anos entre modelos é mostrada na Figura 22.

[000326] Finalmente, o cálculo de AUC foi realizado para a avaliação entre os modelos. De acordo com resultados a seguir, nenhuma proteína melhorou o modelo GDF11. As curvas ROC para cada modelo são mostradas na Figura 23. As curvas ROC para cada modelo foram semelhantes.

- GDF11

```
## ----- -----
##   Cindex lower.95 upper.95
## 1  0.586   0.5579   0.6143
```

- GDF11.WFIKKN1

```
## ----- -----
##   Cindex lower.95 upper.95
## 1 0.5849   0.5572   0.6133
```

- GDF11.WFIKKN2

```
## ----- -----
##   Cindex lower.95 upper.95
## 1 0.5994   0.5717   0.6305
```

- GDF11.WFIKKN1.WFIKKN2

```
## ----- -----
##   Cindex lower.95 upper.95
## 1 0.5988   0.5712     0.63
```

[000327] Em resumo, GASP1 (WFKKN2) pode melhorar o modelo GDF11, mas a melhoria é pequena. GASP2 (WFKKN1) não melhorou o

modelo GDF11.

[000328] O modelo de Cox utilizado neste exemplo é mostrado abaixo.

• GDF11

```
## Call:
## coxph(formula = f, data = x, x = T)
##
##   n= 938, number of events= 465
##
##                    coef exp(coef) se(coef)     z Pr(>|z|)
## GDF11.2765.4.3 -0.3325    0.7171   0.0578 -5.75  8.7e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                exp(coef) exp(-coef) lower .95 upper .95
## GDF11.2765.4.3     0.717       1.39      0.64     0.803
##
## Concordance= 0.602  (se = 0.014 )
## Rsquare= 0.037   (max possible= 0.998 )
## Likelihood ratio test= 35.6  on 1 df,   p=2.42e-09
## Wald test            = 33.1  on 1 df,   p=8.75e-09
## Score (logrank) test = 26.6  on 1 df,   p=2.53e-07
```

• GDF11.WFIKKN1

```
## Call:
## coxph(formula = f, data = x, x = T)
##
##   n= 938, number of events= 465
##
##                      coef exp(coef) se(coef)     z Pr(>|z|)
## GDF11.2765.4.3    -0.3206    0.7257   0.0590 -5.43  5.6e-08 ***
## WFIKKN1.3191.50.2 -0.0409    0.9599   0.0466 -0.88     0.38
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                  exp(coef) exp(-coef) lower .95 upper .95

## GDF11.2765.4.3        0.726       1.38     0.646     0.815
## WFIKKN1.3191.50.2     0.960       1.04     0.876     1.052
##
## Concordance= 0.601  (se = 0.014 )
## Rsquare= 0.038   (max possible= 0.998 )
## Likelihood ratio test= 36.4  on 2 df,   p=1.26e-08
## Wald test            = 34.3  on 2 df,   p=3.55e-08
## Score (logrank) test = 28.8  on 2 df,   p=5.65e-07
```

• GDF11.WFIKKN2

```
## Call:
## coxph(formula = f, data = x, x = T)
##
##   n= 938, number of events= 465
##
##                       coef exp(coef) se(coef)     z Pr(>|z|)
## GDF11.2765.4.3    -0.3369    0.7140   0.0575 -5.86  4.6e-09 ***
## WFIKKN2.3235.50.2  0.2014    1.2232   0.0484  4.16  3.2e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                   exp(coef) exp(-coef) lower .95 upper .95
## GDF11.2765.4.3        0.714      1.401     0.638     0.799
## WFIKKN2.3235.50.2     1.223      0.818     1.112     1.345
##
## Concordance= 0.609  (se = 0.014 )
## Rsquare= 0.055   (max possible= 0.998 )
## Likelihood ratio test= 53  on 2 df,   p=3.18e-12
## Wald test            = 50.7  on 2 df,   p=9.63e-12
## Score (logrank) test = 42.1  on 2 df,   p=7.26e-10
```

- GDF11.WFIKKN1.WFIKKN2

```
## Call:
## coxph(formula = f, data = x, x = T)
##
##   n= 938, number of events= 465
##
##                       coef exp(coef) se(coef)     z Pr(>|z|)
## GDF11.2765.4.3    -0.3294    0.7193   0.0589 -5.60  2.2e-08 ***
## WFIKKN1.3191.50.2 -0.0256    0.9747   0.0466 -0.55     0.58
## WFIKKN2.3235.50.2  0.1989    1.2201   0.0486  4.10  4.2e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                   exp(coef) exp(-coef) lower .95 upper .95
## GDF11.2765.4.3        0.719       1.39     0.641     0.807
## WFIKKN1.3191.50.2     0.975       1.03     0.890     1.068
## WFIKKN2.3235.50.2     1.220       0.82     1.109     1.342
##

## Concordance= 0.609  (se = 0.014 )
## Rsquare= 0.055   (max possible= 0.998 )
## Likelihood ratio test= 53.2  on 3 df,   p=1.62e-11
## Wald test            = 51.4  on 3 df,   p=4.1e-11
## Score (logrank) test = 43.7  on 3 df,   p=1.73e-09
```

REIVINDICAÇÕES

1. Método para triagem de um indivíduo quanto ao risco de um evento cardiovascular (CV) ou para predizer a probabilidade que um indivíduo tenha um evento CV, caracterizado pelo fato de que compreende:

(a) formar um painel de biomarcador compreendendo biomarcadores de proteína MMP12, angiopoietina-2, complemento C7, troponina I cardíaca, proteína relacionada com angiopoietina 4, CCL18/PARC, complexo de alfa-1-antiquimotripsina, GDF11 e alfa-2-antiplasmina;

(b) detectar o nível de cada um dos biomarcadores de proteína do painel em uma amostra do indivíduo; e

(c) comparar os níveis de detecção com níveis de controle.

2. Método de acordo com a reivindicação 1, caracterizado pelo fato de que a probabilidade de o indivíduo ter um evento CV dentro de 4 anos é elevada se o nível de complexo de alfa-1-antiquimotripsina é maior que um nível de controle, e o nível de quatro a seis biomarcadores selecionados a partir do nível de MMP12, angiopoetina-2, complemento C7, troponina I cardíaca, proteína relacionada com angiopoietina 4, CCL18/PARC forem superiores a um nível da respectiva proteína de controle, e se o nível de um biomarcador ou ambos os biomarcadores selecionados a partir de GDF11 e alfa2-antiplasmina for menor do que um nível de controle da respectiva proteína.

3. Método de acordo com a reivindicação de 1 ou 2, caracterizado pelo fato de que o indivíduo tem doença da artéria coronária.

4. Método de acordo com qualquer uma das reivindicações de 1 a 3, caracterizado pelo fato de que o indivíduo não tem um histórico de eventos CV.

5. Método de acordo com qualquer uma das reivindicações 1 a 3, caracterizado pelo fato de que o indivíduo teve pelo menos um evento CV.

6. Método de acordo com qualquer uma das reivindicações de 1 a 5, caracterizado pelo fato de que o evento CV é selecionado a partir de enfarte do miocárdio, acidente vascular cerebral, insuficiência cardíaca congestiva, ataque isquêmico transgênico, e morte.

7. Método de acordo com qualquer uma das reivindicações de 1 a 6, caracterizado pelo fato de que a amostra é selecionada a partir de uma amostra de sangue, uma amostra de soro, uma amostra de plasma, e uma amostra de urina.

8. Método de acordo com qualquer uma das reivindicações de 1 a 7, caracterizado pelo fato de que compreende colocar biomarcadores da amostra do indivíduo em contato com um conjunto de reagentes de captura de biomarcadores, em que cada reagente de captura de biomarcador do conjunto de reagentes de captura de biomarcadores se liga especificamente a um biomarcador diferente a ser detectado.

9. Método de acordo com a reivindicação 8, caracterizado pelo fato de que cada reagente de captura de biomarcador é um anticorpo ou um aptâmero.

10. Método de acordo com a reivindicação 9, caracterizado pelo fato de que cada reagente de captura de biomarcador é um aptâmero.

11. Método de acordo com a reivindicação 10, caracterizado pelo fato de que pelo menos um aptâmero é um aptâmero de taxa de desaceleração lenta, em que

cada aptâmero de taxa de desaceleração lenta se liga, opcionalmente, à sua proteína alvo com uma taxa de desaceleração ($t_{½}$) de ≥ 30 minutos.

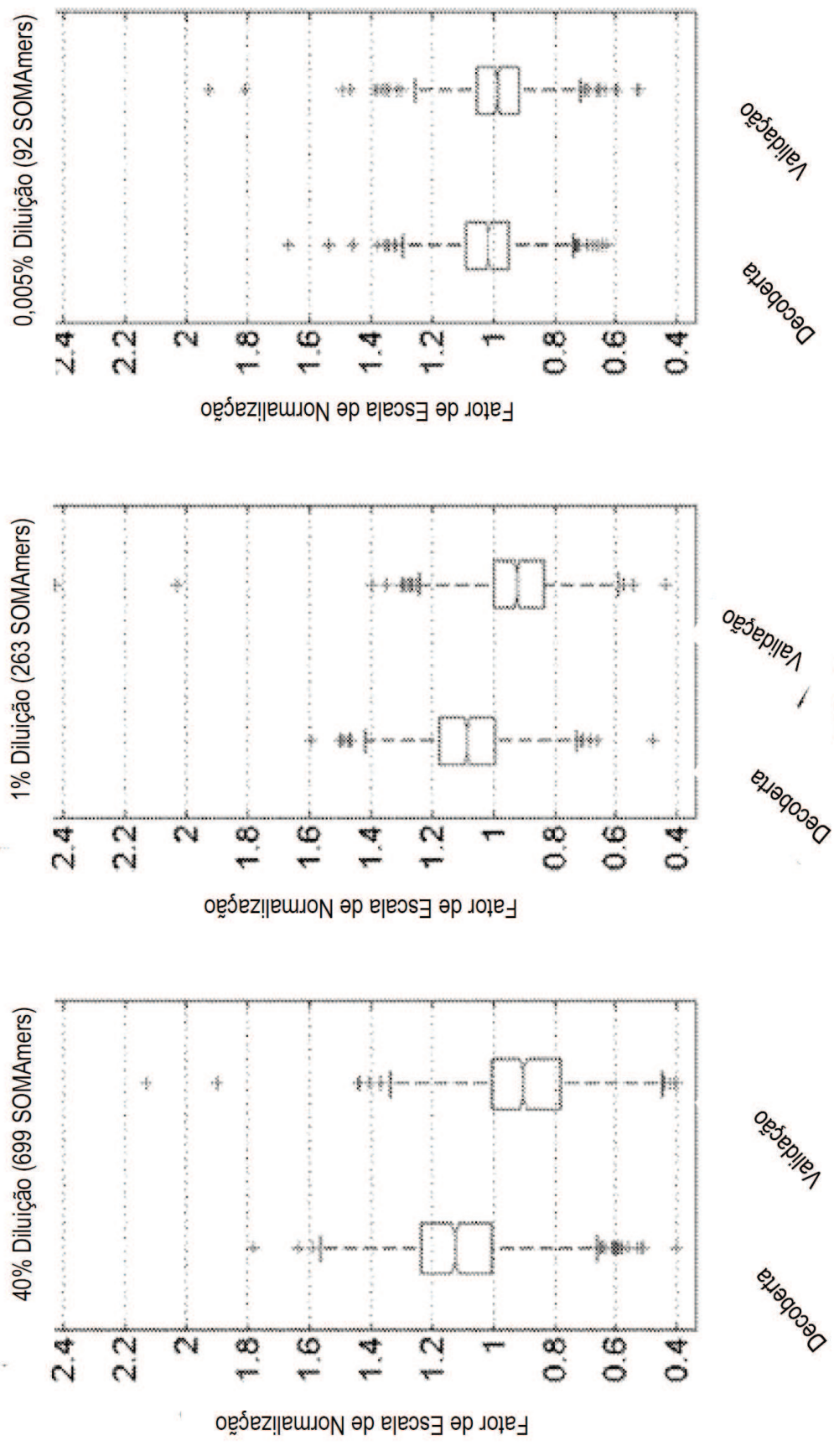

0,005% Diluição (92 SOMAmers)
1% Diluição (263 SOMAmers)
40% Diluição (699 SOMAmers)
Fator de Escala de Normalização
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
Descoberta
Validação

FIG. 1

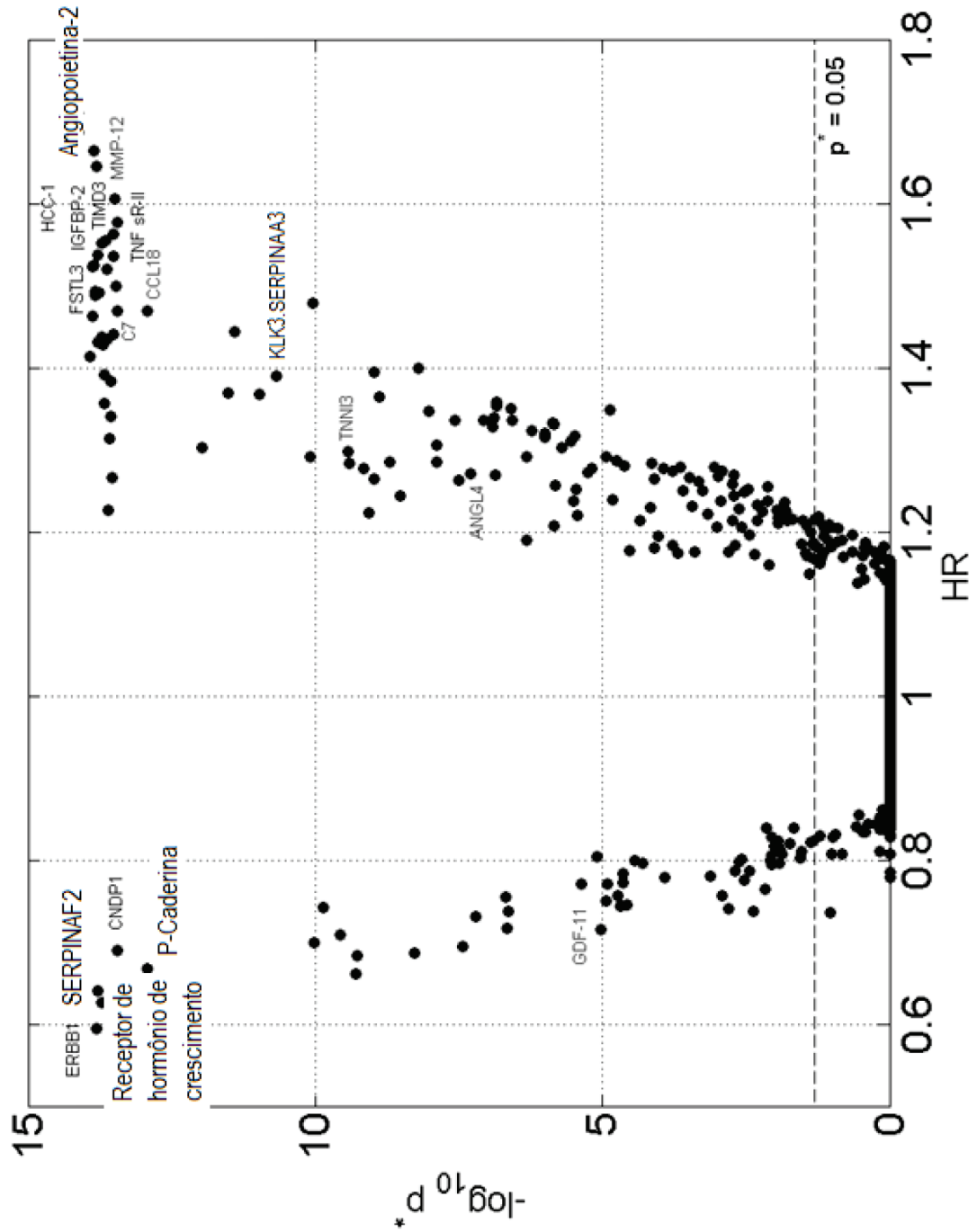

-log10 p*
HR
p* = 0.05
Angiopoietina-2
HCC-1
IGFBP-2
TIMD3
MMP-12
FSTL3
TNF sR-II
CCL18
C7
KLK3.SERPINAA3
TNNI3
ANGL4
GDF-11
ERBB1
SERPINAF2
Receptor de
hormônio de
crescimento
CNDP1
P-Caderina

FIG. 2

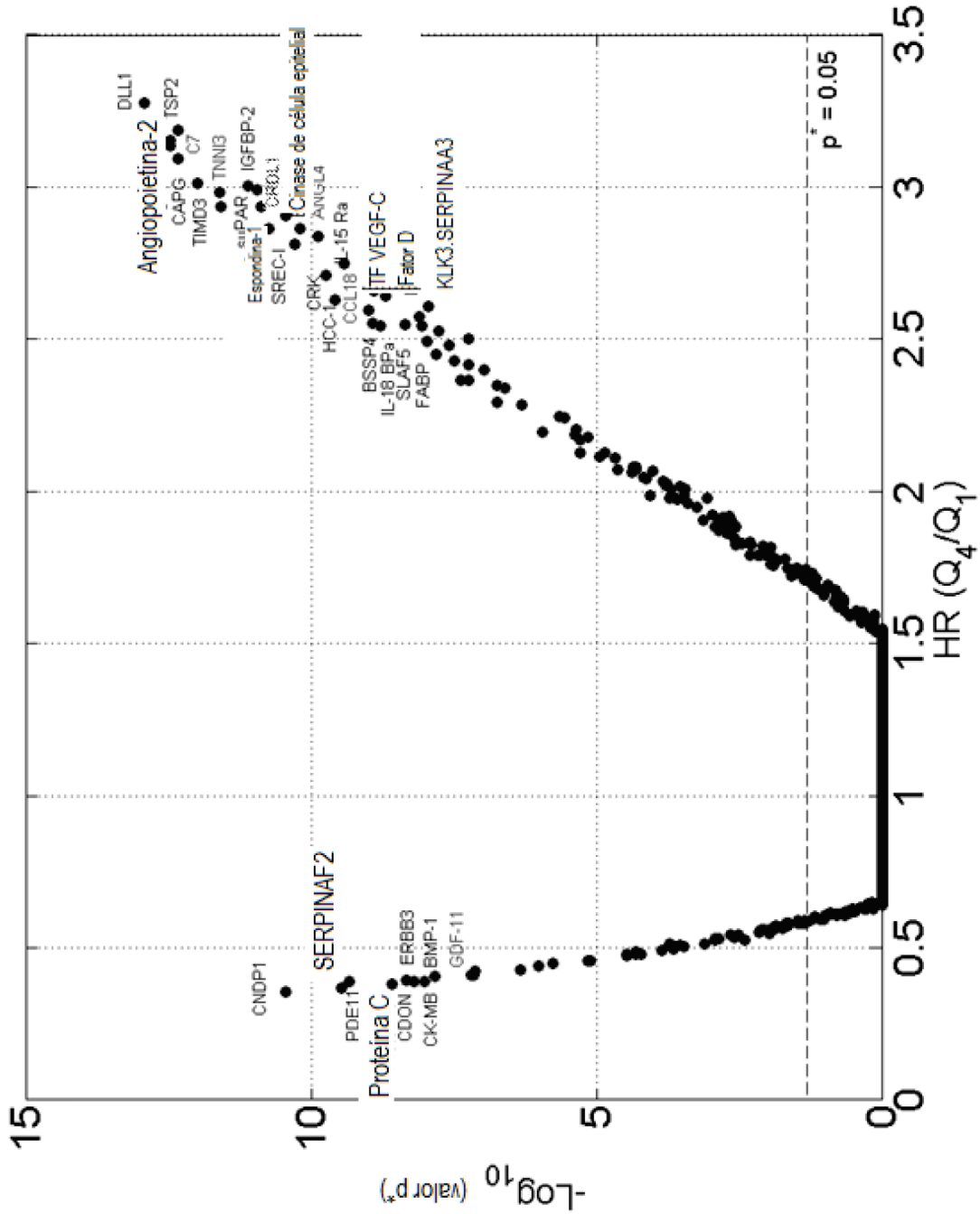


FIG. 2 (continuação)

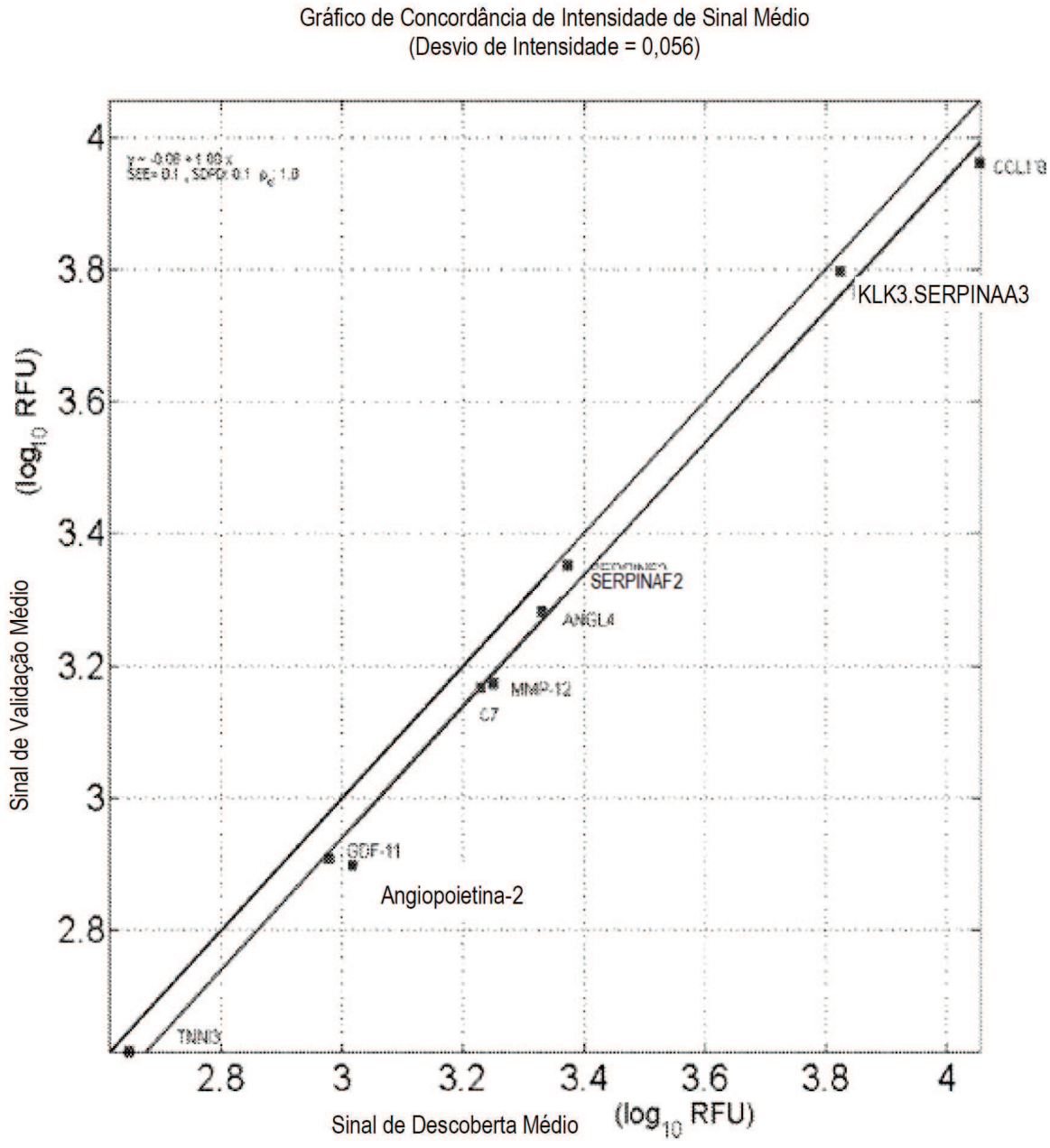

Gráfico de Concordância de Intensidade de Sinal Médio
(Desvio de Intensidade = 0,056)
Sinal de Validação Médio (log10 RFU)
Sinal de Descoberta Médio (log10 RFU)
4
3.8
3.6
3.4
3.2
3
2.8
2.8
3
3.2
3.4
3.6
3.8
4
KLK3.SERPINAA3
SERPINAF2
ANGL4
MMP-12
C7
GDF-11
Angiopoietina-2
TNN3

FIG. 3

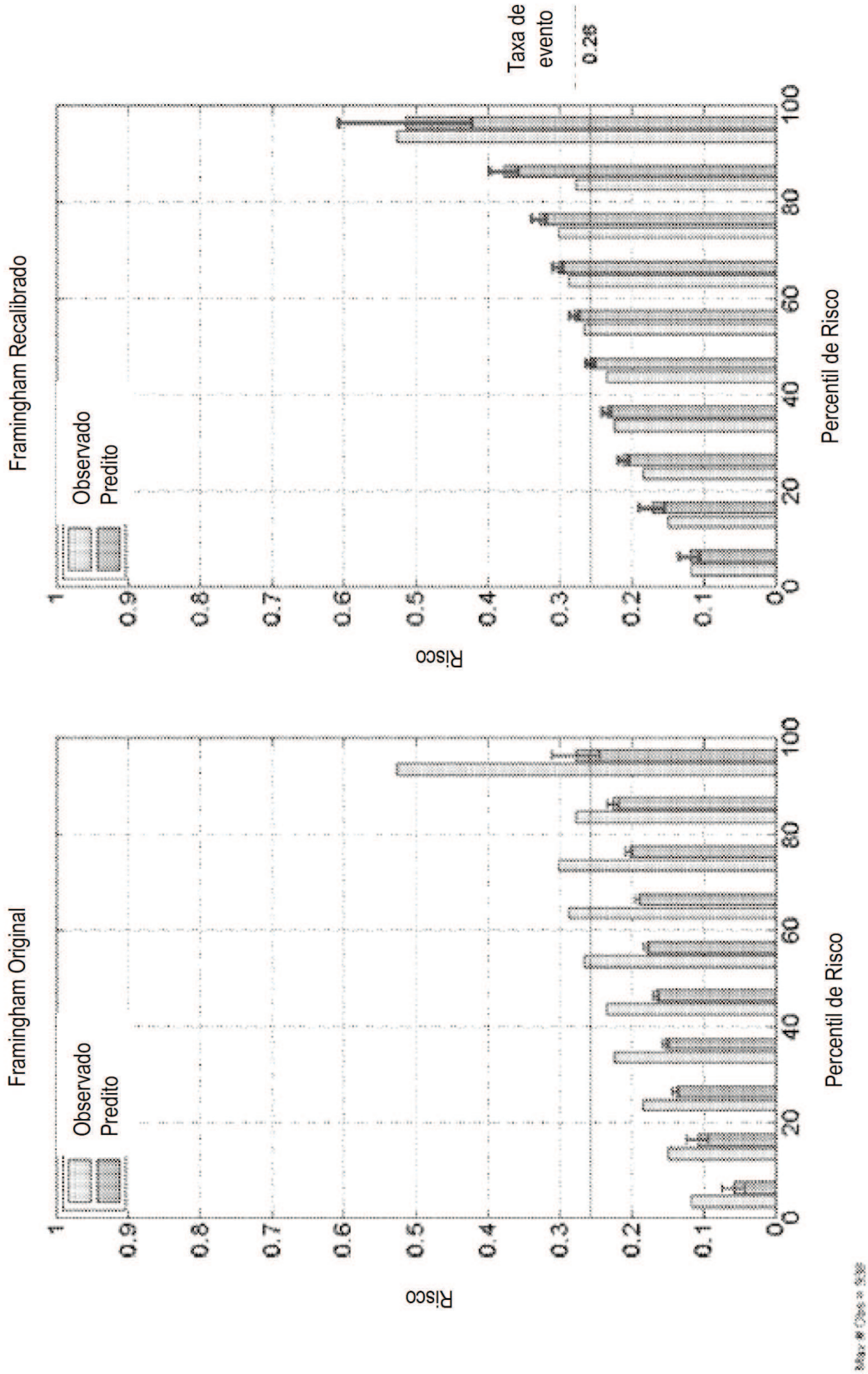

Framingham Original
Framingham Recalibrado
Observado
Predito
Taxa de evento
0.25
Risco
Percentil de Risco
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
20
40
60
80
100

FIG. 4

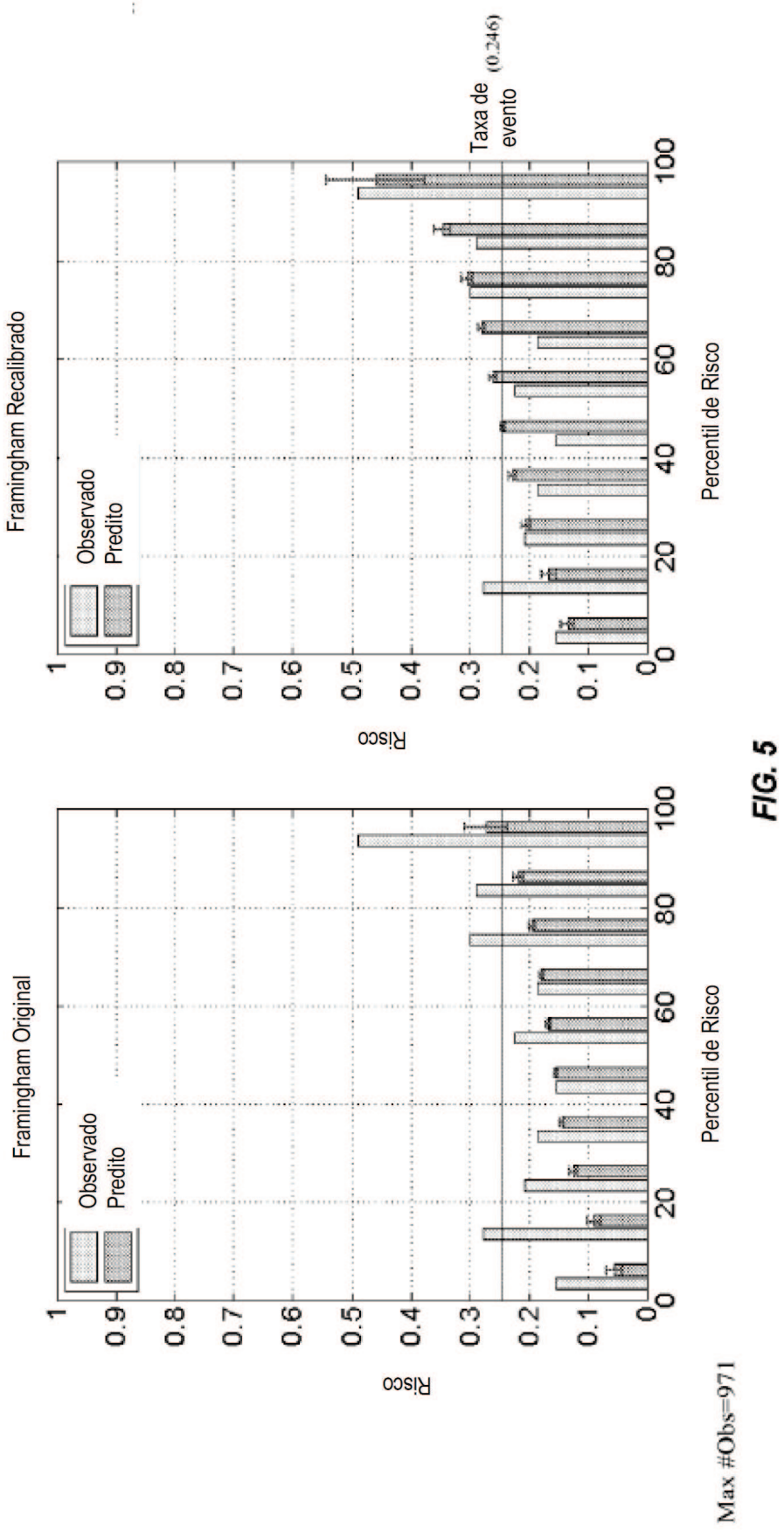

Framingham Recalibrado
Observado
Predito
Taxa de evento (0.246)
Risco
Percentil de Risco
Framingham Original
Observado
Predito
Risco
Percentil de Risco
Max #Obs=971

FIG. 5

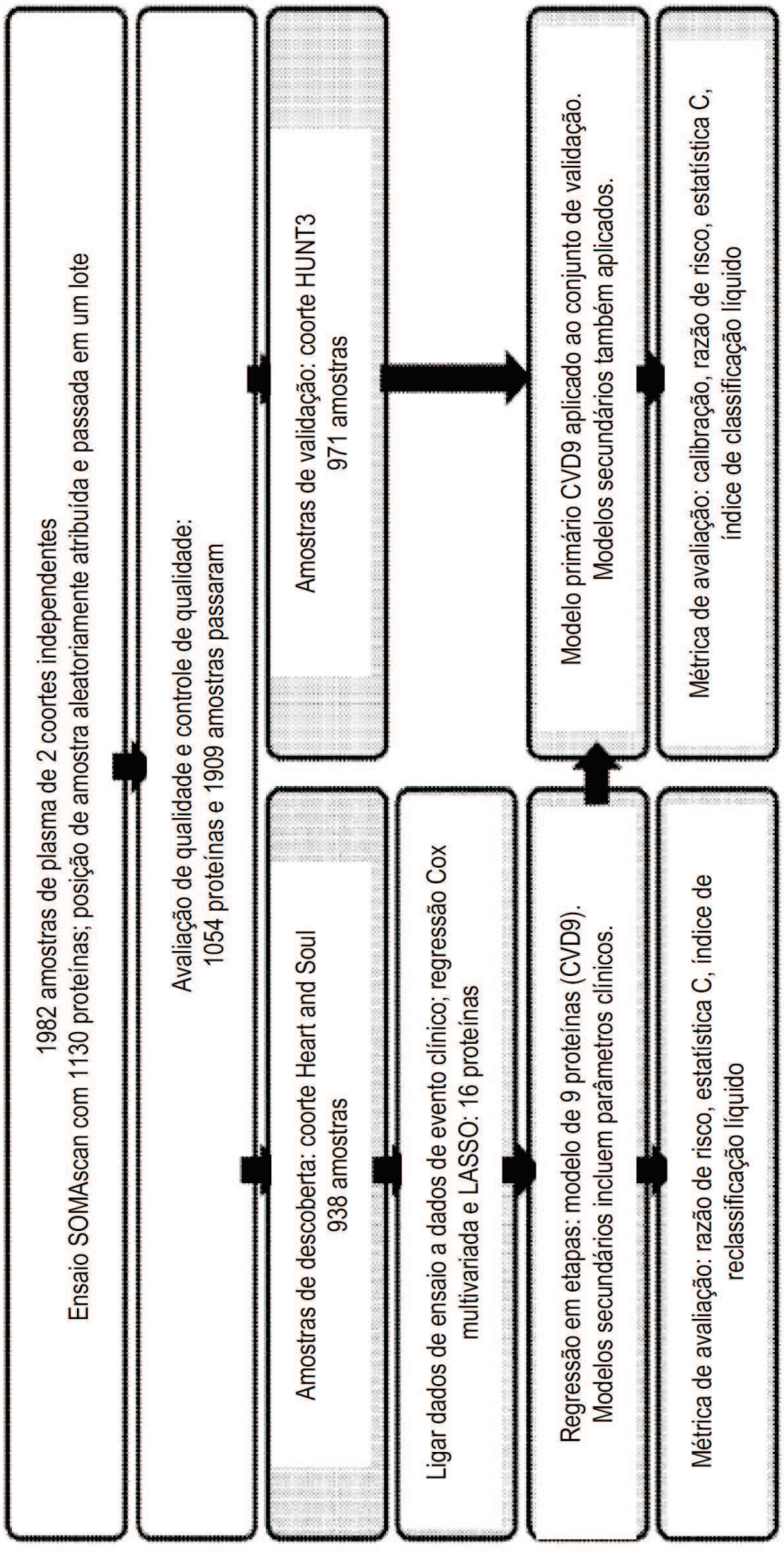

1982 amostras de plasma de 2 coortes independentes
Ensaio SOMAscan com 1130 proteínas; posição de amostra aleatoriamente atribuída e passada em um lote
Avaliação de qualidade e controle de qualidade:
1054 proteínas e 1909 amostras passaram
Amostras de descoberta: coorte Heart and Soul
938 amostras
Amostras de validação: coorte HUNT3
971 amostras
Ligar dados de ensaio a dados de evento clínico; regressão Cox multivariada e LASSO: 16 proteínas
Regressão em etapas: modelo de 9 proteínas (CVD9).
Modelos secundários incluem parâmetros clínicos.
Modelo primário CVD9 aplicado ao conjunto de validação.
Modelos secundários também aplicados.
Métrica de avaliação: razão de risco, estatística C, índice de reclassificação líquido
Métrica de avaliação: calibração, razão de risco, estatística C, índice de classificação líquido

FIG. 6

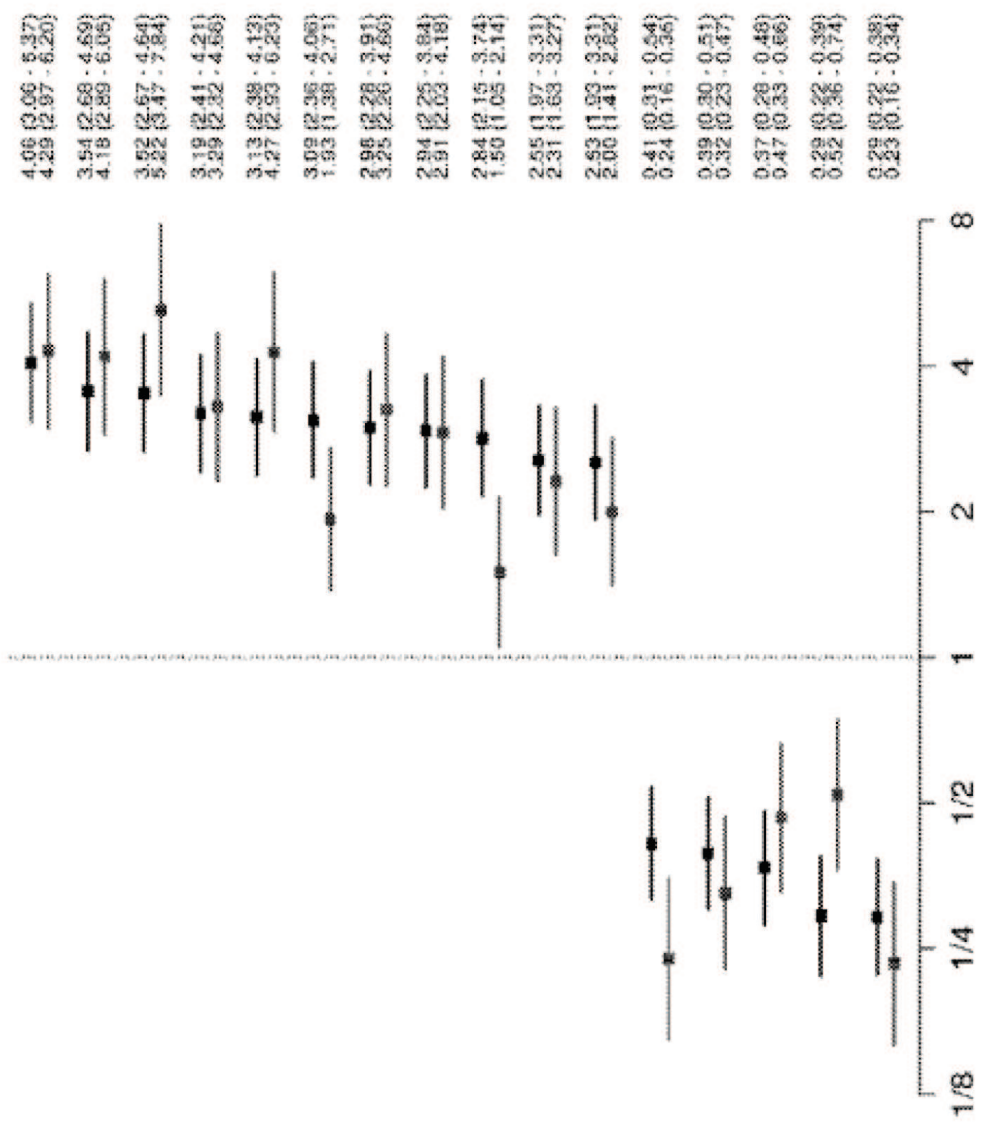

Catepsina H
MMP-7
MMP-12 *
Trombospondina-2
Angiopoietina-2*
Complemento C7*
Proteína de membrana de célula T TIM-3
Troponina I, cardíaca*
Proteína relativa a angiopoietina 4*
CCL18/PARC *
Complexo α1-antiquimotripsina*
GDF-11 *
CDO regulada por oncogene de adesão de célula
α2-Antiplasmina*
Receptor de hormônio de crescimento
Receptor EGF
4.06 (3.06 - 5.37)
4.29 (2.97 - 6.20)
3.54 (2.68 - 4.69)
4.18 (2.89 - 6.05)
3.52 (2.67 - 4.64)
5.22 (3.47 - 7.84)
3.19 (2.41 - 4.21)
3.29 (2.32 - 4.66)
3.13 (2.38 - 4.13)
4.27 (2.93 - 6.23)
3.09 (2.36 - 4.06)
1.93 (1.38 - 2.71)
2.96 (2.26 - 3.91)
3.25 (2.26 - 4.66)
2.94 (2.25 - 3.84)
2.91 (2.03 - 4.18)
2.84 (2.15 - 3.74)
1.50 (1.05 - 2.14)
2.55 (1.97 - 3.31)
2.31 (1.63 - 3.27)
2.53 (1.93 - 3.31)
2.00 (1.41 - 2.82)
0.41 (0.31 - 0.54)
0.24 (0.16 - 0.36)
0.39 (0.30 - 0.51)
0.32 (0.23 - 0.47)
0.37 (0.28 - 0.49)
0.47 (0.33 - 0.66)
0.29 (0.22 - 0.39)
0.52 (0.36 - 0.74)
0.29 (0.22 - 0.38)
0.23 (0.16 - 0.34)
1/8
1/4
1/2
1
2
4
8
Razão de Risco Q4/Q1

FIG. 7

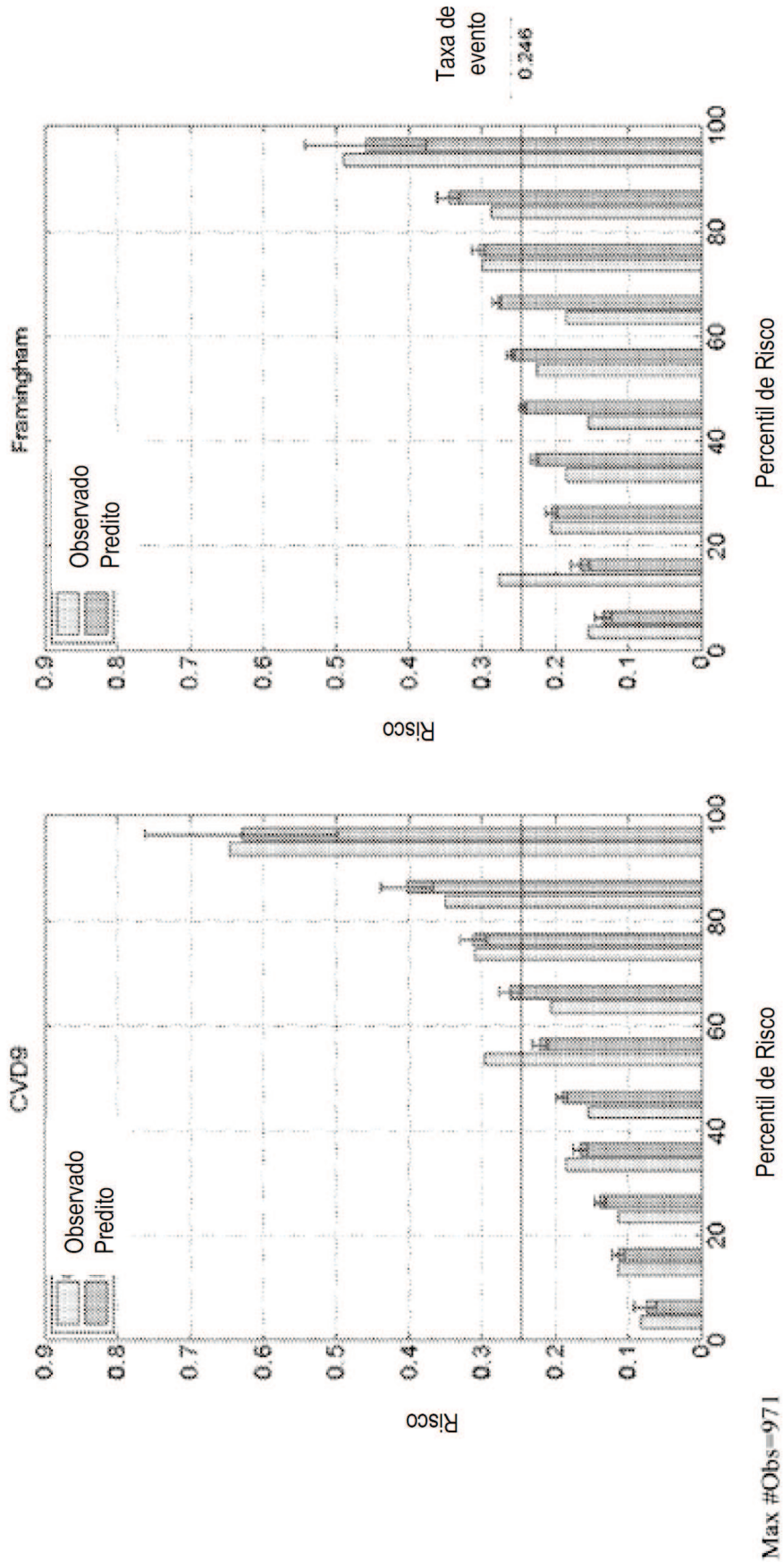

Framingham
Observado
Predito
Taxa de evento
0.246
Risco
Percentil de Risco
CVD9
Observado
Predito
Risco
Percentil de Risco
Max #Obs=971

FIG. 8

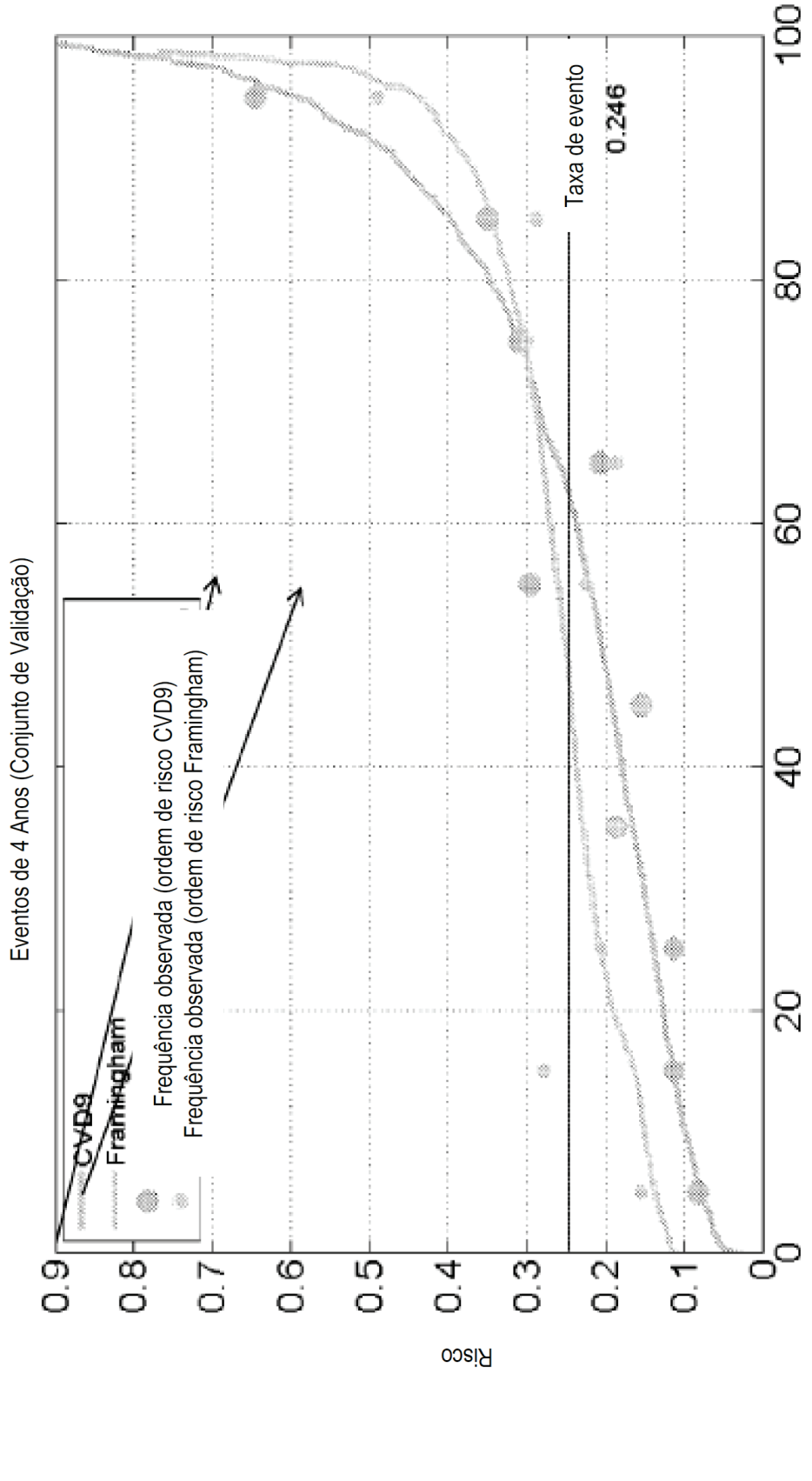

Eventos de 4 Anos (Conjunto de Validação)
Risco
Percentagem de Risco
Max #Obs=971
CVD9
Framingham
Frequência observada (ordem de risco CVD9)
Frequência observada (ordem de risco Framingham)
Taxa de evento
0.246
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0
20
40
60
80
100

FIG. 9

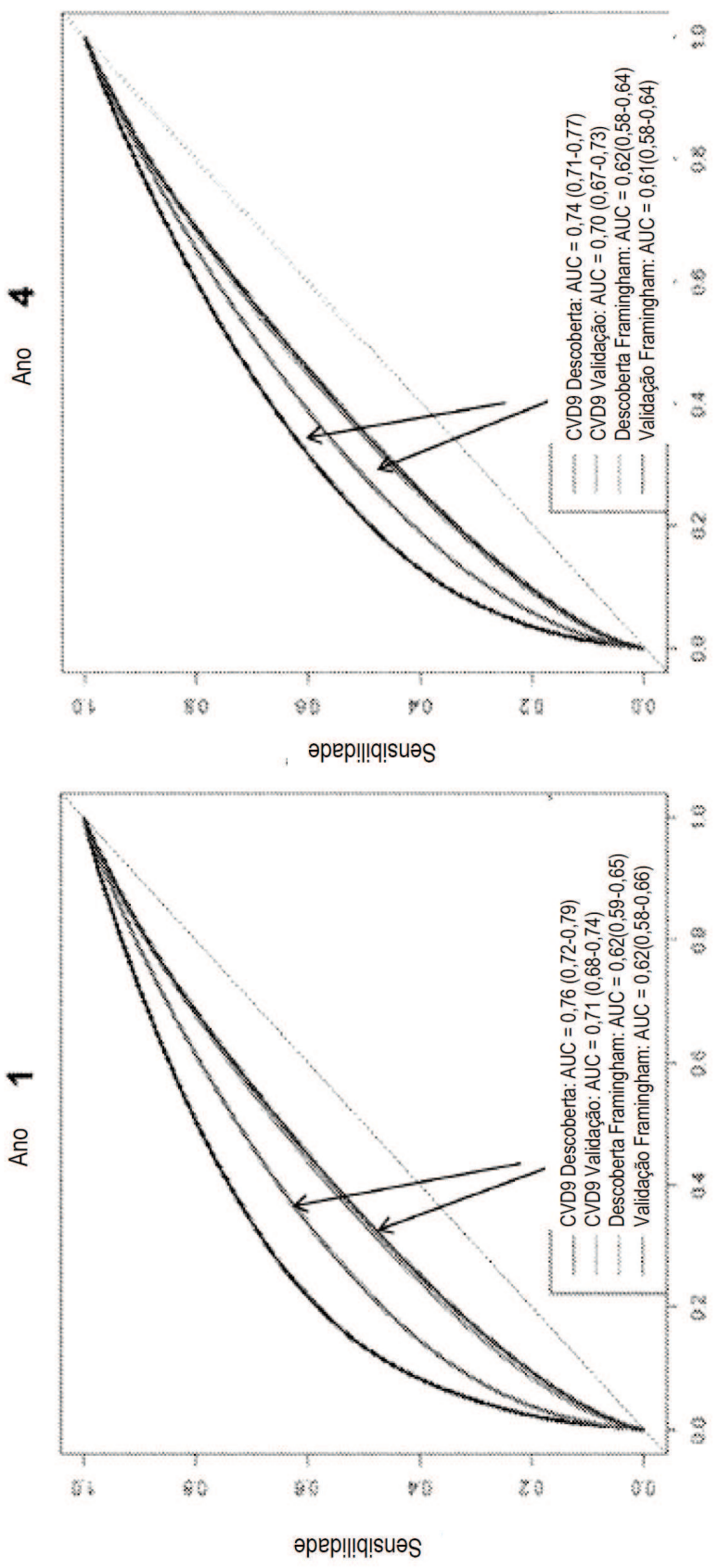

Ano 1
Sensibilidade
1 - Especificidade
CVD9 Descoberta: AUC = 0,76 (0,72-0,79)
CVD9 Validação: AUC = 0,71 (0,68-0,74)
Descoberta Framingham: AUC = 0,62(0,59-0,65)
Validação Framingham: AUC = 0,62(0,58-0,66)
Ano 4
Sensibilidade
1 - Especificidade
CVD9 Descoberta: AUC = 0,74 (0,71-0,77)
CVD9 Validação: AUC = 0,70 (0,67-0,73)
Descoberta Framingham: AUC = 0,62(0,58-0,64)
Validação Framingham: AUC = 0,61(0,58-0,64)

FIG. 10

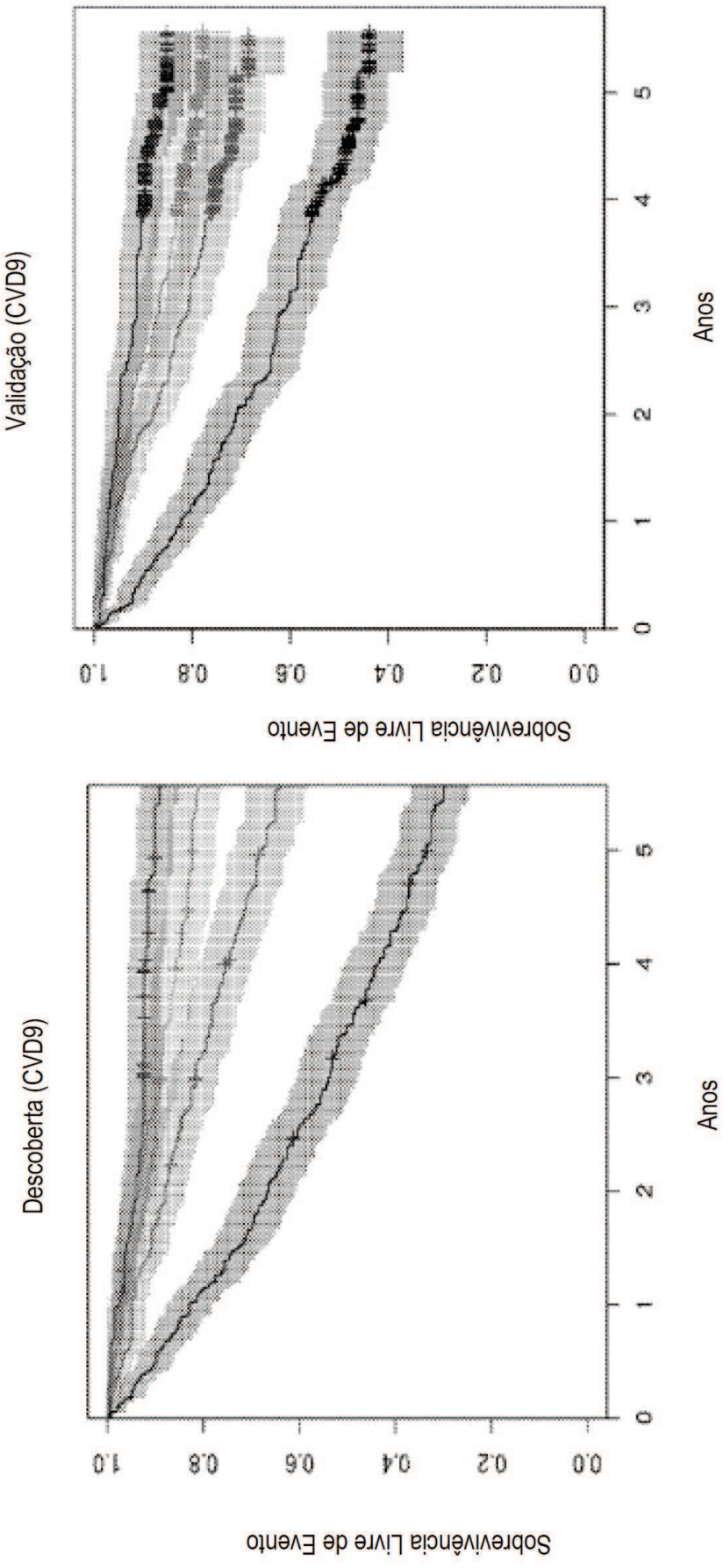

Validação (CVD9)
Descoberta (CVD9)
Sobrevivência Livre de Evento
Anos
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5

FIG. 11

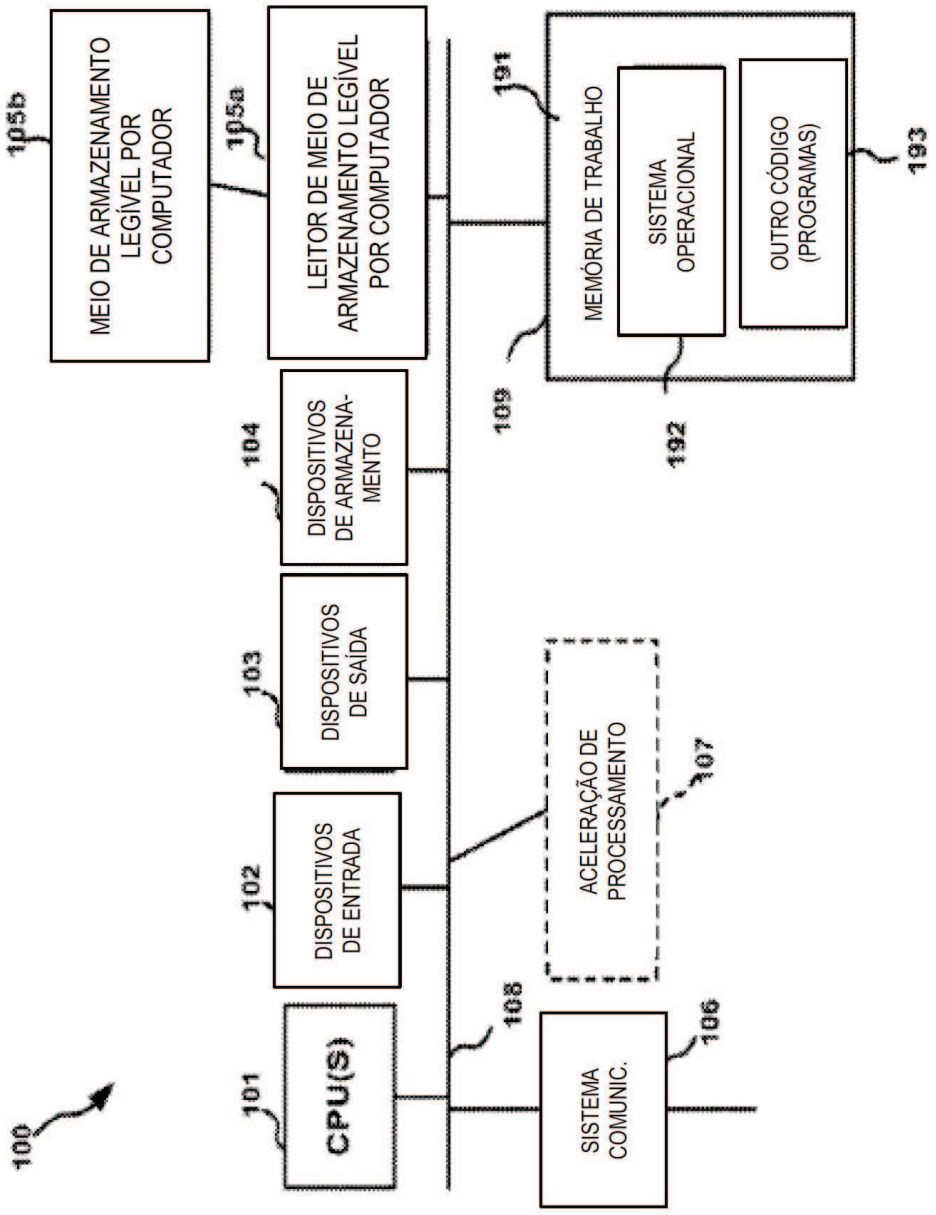

100
101
CPU(S)
102
DISPOSITIVOS DE ENTRADA
103
DISPOSITIVOS DE SAÍDA
104
DISPOSITIVOS DE ARMAZENA-MENTO
105a
LEITOR DE MEIO DE ARMAZENAMENTO LEGÍVEL POR COMPUTADOR
105b
MEIO DE ARMAZENAMENTO LEGÍVEL POR COMPUTADOR
106
SISTEMA COMUNIC.
107
ACELERAÇÃO DE PROCESSAMENTO
108
109
191
MEMÓRIA DE TRABALHO
192
SISTEMA OPERACIONAL
193
OUTRO CÓDIGO (PROGRAMAS)

FIG. 12

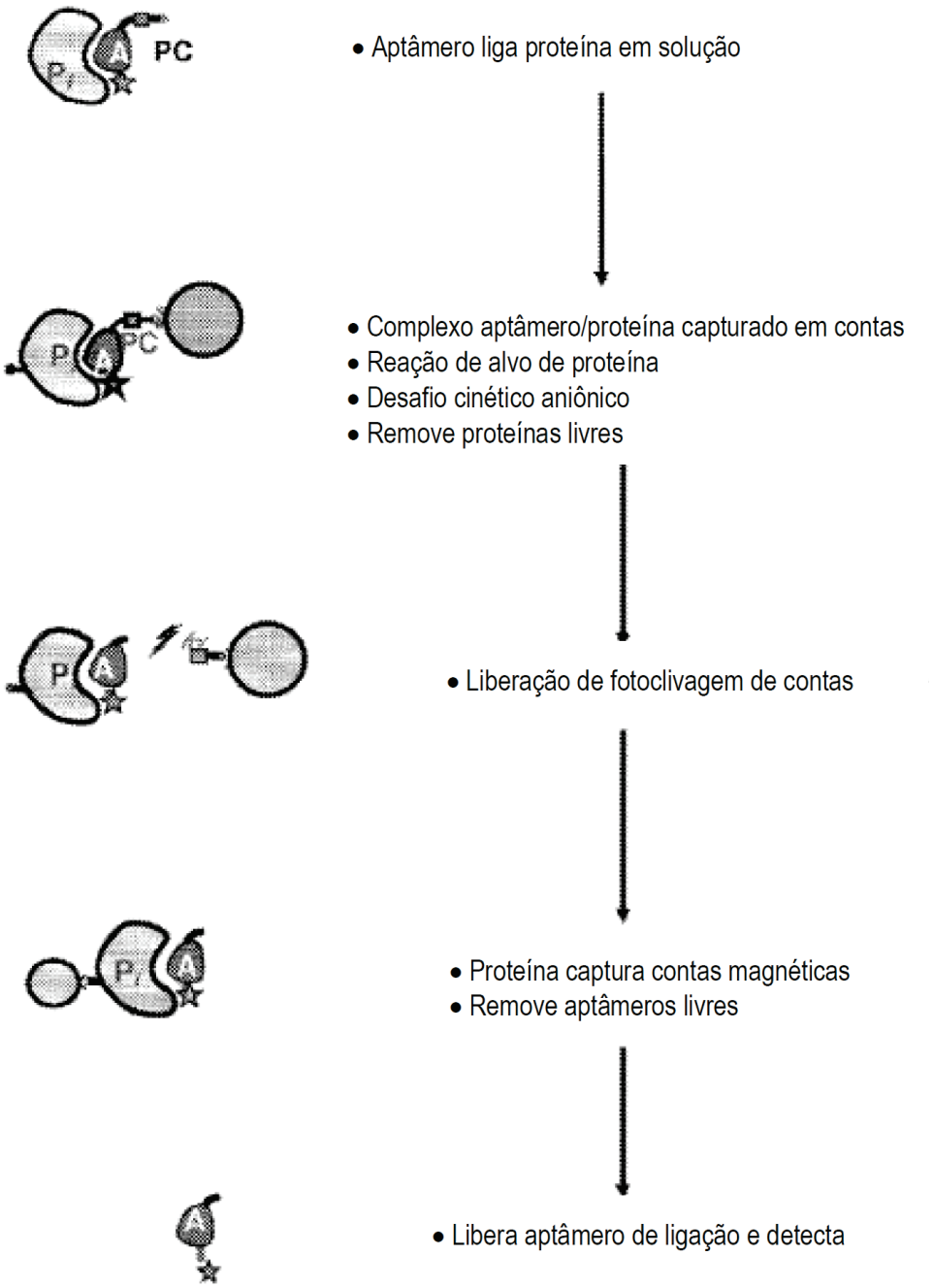

PC
• Aptâmero liga proteína em solução
• Complexo aptâmero/proteína capturado em contas
• Reação de alvo de proteína
• Desafio cinético aniônico
• Remove proteínas livres
• Liberação de fotoclivagem de contas
• Proteína captura contas magnéticas
• Remove aptâmeros livres
• Libera aptâmero de ligação e detecta

FIG. 13

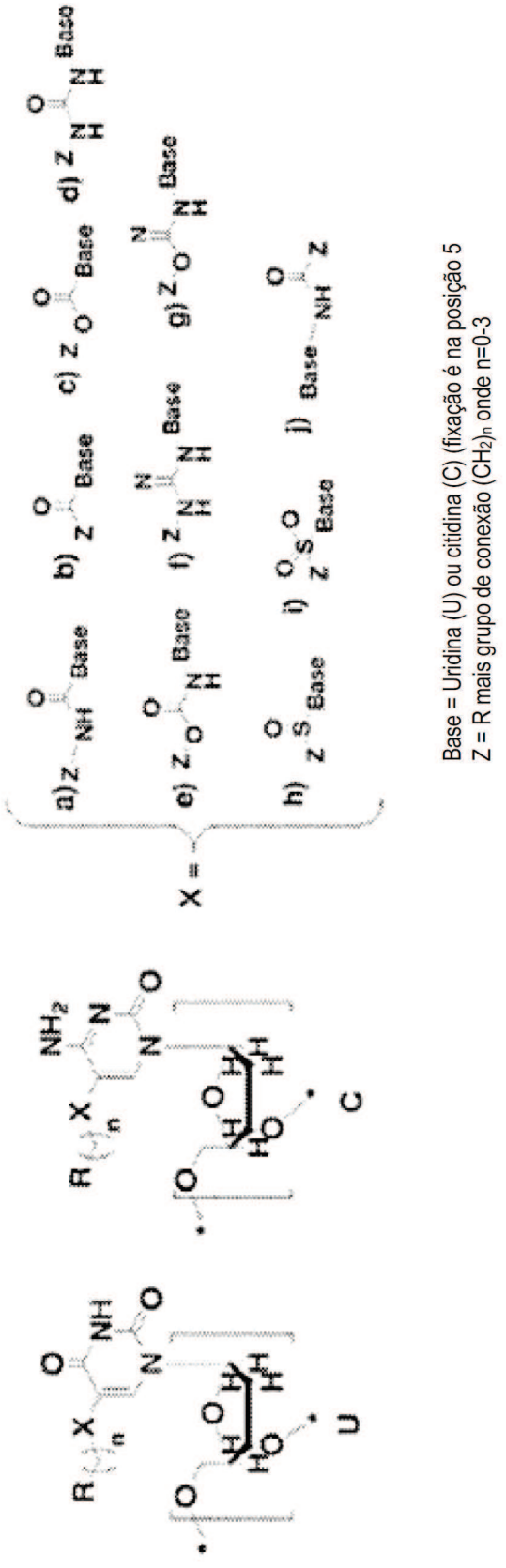

X =
a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
Base
U
C
Base = Uridina (U) ou citidina (C) (fixação é na posição 5
Z = R mais grupo de conexão (CH2)n onde n=0-3

FIG. 14

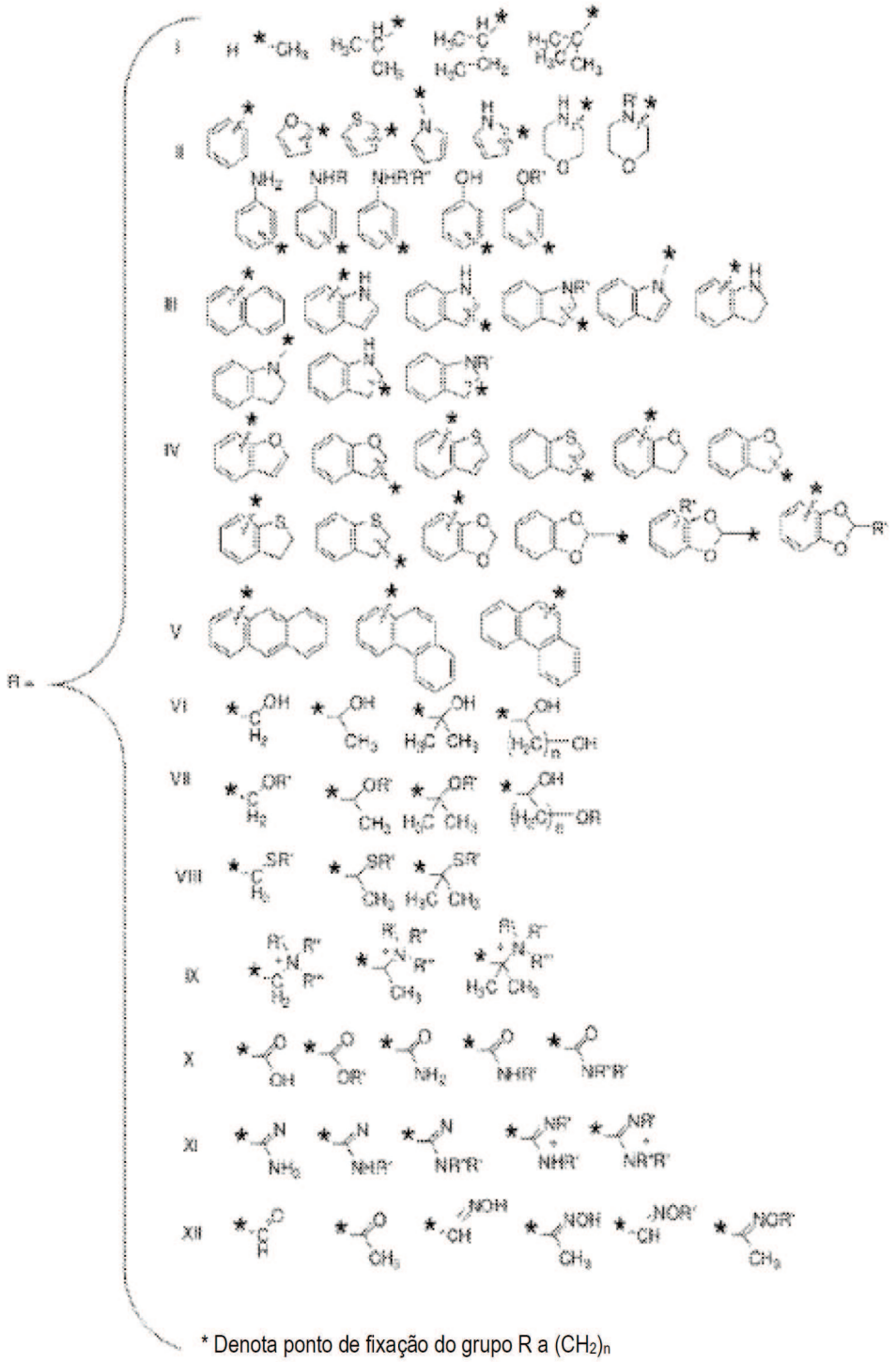

R =
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
* Denota ponto de fixação do grupo R a (CH2)n

FIG. 14 (cont.)

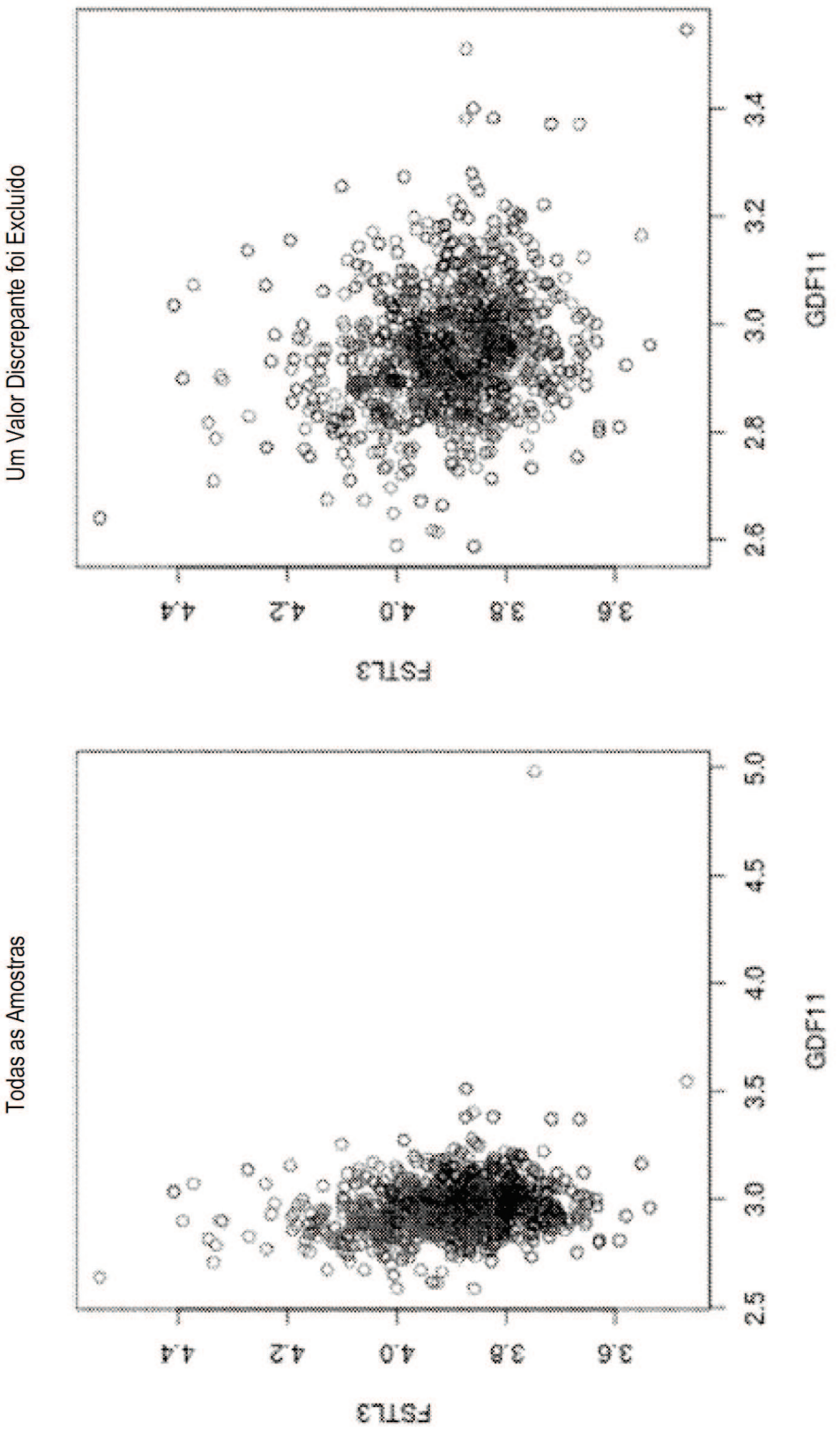

Um Valor Discrepante foi Excluído
Todas as Amostras
FSTL3
GDF11
3.6
3.8
4.0
4.2
4.4
2.6
2.8
3.0
3.2
3.4
2.5
3.0
3.5
4.0
4.5
5.0

FIG. 15

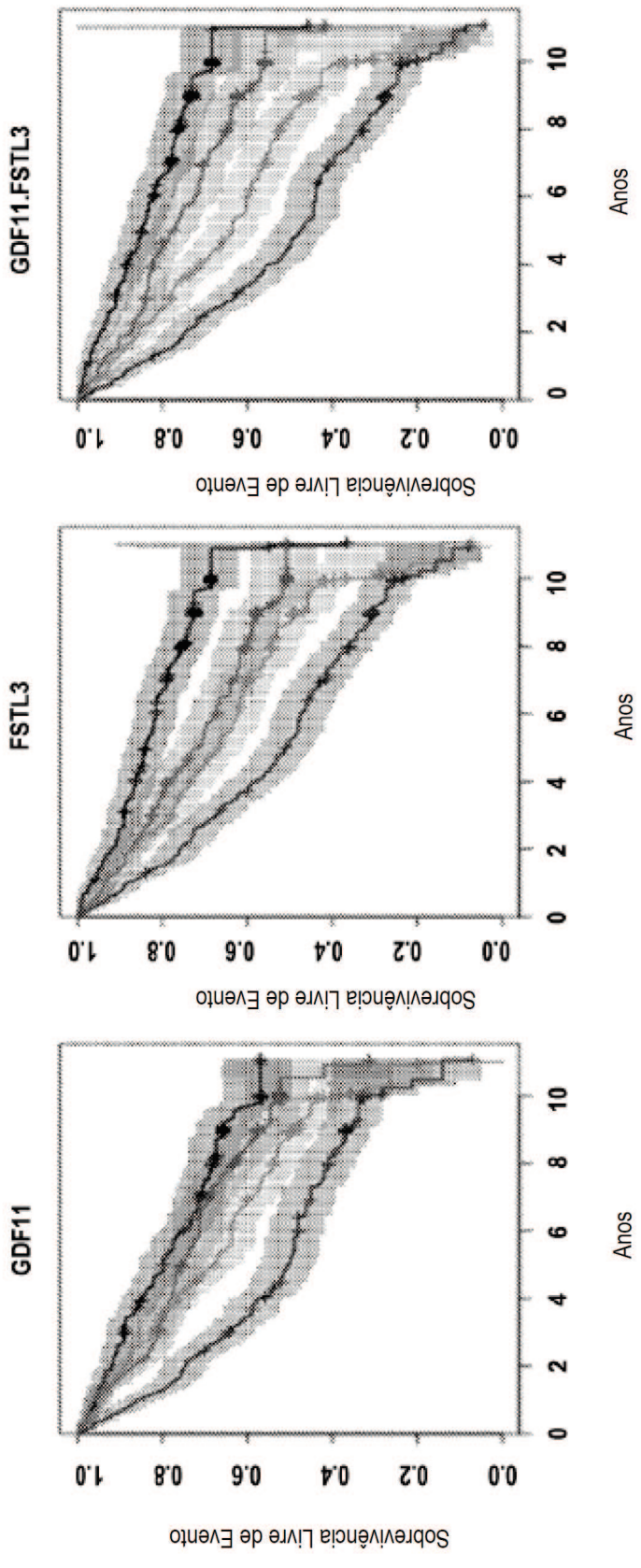

GDF11
FSTL3
GDF11.FSTL3
Sobrevivência Livre de Evento
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
Anos

FIG. 16

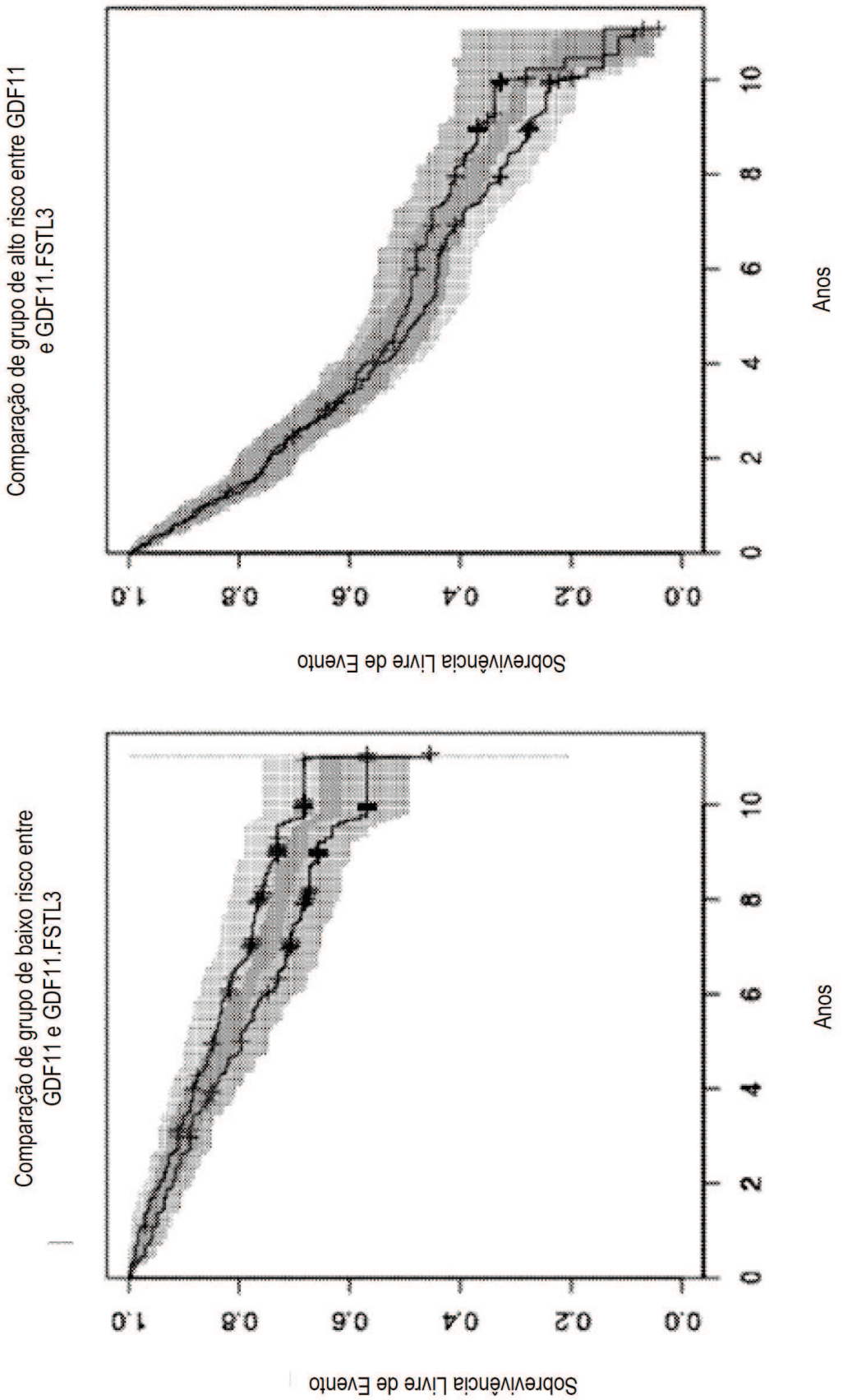

Comparação de grupo de alto risco entre GDF11 e GDF11.FSTL3
Sobrevivência Livre de Evento
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
Anos
Comparação de grupo de baixo risco entre GDF11 e GDF11.FSTL3
Sobrevivência Livre de Evento
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
Anos

FIG. 17

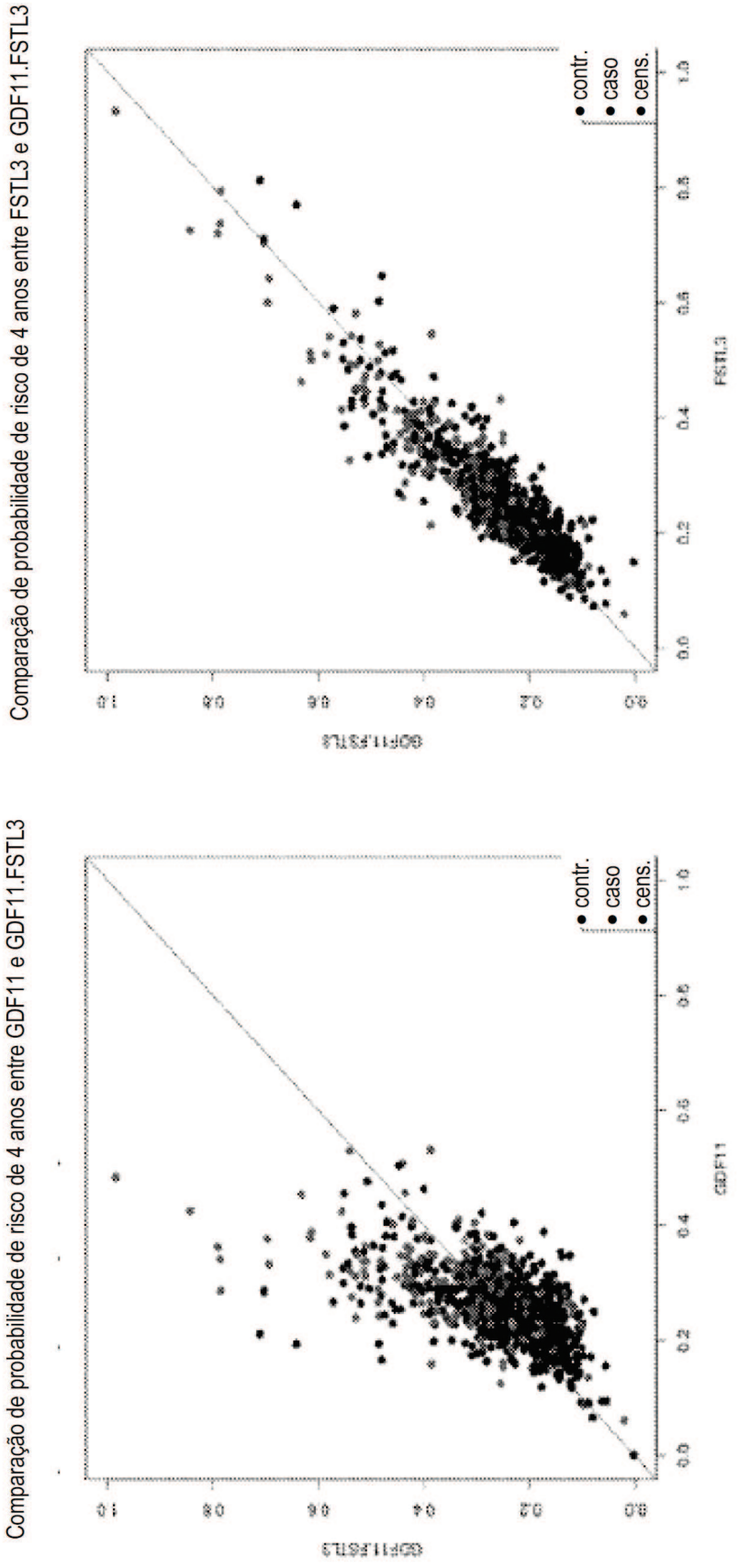

Comparação de probabilidade de risco de 4 anos entre FSTL3 e GDF11.FSTL3
contr.
caso
cens.
GDF11.FSTL3
FSTL3
Comparação de probabilidade de risco de 4 anos entre GDF11 e GDF11.FSTL3
contr.
caso
cens.
GDF11.FSTL3
GDF11

FIG. 18

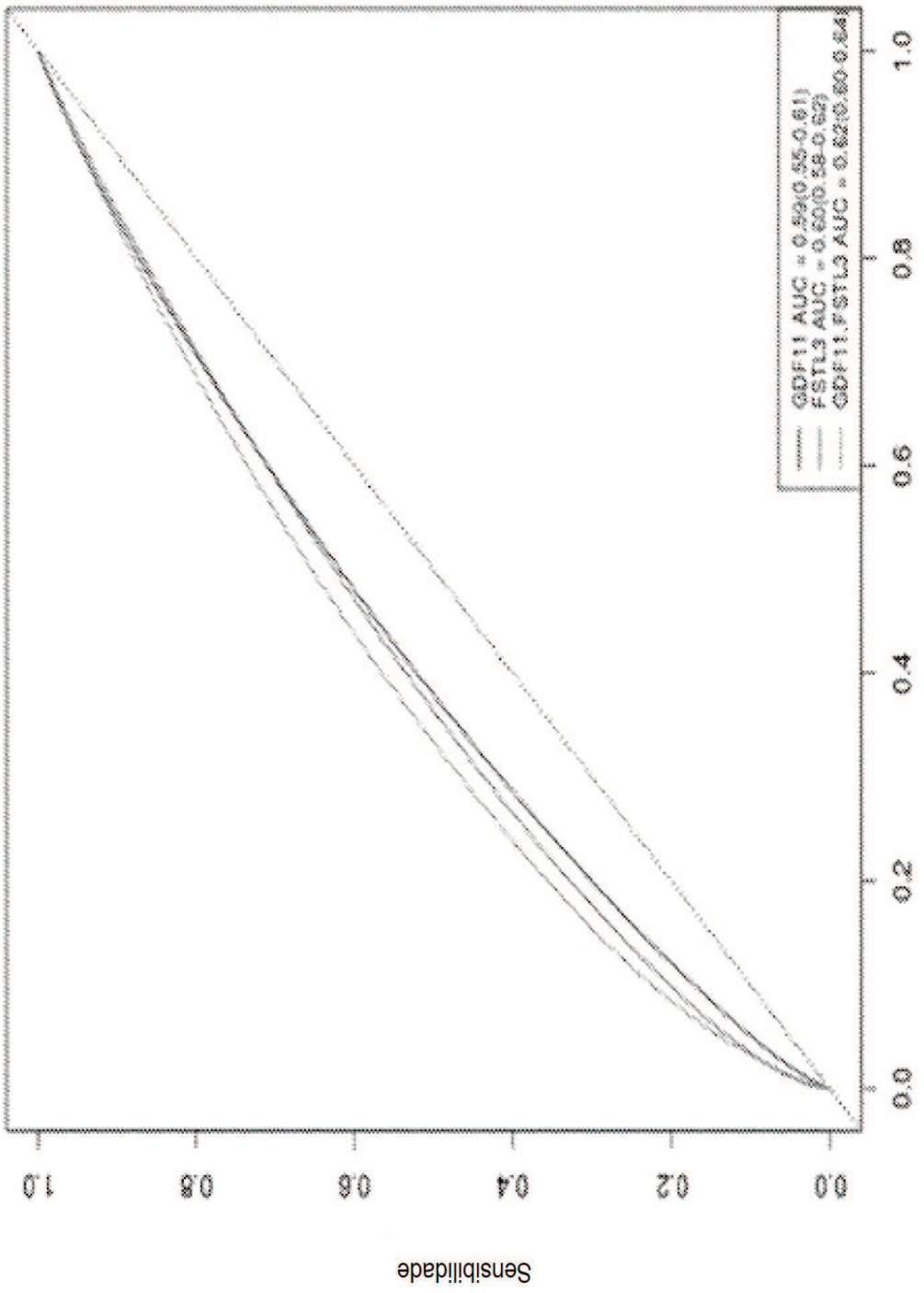

GDF11 AUC = 0.59(0.55-0.61)
FSTL3 AUC = 0.60(0.58-0.62)
GDF11.FSTL3 AUC = 0.62(0.60-0.64)
0.0
0.2
0.4
0.6
0.8
1.0
Sensibilidade
1-Especificidade

FIG. 19

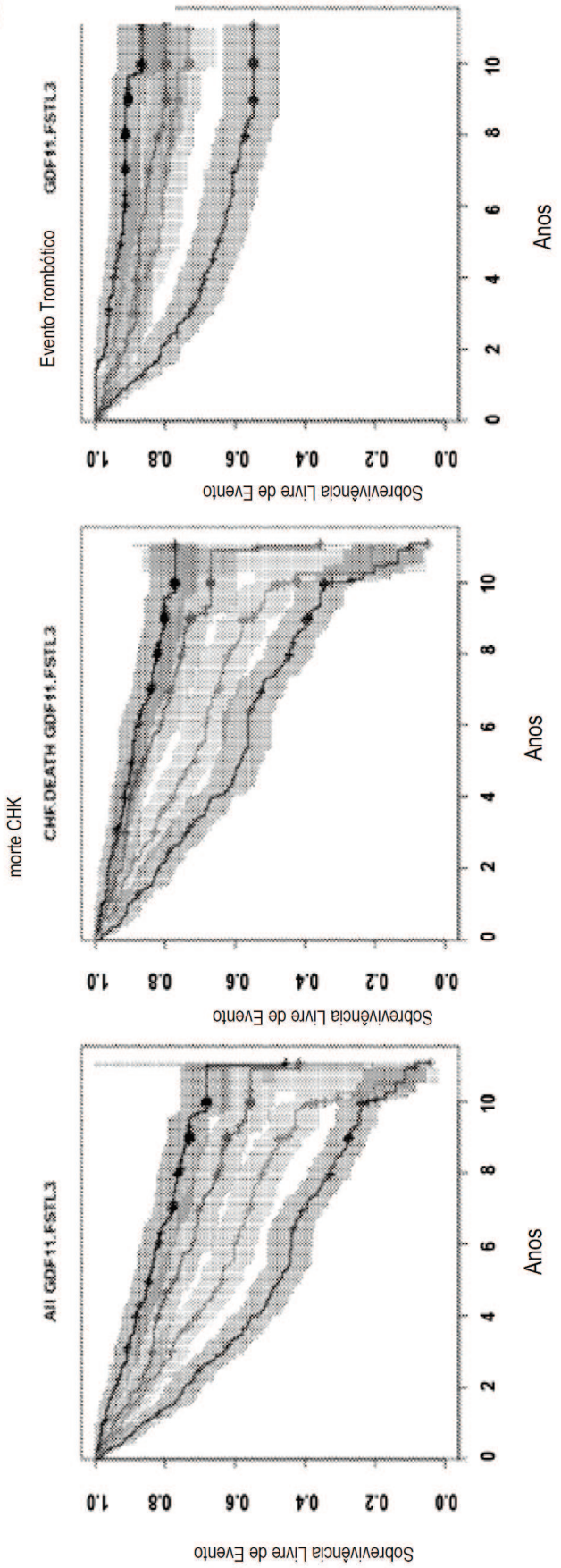

All GDF11.FSTL3
CHF.DEATH GDF11.FSTL3
Evento Trombótico GDF11.FSTL3
Sobrevivência Livre de Evento
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
Anos
morte CHK

FIG. 20

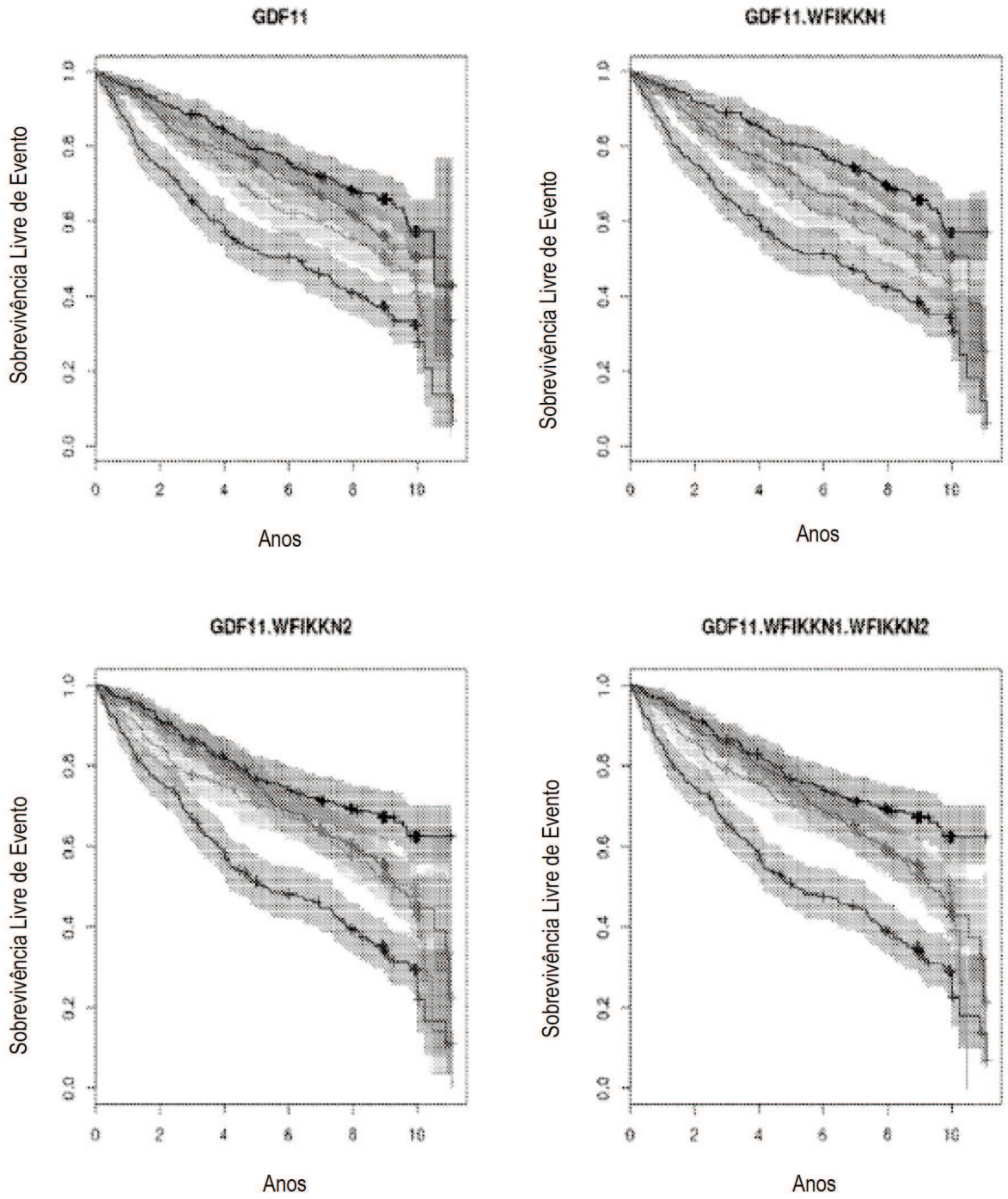

GDF11
GDF11.WFIKKN1
GDF11.WFIKKN2
GDF11.WFIKKN1.WFIKKN2
Sobrevivência Livre de Evento
Anos
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10

FIG. 21

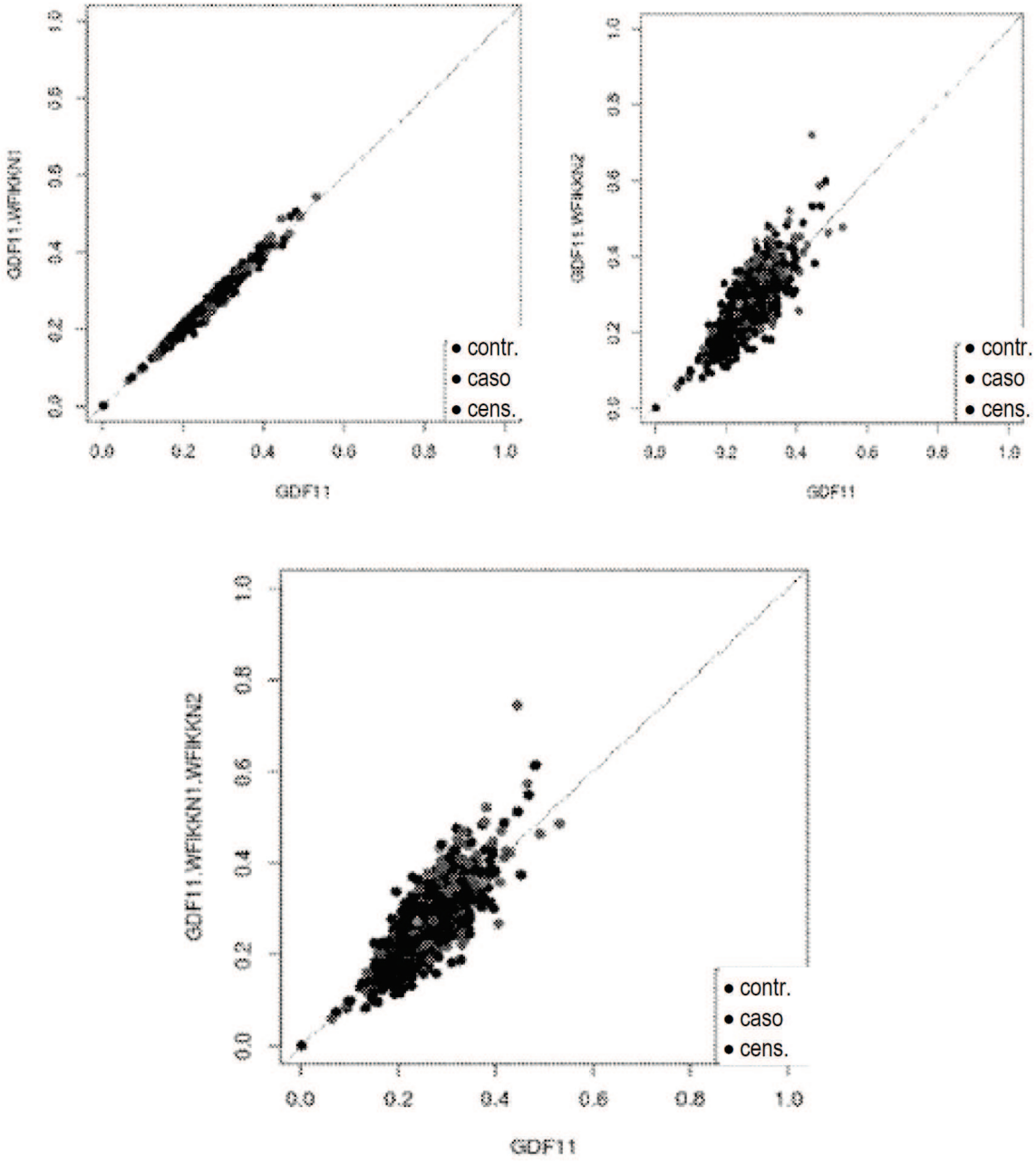

GDF11.WFIKKN1
GDF11
contr.
caso
cens.
GDF11.WFIKKN2
GDF11
contr.
caso
cens.
GDF11.WFIKKN1.WFIKKN2
GDF11
contr.
caso
cens.

FIG. 22

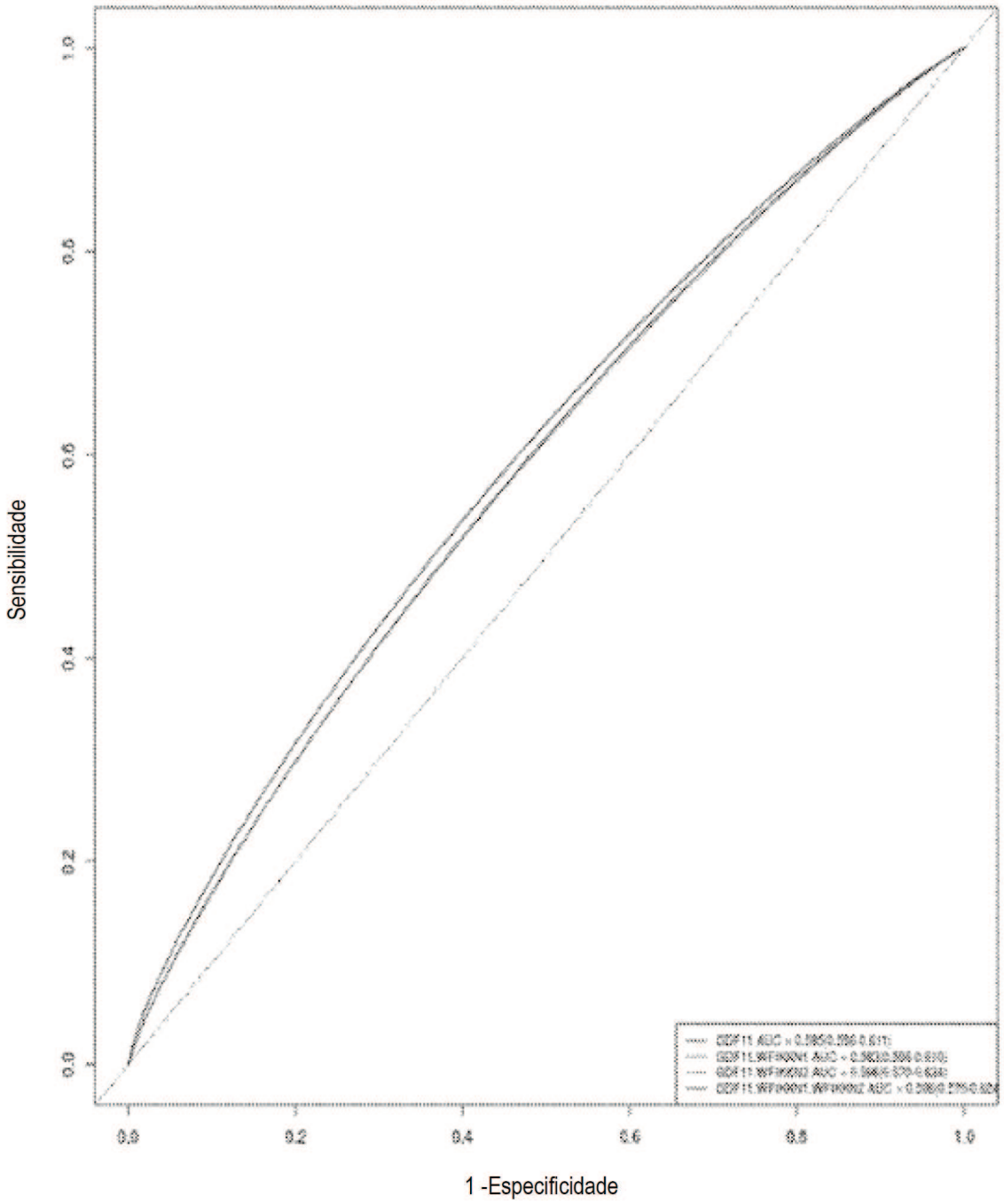


FIG. 23