



República Federativa do Brasil
Ministério da Economia
Instituto Nacional da Propriedade Industrial

(21) BR 112020007449-7 A2



(22) Data do Depósito: 17/10/2018

(43) Data da Publicação Nacional: 20/10/2020

(54) **Título:** PLATAFORMA DE EVIDÊNCIA MOLECULAR PARA OTIMIZAÇÃO AUDITÁVEL E CONTÍNUA DE INTERPRETAÇÃO DE VARIANTES EM TESTES E ANÁLISES GENÉTICAS E GENÔMICAS

(51) **Int. Cl.:** G06F 19/18; G06F 19/26; G06F 19/12.

(30) **Prioridade Unionista:** 17/10/2017 US 62/573,458.

(71) **Depositante(es):** JUNGLA LLC.

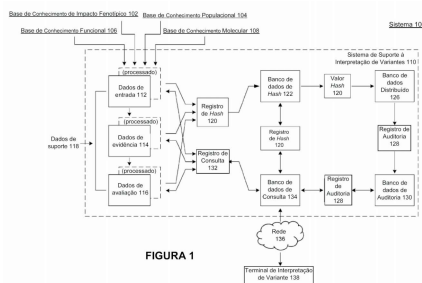
(72) **Inventor(es):** ALEXANDRE COLAVIN; CARLOS L. ARAYA; JASON A. REUTER.

(86) **Pedido PCT:** PCT US2018056304 de 17/10/2018

(87) **Publicação PCT:** WO WO/2019/079464 de 25/04/2019

(85) **Data da Fase Nacional:** 15/04/2020

(57) **Resumo:** A presente invenção refere-se a modalidades de sistema, método e produto de programa de computador para otimizar a determinação de um impacto fenotípico de uma variante molecular identificada em testes, amostras moleculares ou relatórios de sujeitos por meio de incorporação, atualização, monitoramento, validação, seleção e auditoria, executados regularmente, dos modelos de evidência com melhor desempenho para a interpretação de variantes moleculares através de uma pluralidade de classes de evidência.



Relatório Descritivo da Patente de Invenção para **"PLATAFORMA DE EVIDÊNCIA MOLECULAR PARA OTIMIZAÇÃO AUDITÁVEL E CONTÍNUA DE INTERPRETAÇÃO DE VARIANTES EM TESTES E ANÁLISES GENÉTICAS E GENÔMICAS"**.

FUNDAMENTOS

[0001] Os testes moleculares, como testes genéticos e genômicos, são ferramentas de diagnóstico cada vez mais importantes em uma ampla variedade de contextos clínicos, desde o risco de um indivíduo de convulsões neonatais, ritmo cardíaco anormal (por exemplo, arritmia) ou predisposição ao desenvolvimento de cânceres. A determinação do impacto fenotípico (por exemplo, clínico e não clínico, incluindo, entre outros, patogenicidade, funcionalidade ou efeito relativo) de uma variante molecular - como uma variante (de sequência) genotípica - identificada em um sujeito é frequentemente o pilar dos testes moleculares clínicos. A validade e a utilidade dos testes moleculares podem exigir que tais determinações (por exemplo, frequentemente conhecidas como classificações de variantes) sejam baseadas em evidências, objetivas e sistemáticas (Yandell et al. Genome Res. 2011 Set;21(9):1529-42).

[0002] Impulsionados em grande parte pelas tecnologias de sequenciamento de nova geração (NGS), rápidos avanços nas tecnologias genéticas e genômicas levaram a aumentos dramáticos no volume (por exemplo, adoção no mercado), diversidade e escopo (por exemplo, informações de sequência testadas) de testes genéticos e genômicos. Em conjunto, o número de variantes de significado desconhecido aumentou drasticamente, afetando a sensibilidade e a especificidade dos testes genéticos e genômicos clínicos.

[0003] Uma preponderância de variantes moleculares de impacto fenotípico desconhecido (por exemplo, clínico ou não clínico) é uma característica de quase todos os genes e todas as populações, inclu-

indo muitos genes clinicamente significativos. Mesmo nas condições e genes clínicos mais estudados, o conhecimento existente sobre o significado clínico de variantes moleculares geralmente permanece escasso. Por exemplo, no caso do gene *BRCA1*, um grande consórcio internacional de geneticistas clínicos, patologistas moleculares e especialistas em *BRCA1* definiram classificações para 108 variantes moleculares não sinônimas, fornecendo suporte de significância clínica para ~0,7% dos ~16.200 possíveis variantes (de sequência) genotípicas de nucleotídeo único não-sinônimas no *BRCA1* (*BRCA Exchange*). Como consequência, a grande maioria das variantes moleculares identificadas nos modernos painéis genéticos e testes genômicos não têm impacto fenotípico ou significado clínico conhecido. Por exemplo, relatórios recentes indicam que os testes modernos do painel genético do câncer hereditário podem encontrar até noventa e cinco variantes de significância desconhecida por variante causadora de doença conhecida (razão de 95:1) (Maxwell et al., 2016).

[0004] Além de sua disponibilidade limitada, o conhecimento e as classificações existentes sobre o impacto fenotípico (clínico ou não clínico) das variantes moleculares estão em constante evolução. Por exemplo, ~50% das variantes (de sequência) genotípicas de nucleotídeo não-sinônimas de *BRCA1* em um grande repositório público de classificações de significância clínico (*ClinVar*) têm classificações conflitantes, e uma definição de verdade baseada em consenso pode levar a uma instabilidade de classificação de ~11% em uma janela de 12 meses (Landrum et al., 2015). Em muitos genes e para muitas condições, o crescimento de classificações conflitantes pode ultrapassar o crescimento de novas classificações derivadas de consenso (Landrum et al., 2015). Além disso, avanços consistentes no entendimento da variação genômica, etiologia da doença e patologia e epidemiologia molecular, entre outras características, levaram a uma evolução con-

sistente do “conjunto verdade” correspondente de impactos e classificações variantes.

[0005] Durante a interpretação e revisão de variantes, um fornecedor de testes genéticos ou genômicos pode solicitar acesso a evidências em torno de uma variante, gene, condição e caso por meio de um sistema de suporte à interpretação de variantes. Devido ao alto número de variantes genéticas de significância clínica desconhecida, os fornecedores de testes genéticos e genômicos recorrem rotineiramente a uma variedade de tipos de evidências para determinar o impacto fenotípico (por exemplo, clínico ou não clínico) de variantes moleculares de efeito desconhecido identificado em sujeitos e testes. Um sistema de suporte à interpretação de variantes pode incluir uma ou mais linhas de evidência de suporte, incluindo, entre outros, dados de preditores computacionais, *hotspots* mutacionais, ensaios funcionais e métricas populacionais, entre outros. No entanto, devido ao crescimento consistente e à natureza instável das classificações de variantes - que formam a base de "conjuntos verdade" para a avaliação de modelos de evidência - as métricas de desempenho calculadas (por exemplo, diagnóstico, classificação, precisão de regressão etc.) para qualquer modelo de evidência são frequentemente desatualizados. Além disso, a dependência de uma ampla variedade de modelos de evidência desenvolvidos (por exemplo, calculados, analisados ou agregados) e avaliados em contextos distintos (por exemplo, com definições de conjuntos verdade frequentemente separados) geralmente resulta em métricas de avaliação incoerentes entre os modelos de evidência. Juntos, esses fatores complicam a avaliação e o uso de modelos de evidência como suporte para a interpretação de variantes. Como consequência, um sistema de suporte à interpretação de variantes não pode ser capaz de comparar de maneira confiável o desempenho de modelos de evidência, nos quais as avaliações são baseadas em dados di-

ferentes, dentro ou entre suas diferentes classes.

[0006] Além disso, o sistema de suporte à interpretação de variantes pode conter modelos de evidência que foram avaliados com diferentes métricas de desempenho (por exemplo, diagnóstico, classificação, precisão de regressão etc.) de interesse. Assim, o sistema de suporte à interpretação de variantes pode ser incapaz de comparar sistemática e objetivamente o desempenho dos diferentes modelos de evidência. Embora o trabalho científico e as publicações continuem gerando rotineiramente novos modelos de evidência, a falta de definições uniformes do "conjunto verdade", a falta de atualização síncrona e os vieses em sua avaliação de desempenho (por exemplo, como pode surgir de interesses de autoria), podem limitar a qualidade inerente e utilidade das evidências geradas e suas métricas de desempenho associadas. Como consequência, um sistema de suporte à interpretação de variantes não pode comparar de forma confiável o desempenho de modelos de evidência que foram avaliados com diferentes métricas de desempenho, dentro ou entre suas diferentes classes.

[0007] Além dessas questões com avaliação de evidências, o crescimento consistente e a natureza mutável das classificações existentes (por exemplo, e, portanto, conjuntos verdade) afetam a robustez dos modelos de evidência, que geralmente exigem uma etapa de aprendizagem supervisionada para geração. À medida que os conjuntos verdade evoluem continuamente, tanto a avaliação quanto a geração de evidências podem exigir atualização. Como tal, o sistema de suporte à interpretação de variantes não pode ter acesso aos modelos de evidência mais atualizados possíveis. Por exemplo, o sistema de suporte à interpretação de variantes pode conter um preditor computacional que produz uma previsão para uma variante genética que é inconsistente com o impacto fenotípico conhecido da variante, conforme aprendido após a geração do preditor.

[0008] Finalmente, o sistema de suporte à interpretação de variantes pode ser incapaz de confirmar que um modelo de evidência foi gerado em um determinado momento no tempo ou gerado com um determinado conjunto de dados. Um fornecedor de testes genéticos e genômicos que obtém evidências de suporte do sistema de suporte à interpretação de variantes pode, portanto, não conseguir garantir que as métricas de desempenho (por exemplo, diagnóstico, classificação, precisão de regressão etc.) do modelo de evidência estejam atualizadas, robustas e calculadas exclusivamente em dados separados, por exemplo, com base em variantes não usadas (ou disponíveis) na geração do modelo.

[0009] Por conseguinte, há necessidade de novos e melhorados sistemas de suporte à interpretação de variantes que superem as deficiências dos sistemas atualmente disponíveis.

BREVE SUMÁRIO

[0010] A presente divulgação fornece um método implementado por computador, o método compreendendo (i) registrar um modelo de evidência compreendendo dados de evidência, em que os dados de evidência descrevem um impacto fenotípico previsto de uma variante molecular para uma entidade alvo; (ii) avaliar dados de desempenho de validação para o modelo de evidência com base em dados de produção; (iii) gerar um valor *hash* de dados de suporte para o modelo de evidência, em que os dados de suporte compreendem os dados de evidência, e a geração do valor *hash* permite a avaliação prospectiva dos dados de evidência em resposta ao recebimento de dados de teste para o modelo de evidência; (iv) em resposta ao recebimento dos dados de teste para o modelo de evidência, avaliar os dados de desempenho do teste para o modelo de evidência com base nos dados de evidência e nos dados de teste; (v) classificar o modelo de evidência em um conjunto de modelos de evidência para a entidade alvo com

base nos dados de desempenho da validação ou nos dados de desempenho do teste; e (vi) em resposta a uma consulta para o impacto fenotípico previsto da variante molecular para a entidade alvo a partir de um terminal de interpretação de variantes, fornecer o impacto fenotípico previsto usando um modelo de evidência com melhor desempenho para a entidade alvo com base na classificação.

[0011] Em alguns aspectos, a entidade alvo compreende um elemento funcional, molécula ou variante molecular e um fenótipo de interesse.

[0012] Em alguns aspectos, o registro compreende ainda gerar o modelo de evidência com base nos dados de produção usando uma técnica de aprendizagem de máquina. Em alguns aspectos, o registro compreende ainda a importação do modelo de evidência ou dos dados de evidência. Em alguns aspectos, o método compreende ainda gerar os dados de suporte de pelo menos um dos dados de evidência, dados de produção, dados de teste, dados de desempenho de validação ou dados de desempenho de teste.

[0013] Em alguns aspectos, a geração do valor *hash* permite a avaliação do conteúdo dos dados de suporte e o tempo de criação dos dados de suporte. Em alguns aspectos, o método compreende ainda o recebimento dos dados de produção de uma base de conhecimento clínico.

[0014] Em alguns aspectos, a avaliação dos dados de desempenho de validação compreende ainda (i) calcular, usando o modelo de evidência e uma técnica de validação de modelo, uma pontuação de impacto fenotípico para a variante molecular da entidade alvo nos dados de produção; e (ii) gerar os dados de desempenho de validação com base na pontuação de impacto fenotípico usando uma métrica de desempenho de interesse.

[0015] Em alguns aspectos, a avaliação dos dados de desempe-

nho do teste compreende ainda (i) calcular, usando o modelo de evidência e uma técnica de validação de modelo, uma pontuação de impacto fenotípico para a variante molecular da entidade alvo nos dados de teste; e (ii) gerar os dados de desempenho do teste com base na pontuação de impacto fenotípico usando uma métrica de desempenho de interesse.

[0016] Em alguns aspectos, o método compreende ainda armazenar o valor *hash* dos dados de suporte em um banco de dados, em que o banco de dados associa o valor *hash* aos dados de suporte. Em alguns aspectos, o método compreende ainda a inserção do valor *hash* em uma estrutura de dados distribuídos.

[0017] Em alguns aspectos, o método compreende ainda fornecer um registro de auditoria a um terminal de interpretação de variantes, em que o registro de auditoria faz referência a uma entrada para os dados de suporte na estrutura de dados distribuídos, e o registro de auditoria permite que o terminal de interpretação de variantes examine o conteúdo dos dados de suporte e um momento da criação dos dados de suporte. Em alguns aspectos, a estrutura de dados distribuídos é uma estrutura de dados *blockchain*. Em alguns aspectos, a estrutura de dados distribuídos é um *feed* distribuído.

[0018] A presente divulgação também fornece um sistema terminal de interpretação de variantes, compreendendo: uma memória; e pelo menos um processador acoplado à memória e configurado para: enviar uma consulta de suporte a um sistema de interpretação de variantes para dados de suporte de um modelo de evidência que atenda a um conjunto de métricas de desempenho para uma entidade alvo; receber os dados de suporte e um registro de auditoria associado para os dados de suporte do sistema de interpretação de variantes; enviar uma consulta de auditoria para uma estrutura de dados distribuídos, em que a consulta de auditoria compreende o registro de auditoria pa-

ra os dados de suporte; receber um certificado de validação para o registro de auditoria do banco de dados distribuído em resposta ao envio da consulta de auditoria; e determinar um estado dos dados de suporte em um momento específico com base no registro de auditoria.

[0019] Em alguns aspectos, o pelo menos um processador está configurado para: calcular um valor *hash* dos dados de suporte para o modelo de evidência; e determinar se o valor *hash* corresponde a um valor *hash* no registro de auditoria para os dados de suporte para o modelo de evidência. Em alguns aspectos, a entidade alvo compreende um elemento funcional, molécula ou variante molecular e um fenótipo de interesse.

BREVE DESCRIÇÃO DOS DESENHOS

[0020] Os desenhos anexos estão incorporados neste documento e fazem parte do relatório descritivo.

[0021] A **FIGURA 1** é um diagrama de blocos de um sistema para fornecer um conjunto ideal de modelos de evidência para descrever ou prever os impactos fenotípicos de variantes moleculares para um ou mais elementos funcionais (ou moléculas), fenótipos, contextos ou conjunto de variantes de interesse em um determinado tempo, de acordo com alguns aspectos.

[0022] A **FIGURA 2** é um diagrama de exemplo de um sistema que fornece um conjunto ideal de modelos de evidência para descrever ou prever os impactos fenotípicos de variantes moleculares para um ou mais elementos funcionais (ou moléculas), fenótipos, contextos ou conjunto de variantes de interesse em um determinado momento, de acordo com alguns aspectos.

[0023] A **FIGURA 3** é um fluxograma que ilustra um processo que fornece um conjunto ideal de modelos de evidência para descrever ou prever os impactos fenotípicos de variantes moleculares para um ou mais elementos funcionais (ou moléculas), fenótipos, contextos

ou conjunto de variantes de interesse em um determinado momento, de acordo com alguns aspectos.

[0024] A **FIGURA 4** é um fluxograma que ilustra um processo para examinar um modelo de evidência para descrever ou prever os impactos fenotípicos de variantes moleculares para um ou mais elementos funcionais (ou moléculas), fenótipos, contextos ou conjunto de variantes de interesse em um determinado momento, de acordo com alguns aspectos.

[0025] A **FIGURA 5** é um sistema de computador exemplar útil para implementar vários aspectos das invenções descritas neste documento.

[0026] A **FIGURA 6** é uma figura para um sistema 600 para gerar uma árvore de Merkle a partir de uma lista de evidências e entradas de dados de suporte e enviá-la para um registro público com marcador temporal, de acordo com um aspecto.

[0027] Nos desenhos, números de referência semelhantes geralmente indicam elementos idênticos ou similares. Além disso, geralmente, o(s) dígito(s) mais à esquerda de um número de referência identifica(m) o desenho no qual o número de referência aparece pela primeira vez.

DESCRIÇÃO DETALHADA

[0028] São fornecidos neste documento aspectos do sistema, aparelho, dispositivo, método e/ou produto do programa de computador e/ou combinações e subconjuntos destes, para otimizar a determinação do impacto (por exemplo, clínico ou não clínico) fenotípico (por exemplo, patogenicidade, funcionalidade ou efeito relativo) de variantes moleculares identificadas em testes moleculares, amostras ou relatórios de sujeitos - como variantes (de sequência) genotípicas identificadas em testes genéticos e genômicos, amostras ou relatórios - por meio de incorporação, atualização, monitoramento, validação, seleção

e auditoria dos modelos de evidência de suporte com melhor desempenho para a interpretação de variantes moleculares em uma pluralidade de classes de evidência.

[0029] Em alguns aspectos, cada modelo de evidência pode constituir um sistema de variantes moleculares únicas e seu impacto fenotípico associado (por exemplo, clínico ou não clínico) (por exemplo, patogenicidade, funcionalidade ou efeito relativo). Como seria percebido por uma pessoa versada na técnica, os modelos de evidência podem ser derivados usando uma variedade de métodos, técnicas e fontes de dados, incluindo modelos computacionais e experimentais ou combinações destes (por exemplo, treinamento de preditores computacionais, computação). *hotspots* mutacionais, definindo limites de frequência de alelos populacionais ou medindo o impacto funcional de variantes em ensaios moleculares ou celulares). Por exemplo, pontuações ou previsões de variantes de um preditor computacional podem ser acessadas para determinar o provável impacto fenotípico (por exemplo, clínico ou não clínico) de uma variante (de sequência) genotípica de significância desconhecida. Por exemplo, ao interpretar a significância clínica de variantes genotípicas (de sequência) em *MSH2* (por exemplo, um gene que codifica uma proteína supressora de tumor estabelecida), um preditor computacional pode determinar que certas variantes (de sequência) genotípicas (ou variantes moleculares) de significância clínica desconhecida provavelmente pode aumentar o risco de vida da Síndrome de Lynch em sujeitos portadores da variante.

[0030] Em alguns aspectos, um sistema de suporte à interpretação de variantes pode integrar e utilizar muitas linhas de evidências diferentes (por exemplo, modelos de evidências) para determinar o impacto fenotípico (por exemplo, clínico ou não clínico) das variantes moleculares identificadas em testes de diagnóstico molecular, amostras ou relatórios de sujeitos. Entretanto, em alguns aspectos, um sis-

tema de suporte à interpretação de variantes pode não ter métricas de desempenho comparáveis (por exemplo, precisão bruta, precisão equilibrada - como o Coeficiente de Correlação de Matthew (MCC), taxa de verdadeiros positivos (TPR) ou sensibilidade, taxa de verdadeiros negativos (TNR) ou especificidade, valor preditivo positivo (VPP) e valor preditivo negativo (VPN)) que são específicos ao contexto diagnóstico do teste para cada um e todos os modelos de evidência. Isso ocorre porque o sistema de suporte à interpretação de variantes pode conter modelos de evidência que foram gerados ou avaliados usando diferentes conjuntos verdade em diferentes momentos ou avaliados usando diferentes métricas de desempenho.

[0031] Além disso, devido ao crescimento contínuo e às mudanças no conhecimento e na classificação dos impactos fenotípicos (por exemplo, clínicos ou não clínicos) das variantes moleculares, em alguns aspectos, o sistema de suporte à interpretação de variantes não pode conter os modelos de evidência mais atualizados possíveis e pode conter modelos de evidência que estejam em conflito com os impactos fenotípicos (por exemplo, clínicos ou não clínicos) aprendidos após a geração dos modelos de evidência. Assim, o sistema de suporte à interpretação de variantes não pode fornecer de forma confiável os modelos de evidência mais apropriados e atualizados para um fornecedor de testes genéticos e genômicos.

[0032] Portanto, o que é necessário é uma maneira de um sistema de suporte à interpretação de variantes incorporar regularmente (por exemplo, gerar ou importar), monitorar, atualizar, validar, selecionar, distribuir e examinar diferentes modelos de evidência para garantir o uso dos modelos de evidência com melhor desempenho para a interpretação de variantes moleculares no contexto de indivíduos, famílias ou populações específicas, em um determinado momento no tempo. Em alguns aspectos, o sistema de suporte à interpretação de vari-

antes pode fornecer avaliação(ões) consistente(s), sistematicamente (1) definindo conjuntos verdade que descrevem os impactos fenotípicos (por exemplo, marcadores) de variantes moleculares e (2) pontuando modelos de evidência em um conjunto de métricas de desempenho usando dados de teste dos impactos fenotípicos (por exemplo, clínicos ou não clínicos) das variantes moleculares. Como seria percebido por uma pessoa versada na técnica, os dados de teste podem se referir a dados separados, especificando os impactos fenotípicos (por exemplo, clínicos ou não clínicos) de variantes moleculares não utilizadas nos dados para geração dos modelos de evidência (por exemplo, os dados de produção). O sistema de suporte à interpretação de variantes também pode precisar fornecer uma maneira independente de verificar se um modelo de evidência fornecido foi gerado com ou sem o uso de dados específicos, (1) confirmando diretamente a presença ou ausência dos dados nos dados de produção, ou (2) inferindo a ausência de dados específicos nos dados de produção comparando os marcadores temporais da incorporação do modelo de evidência (por exemplo, geração ou importação) com os marcadores temporais conhecidos ou aceitos para a disponibilidade de dados específicos.

[0033] Em alguns aspectos - em um momento específico - a seleção objetiva de modelos de evidência pode atender a critérios de desempenho específicos para uso na interpretação de impactos fenotípicos (por exemplo, clínicos ou não clínicos) de variantes moleculares - como variantes (de sequência) genotípicas - em um ou mais elementos funcionais (por exemplo, codificadores ou não codificadores) (por exemplo, genes codificadores de proteínas, genes não codificadores, domínios moleculares como domínios de proteínas ou RNA, promotores, potenciadores, silenciadores, sítios reguladores de ligação, origens de replicação etc.) no(s) genoma(s) (por exemplo, nuclear, mitocondrial, etc.) ou em suas moléculas derivadas. Como seria percebido

por uma pessoa versada na técnica, uma variante (de sequência) genotípica pode ser uma variante de nucleotídeo único (SNV), uma variante de número de cópias (CNV) ou uma inserção ou deleção que afeta uma sequência codificadora ou não codificadora (ou ambas) no genoma. Como seria percebido por uma pessoa versada na técnica, uma variante molecular pode ser uma substituição de um único aminoácido em uma molécula de proteína, uma substituição de um único nucleotídeo em uma molécula de RNA, uma substituição de um único nucleotídeo em uma molécula de DNA, ou qualquer outra alteração molecular na sequência cognata de um polipeptídeo biológico. Como seria percebido por uma pessoa versada na técnica, um fenótipo pode ser uma ou mais características observáveis clínicas ou não clínicas e pode ser avaliado no contexto de populações específicas, grupos etários, gêneros, tecidos ou tipos de mutação (por exemplo, somática, herdada por linhagem germinativa, *de novo* de linhagem germinativa). Especificamente, em alguns aspectos, um sistema de suporte à interpretação de variantes pode incorporar regularmente (por exemplo, gerar ou importar) ou atualizar modelos de evidência para a interpretação de variantes moleculares em elementos funcionais (por exemplo, codificadores ou não codificadores) em genomas ou moléculas derivadas no contexto de fenótipos específicos ou coleções de fenótipos. Em alguns aspectos, o sistema de suporte à interpretação de variantes pode avaliar regularmente o desempenho comparativo de modelos de evidência contra novos dados (por exemplo, novos ou alterados) dos impactos fenotípicos (por exemplo, clínicos ou não clínicos) de variantes moleculares. Em alguns aspectos, o sistema de suporte à interpretação de variantes pode garantir com segurança que a seleção objetiva dos modelos de evidência atenda aos critérios de desempenho específicos para um ou mais elementos funcionais e fenótipos (por exemplo, codificadores ou não codificadores) no momento da interpretação da vari-

ante.

[0034] A **FIGURA 1** é um diagrama de blocos de um sistema 100 para fornecer um conjunto ideal de modelos de evidência 114 para descrever ou prever os impactos fenotípicos de variantes moleculares para um ou mais elementos funcionais (ou moléculas), fenótipos, contextos ou conjunto de variantes de interesse em um determinado tempo, de acordo com alguns aspectos. O sistema 100 inclui dados de várias bases de conhecimento (por exemplo, base de conhecimento de impacto fenotípico 102, base de conhecimento da população 104, base de conhecimento funcional 106 ou base de conhecimento molecular 108), um sistema de suporte à interpretação de variantes 110, uma rede 136 e terminal de interpretação de variantes 138. Como seria percebido por uma pessoa versada na técnica, alguns aspectos do sistema 100 podem fornecer uma multiplicidade de modelos de evidência para um ou mais elementos funcionais e fenótipos, bem como modelos de evidência para contextos específicos. Por exemplo, em alguns aspectos, o sistema 100 pode usar um ou mais modelos de evidência de diversas classes (por exemplo, preditores computacionais, *hotspots* mutacionais, ensaios funcionais, simulações biofísicas, limites de frequência de alelos populacionais ou outros).

[0035] Em alguns aspectos, uma base de conhecimento de impacto fenotípico 102 (por exemplo, clínica ou não clínica) inclui um ou mais bancos de dados de informações de variantes moleculares. Em alguns aspectos, um banco de dados de informações de variantes moleculares pode incluir informações de variantes moleculares e seus fenótipos associados ou impactos fenotípicos. As associações de impacto fenotípico e fenotípico de variantes moleculares podem ser derivadas da observação de variantes moleculares em indivíduos, famílias e populações afetadas e não afetadas ou em modelos experimentais representativos. Por exemplo, o teste clínico pode estabelecer que

uma variante molecular é patogênica ou benigna com base em uma taxa de observação estatisticamente significativa em indivíduos afetados ou não afetados, respectivamente.

[0036] Em alguns aspectos, as bases de conhecimento (por exemplo, base de conhecimento de impacto fenotípico 102) usadas para gerar dados de entrada 112 podem ser bancos de dados públicos, nos quais as informações são abertas ao público. Em alguns aspectos, uma base de conhecimento pode ser um banco de dados privado (por exemplo, proprietário), no qual as informações são acessíveis apenas à empresa ou entidade que criou o banco de dados ou àquelas com permissão para acessar o banco de dados.

[0037] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 inclui um banco de dados de dados de entrada 112, um banco de dados de dados de evidência 114, um banco de dados de dados de avaliação 116, um banco de dados de registros *hash* 120 (por exemplo, um banco de dados de *hash* 122) e um banco de dados de registros de auditoria 128 (por exemplo, um banco de dados de auditoria 130). Em alguns aspectos, os dados de entrada 112 (por exemplo, dados de variantes, resíduos (por exemplo, posições) e moléculas de elementos funcionais (por exemplo, codificadores ou não codificadores) no genoma) são importados para o sistema de suporte à interpretação de variantes 110. Os dados de entrada 112 podem ser usados diretamente como dados de evidência 114, ou podem ser utilizados na geração de dados de evidência 114 (por exemplo, modelos de evidência). Os dados de evidência 114 podem descrever um conjunto de variantes moleculares e seu impacto fenotípico associado (por exemplo, clínico ou não clínico) (por exemplo, patogenicidade, funcionalidade ou efeito relativo). Em alguns aspectos, o desempenho (por exemplo, precisão) dos dados de evidência 114 (por exemplo, direto ou processado) pode ser registrado e armazenado em dados de avali-

ação 116. Em alguns aspectos, o valor *hash* 124 dos dados de entrada individuais brutos ou processados 112, dados de evidência 114 ou dados de avaliação 116 (ou combinações dos mesmos), denominados coletivamente os dados de suporte 118 para um modelo de evidência, são computados e armazenados em um registro *hash* 120 em um banco de dados de *hash* 122, permitindo a associação única e a verificação de identidade dos valores *hash* 124 com dados de entrada brutos ou processados 112, dados de evidência 114 ou dados de avaliação 116. Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode gerar um registro de auditoria 128 armazenando o valor *hash* 124 dos dados de suporte 118 de um modelo de evidência (ou conjunto de modelos de evidência) em um banco de dados distribuído 126 (por exemplo, um *blockchain*, um *feed* público ou privado (por exemplo, *feed* do Twitter®) ou várias outras estruturas de dados que seriam percebidas por uma pessoa versada na técnica) acessível pelo terminal de interpretação de variantes 138. Em alguns aspectos, os registros de auditoria 128 são armazenados em um banco de dados de auditoria 130 que associa valores *hash* 124 e registros de auditoria 128. Em alguns aspectos, um banco de dados de consulta 134 é acessado através da rede 136 e fornece acesso às informações do registro de consulta 132 a partir de dados de suporte 118 em resposta a solicitações de variantes moleculares, elementos funcionais (ou moléculas), fenótipos, contextos, etc. Em alguns aspectos, o banco de dados de consulta 134 fornece os registros de *hash* associados 120 e as informações dos registros de auditoria 128 com os dados de suporte solicitados 118. Em alguns aspectos, um banco de dados de consulta 134 é acessível através de uma interface de programa de aplicativo (API). Esse recurso do sistema de suporte à interpretação de variantes 110 pode permitir que o terminal de interpretação de variantes 138 examine a disponibilidade, data de criação e conteúdo dos dados de

suporte 118 - por exemplo, dados de entrada 112, dados de evidência 114 (por exemplo, modelos de evidência) ou dados de avaliação 116 - relacionados a qualquer modelo de evidência que descreva os impactos fenotípicos de variantes moleculares para um ou mais elementos funcionais (ou moléculas), fenótipos, contextos ou conjunto de variantes de interesse em um determinado momento. Como seria percebido por uma pessoa versada na técnica, os dados de suporte 118 podem se referir a quaisquer dados de entrada 112, dados de evidência 114 ou dados de avaliação 116 ou seus derivados.

[0038] Em alguns aspectos, após incorporar (por exemplo, gerar ou importar), monitorar, atualizar e validar modelos de evidência, o sistema de suporte à interpretação de variantes 110 pode seguir um processo para selecionar e distribuir suporte à interpretação de variantes de modelos de evidência que garantam desempenho, precisão, e confiabilidade dos dados de suporte 118 fornecidos a um terminal de interpretação de variantes 138. Primeiro, esse processo pode garantir uma avaliação comparativa confiável de diferentes modelos de evidência. Segundo, esse processo pode reduzir o tempo substancial envolvido para o sistema revisar, selecionar e fornecer o(s) modelo(s) de evidência com melhor desempenho. Terceiro, esse processo pode permitir a validação independente e automatizada dos dados usados para o(s) modelo(s) de evidência fornecido(s). Finalmente, esse processo pode garantir que o(s) modelo(s) de evidência com os critérios de desempenho desejados sejam selecionados para uso na interpretação de variantes no terminal de interpretação de variantes 138, no momento da solicitação. Assim, esse processo oferece uma solução tecnológica aprimorada para a prática convencional da indústria de seleção e uso de modelos de evidências, que geralmente depende da agregação de modelos de evidências de fontes diferentes, geradas a partir de dados diferentes e avaliadas em relação a dados diferentes

ou usando métricas de desempenho diferentes. Além disso, esse sistema fornece um processo alternativo para classificar e selecionar automaticamente o modelo de evidência com melhor desempenho para elementos funcionais e fenótipos específicos.

[0039] Essa solução tecnológica aprimorada está necessariamente enraizada na tecnologia de incorporar (por exemplo, gerar e importar), avaliar, examinar e distribuir modelos de evidência, como preditores computacionais, para a interpretação (por exemplo, classificação) de variantes moleculares. Especificamente, o sistema de suporte à interpretação de variantes 110 pode seguir uma série de etapas imediatamente após a incorporação de um modelo de evidência. Para os modelos de evidência gerados e importados, essas etapas podem incluir o cálculo de uma série de métricas de controle de desempenho e qualidade (por exemplo, os dados de avaliação 116), gerar registros de *hash* 120 para dados de suporte 118 e gerar um registro de auditoria 128 dos dados de suporte 118 em um banco de dados distribuído 126 para fins de auditoria futura. Essas etapas podem estabelecer uma base de referência para a avaliação comparativa do desempenho de (por exemplo, diversos) modelos de evidência em função da crescente e variável base de conhecimento de impactos fenotípicos (por exemplo, clínicos ou não clínicos) para variantes moleculares de impacto previamente indeterminado (por exemplo, significado clínico). Em outras palavras, essas etapas podem permitir que o sistema de suporte à interpretação de variantes 110 avalie o desempenho de modelos de evidência (e dados de suporte associados 118 e métodos para sua geração), tendo em vista apenas os impactos fenotípicos (por exemplo, clínicos ou não clínicos) de variantes moleculares com novas associações.

[0040] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode utilizar ou acionar módulos independentes (por

exemplo, pré-programados) para importar ou processar diretamente dados de entrada 112. Os dados de entrada 112 podem se referir a dados de variantes, resíduos (por exemplo, posição) e moléculas de elementos funcionais (por exemplo, codificadores ou não codificadores) no genoma. Os dados variantes podem especificar os impactos fenotípicos (por exemplo, clínicos ou não clínicos) das variantes moleculares e podem ser derivados de observações clínicas ou não clínicas nos indivíduos, famílias e populações afetadas e não afetadas ou pontuações de variantes derivadas de preditores computacionais, modelos ou simulações, pontuações de variantes derivadas de ensaios e medições funcionais e pontuações de variantes derivadas de frequências de alelos da população. Os dados de resíduos podem incluir dados que descrevem propriedades evolutivas e relações de, entre e dentre resíduos em elementos funcionais, propriedades fisicoquímicas e relações de, entre e dentre resíduos em elementos funcionais, propriedades funcionais e relações de, entre e dentre resíduos em elementos funcionais, propriedades estruturais e relações de, entre e dentre resíduos em elementos funcionais e propriedades dinâmicas e relações de, entre e dentre resíduos em elementos funcionais. Os dados moleculares podem incluir dados que descrevem informações funcionais, evolutivas, estruturais e dinâmicas dos elementos funcionais. Um exemplo de dados de entrada 112 são dados (ou bancos de dados) de frequências de alelos ou variantes observados na população geral ou em populações específicas (por exemplo, dados do Exome Aggregation Consortium (Lek et al., 2016)). Um exemplo de dados derivados dos dados de entrada 112 é uma lista de variantes (de sequência) genotípicas que provavelmente são benignas devido a sua alta frequência na população geral ou em populações específicas. Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 também pode armazenar um identificador de dados de entrada que identifica

exclusivamente dados de entrada 112 (por exemplo, dados brutos ou processados).

[0041] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode acionar módulos independentes (por exemplo, pré-programados) para importar ou gerar diretamente dados de evidência 114 (por exemplo, previsões dos impactos fenotípicos de variantes moleculares) dos dados de entrada 112, conforme gerado por modelos de evidência como preditores computacionais desenvolvidos usando métodos de aprendizagem de máquina. Em alguns aspectos, os dados de evidência 114 podem indicar as previsões específicas dos impactos fenotípicos de variantes moleculares. Em alguns outros aspectos, os dados de evidência 114 podem se referir a objetos, algoritmos e funções que produzem previsões dos impactos fenotípicos de variantes moleculares.

[0042] Em alguns aspectos, um modelo de evidência pode ser gerado (por exemplo, treinado) para prever os impactos fenotípicos (por exemplo, clínicos ou não clínicos) de variantes moleculares usando uma diversidade de métodos e técnicas de aprendizagem de máquina. Em alguns aspectos, um modelo de evidência (por exemplo, um preditor computacional) para um determinado elemento ou molécula funcional pode ser específico para um fenótipo e/ou contexto. Em alguns aspectos, um modelo de evidência (ou dados de evidência 114 do modelo de evidência) pode ser específico para um determinado domínio funcional, subconjunto específico de resíduos ou subconjunto específico de variantes moleculares de um elemento funcional, como o conjunto de variantes (de sequência) genotípicas de nucleotídeo único não-sinônimas (isto é, mutações missense acessíveis ao SNV) em um domínio proteico específico de um gene que codifica a proteína. Em alguns outros aspectos, um modelo de evidência pode ser específico para um grupo de elementos funcionais relacionados, como um con-

junto de proteínas de estrutura e função homólogas. Um exemplo de dados de evidência 114 (por exemplo, dados brutos) gerados por um modelo de evidência é uma tabela das probabilidades da patogenicidade de todas as possíveis variantes (de sequência) genótípicas de nucleotídeo único não-sinônimas em um gene codificador de proteína para um fenótipo clínico específico. Um exemplo de dados de evidência 114 (por exemplo, processados) derivados de um modelo de evidência é uma tabela das classificações patogênicas ou benignas previstas de 50% das previsões de maior confiança do modelo de evidência. Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 também pode armazenar um identificador de modelo de evidência que identifica exclusivamente um modelo de evidência e seus dados de entrada associados 112, dados de evidência 114 e dados de avaliação 116.

[0043] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 armazena metadados de desempenho e controle de qualidade (por exemplo, os dados de avaliação 116) relacionados a um modelo de evidência em um banco de dados de avaliação. Por exemplo, o sistema de suporte à interpretação de variantes 110 pode computar e/ou armazenar em dados de avaliação 116 dados de desempenho de validação correspondentes a conjuntos uniformes de métricas de desempenho (por exemplo, diagnóstico, classificação, precisão de regressão, etc.) computados usando dados de produção. Em alguns aspectos, o sistema de interpretação de variantes 110 pode alavancar um esquema de validação cruzada para calcular métricas de desempenho usando conjuntos separados de variantes moleculares disponíveis nos dados de produção, mas retidos na geração de modelos de evidência durante os dados de treinamento. Da mesma forma, o sistema de suporte à interpretação de variantes 110 pode computar e/ou armazenar dados de avaliação 116 na forma de dados de de-

sempenho de teste correspondentes a conjuntos uniformes de métricas de precisão de diagnóstico para dados de teste (por exemplo, variantes moleculares separadas indisponíveis nos dados de produção) em um momento posterior. Como seria percebido por uma pessoa versada na técnica, a avaliação das métricas de desempenho calculadas entre previsões de impacto fenotípico de modelos de evidência e os impactos fenotípicos determinados (ou disponibilizados) em um momento após a geração do modelo de evidência podem permitir uma avaliação prospectiva robusta do desempenho de diversos modelos de evidências sob definições sistemáticas de conjuntos de verdade e métricas de desempenho. Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 também pode armazenar um identificador de dados de avaliação que identifica exclusivamente dados de avaliação 116 brutos ou processados.

[0044] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode avaliar os dados de desempenho de validação de um modelo de evidência. Como seria percebido por uma pessoa versada na técnica, o sistema de suporte à interpretação de variantes 110 pode avaliar os dados de desempenho de validação do modelo de evidência, a fim de fornecer uma estimativa imparcial do desempenho preditivo (por exemplo, precisão) do modelo de evidência para a interpretação dos impactos fenotípicos (por exemplo, clínicos ou não clínicos) das variantes moleculares em um determinado momento. Isso pode superar o problema de um fornecedor de testes genéticos não poder avaliar o desempenho preditivo de um modelo de evidência específico, como um preditor computacional, devido ao crescimento contínuo e à mudança na base de conhecimento de impactos fenotípicos para variantes genéticas.

[0045] Como seria percebido por uma pessoa versada na técnica, o sistema de suporte à interpretação de variantes 110 pode avaliar

os dados de desempenho de validação do modelo de evidência usando várias técnicas de validação de modelo, incluindo, por exemplo, diversas técnicas que são padrão nos campos de aprendizagem de máquina e ciência de dados. Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode aplicar um esquema de treinamento/validação de validação cruzada (por exemplo, estimativa de rotação) usando os dados de produção do modelo de evidência como uma técnica de validação de modelo para avaliar como os dados de desempenho da validação de uma análise estatística (por exemplo, computados em subconjuntos dos dados de produção) generalizarão para conjuntos independentes de variantes moleculares.

[0046] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode gerar um registro de *hash* 120, gerando um valor *hash* 124 dos dados de suporte 118 (por exemplo, dados de entrada 112, dados de evidência 114 ou dados de avaliação 116) relacionados a um modelo de evidência (ou conjunto de modelos de evidência) e armazenar o registro de *hash* 120 no banco de dados de *hash* 122. Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode criar um registro de *hash* 120 da identidade e marcadores de variantes moleculares nos dados de produção. Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode gerar um registro de *hash* 120 das pontuações de impacto fenotípico, probabilidades, previsões e/ou estimativas de confiança associadas, conforme gerado por um modelo de evidência. Em alguns aspectos, o registro de *hash* 120 pode ser armazenado em um banco de dados de *hash* 122 dentro do sistema de suporte à interpretação de variantes 110 que relaciona o valor *hash* 124 aos dados, ou combinação de dados, a partir dos quais o valor *hash* 124 foi calculado, bem como a função de *hash* usada para calcular o valor *hash* 124 dos dados. Em alguns aspectos, o sistema de suporte à interpretação de variantes 110

pode gerar um registro de auditoria 128 de qualquer dado de suporte 118 usado para gerar, monitorar ou validar uma ou mais previsões para uma ou mais variantes moleculares para permitir que o terminal de interpretação de variantes 138 examine diversas características do modelo de evidência. Como seria percebido por uma pessoa versada na técnica, o valor *hash* 124 dos dados sempre pode ser regenerado, dados os dados originais e a função de *hash*. Como seria percebido por uma pessoa versada na técnica, uma função de *hash* estatisticamente resistente a colisões pode ser usada para gerar valor *hash* 124 a partir de dados de suporte 118 que identificam exclusivamente dados de suporte 118. Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode gerar um único valor *hash* 124 a partir de uma combinação de valores *hash* para armazenamento no banco de dados de *hash* 122, tal como calculando o valor *hash* 124 como a raiz da árvore de Merkle com outros valores *hash* como folhas na árvore. Como seria percebido por uma pessoa versada na técnica, várias funções *hash* podem ser usadas para gerar o valor *hash* 124.

[0047] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode gerar um registro *hash* 120 para um conjunto de dados de suporte 118 a partir de um ou mais modelos de evidência (1) calculando o valor *hash* 124 e armazenando o registro *hash* 120 para um único objeto de dados (por exemplo, um objeto de dados compactados) contendo todos os dados de suporte ou (2) calculando o valor *hash* 124 do conjunto de valores *hash* 124 associado a um ou mais dados de suporte no banco de dados de *hash* 122.

[0048] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode gerar um registro de auditoria 128 armazenando o valor *hash* 124 dos dados de suporte do modelo de evidência 118 em um banco de dados distribuído 126 (por exemplo, um *blockchain*, um *feed* público ou privado (por exemplo, *feed* do Twitter®), ou várias

outras estruturas de dados que seriam percebidas por uma pessoa versada na técnica) para permitir que o terminal de interpretação de variantes 138 examine os dados de suporte associados ao modelo de evidência 118 (por exemplo, dados de entrada 112, dados de evidência 114 e/ou avaliação dados 116). Em alguns aspectos, o banco de dados distribuído 126 pode ser imutável. Em outros aspectos, o banco de dados distribuído 126 pode estar atrás de um firewall para impedir que a entidade que controla o sistema de suporte à interpretação de variantes 110 modifique os registros de auditoria 128. O registro de auditoria 128 pode incluir um registro tempo que representa a data e hora em que o valor *hash* 124 foi inserido no banco de dados distribuído 126. Em alguns aspectos, o registro tempo é adicionado automaticamente pelo banco de dados distribuído 126, refletindo a data e hora precisas em que o valor *hash* 124 foi armazenado. O registro de auditoria 128 também pode incluir identificadores que identificam exclusivamente o valor *hash* associado 124 dentro do banco de dados distribuído 126. Os identificadores do registro de auditoria 128 também podem identificar exclusivamente os dados correspondentes no sistema de suporte à interpretação de variantes 110. Os registros de auditoria 128 podem ser armazenados em um banco de dados de auditoria 130 dentro do sistema de suporte à interpretação de variantes 110. O valor *hash* 124 relaciona o registro de auditoria 128 no banco de dados de auditoria 130 com o registro *hash* 120 correspondente no banco de dados *hash* 122, e o registro *hash* 120 associado relaciona cada valor *hash* 124 com os dados de suporte 118 e função(ões) *hash* correspondentes ou associadas

[0049] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode inserir o valor *hash* 124 em uma estrutura de dados de *blockchain*, registrando um registro de auditoria correspondente 128 contendo todas as informações necessárias para identificar

a entrada na estrutura de dados. Como seria percebido por alguém versado na técnica, uma estrutura de dados de *blockchain* pode ser um banco de dados distribuído que mantém uma lista continuamente crescente de blocos ordenados (por exemplo, que podem ser identificados com os registros de auditoria 128). Além disso, como seria percebido por uma pessoa versada na técnica, uma estrutura de dados de *blockchain* é inerentemente resistente à modificação de seus dados. Uma vez gravados, os dados em um bloco não podem ser alterados retroativamente. Assim, um registro de auditoria baseado em *blockchain* 128 pode ser usado para confirmar a disponibilidade de dados específicos dentro do sistema de suporte à interpretação de variantes 110 em uma data e hora específicas. Em alguns aspectos, a disponibilidade de dados específicos é inferida a partir da associação exclusiva entre um valor *hash* específico e esses dados específicos.

[0050] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode inserir o valor *hash* 124 em uma estrutura de dados segura, remota, independente ou de terceiros (por exemplo, *feed* do Twitter®), registrando o registro de auditoria correspondente 128 contendo as informações necessárias para identificar a entrada (e sua data de criação). Além disso, como seria percebido por uma pessoa versada na técnica, uma estrutura de dados segura, remota, independente ou de terceiros pode ser inerentemente resistente à modificação de seus dados. Assim, um registro de auditoria 128 associado a um valor *hash* 124 armazenado em uma estrutura de dados segura, remota, independente ou de terceiros (por exemplo, *feed* do Twitter®) pode ser usado para confirmar a disponibilidade de dados específicos no sistema de suporte à interpretação de variantes 110 em uma data e hora específicas.

[0051] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode receber novos dados acerca dos impactos feno-

típicos (por exemplo, clínicos ou não clínicos) de variantes moleculares, em alguns aspectos, da base de conhecimento do impacto fenotípico 102. Esses dados podem incluir impactos fenotípicos (por exemplo, clínicos ou não clínicos) para variantes moleculares de impactos fenotípicos desconhecidos no momento da geração do modelo de evidência ou indisponíveis no momento da geração do modelo de evidência. Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode avaliar as previsões de impacto fenotípico (por exemplo, dados de evidência 114) do modelo de evidência contra novos impactos fenotípicos usando um conjunto uniforme de métricas de desempenho (por exemplo, diagnóstico, classificação, precisão de regressão, etc.) para determinar os dados de desempenho de teste do modelo de evidência ou dados de evidência associados 114. Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode registrar dados de desempenho de teste do modelo de evidência atualizando os dados de avaliação 116.

[0052] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode comparar dados de desempenho de teste e dados de desempenho de validação ou suas estimativas de dispersão associadas (por exemplo, intervalos de confiança), para determinar se um modelo de evidência, ou seus dados associados de evidência 114, atende ao desempenho esperado (ou necessário) (ou está dentro do intervalo de desempenhos esperado). Por exemplo, o sistema de suporte à interpretação de variantes 110 pode examinar se as métricas de desempenho obtidas nos dados de desempenho de teste atendem aos requisitos de desempenho esperados (ou necessários) (ou estão dentro do intervalo de desempenhos esperado) determinados nos dados de desempenho de validação e na análise associada de generalização.

[0053] Em alguns aspectos do sistema de suporte à interpretação

de variantes 110, comparações de dados de desempenho de teste e dados de desempenho de validação aplicam controles de qualidade de inversão de marcador (por exemplo, registrados) nos dados de avaliação 116 para normalizar os dados de desempenho de teste. Esse recurso permite a avaliação dos dados de desempenho do teste para levar em conta a volatilidade (por exemplo, observada ou esperada) dos marcadores nos impactos fenotípicos devido à natureza crescente e variável dos impactos fenotípicos (por exemplo, clínicos ou não clínicos) na base de conhecimento antes das comparações com os dados de desempenho da validação.

[0054] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode atualizar modelos de evidência em resposta a novos dados, bem como calcular métricas de desempenho para o conjunto separado de (por exemplo, novas) variantes moleculares. Como seria percebido por uma pessoa versada na técnica, várias técnicas de validação de modelo podem ser usadas. Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode determinar um resultado de desempenho de teste com base nas pre

ompreender ambos a precisão (por exemplo, qualidade) das previsões, bem como a cobertura (por exemplo, quantidade) das possíveis variantes moleculares em um elemento funcional (ou molécula) de interesse. Em alguns outros aspectos, o sistema de suporte à interpretação de variantes 110 pode determinar um resultado atualizado do desempenho do teste com base nas previsões de impacto fenotípico atualizadas dos modelos de evidência atualizados usando uma ou mais métricas de desempenho.

[0055] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode avaliar os dados de desempenho de validação

e testar os dados de desempenho do modelo de evidência de acordo com uma ou mais métricas de desempenho (por exemplo, diagnóstico, classificação, precisão de regressão, etc.), que podem considerar tanto a precisão (por exemplo, qualidade) das previsões, bem como a cobertura (por exemplo, quantidade) das possíveis variantes moleculares em um elemento funcional (ou molécula) de interesse. Por exemplo, em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode avaliar uma ou mais métricas de desempenho relacionadas à precisão do diagnóstico para uma ou mais previsões de dados de evidência 114. Como seria percebido por uma pessoa versada na técnica, várias métricas de desempenho podem ser usadas. Por exemplo, as métricas de diagnóstico podem incluir, mas não estão limitadas a, um ou mais dos seguintes itens:

- Precisão (bruta): a proporção de resultados verdadeiros entre o número total de casos examinados.
- Precisão Equilibrada: uma medida de verdadeiros e falsos positivos e negativos na classificação binária que pode ser usada quando a representação de classe binária está desequilibrada (são de tamanhos marcadamente diferentes), como o Coeficiente de Correlação de Matthew.
- Taxa de Verdadeiros Positivos (TPR): mede a proporção de sujeitos com a característica ou condição corretamente identificada como tal.
- Taxa de Verdadeiros Negativos (TNR): mede a proporção de sujeitos que não possuem a característica ou condição corretamente identificada como tal.
- Valor Preditivo Positivo (VPP): representa a probabilidade de ter a característica ou condição entre aqueles que dão positivo.
- Valor Preditivo Negativo (VPN): representa a probabili-

dade de não ter a característica ou condição entre aqueles que apresentam resultado negativo.

- Verdadeiro Positivo (TP): um resultado de teste que detecta a condição quando ela está presente.
- Verdadeiro Negativo (TN): um resultado de teste que não detecta a condição quando ela está ausente.
- Falso Positivo (FP): um resultado de teste que detecta a condição quando ela está ausente.
- Falso Negativo (FN): um resultado de teste que não detecta a condição quando ela está presente.
- Cobertura (CVG): Fração das possíveis variantes moleculares em um elemento funcional (ou molécula) de interesse.

[0056] Em alguns aspectos, avaliar modelos de evidência usando conjuntos uniformes de métricas de desempenho em conjuntos separados de variantes moleculares (por exemplo, dados de desempenho de validação e dados de desempenho de teste) pode superar o problema de não conseguir comparar efetivamente o desempenho preditivo dos modelos de evidência. Como seria percebido por uma pessoa versada na técnica, várias métricas de desempenho podem ser usadas, bem como conjuntos separados distintos (por exemplo, uniformes e não uniformes) de variantes moleculares. Os sistemas de suporte à interpretação de variantes existentes podem ser incapazes de avaliar o desempenho preditivo de um modelo de evidência porque a métrica de diagnóstico usada para medir o desempenho do modelo de evidência variou em toda a gama diversificada de genes e distúrbios nos testes genéticos clínicos. Além disso, os sistemas de suporte à interpretação de variantes existentes não conseguem avaliar o desempenho preditivo de um modelo de evidência, porque a métrica de diagnóstico solicitada de interesse usada durante a seleção difere da métrica de diagnóstico de interesse usada pelo criador do modelo de evidência duran-

te a avaliação. Assim, porque o sistema de suporte à interpretação de variantes 110 avalia os dados de desempenho de validação e dados de desempenho de teste para um modelo de evidência (ou dados de evidência associados 114) usando uma ou mais métricas de desempenho que são consistentes dentre as variantes moleculares que têm impactos fenotípicos conhecidos para um conjunto de consultas de elementos funcionais, fenótipos e contextos, o sistema de suporte à interpretação de variantes 110 pode fornecer dados de desempenho de validação objetivos e facilmente comparáveis e dados de desempenho de teste para os modelos de evidência e dados de evidência associados 114, a qualquer momento, ao contrário dos sistemas de suporte à interpretação de variantes existentes usados por fornecedores de testes genéticos clínicos.

[0057] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode gerar um modelo de evidência ou dados de evidência associados 114, de acordo com um modelo de aprendizagem de máquina. Um modelo de aprendizagem de máquina pode ser um programa com parâmetros ajustáveis que podem ser ajustados em resposta a dados recebidos anteriormente, a fim de melhorar o comportamento de previsão do modelo. Em alguns outros aspectos, o sistema de suporte à interpretação de variantes 110 adquire um modelo de evidência, ou dados de evidência 114, de uma fonte externa (por exemplo, um banco de dados público contendo previsões de impactos fenotípicos de variantes moleculares, conforme gerado a partir de um preditor computacional publicado).

[0058] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode gerar um modelo de evidência usando vários dados de entrada 112 (por exemplo, dados clínicos, funcionais, bioquímicos, biofísicos, evolutivos, genéticos e outros dados que seriam percebidos por uma pessoa versada na técnica). Por exemplo, o sis-

tema de suporte à interpretação de variantes 110 pode aplicar técnicas de aprendizagem de máquina não supervisionadas, semissupervisionadas e supervisionadas (ou combinações destes) para gerar (por exemplo, treinar) um modelo de evidência - dados de evidência 114 associados - como um preditor computacional, que associa dados de entrada 112 brutos e/ou processados de variantes, resíduos ou características moleculares com os dados de entrada brutos 112 e/ou processados de impactos fenotípicos marcados (por exemplo, a patogenicidade ou neutralidade de variantes genéticas de significância clínica conhecida), conforme pode ser determinado da base de conhecimento de impacto fenotípico 102. O sistema de suporte à interpretação de variantes 110 pode treinar um ou mais modelos de aprendizagem de máquina para gerar um modelo de evidência para aprender uma série de regras gerais que preveem os impactos fenotípicos (por exemplo, marcadores) de variantes moleculares (por exemplo, os impactos fenotípicos) com base das características de variantes, resíduos ou moléculas (por exemplo, características) das variantes moleculares. O sistema de suporte à interpretação de variantes 110 pode determinar essas regras gerais ajustando os parâmetros de um ou mais modelos de aprendizagem de máquina. Como seria percebido por uma pessoa versada na técnica, o modelo de evidência pode representar um ou mais modelos de evidência gerados ou importados.

[0059] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode treinar novamente (ou atualizar) o modelo de aprendizagem de máquina de um modelo de evidência, como um preditor computacional ou *hotspot* mutacional, com base em seus dados de avaliação 116 associados, como dados de desempenho de validação brutos ou processados ou teste dados de desempenho. Como seria percebido por uma pessoa versada na técnica, o sistema de suporte à interpretação de variantes 110 pode iterar através de estratégias

ou processos de avaliação de treinamento até que o modelo de evidência, por exemplo, o preditor computacional, atinja um nível limite de desempenho em seus dados de desempenho de validação, teste dados de desempenho ou ambos. Como seria percebido por uma pessoa versada na técnica, o sistema de suporte à interpretação de variantes 110 pode especificar os níveis limite de desempenho com base em uma multiplicidade de fatores, incluindo um ou mais limites para uma ou mais métricas de desempenho (por exemplo, diagnóstico, classificação, precisão de regressão etc.).

[0060] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 incorpora regularmente (por exemplo, gera ou importa), atualiza, avalia e valida modelos de evidência, como preditores computacionais. Esses novos modelos de evidência podem ser baseados e gerados em resposta a novos dados sobre os impactos fenotípicos dos dados de variantes moleculares recebidos pelo sistema de suporte à interpretação de variantes 110.

[0061] Após calcular e incorporar dados de desempenho de validação ou dados de desempenho de teste para modelos de evidência no banco de dados de avaliação, o sistema de suporte à interpretação de variantes 110 pode classificar o modelo de evidência, ou dados de evidência associados 114, entre outros modelos de evidência (ou dados de evidência 114) com base em seus dados de desempenho de validação, dados de desempenho de teste, histórico de dados de desempenho de validação (por exemplo, histórico ou simulado), histórico de dados de desempenho de teste (por exemplo, histórico ou simulado) ou combinações destes. Como seria percebido por uma pessoa versada na técnica, o sistema de suporte à interpretação de variantes 110 pode classificar o modelo de evidência, ou dados de evidência associados 114, com base em uma ou mais métricas de desempenho.

[0062] Em alguns aspectos, o sistema de suporte à interpretação

de variantes 110 pode incorporar (por exemplo, gerar ou importar), monitorar, atualizar, validar, selecionar, distribuir e examinar um modelo de evidência ou seus dados de suporte associados 118. Como seria percebido por uma pessoa versada na técnica, em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode executar os mesmos procedimentos (ou relacionados) para incorporar (por exemplo, gerar ou importar), monitorar, atualizar, validar, selecionar, distribuir e examinar em relação a diversas classes de modelos de evidência (por exemplo, *hotspots* mutacionais, preditores computacionais ou ensaios funcionais). O sistema de suporte à interpretação de variantes 110 pode gerar ou importar regularmente novos modelos de evidência, ou dados de evidência associados 114, para determinadas variantes moleculares, elementos funcionais (ou moléculas), fenótipos, contextos e métricas de desempenho de interesse. O sistema de suporte à interpretação de variantes 110 pode registrar e monitorar o desempenho de um modelo de evidência ou de seus dados de suporte associados 118. O sistema de suporte à interpretação de variantes 110 pode atualizar modelos de evidência em resposta a novos dados de entrada 112 ou dados de avaliação 116. O sistema de suporte à interpretação de variantes 110 também pode validar um modelo de evidência em resposta a receber dados separados (por exemplo, novos) para variantes moleculares de base de conhecimento de impacto fenotípico 102. O sistema de suporte à interpretação de variantes 110 pode selecionar suporte a partir de modelos de evidência que atendem a requisitos de desempenho específicos com base em dados de desempenho de validação ou dados de desempenho de teste ou combinações destes. O sistema de suporte à interpretação de variantes 110 pode distribuir previsões dos impactos fenotípicos de variantes moleculares a partir de modelos de evidência selecionados, como dados de evidência 114, em resposta a uma consulta do terminal de interpretação de

variantes 138. Finalmente, o sistema de suporte à interpretação de variantes 110 pode permitir a auditoria da disponibilidade, data de criação ou conteúdo dos dados de suporte 118 - incluindo dados de entrada 112 (por exemplo, dados classificados usados no treinamento), dados de evidência 114 (por exemplo, previsões de modelos de evidência de impactos fenotípicos) ou dados de avaliação 116 - de modelos de evidência selecionados em resposta a uma solicitação de auditoria de um terminal de interpretação de variantes 138. Como tal, o sistema de suporte à interpretação de variantes 110 pode distribuir e examinar dados de suporte à interpretação de variantes 118 fornecidos por um modelo de evidência e relacionados aos impactos fenotípicos de variantes moleculares para um ou mais elementos funcionais (ou moléculas), fenótipos, contextos ou conjuntos de variantes de interesse em um determinado momento, em resposta a consultas e solicitações de um terminal de interpretação de variantes 138.

[0063] Em resposta à entrada do usuário ou solicitações automatizadas, um terminal de interpretação de variantes 138 pode consultar o sistema de suporte de interpretação de variantes 110 para o suporte de interpretação de variantes de modelos de evidência e dados de evidência associados 114, que atendem aos requisitos de desempenho desejados para uma dada variante molecular - como uma variante (de sequência) genotípica definida pelo cromossomo, posição, nucleotídeo de referência e mutação ou alelo em um genoma de referência - para um determinado fenótipo ou conjunto de fenótipos de interesse e para uma dada estratégia de otimização de diagnóstico (por exemplo, que priorize a taxa de verdadeiros positivos acima da precisão equilibrada). Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 recebe consultas de um terminal de interpretação de variantes 138 solicitando suporte de interpretação de variantes para uma ou mais variantes moleculares. Em alguns aspectos, as consultas de

um terminal de interpretação de variantes 138 podem delinear os elementos funcionais, fenótipos, contexto ou métricas de desempenho de interesse. Em alguns aspectos, as consultas de um terminal de interpretação de variantes 138 podem delinear os requisitos de desempenho para o suporte à interpretação de variantes. Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 responde com as previsões de impacto fenotípico correspondentes para os dados de evidência 114 com classificação mais alta do conjunto de modelos de evidência para uma dada variante molecular, elemento funcional (ou molécula), fenótipo ou conjunto de fenótipos, e métricas de desempenho de interesse, juntamente com metadados para examinar os referidos modelos de evidência e seus dados de suporte associados 118. Em alguns aspectos, os modelos de evidência foram classificados e selecionados com base em dados de avaliação específicos 116 (por exemplo, dados de desempenho de validação ou dados de desempenho de teste) ou uma combinação destes. Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode fornecer dados de entrada associados 112 (por exemplo, dados de produção ou dados de teste), dados de evidência 114 (por exemplo, previsões de impacto fenotípico associadas), dados de avaliação 116 (por exemplo, dados de desempenho de validação ou dados de desempenho de teste) e informações de auditoria - incluindo um registro de auditoria 128 e/ou marcador temporal - para validar a disponibilidade, data de criação e conteúdos dos dados de entrada 112, dados de evidência 114 ou dados de avaliação 116 para os modelos de evidência selecionados. Como seria percebido por uma pessoa versada na técnica, uma porção ou todos esses diversos itens de dados podem ser fornecidos.

[0064] Em alguns aspectos, as informações de auditoria podem incluir uma referência ao banco de dados distribuído 126 que contém os registros de auditoria 28, juntamente com todos os identificadores

de registros de auditoria associados. Por exemplo, as informações de auditoria podem incluir uma referência a uma estrutura de dados *blockchain* que contém os registros de auditoria 28.

[0065] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode se comunicar com o terminal de interpretação de variantes 138 através de uma rede 136. A rede 136 pode ser qualquer rede ou combinação de redes, incluindo a Internet, uma rede de área local (LAN), uma rede de área ampla (WAN), uma rede sem fio, uma rede celular ou vários outros tipos de redes, como seria percebido por uma pessoa versada na técnica. Por exemplo, o terminal de interpretação de variantes 138 pode ser um terminal remoto que consulta o sistema de suporte de interpretação de variantes 110 pela rede 136 para o modelo de evidência mais preciso ou dados de evidência associados 114, para uma dada variante molecular, elemento funcional (ou molécula), fenótipo, ou contexto. Embora o aspecto de um terminal remoto seja usado doravante para ilustração e explicação, o terminal de interpretação de variantes 138 não precisa estar distante do sistema de suporte de interpretação de variantes 110, mas pode ser local para o sistema de suporte de interpretação de variantes 110, de modo que o terminal de interpretação de variantes 138 se comunique diretamente com o sistema de interpretação de variantes 110.

[0066] Em alguns aspectos, um auditor pode obter prova da data e hora de entrada dos dados de suporte 118 (por exemplo, brutos ou processados) para cada modelo de evidência, incluindo dados de entrada 112, dados de evidência 114, dados de avaliação 116. Por exemplo, um auditor pode consultar o banco de dados distribuído 126 (por exemplo, a estrutura de dados *blockchain*) com os identificadores de registro de auditoria (por exemplo, um recibo *blockchain*) correspondentes aos dados relacionados a um modelo de evidência. Em resposta, o auditor pode receber uma confirmação de que um valor

hash 124 específico correspondente aos dados de suporte 118 para auditoria e registro de auditoria 128 de interesse estava disponível em uma determinada data e hora.

[0067] Isso permite que o auditor determine (1) que um modelo de evidência foi incorporado (por exemplo, gerado) usando dados disponíveis no momento da entrada de seu registro de auditoria 128 e (2) que as previsões de impacto fenotípico geradas por um modelo de evidência, ou dados de evidência associados 114, estavam disponíveis em uma determinada data e hora.

[0068] Em resposta à entrada do usuário ou a uma solicitação automatizada, o terminal de interpretação de variantes 138 pode examinar um modelo de evidência ou seus dados de suporte associados 118, a fim de garantir qualquer um ou uma combinação dos seguintes itens: (a) que o modelo de evidência ou os dados de suporte associados 118 estavam disponíveis em uma determinada data e hora, (b) que o modelo de evidência ou dados de evidência associados 114 foram gerados (por exemplo, treinados) usando dados de entrada específicos 112 que estavam disponíveis na data e hora da criação do modelo de evidência, (c) que o modelo de evidência ou dados de evidência associados 114 foram gerados (por exemplo, treinados) sem o uso de dados de entrada específicos 112 que não estavam disponíveis na data e hora da criação do modelo de evidência, (d) que o modelo de evidência ou dados de evidência associados 114 contém previsões de impacto fenotípico específicas (que podem ter sido fornecidas ao terminal de interpretação de variantes 138) ou (e) que o modelo de evidência ou dados de evidência associados 114 atingem o desempenho esperado na base de dados de desempenho de validação ou dados de desempenho de teste relatados nos dados de avaliação 116 em conjuntos de dados separados. Isso pode fornecer confiança ao terminal de interpretação de variantes 138 de que o suporte à interpretação de

variantes fornecido é baseado em modelos de evidência ou dados de evidência associados 114 que atendem aos requisitos de desempenho especificados e que o modelo de evidência ou dados de evidência associados 114 não foram manipulados. Essa auditoria também pode fornecer aos pacientes e médicos confiança adicional de que seus resultados genéticos clínicos foram determinados usando modelos de evidência e dados de suporte 118 robustos e transparentes.

[0069] Em alguns aspectos, o terminal de interpretação de variantes 138 pode obter prova da disponibilidade, conteúdo e data e hora da criação dos dados de suporte 118 - incluindo dados de entrada 112, dados de evidência 114, dados de avaliação 116 - usados para gerar um determinado modelo de evidência (por exemplo, um preditor computacional) ou conjunto de modelos de evidência (por exemplo, uma combinação de preditores computacionais, *hotspots* mutacionais e ensaios funcionais), que podem ser fornecidos ao usuário. Por exemplo, um auditor pode instruir o terminal de interpretação de variantes 138 para examinar um preditor computacional no banco de dados distribuído 126 (por exemplo, o *blockchain*) com os identificadores do registro de auditoria 128 para dados de suporte 118 associados ao preditor computacional de interesse. Em resposta, o terminal de interpretação de variantes 138 pode receber um certificado de validação do banco de dados distribuído 126, incluindo o valor *hash* 124 dos dados de suporte 118. Em alguns aspectos, o certificado de validação pode ser um certificado de recibo fornecido por terceiros ou dados mantidos pelo sistema de suporte à interpretação de variantes 110. Por exemplo, em alguns aspectos, o certificado de validação pode ser um certificado de recebimento fornecido a partir de uma estrutura de dados *blockchain* contendo o valor *hash* 124, data e hora da criação da entrada associada ao registro de auditoria 128 dos dados de suporte 118. O terminal de interpretação de variantes 138 pode confirmar que o certificado de

validação corresponde aos dados de suporte 118 para o preditor computacional sob auditoria, confirmando que o código de validação (por exemplo, valor *hash* calculado) corresponde a (1) o valor *hash* 124 do registro de *hash* 120 de dados de suporte específicos 118 (por exemplo, os dados de evidência 114 correspondentes às previsões de impactos fenotípicos para um preditor computacional individual), (2) o valor *hash* 124 do registro de *hash* 120 de um único objeto compactado dos dados de suporte 118 (para entradas de banco de dados em massa) ou (3) o valor *hash* 124 dos registros de *hash* 120 de um conjunto de dados de suporte 118 (por exemplo, para entradas de banco de dados em massa).

[0070] A **FIGURA 2** é um diagrama exemplar do sistema 200 para fornecer o modelo de evidência de mais alto desempenho objetivamente para um gene e fenótipo codificador de proteínas, de acordo com alguns aspectos. A **FIGURA 2** é discutida com referência à **FIGURA 1**. Na **FIGURA 2**, os dados de produção 206 podem representar um conjunto de variantes moleculares com impactos fenotípicos associados (por exemplo, efeitos moleculares), como derivados de diversos dados de entrada 112, como dados clínicos 202 e dados da população 204. Na **FIGURA 2**, os dados de teste 212 podem representar um conjunto de variantes moleculares com impactos fenotípicos associados (por exemplo, efeitos moleculares), como derivados de novos dados de entrada 112, como dados clínicos 208 e dados populacionais 210, para variantes que são separadas das contidas em dados de produção 206. Como exemplo, os impactos fenotípicos indicando se as variantes moleculares associadas em um determinado gene codificador de proteína são considerados patogênicos ou benignos (por exemplo, neutros) em condições clínicas específicas.

[0071] Em alguns aspectos, após gerar ou importar dados de produção 206, o sistema de suporte à interpretação de variantes 110

utiliza um módulo de geração de evidências 214 para gerar dados de evidência 220 a partir de um modelo de evidência, como um preditor computacional, e calcular os dados de desempenho de validação associados 222 do modelo usando dados de produção 206 para um dado gene e fenótipo codificador de proteínas. Por exemplo, nos dados de evidência 220 gerados pelo módulo de geração de evidências 214, prevê-se que a variante molecular *H41R* (por exemplo, variante (de sequência) genotípica 17:43115738T>C) pertença à classe *Patogênica*, com probabilidade de ser classificada como *Patogênica* igual a $P(\text{Patogênico})$, conforme estimado no conjunto de partes da validação cruzada (iterações de treinamento/validação) nas quais a variante molecular *H41R* foi excluída no treinamento. O sistema de suporte à interpretação de variantes 110 pode ajustar os parâmetros do modelo de evidência gerando os dados de evidência 220.

[0072] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode avaliar os dados de desempenho de validação 222 de um modelo de evidência 220. Como seria percebido por uma pessoa versada na técnica, os dados de evidência 220 podem ser gerados por uma diversidade de modelos de evidência, incluindo preditores computacionais, e podem ser gerados usando uma variedade diversificada de técnicas e métodos, incluindo técnicas e métodos não supervisionados, semissupervisionados ou supervisionados de aprendizagem de máquina. Como seria percebido por uma pessoa versada na técnica, um modelo de evidência pode representar um ou mais (um conjunto de) modelos de evidência gerados usando os dados de produção 206.

[0073] Na **FIGURA 2**, o sistema de suporte à interpretação de variantes 110 pode utilizar um módulo de geração de evidências 214 que determina os dados de desempenho de validação 222 do modelo de evidência subjacente aos dados de evidências 220, usando um es-

quema de validação cruzada, como um esquema de validação e treinamento de validação cruzada *leave-one-out* (LOOCV). Por exemplo, nas estatísticas de resumo das métricas de desempenho diagnóstica da **FIGURA 2** em todo o esquema completo de validação cruzada são agregadas e fornecidas nos dados de desempenho de validação 222. Este processo de gravação estabelece uma linha de base de desempenho para dados de evidência 220.

[0074] Na **FIGURA 2**, o sistema de suporte à interpretação de variantes 110 pode utilizar um módulo de avaliação de evidência 216 que determina os dados de desempenho de teste 224 dos dados de evidência 220 com base nos dados de teste 212.

[0075] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode alavancar um ou uma combinação de módulos de geração de evidências 214 e módulos de avaliação de evidências 216 para gerar e avaliar uma pluralidade de modelos de evidência utilizando os dados de produção 206 (por exemplo, treinando múltiplos preditores computacionais e *hotspots* mutacionais).

[0076] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 irá gerar, enviar e armazenar registros *hash* 120 e registros de auditoria 128 dos dados de produção 206, dados de teste 212, dados de evidência 220, dados de desempenho de validação 222 e/ou dados de desempenho de teste 224 associados com um modelo de evidência (por exemplo, etapas 228, 230, 232, 234 e/ou 236, respectivamente).

[0077] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode gerar, enviar e armazenar registros de *hash* 120 dos dados de produção 206, dados de teste 212, dados de evidência 220, dados de desempenho de validação 222 e/ou dados de desempenho de teste 224 seguindo um processo de cálculo do valor *hash* 124 dos dados correspondentes e armazenar um registro de *hash* cor-

respondente 120 em um banco de dados de *hash* 122, como mostrado na **FIGURA 1**.

[0078] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode gerar, enviar e armazenar registros de auditoria 128 dos dados de produção 206, dados de teste 212, dados de evidência 220, dados de desempenho de validação 222 e/ou dados de desempenho de teste 224 seguindo um processo de acesso do valor *hash* 124 dos dados correspondentes no registro de *hash* 120 de um banco de dados de *hash* 122, inseri-lo em um banco de dados distribuído 126 e armazenar o registro de auditoria associado 128 da entrada em um banco de dados de auditoria 130 mostrado na **FIGURA 1**. O registro de auditoria 128 pode incluir um marcador temporal que representa quando os dados correspondentes foram estabelecidos e um identificador de registro que identifica exclusivamente a entrada no banco de dados distribuído 126.

[0079] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode calcular um valor *hash* alvo de um dado de subconjunto alvo de um ou mais dados de suporte 118, como o impacto fenotípico de uma variante molecular individual, conforme previsto por um modelo de evidência, uma forma granular de dados de evidência 220. Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode registrar o valor *hash* alvo (por exemplo, calculado a partir de dados do subconjunto alvo) em um registro de *hash* alvo no banco de dados de *hash*. Em alguns aspectos, o registro de *hash* alvo inclui informações adicionais sobre o valor *hash*, incluindo, por exemplo, um valor *hash* principal que pode ser recalculado (ou usado para validar) os registros de *hash* alvo de uma pluralidade de dados de subconjunto alvo. Por exemplo, em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode calcular o valor *hash* principal como a raiz de um conjunto de valores *hash* alvo (por exemplo,

folhas) usando uma estrutura de árvore de Merkle. Juntamente com os sistemas, métodos e aplicativos descritos e permitidos neste documento, esse processo de gravação pode garantir que haja uma maneira eficaz e objetiva de examinar a disponibilidade, a data de criação e o conteúdo de uma pluralidade de formas granulares de dados de suporte 118 associadas a um único registro de auditoria 128 derivado de um valor *hash* principal.

[0080] Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 pode classificar dados de evidência 220 entre outros dados de evidência que descrevem os impactos fenotípicos de variantes moleculares para um ou mais elementos funcionais (ou moléculas), fenótipos, contextos ou conjunto de variantes de interesse em um determinado momento. Por exemplo, o sistema de suporte à interpretação de variantes 110 pode classificar os dados de evidência 220 do preditor computacional entre outros modelos de evidência (por exemplo, preditores computacionais e ensaios funcionais) com base em seus dados de desempenho de validação 222 ou dados de desempenho de teste 224. Como seria percebido por uma pessoa versada na técnica, os dados de desempenho de validação 222 ou dados de desempenho de teste 224 para dados de evidência 220 podem ser comparados com outros resultados de desempenho com base em uma ou mais métricas de desempenho de interesse. O uso de várias métricas de desempenho pode permitir que os modelos de evidência sejam classificados sob diversas heurísticas, otimizadas para o contexto clínico de interesse. Como seria percebido por uma pessoa versada na técnica, o contexto clínico de interesse pode exigir a otimização de estratégias de diagnóstico com relação a métricas de desempenho específicas.

[0081] Em alguns aspectos, para determinar a precisão de um modelo de evidência, o sistema de suporte à interpretação de varian-

tes 110 valida o desempenho do modelo de evidência usando um ou mais conjuntos de dados de desempenho de validação 222, um ou mais conjuntos de dados de desempenho de teste 224 ou combinações de validação dados de desempenho 222 e dados de desempenho de teste 224. Por exemplo, o sistema de suporte à interpretação de variantes 110 pode validar o desempenho dos dados de evidência 220 confirmando que os dados de desempenho de teste 224 se enquadram em um intervalo de confiança específico (ou intervalo de dispersão) para uma ou mais métricas de desempenho estimadas com base nos dados de desempenho de validação 222 ou dados de desempenho de teste anteriores. Como seria percebido por uma pessoa versada na técnica, várias técnicas de validação de modelo podem ser usadas para validar a conformidade dos dados de desempenho de teste 224 em relação aos dados de desempenho de validação 222 ou dados de desempenho de teste anteriores 224.

[0082] Por exemplo, na **FIGURA 2**, a variante *C44S* é uma variante molecular descoberta como patogênica (e anotada como tal nos dados clínicos 208) em um momento após a criação dos dados de evidência 220. Embora essa variante não estivesse nos dados de produção 206 para os dados de evidência 220, os dados de evidência 220 preveem com precisão seu impacto fenotípico como patogênico.

[0083] Em alguns aspectos, após o cálculo dos dados de desempenho de validação 222 e/ou dados de desempenho de teste 224, o sistema de suporte à interpretação de variantes 110 pode classificar o modelo de evidência 220 com base em seus dados de desempenho de validação 222 e/ou dados de desempenho de teste 224, ou uma combinação destes.

[0084] Em uma consulta para evidência 226, o terminal de interpretação de variantes 138 pode solicitar um conjunto ideal de dados de evidência 220 para descrever ou prever os impactos fenotípicos de

variantes moleculares para um ou mais elementos funcionais (ou moléculas), fenótipos, contextos ou conjunto de variantes de interesse em um determinado momento. Por exemplo, a consulta pode solicitar os dados de produção 206, os dados de desempenho de validação 222, os dados de desempenho de teste 224 e o impacto fenotípico previsto para os dados de evidência 220 com a mais alta precisão equilibrada, conforme medido pelo Coeficiente de Correlação de Matthew (MCC) nos dados de desempenho de teste 224 para uma variante molecular em consideração para um fenótipo específico de interesse. Em resposta, manipulando consultas por meio de um modelo de distribuição de evidências 218, o sistema de suporte à interpretação de variantes 110 pode retornar os dados de suporte solicitados 118 (por exemplo, dados de produção 206, dados de desempenho de validação 222, dados de desempenho de teste 224 e previsão de impacto fenotípico) para os dados de evidência 220 com a mais alta precisão equilibrada para a variante molecular e fenótipo em consideração. A consulta também pode incluir as métricas de desempenho de interesse para otimização. Neste caso, o sistema de suporte à interpretação de variantes 110 pode retornar o modelo de evidência mais preciso para a(s) métrica(s) de diagnóstico particulares de interesse.

[0085] Em alguns aspectos, o módulo de distribuição de evidências 218 do sistema de suporte à interpretação de variantes 110 pode incluir registros de *hash* 120 e registros de auditoria 128 para os dados de suporte 118 solicitados para confirmar o conteúdo, disponibilidade ou data de criação dos dados de suporte fornecidos 118. Em alguns aspectos, o módulo de distribuição de evidências 218 pode acessar informações no banco de dados de *hash* 122 para identificar os registros de *hash* 120 dos dados de suporte fornecidos 118 e usa o valor *hash* 124 dos registros de *hash* 120 para recuperar os registros de auditoria 128 dos dados de suporte 118 associados. Em alguns aspec-

tos, para permitir a auditoria dos dados de suporte fornecidos 118, o módulo de distribuição de evidências 218 retorna os registros de consulta 132 com os dados de suporte desejados 118, os registros de *hash* 120 associados (por exemplo, incluindo o valor *hash* 124 e a função *hash*), e os registros de auditoria associados 128.

[0086] Em alguns aspectos, o terminal de interpretação de variantes 138 pode aplicar esses métodos para obter prova da disponibilidade, conteúdo e data e hora da criação dos dados de suporte 118. Por exemplo, um auditor pode instruir o terminal de interpretação de variantes 138 a examinar os dados de evidência 220 associados aos impactos fenotípicos fornecidos em resposta a uma consulta. Usando o registro de auditoria 128 dos dados de evidência 220 na resposta, o terminal de interpretação de variantes pode certificar a entrada no banco de dados distribuído 126 (por exemplo, o *blockchain*) recebendo um certificado de validação do banco de dados, incluindo o valor *hash* 124 armazenado em banco de dados distribuído 126. Em alguns aspectos, o certificado de validação pode ser um certificado de recibo fornecido por terceiros ou dados mantidos pelo sistema de suporte à interpretação de variantes 110. Por exemplo, em alguns aspectos, o certificado de validação pode ser um certificado de recibo fornecido a partir de uma estrutura de dados *blockchain* contendo o valor *hash* 124, data e hora da criação da entrada associada ao registro de auditoria 128 dos dados de suporte 118. O terminal de interpretação de variantes 138 pode confirmar que o certificado de validação corresponde aos dados de suporte 118 para o modelo de evidência sob auditoria, confirmando que o código de validação (por exemplo, valor *hash* calculado) corresponde a (1) o valor *hash* 124 do registro de *hash* 120 de dados de suporte específicos 118 (por exemplo, os dados de evidência 220 correspondentes às previsões de impactos fenotípicos para um preditor computacional individual), (2) o valor *hash* 124 do registro

de *hash* 120 de um único objeto compactado dos dados de suporte 118 (para entradas de banco de dados em massa) ou (3) o valor *hash* 124 dos registros de *hash* 120 de um conjunto de dados de suporte 118 (por exemplo, para entradas de banco de dados em massa).

[0087] A **FIGURA 3** é um fluxograma para método 300 para fornecer um conjunto ideal de modelos de evidência para descrever ou prever os impactos fenotípicos de variantes moleculares para um ou mais elementos funcionais (ou moléculas), fenótipos, contextos ou conjunto de variantes de interesse em um determinado momento, de acordo com alguns aspectos. O método 300 pode ser executado por lógica de processamento com hardware (por exemplo, circuitos, lógica dedicada, lógica programável, microcódigo, etc.), software (por exemplo, instruções executadas em um dispositivo de processamento) ou uma combinação destes. Deve ser percebido que nem todas as etapas podem ser necessárias para realizar a divulgação fornecida neste documento. Além disso, algumas das etapas podem ser executadas simultaneamente ou em uma ordem diferente da mostrada na **FIGURA 3**, como será entendido por uma pessoa versada na técnica.

[0088] O método 300 deve ser descrito com referência à **FIGURA 1** e **FIGURA 2**. No entanto, o método 300 não está limitado aos referidos aspectos de exemplo.

[0089] Em 302, o sistema de suporte à interpretação de variantes 110 recebe dados de entrada 112, incluindo dados clínicos 202 e dados populacionais 204, permitindo a geração (ou importação) de um modelo de evidência. Em alguns aspectos, o sistema de suporte à interpretação de variantes 110 deriva impactos fenotípicos (por exemplo, marcadores) descritos nos dados de produção 206 a partir de dados clínicos 202 e dados populacionais 204.

[0090] Em 304, o sistema de suporte à interpretação de variantes 110 gera, envia e armazena registros de *hash* 120 e registros de audi-

toria 128 para dados de produção 206. Como seria percebido por uma pessoa versada na técnica, o sistema de suporte à interpretação de variantes 110 pode gerar ou adquirir uma pluralidade de dados de entrada 112, incluindo dados de um conjunto diversificado de bases de conhecimento (102, 104, 106, 108) e de forma similar gere, envie e armazene registros de *hash* 120 e registros de auditoria 128 para esses dados.

[0091] Em 306, o sistema de suporte à interpretação de variantes 110 gera (por exemplo, treina) um preditor computacional (por exemplo, modelo de evidência) usando os impactos fenotípicos (por exemplo, marcadores) de variantes moleculares, conforme descrito nos dados de produção 206.

[0092] Em 308, o sistema de suporte à interpretação de variantes 110 gera, envia e armazena registros de *hash* 120 e registros de auditoria 128 para dados de evidência 220 gerados pelo preditor computacional (por exemplo, modelo de evidência).

[0093] Em 310, o sistema de suporte à interpretação de variantes 110 avalia os dados de desempenho de validação 222 do preditor computacional (por exemplo, modelo de evidência) conforme calculado usando um esquema de validação/treinamento de validação cruzada *leave-one-out*. Como seria percebido por uma pessoa versada na técnica, uma pluralidade de esquemas e técnicas de validação nos campos de aprendizagem de máquina e ciência de dados pode ser usada para derivar os dados de desempenho de validação 222 de preditores computacionais treinados em dados de produção 206.

[0094] Em 312, o sistema de suporte à interpretação de variantes 110 gera, envia e armazena registros de *hash* 120 e registros de auditoria 128 para dados de produção 222.

[0095] Em 314, o sistema de suporte à interpretação de variantes 110 recebe novos dados de entrada 112 (por exemplo, dados clínicos

208 e dados de população 210) e gera dados de teste 212 que descrevem os impactos fenotípicos de variantes moleculares não incluídas nos dados de produção 206.

[0096] Em 316, o sistema de suporte à interpretação de variantes 110 avalia os dados de desempenho de teste 224 do preditor computacional (por exemplo, modelo de evidência) conforme calculado com base no conjunto separado de variantes moleculares descritas nos dados de teste 212.

[0097] Em 318, o sistema de suporte à interpretação de variantes 110 gera, envia e armazena registros de *hash* 120 e registros de auditoria 128 para dados de produção 224.

[0098] Em 320, o sistema de suporte à interpretação de variantes 110 filtra, classifica e/ou seleciona um conjunto ideal de modelos de evidência, incluindo, por exemplo, o preditor computacional da etapa 306, com base em seus dados de desempenho de validação 222 e dados de desempenho de teste 224, classificando e selecionando o preditor computacional (da etapa 306) como preditor computacional com uma precisão equilibrada (por exemplo, Coeficiente de Correlação de Matthew (MCC)) nos dados de desempenho do teste 224 dentro do intervalo esperado (por exemplo, intervalo de confiança de 95%) das estimativas de precisão equilibrada dos dados de desempenho de validação 222 e a maior precisão equilibrada (MCC). Neste exemplo, o sistema de suporte à interpretação de variantes 110 pode limitar a seleção de modelos de evidência (ou dados de evidência associados 114) àqueles nos quais os dados de desempenho de teste estão dentro do intervalo esperado dos dados de desempenho de validação 222 e, subsequentemente, selecionam o modelo de evidência (ou dados de evidência 114) com o desempenho objetivamente mais alto. Como seria percebido por uma pessoa versada na técnica, o conjunto ideal de modelos de evidência pode incluir um único modelo de evidência

ou vários modelos de evidência.

[00099] Em 322, o sistema de suporte à interpretação de variantes 110 recebe uma consulta do terminal de interpretação de variantes 138 para o impacto fenotípico previsto de uma variante molecular específica de interesse que maximiza a precisão equilibrada (MCC) da interpretação.

[00100] Em 324, o sistema de suporte à interpretação de variantes 110 responde com o impacto fenotípico previsto do preditor computacional (da etapa 306).

[00101] A **FIGURA 4** é um fluxograma para um método 400 para examinar um preditor computacional para uma dada variante molecular, elemento (ou molécula) funcional, fenótipo ou contexto, de acordo com um aspecto. O método 400 pode ser executado por lógica de processamento com hardware (por exemplo, circuitos, lógica dedicada, lógica programável, microcódigo, etc.), software (por exemplo, instruções executadas em um dispositivo de processamento) ou uma combinação destes. Deve ser percebido que nem todas as etapas podem ser necessárias para realizar a divulgação fornecida neste documento. Além disso, algumas das etapas podem ser executadas simultaneamente ou em uma ordem diferente da mostrada na **FIGURA 4**, como será entendido por uma pessoa versada na técnica.

[00102] O método 400 deve ser descrito com referência à **FIGURA 1** e **FIGURA 2**. No entanto, o método 400 não está limitado aos referidos aspectos de exemplo.

[00103] Em 402, o terminal de interpretação de variantes 138 envia uma consulta ao sistema de suporte de interpretação de variantes 110 para o modelo de evidência com a maior precisão equilibrada (MCC) para um elemento funcional e fenótipo de interesse clínico. Em alguns aspectos, a consulta também é para um contexto específico de interesse.

[00104] Em 404, o terminal de interpretação de variantes 138 recebe dados de suporte 118 do modelo de evidência cujos dados 220 exibem os dados mais altos de desempenho de teste 224 precisão equilibrada (por exemplo, MCC) e cuja precisão equilibrada dos dados de desempenho de teste 224 estão dentro do intervalo esperado (por exemplo, intervalo de confiança de 95%) das estimativas de precisão equilibrada dos dados de desempenho de validação correspondentes do modelo de evidência 222, entre modelos de evidência para o elemento funcional e fenótipo de interesse. O terminal de interpretação variante 138 ainda recebe registros de auditoria 128 associados aos dados de suporte fornecidos 118 para o elemento funcional e fenótipo clínico de interesse. Os dados de suporte 118 podem incluir os dados de produção 206, os dados clínicos 202, os dados clínicos 204, os dados de desempenho de validação 222, os dados de desempenho de teste 224 e os dados de evidência 220.

[00105] Em 406, o terminal de interpretação de variantes 138 envia uma consulta ao banco de dados distribuído 126 que contém os registros de auditoria recebidos 128 e recebe um certificado de validação para o registro de auditoria 128 de cada dado de suporte 118 do banco de dados distribuído 126. O certificado de validação pode incluir o valor *hash* 124 e o marcador temporal do registro de auditoria correspondente 128 no banco de dados distribuído 126. Como seria percebido por uma pessoa versada na técnica, o certificado de validação pode ser um certificado de recibo fornecido por terceiros ou pode ser mantido por dados pelo sistema de suporte à interpretação de variantes 110.

[00106] Em 408, o terminal de interpretação de variantes 138 confirma que os dados de suporte fornecidos 118 correspondem ao registro de auditoria 128, avaliando a equivalência do código de validação (por exemplo, valor *hash* calculado dos dados de suporte 118) e o va-

lor *hash* 124 dos registros de auditoria 128 associados com dados de suporte 118. A correspondência do código de validação e o valor *hash* 124 confirmam que os dados de suporte fornecidos 118 foram gerados no ou antes do marcador temporal fornecido na etapa 406. Por exemplo, combinar o código de validação calculado a partir dos dados de evidência 220 fornecidos como dados de suporte 118 na consulta e o valor *hash* 124 dos dados de evidência 220 pode permitir que um terminal de interpretação de variante de auditoria 138 confirme que os impactos fenotípicos previstos específicos fornecidos pelos dados de evidência 220 foram gerados no ou antes do marcador temporal do registro de auditoria 128. Por exemplo, o terminal de interpretação de variantes 138 pode comparar o marcador temporal na etapa 406 com um marcador temporal de interesse para verificar se os dados de evidência 220 foram gerados antes de um marcador temporal de interesse. O terminal de interpretação de variante 138 pode ainda comparar o valor *hash* 124 na etapa 406 com o código de validação derivado dos dados de produção 206 para determinar que os dados de evidência 220 foram gerados usando os dados de produção fornecidos 206.

[00107] Vários aspectos podem ser implementados, por exemplo, usando um ou mais sistemas de computador, como o sistema de computador 500 mostrado na **FIGURA 5**. O sistema de computador 500 pode ser usado, por exemplo, para implementar o método 300 da **FIGURA 3**. Por exemplo, o sistema de computador 500 pode gerar um registro de validação para um preditor computacional treinado. O sistema de computador 500 pode ainda ser utilizado, por exemplo, para implementar o método 400 da **FIGURA 4**. Por exemplo, o sistema de computador 500 pode fornecer o preditor computacional mais preciso para um usuário em um provedor de testes genéticos, juntamente com os metadados associados ao preditor computacional mais preciso. O sistema de computador 500 pode ainda mapear uma pluralidade de

tons para um bloco de recursos com base na alocação de bloco de recursos determinada, de acordo com alguns aspectos. O sistema de computador 500 pode ser qualquer computador capaz de executar as funções descritas neste documento.

[00108] O sistema de computador 500 pode ser qualquer computador conhecido capaz de executar as funções descritas neste documento.

[00109] O sistema de computador 500 inclui um ou mais processadores (também chamados de unidades centrais de processamento, ou CPUs), como um processador 504. O processador 504 está conectado a uma infraestrutura de comunicação ou barramento 506.

[00110] Um ou mais processadores 504 podem, cada um, ser uma unidade de processamento gráfico (GPU). Em um aspecto, uma GPU é um processador que é um circuito eletrônico especializado projetado para processar aplicativos matematicamente intensivos. A GPU pode ter uma estrutura paralela eficiente para o processamento paralelo de grandes blocos de dados, como dados matematicamente intensivos comuns a aplicativos de computação gráfica, imagens, vídeos, etc.

[00111] O sistema de computador 500 também inclui dispositivos de entrada/saída de usuário 503, como monitores, teclados, dispositivos apontadores, etc., que se comunicam com a infraestrutura de comunicação 506 através da(s) interface(s) de entrada/saída de usuário 502.

[00112] O sistema de computador 500 também inclui uma memória principal ou primária 508, como memória de acesso aleatório (RAM). A memória principal 508 pode incluir um ou mais níveis de cache. A memória principal 508 armazenou nela a lógica de controle (ou seja, software de computador) e/ou dados.

[00113] O sistema de computador 500 também pode incluir um ou mais dispositivos de armazenamento secundário ou memória 510. A

memória secundária 510 pode incluir, por exemplo, uma unidade de disco rígido 512 e/ou um dispositivo ou disco de armazenamento removível 514. O disco de armazenamento removível 514 pode ser uma unidade de disquete, uma unidade de fita magnética, uma unidade de disco compacto, um dispositivo de armazenamento óptico, dispositivo de backup em fita e/ou qualquer outro dispositivo/unidade de armazenamento.

[00114] O disco de armazenamento removível 514 pode interagir com uma unidade de armazenamento removível 518. A unidade de armazenamento removível 518 inclui um dispositivo de armazenamento utilizável ou legível por computador que armazena o software do computador (lógica de controle) e/ou dados. A unidade de armazenamento removível 518 pode ser um disquete, fita magnética, disco compacto, DVD, disco de armazenamento óptico e/qualquer outro dispositivo de armazenamento de dados do computador. O disco de armazenamento removível 514 lê e/ou grava na unidade de armazenamento removível 518 de uma maneira bem conhecida.

[00115] De acordo com um aspecto exemplar, a memória secundária 510 pode incluir outros meios, instrumentos ou outras abordagens para permitir que programas de computador e/ou outras instruções e/ou dados sejam acessados pelo sistema de computador 500. Tais meios, instrumentos ou outras abordagens podem incluir, por exemplo, uma unidade de armazenamento removível 522 e uma interface 520. Exemplos da unidade de armazenamento removível 522 e da interface 520 podem incluir um cartucho de programa e uma interface de cartucho (como a encontrada em dispositivos de videogame), um chip de memória removível (como um EPROM ou PROM) e um soquete associado, um cartão de memória e Porta USB, cartão de memória e slot para cartão de memória associado e/ou qualquer outra unidade de armazenamento removível e interface associada.

[00116] O sistema de computador 500 pode incluir ainda uma interface ou rede de comunicação 524. A interface de comunicação 524 permite que o sistema de computador 500 se comunique e interaja com qualquer combinação de dispositivos remotos, redes remotas, entidades remotas, etc. (referenciados individual e coletivamente pelo número de referência 528). Por exemplo, a interface de comunicação 524 pode permitir que o sistema de computador 500 se comunique com os dispositivos remotos 528 pelo caminho de comunicações 526, que pode ser cabeado e/ou sem fio, e que pode incluir qualquer combinação de LANs, WANs, Internet, etc. Lógica de controle e/ou dados podem ser transmitidos para e do sistema de computador 500 através do caminho de comunicação 526.

[00117] Em um aspecto, um aparelho tangível ou artigo de fabricação compreendendo um meio tangível utilizável ou legível por computador com lógica de controle (software) armazenada no mesmo também é referido neste documento como um produto de programa de computador ou dispositivo de armazenamento de programa. Isso inclui, mas não está limitado a, sistema de computador 500, memória principal 508, memória secundária 510 e unidades de armazenamento removíveis 518 e 522, bem como artigos tangíveis de fabricação que incorporam qualquer combinação dos itens anteriores. Essa lógica de controle, quando executada por um ou mais dispositivos de processamento de dados (como o sistema de computador 500), faz com que esses dispositivos de processamento de dados operem conforme descrito neste documento.

[00118] Com base nos ensinamentos contidos nesta divulgação, será evidente para as pessoas versadas na(s) técnica(s) relevante(s) como produzir e usar aspectos desta divulgação usando dispositivos de processamento de dados, sistemas de computador e/ou arquiteturas de computador que não sejam mostrados na **FIGURA 5**. Em parti-

cular, aspectos podem operar com implementações de software, hardware e/ou sistema operacional diferentes dos descritos neste documento.

[00119] Deve ser percebido que a seção Descrição Detalhada, e não qualquer outra seção, se destina a ser usada para interpretar as reivindicações. Outras seções podem estabelecer um ou mais aspectos exemplares, mas não todos, como contemplados pelo(s) inventor(es) e, portanto, não se destinam a limitar esta divulgação ou as reivindicações anexas de qualquer forma.

[00120] Embora esta divulgação descreva aspectos exemplares para campos e aplicações exemplares, deve ser entendido que a divulgação não está limitada a eles. Outros aspectos e modificações são possíveis e estão dentro do escopo e espírito desta divulgação. Por exemplo, e sem limitar a generalidade deste parágrafo, os aspectos não se limitam ao software, hardware, firmware e/ou entidades ilustrados nas figuras e/ou descritos neste documento. Além disso, aspectos (descritos ou não explicitamente neste documento) têm utilidade significativa para campos e aplicações além dos exemplos descritos neste documento.

[00121] Aspectos foram descritos neste documento com a ajuda de blocos de construção funcionais que ilustram a implementação de funções especificadas e seus relacionamentos. Os limites desses blocos de construção funcionais foram aqui arbitrariamente definidos para a conveniência da descrição. Limites alternativos podem ser definidos desde que as funções e os relacionamentos especificados (ou equivalentes destes) sejam executados adequadamente. Além disso, aspectos alternativos podem executar blocos funcionais, etapas, operações, métodos etc. usando pedidos diferentes daqueles descritos neste documento.

[00122] As referências neste documento a "um (numeral) aspec-

to", "um (artigo indefinido) aspecto", "um exemplo de aspecto" ou frases semelhantes indicam que o aspecto descrito pode incluir um recurso, estrutura ou característica específica, mas todos os aspectos podem não necessariamente incluir o recurso, estrutura ou característica específica. Além disso, essas frases não estão necessariamente se referindo ao mesmo aspecto. Além disso, quando uma característica, estrutura ou característica específica é descrita em conexão com um aspecto, seria do conhecimento das pessoas versadas na(s) técnica(s) relevante(s) incorporar tal recurso, estrutura ou característica em outros aspectos, independentemente de ser ou não mencionados neste documento ou descritos explicitamente. Além disso, alguns aspectos podem ser descritos usando a expressão "acoplado" e "conectado" junto com seus derivados. Esses termos não são necessariamente concebidos como sinônimos um para o outro. Por exemplo, alguns aspectos podem ser descritos usando os termos "conectado" e/ou "acoplado" para indicar que dois ou mais elementos estão em contato físico ou elétrico direto um com o outro. O termo "acoplado", no entanto, também pode significar que dois ou mais elementos não estão em contato direto uns com os outros, mas ainda assim cooperam ou interagem uns com os outros.

[00123] A amplitude e o escopo da presente divulgação não devem ser limitados por nenhuma das modalidades exemplificativas descritas acima, mas devem ser definidos apenas de acordo com as reivindicações a seguir e seus equivalentes.

[00124] Em alguns aspectos, os métodos usados para gerar Dados de Produção 204 e Dados de Teste 212, incluindo a definição específica de conjuntos de verdade que descrevem os impactos fenotípicos podem ser gerados de várias maneiras a partir de uma ou várias bases de conhecimento 102, 104, 106 e 108. Em outros aspectos, várias definições de conjuntos de verdade podem ser definidas a partir

de uma variedade de fontes de interpretações de variantes clínicas que variam em uma ou mais propriedades, como qualidade ou escopo ou qualidade ao longo do tempo. Em outros aspectos, uma variedade de definições de conjuntos de verdade pode ser necessária em alguns contextos para gerar dados suficientes antes de continuar a gerar dados de evidência. Em outros aspectos, vários métodos apropriados para converter dados da base de conhecimento 102, 104, 106 e 108 em dados de entrada apropriados 112 podem existir. Por exemplo, a interpretação da Base de Conhecimento Populacional 104 como um conjunto de verdade pode exigir um ou mais pontos de corte de frequência populacional apropriados com base nas características do fenótipo descrito pelo conjunto de verdade, como penetrância, prevalência, idade de início ou padrão de herança.

[00125] Em outros aspectos, os Dados de Desempenho de Validação 222 podem ser avaliados usando Dados de Produção 206 derivados de vários métodos. Por exemplo, os Dados de Evidência 220 podem ser derivados usando Dados de Produção 206 derivados da inclusão de todos os Dados Clínicos 202, enquanto os Dados de Desempenho de Validação 222 podem ser derivados avaliando os Dados de Evidência 220 com Dados de Produção 206 derivados de um conjunto limitado de Dados Clínicos 202.

[00126] Em alguns aspectos, o método ou métodos específicos associados aos Dados de Evidência 202 usados para gerar dados de produção 206 e Dados de Validação 222 podem ser distintos do método ou dos métodos específicos usados para gerar Dados de Teste 212. Em alguns aspectos, os Dados de Desempenho de Teste 224 podem levar em consideração os métodos usados nos Dados de Produção 206 e Dados de Validação 222 para criar um conjunto de verdade separado, sem conhecimento incluído em zero, um ou mais métodos usados nos Dados de Produção 206. Em outros aspectos, vários

métodos usados para gerar Dados de Desempenho de Teste 224 podem ser usados. Por exemplo, os Dados de Desempenho do Teste 224 podem gerar duas ou mais pontuações, variando os Dados Clínicos 202 usados para gerar os Dados de Desempenho do Teste 224.

[00127] Em alguns aspectos, os métodos específicos usados para gerar Dados de Produção 206 e Dados de Teste 212 podem ser avaliados com base em vários métodos. Em alguns aspectos, a qualidade relativa dos métodos utilizados para gerar os Dados de Produção 206 e os Dados de Teste 212 que são gerados pode ser avaliada com base na autoconsistência dos métodos ao longo do tempo. Em outros aspectos, a qualidade relativa dos métodos usados para gerar Dados de Produção 206 pode ser avaliada com base nos Dados de Desempenho de Validação 222 e nos Dados de Desempenho de Teste 224 dos Dados de Evidência 220 gerados a partir de Dados de Produção 206 decorrentes de cada método. Em outros aspectos, a classificação relativa dos métodos usados para gerar Dados de Produção 206 e Dados de Teste 212 pode mudar ao longo do tempo. Em outros aspectos, a classificação relativa dos métodos usados para gerar Dados de Produção 206 e Dados de Teste 212 pode ser determinada através de múltiplos Dados de Evidência 220.

[00128] Em alguns aspectos, os Dados de Evidência 220 para um único fenótipo são gerados com base em uma ou mais definições específicas dos Dados de Produção 206. Da mesma forma, os Dados de Desempenho de Validação 222 para Dados de Evidência 220 podem ser avaliados em relação a uma ou mais definições dos Dados de Produção 206. Da mesma forma, os Dados de Desempenho do Teste 224 para Dados de Evidência 220 podem ser avaliados em relação a uma ou mais definições dos Dados de Teste 212. Por exemplo, os Dados de Desempenho de Teste 224 ou Dados de Desempenho de Validação 222 para Dados de Evidência 220 podem ser gerados usando to-

dos os Dados Clínicos 208 e Dados Clínicos 202 respectivamente, ou um subconjunto dos Dados Clínicos mais relevantes para um fenótipo específico.

[00129] Em alguns aspectos, os Dados de Evidência 220 podem ser calibrados após serem gerados. Em alguns aspectos, os Dados de evidência 220 podem ser ajustados para maximizar a concordância com uma definição diferente de Dados de Produção 206 ou uma versão análoga de Dados de Produção 206 gerada com Dados Clínicos 202 a partir de uma data diferente. Em outros aspectos, vários Dados de Evidência 220 podem ser agrupados em virtude de serem gerados a partir da mesma definição ou de definição semelhante de Dados de Produção 206 e calibrados juntos. Por exemplo, um ou mais Dados de Evidência 220 gerados a partir de uma definição específica de Dados de Produção 206 conhecida por superestimar as probabilidades do fenótipo Patogênico podem ser calibrados com relação a uma definição de Dados de Produção 206 que é mais concordante com as probabilidades do mundo real de um fenótipo Patogênico. Em alguns aspectos, os métodos de calibração podem aplicar uma transformação às probabilidades de patogenicidade nos Dados de Evidência 220 para manter a classificação de probabilidades e, ao mesmo tempo, fornecer uma melhor probabilidade de patogenicidade no mundo real. Em outros aspectos, os métodos de calibração podem determinar um ponto de corte de probabilidade ideal acima e abaixo do qual as variantes moleculares podem ser classificadas de maneira ideal de acordo com os requisitos de desempenho específicos estabelecidos. Como pode ser percebido por uma pessoa versada na técnica, uma variedade de técnicas de calibração pode ser usada para otimizar a concordância entre os Dados de Evidência 220 e Dados de Produção 206 ou Dados de Teste 212. Em alguns aspectos, os métodos de calibração são implementados como parte do Módulo de Geração de Evidências 214 ou

do Módulo de Avaliação de Evidências 216. Em outros aspectos, os parâmetros e métodos específicos dos métodos de calibração são armazenados como parte dos Dados de Validação 222 e Dados de Desempenho de Teste 224.

[00130] Como pode ser percebido por uma pessoa versada na técnica, o equilíbrio específico ou a qualidade dos dados utilizados para avaliar previsões podem alterar o desempenho medido das previsões. Por exemplo, o desempenho PPV e NPV de um diagnóstico pode variar drasticamente, dependendo da distribuição de verdadeiros positivos e verdadeiros negativos na população de teste. Em alguns aspectos, os Dados de Produção 206 e os Dados de Teste 212 podem ser modificados antes do cálculo dos Dados de Desempenho de Validação 222 e Dados de Desempenho de Teste 224, respectivamente, para obter um resultado específico. Em alguns aspectos, os Dados de Desempenho de Validação 222 e os Dados de Desempenho de Teste 224 podem ser calculados após casos positivos e negativos serem equilibrados. Por exemplo, um Dados de Produção 206 com 100 variantes e 10 variantes Benignas pode ser amostrado novamente para obter um número igual de variantes Benignas e Patogênicas antes do cálculo dos Dados de Desempenho de Validação 222. Em outros aspectos, as distribuições de fenótipo específicas dos Dados de Produção 206 ou Dados de Teste 212 podem ser amostradas novamente para obter paridade com a população de testes definida na Consulta para Evidência 226. Em outros aspectos, os Dados de Validação 222 e os Dados de Desempenho de Teste 224 podem ser recalculados diretamente para representar uma distribuição fenotípica específica. Em outros aspectos, os Dados de Desempenho de Teste 224 podem alterar a distribuição dos Dados de Teste 212 para obter paridade com um Dado de Produção específico 206. Como pode ser percebido por uma pessoa versada na técnica, a modificação dos Dados de Desempenho

de Validação 222 e Dados de Desempenho de Teste 224 para alterar a distribuição aparente ou a qualidade dos Dados de Produção 206 e Dados de Teste 212, respectivamente, pode ocorrer no momento da geração dos dados de desempenho, ou em resposta a uma Consulta para Evidência 226 específica. Em alguns aspectos, ao alcançar a paridade entre os Dados de Desempenho de Validação 222 e os Dados de Desempenho de Teste 224, o Módulo de Distribuição de Evidência 218 pode comparar e classificar mais diretamente as métricas sensíveis à população, como NPV e PPV.

[00131] Em alguns aspectos, os Dados de Evidência 220 podem ser otimizados para alcançar métricas de desempenho específicas nos Dados de Desempenho de Validação 222 ou Dados de Desempenho de Teste 224, restringindo os Dados de Evidência 220 a essas variantes com previsões mais confiáveis. Por exemplo, os Dados de Evidência 220 não podem atingir um limite de métrica de desempenho específico quando todas as variantes são consideradas, mas pode atingir o mesmo limite de métrica de desempenho se apenas as previsões mais confiáveis de 80% forem consideradas. Em alguns aspectos, os parâmetros e métodos de otimização, como o limite de confiança ideal, são armazenados nos Dados de Desempenho de Validação 222 e Dados de Desempenho de Teste 224.

[00132] Em alguns aspectos, os parâmetros de calibração e otimização dos Dados de Desempenho de Validação 222 e Dados de Desempenho de Teste 224 são aplicados às previsões dos Dados de Evidência 220 no Módulo de Distribuição de Evidência 218. Em algumas modalidades, o Módulo de Distribuição de Evidência fornece os Dados de Evidência 220 originais e os Dados de Evidência 220 otimizados, fornecidos como Consulta para Evidência 226.

[00133] Em alguns aspectos, várias técnicas de calibração e otimização para os Dados de Desempenho de Validação 222 e Dados de

Desempenho de Teste 224 são geradas para atingir requisitos de desempenho específicos. O Módulo de Distribuição de Evidência 218 recuperará a técnica específica de calibração e otimização que melhor atende aos requisitos da Consulta por Evidência 226.

[00134] Em alguns aspectos, vários Dados de Evidência 220 atendem aos requisitos da Consulta por Evidência 226. Em outros aspectos, várias técnicas de calibração ou otimização dos Dados de Evidência 220 atendem aos requisitos de Consulta para Evidência 226. Nesses casos, o Módulo de Distribuição de Evidência 218 deve escolher um ou mais Dados de Evidência 220 dos quais forneça uma resposta à Consulta de Evidência 226. Em alguns aspectos, os Dados de Evidência disponíveis 220 são ordenados por seus Dados de Desempenho de Validação 222 disponíveis e Dados de Desempenho de Teste 224 de acordo com os requisitos de desempenho da Consulta para Evidência 226. Em outros aspectos, o Modelo de Distribuição de Evidências 218 pode ordenar os Dados de Evidência 220 de acordo com métricas que não são específicas para nenhum Dado de Evidência 220 específico. Em outras modalidades, o Modelo de Distribuição de Evidência 218 pode combinar dois ou mais Dados de Evidência ordenados. Por exemplo, a primeira ordem do Módulo de Distribuição de Evidência 218 pode ordenar os Dados de Evidência 220 por uma métrica específica dos Dados de Desempenho de Validação 220 e, em seguida, aumentar pelo desempenho em massa dos Dados de Evidência 220 gerados a partir de Dados de Produção equivalentes 206 através de muitos Dados de Evidência 220.

[00135] Em alguns aspectos, após a classificação dos Dados de Evidência 220, o Módulo de Distribuição de Evidência 218 pode não encontrar um Dados de Evidência 220 com desempenho suficiente para atender aos requisitos da Consulta para Evidência 226. Em alguns aspectos, o Módulo de Distribuição de Evidência 218 não pode

fornecer os Dados de Evidência 220 mais bem classificados. Em alguns aspectos, se a Consulta para Evidência 226 solicitar o modelo de classificação mais alta, o Módulo de Distribuição de Evidência 218 ainda poderá fornecer os Dados de Evidência 220, mesmo que não atenda aos requisitos da Consulta para Evidência 226. Em alguns aspectos, a Consulta para Evidência 226 não pode ter limites de desempenho específicos, mas pode solicitar os Dados de Evidência 220 que atingem a classificação mais alta de acordo com uma classificação definida pela Consulta para Evidência 226 e executada pelo Módulo de Distribuição de Evidência 218.

[00136] Em alguns aspectos, os Dados de Evidência 220 podem fornecer e ser avaliados com relação a técnicas de interpretação e validação não baseadas em classificação. Por exemplo, os Dados de Evidência 220 podem prever a classificação fenotípica (por exemplo, Patogênica ou Benigna), probabilidades (por exemplo, 22% de chance de patogenicidade) ou fenótipos de dimensão superior. Por exemplo, os Dados de Desempenho de Validação 222 podem armazenar as métricas não baseadas em classificação dos Dados de Evidência 222 em relação aos Dados de Produção 206, como a Área sob a Curva de Operação do Receptor ou Coeficiente de Correlação de Spearman. Em alguns aspectos, o Módulo de Distribuição de evidências 218 classificará os Dados de Evidência 220 adequados, usando métricas não baseadas em classificação isoladamente ou ao lado de outras métricas. Em outros aspectos, o Módulo de Distribuição de Evidências 218 pode fornecer pontuações contínuas e baseadas em classificação a partir dos dados de evidências 220.

[00137] Como pode ser percebido em uma pessoa versada na técnica, algumas aplicações do Sistema de Suporte à Interpretação de Variantes podem existir no campo de classificação de uma lista de variantes em ordem de probabilidade de um ou vários fenótipos. Por

exemplo, alguns usuários do Sistema de Suporte à Interpretação de Variantes podem estar interessados em identificar a variante causal mais provável em um paciente que exibe um fenótipo específico. Em alguns aspectos, a Consulta para Evidência 226 solicitará as probabilidades relativas de fenótipo de uma lista de variantes. Nesses casos, o Módulo de Distribuição de Evidência classificará e coletará os Dados de Evidência 220 com melhor desempenho para cada variante na Consulta por Evidência 206. Em alguns aspectos, o Módulo de Distribuição de Evidência 218 retornará uma lista de todas as variantes acima de uma probabilidade específica de fenótipo. Em alguns aspectos, a probabilidade particular de fenótipo é estabelecida internamente. Em outros aspectos, a probabilidade específica de fenótipo é estabelecida de acordo com os requisitos estabelecidos na Consulta para Evidência 226. Em outros aspectos, o Módulo de Distribuição de Evidência 218 fornecerá a lista de variantes da Consulta para Evidência em ordem de classificação, de acordo com suas probabilidades relativas para um ou mais fenótipos. Em outros aspectos, o Módulo de Distribuição de Evidências incluirá as probabilidades relativas associadas a cada fenótipo de cada variante na Consulta para Evidência 226. Em outros aspectos, apenas um determinado número ou porcentagem de evidência na Consulta para Evidência é retornada pelo Módulo de Distribuição de Evidência.

[00138] Em alguns aspectos, o Sistema de Suporte à interpretação de Variantes pode rastrear a Consulta para Evidência 226 de determinadas entidades de consulta em um banco de dados distinto. Posteriormente, o Sistema de Suporte à Interpretação de Variantes pode usar a Consulta para Evidência 226 registrada como parte dos Dados de Entrada para refinar ainda mais os Dados de Produção 206 ou Dados de Teste 212. Por exemplo, os Dados Clínicos 208 podem ter sido parcialmente derivados de uma Consulta para Evidência 226 e

Dados de Evidência 220, nesse momento o sistema de Suporte à interpretação de Variante pode optar por excluir Dados Clínicos 208 específicos para evitar conclusões tautológicas nos Dados de Desempenho do Teste 224.

[00139] O Sistema 600 (**FIGURA 6**) pode ser executado por lógica de processamento com hardware (por exemplo, circuitos, lógica dedicada, lógica programável, microcódigo, etc.), software (por exemplo, instruções executadas em um dispositivo de processamento) ou uma combinação destes. Deve ser percebido que nem todas as etapas podem ser necessárias para realizar a divulgação fornecida neste documento. Além disso, algumas das etapas podem ser executadas simultaneamente ou em uma ordem diferente da mostrada na **FIGURA 6**, como será entendido por uma pessoa versada na técnica.

[00140] Os sistemas 600 e 601 devem ser descritos com referência à **FIGURA 1** e **FIGURA 2**. No entanto, o sistema 600 não está limitado aos referidos aspectos de exemplo.

[00141] Os dados de evidência e de suporte 602 podem ser agregados para um conjunto de previsões dos Dados de Evidência 220 e quaisquer dados de suporte, como Dados de Produção 206, Dados de Desempenho de Validação 222 ou qualquer outro Dado de Entrada 122. Os dados de suporte para cada variante podem ser separados em um Dado de Suporte Específico de Variante 604, incluindo informações de identificação (como variante genotípica e variante molecular). Em alguns aspectos, os Dados de Suporte Específicos da Variante 604 incluem informações específicas para as previsões do Modelo de Evidência 220 ou Dados de Desempenho de Validação 222 específicos para a variante. Em outros aspectos, os Dados de Suporte Específicos da Variante 604 incluem informações sobre os Dados de Entrada 122 ou parâmetros específicos do Modelo de Evidência 214. Os Dados de Suporte Específicos da Variante 5604 podem abranger um

Escopo de Variante Molecular predefinido 603, por exemplo: todas as previsões para variantes missense de um Dado de Evidência 220 específico.

[00142] Para cada Dado de Suporte Específico da Variante 604, uma Função *Hash* Resistente à Colisão 606 pode ser definida para converter deterministicamente os Dados de Suporte Específicos da Variante 604 em um valor *hash* 607 para cada Dado de Suporte Específico da Variante 604. Em alguns aspectos, a Função *Hash* Resistente à Colisão 606 específica pode ser definida nos Dados de Suporte Específicos da Variante 604. Em outros aspectos, um valor aleatório também pode ser adicionado aos Dados de Suporte Específicos da Variante 222, por exemplo, para reduzir a probabilidade de descriptografia indesejada.

[00143] Os Valores *Hash* 607 formam os Nós Folhas da Árvore de Merkle 610 da Arquitetura da Árvore de Merkle 608. Os Nós Folhas da Árvore de Merkle podem ser ordenados de maneira reproduzível e predefinida. Pares de Nós Folhas da Árvore de Merkle 610 podem ser concatenados e o valor pode ser transformado em *hash* nos Nós Não-Folhas de Árvore de Merkle 612 por uma Função de *Hash* Resistente à Colisão. O processo pode ser repetido até que uma única Raiz de Árvore de Merkle 614 seja calculada. A Raiz da Árvore de Merkle 614 pode então ser submetida a um Registro Público Distribuído com Marcador temporal 618.

[00144] Como seria percebido por uma pessoa versada na técnica, o número de entradas individuais de Dados de Suporte Específicos de Variante 604 que podem ser incluídas em uma Arquitetura da Árvore de Merkle 608 pode ser restringido pela Profundidade da Árvore de Merkle 616. Por exemplo, uma Profundidade de Árvore de Merkle 616 de 5 permitiria apenas 32 Nós Folhas da Árvore de Merkle. Portanto, ao restringir a Profundidade da Árvore de Merkle 616, a Arquitetura da

Árvore de Merkle 608 naturalmente restringe o número de Dados de Suporte Específicos de Variantes 604 que podem ser atribuídos a uma única Raiz da Árvore de Merkle 614 em um Registro Público Distribuído com Marcador temporal 618. Como resultado, uma Arquitetura de Árvore de Merkle com uma Profundidade de Árvore de Merkle fixa e divulgada 616 limita passivamente os maus atores, por exemplo, de tentar armazenar todas as combinações e/ou enumerações dos Dados de Suporte Específicos Variantes 604. Por exemplo, um ator ruim poderia armazenar uma entrada de Dados de Suporte Específicos da Variante 604 alegando que a Variante Molecular P871L é Patogênica e outra entrada de Dados de Suporte Específicos da Variante 604 alegando que a mesma variante molecular é Benigna. Uma Profundidade da Árvore de Merkle limitada para um Escopo de Variante 603 definido específico pode impedir que um ator ruim atribua todas as previsões possíveis para variantes específicas ao registro público sob uma única Raiz da Árvore de Merkle 614. Como seria percebido por uma pessoa versada na técnica, o número de variantes possíveis para um Escopo de Variante 603 poderia ser determinado por terceiros, que poderiam confirmar que a Profundidade da Árvore de Merkle não excede o Escopo de Variante 603 definido.

[00145] Sistema 601 para fornecer uma prova de segurança de *hash* demonstrando a existência de Dados de Suporte Específicos de Variante 620 em um registro de data/hora específico em um registro público, de acordo com um aspecto. O Sistema 601 pode ser executado por lógica de processamento com hardware (por exemplo, circuitos, lógica dedicada, lógica programável, microcódigo, etc.), software (por exemplo, instruções executadas em um dispositivo de processamento) ou uma combinação destes. Deve ser percebido que nem todas as etapas podem ser necessárias para realizar a divulgação fornecida neste documento. Além disso, algumas das etapas podem ser execu-

tadas simultaneamente ou em uma ordem diferente da mostrada na **FIGURA 6**, como será entendido por uma pessoa versada na técnica.

[00146] No Sistema 601, o marcador temporal associado ao conhecimento incorporado nos Dados de Suporte Específicos de Variante 620 (que podem ser uma entrada específica dos Dados de Suporte Específicos de Variante 604), pode ser demonstrado com uma pequena fração das informações usadas para gerar a Arquitetura da Árvore de Merkle 608 original. Por exemplo, o Sistema 601 fornece uma Prova de Árvore de Merkle 624 para provar a associação de Dados de Suporte Específicos de Variante 620 ilustrando matematicamente como transformar Dados de Suporte Específicos de Variante 620 usando uma Função de *Hash* Resistente à Colisão 622 e combinando o *hash* resultante com outros *hash* para recuperar a Raiz da Árvore de Merkle 614, armazenada no registro público com um registro de data/hora. As informações retornadas como prova para uma Consulta para Prova de Auditoria 624 podem ser compostas de quatro partes - a Raiz da Árvore de Merkle 616, os Nós Não-Folha da Árvore de Merkle 612 adjacentes ao caminho que leva aos Nós Folha da Árvore de Merkle 610 associados aos Dados de Suporte Específicos da Variante 620, a Função *Hash* Resistente à Colisão 622 usada para converter os Dados de Suporte Específicos da Variante 620 e, finalmente, o conteúdo dos Dados de Suporte Específicos da Variante 620. Com essas quatro partes, uma pessoa versada na técnica pode gerar rapidamente uma prova matemática de que os Dados de Suporte Específicos da Variante 620 estão criptograficamente associados à Raiz da Árvore de Merkle 614 publicada no Registro Público Distribuído com Marcador Temporal 618. Portanto, um auditor que faz uma Consulta para Prova de Auditoria 624 ao Sistema 601 pode determinar rapidamente que um Dado de Suporte Específico da Variante 620 para uma ou mais variantes foi gerado no ou antes do marcador temporal associado à Raiz da Árvore de

Merkle 614.

[00147] As descrições ou previsões de impactos fenotípicos de variantes moleculares para um ou mais elementos funcionais (ou moléculas), fenótipos, contextos ou conjunto de variantes de interesse em um determinado momento, obtidos usando os métodos e sistemas divulgados neste documento (por exemplo, os métodos de interpretação de variantes, sistemas de suporte à interpretação de variantes e sistemas terminais de interpretação de variantes da presente divulgação) podem ser utilizados, por exemplo, como parte de diagnósticos ou tratamentos. Os sistemas, aparelhos, dispositivos, métodos e/ou produtos do programa de computador fornecidos neste documento e/ou combinações e subcombinações destes, podem ser usados para otimizar a determinação do impacto fenotípico (por exemplo, clínico ou não clínico) (por exemplo, patogenicidade, funcionalidade ou efeito relativo) de variantes moleculares identificadas em testes moleculares, amostras ou relatórios de sujeitos - como variantes (de sequência) genotípicas identificadas em testes, amostras ou relatórios genéticos e genômicos - por meio de incorporação, atualização, monitoramento, validação, seleção e auditoria dos modelos de evidência de suporte com melhor desempenho para a interpretação de variantes moleculares em uma pluralidade de classes de evidência. Essas informações podem ser usadas posteriormente, por exemplo, para decidir se um paciente deve ser tratado, cessar o tratamento de um paciente, selecionar um paciente para tratamento, prever o prognóstico de um paciente, selecionar um determinado agente terapêutico etc.

[00148] Deve-se entender que os métodos divulgados abaixo não se limitam ao tratamento clínico e podem estar relacionados a decisões de estilo de vida. Por exemplo, em resposta a uma determinação sobre um potencial impacto fenotípico, o sujeito pode efetuar alterações na dieta ou no estilo de vida.

[00149] Deve-se entender que os métodos divulgados abaixo não se limitam à interpretação de variantes únicas em indivíduos únicos e podem estar relacionados a múltiplas variantes em um ou mais indivíduos. Por exemplo, classificando o impacto fenotípico relativo de variantes em uma coorte de pacientes em um ensaio clínico.

[00150] Deve-se entender que os métodos divulgados abaixo não se limitam à interpretação de variantes em indivíduos vivos. Por exemplo, uma interpretação *post-mortem* das variantes de um indivíduo pode informar o risco de herdabilidade para seus parentes.

[00151] Como utilizado neste documento, os termos "tratar", "tratamento" ou "tratamento de" se referem a reduzir o potencial de uma doença, distúrbio ou fenótipo, reduzindo a ocorrência de uma doença, distúrbio ou fenótipo e/ou uma redução na gravidade da doença ou distúrbio, de preferência, na medida em que o sujeito não sofra mais desconforto e/ou função alterada devido a ele. Por exemplo, o tratamento pode se referir à capacidade de uma terapia quando administrada a um sujeito, para impedir que uma doença ou distúrbio ocorra e/ou curar ou aliviar sintomas, sinais ou causas de uma doença. O tratamento também se refere à mitigação ou diminuição de pelo menos um sintoma clínico e/ou inibição ou retardo na progressão da condição e/ou prevenção ou atraso no aparecimento de uma doença ou enfermidade. O tratamento também pode se referir à mitigação ou tratamento dos efeitos indiretos de uma doença, distúrbio ou fenótipo, como por exemplo, informar decisões de planejamento familiar. Assim, os termos "tratar", "tratamento" ou "tratamento de" (ou termos gramaticalmente equivalentes) referem-se a regimes de tratamento profilático e terapêutico.

[00152] Os métodos e sistemas da presente divulgação podem fornecer um benefício no diagnóstico e/ou tratamento de uma doença, distúrbio ou fenótipo. Um benefício não é necessariamente uma cura

para uma doença ou distúrbio específico, mas engloba um resultado que geralmente inclui o alívio da doença, distúrbio ou fenótipo, ou aumento da sobrevida, eliminação da doença ou distúrbio, redução de um sintoma associado à doença ou distúrbio, prevenção ou alívio de uma doença secundária, distúrbio ou condição resultante da ocorrência de uma doença primária, distúrbio ou fenótipo e/ou prevenção da doença, distúrbio ou fenótipo.

[00153] Os termos "sujeito" ou "paciente", conforme utilizados neste documento, referem-se a qualquer sujeito para quem o diagnóstico, prognóstico ou terapia de uma doença, distúrbio ou fenótipo é desejado. Conforme usado neste documento, os termos "sujeito" ou "paciente" incluem qualquer organismo humano ou não humano. O termo "organismo não humano" inclui todos os organismos, como primatas não humanos, ovelhas, cães, gatos, cavalos, vacas, ursos, galinhas, anfíbios, répteis, peixes, insetos, bactérias, etc.

[00154] Em certos aspectos, os métodos e sistemas divulgados neste documento podem ser usados para tomar decisões relacionadas à administração de um agente terapêutico, que pode ser um agente usado para prevenir, tratar, gerenciar ou melhorar uma doença ou condição.

[00155] O termo "terapia", como utilizado neste documento, inclui qualquer meio para curar, mitigar ou prevenir uma doença ou distúrbio, incluindo, por exemplo, agentes terapêuticos, instrumentação, medidas de suporte e procedimentos cirúrgicos ou de reabilitação. A este respeito, o termo terapia abrange qualquer protocolo, método e/ou terapêutico ou diagnóstico que possa ser utilizado na prevenção, controle, tratamento e/ou melhora de uma doença ou distúrbio.

[00156] O termo "agente terapêutico", como utilizado neste documento, refere-se a qualquer substância terapeuticamente ativa que é administrada a um sujeito com uma doença ou distúrbio para produzir

um efeito desejado, geralmente benéfico. O termo agente terapêutico inclui, por exemplo, agentes terapêuticos clássicos de baixo peso molecular comumente referidos como fármacos e produtos biológicos de pequenas moléculas, incluindo mas não se limitando a: anticorpos ou seus fragmentos ativos, peptídeos, lipídios, fármacos proteicos, fármacos conjugados com proteínas, enzimas, oligonucleotídeos, ribozimas, material genético, príons, vírus, bactérias e células eucarióticas. Um agente terapêutico também pode ser um pró-fármaco, que metaboliza na substância terapeuticamente ativa desejada quando administrado a um sujeito. Em alguns aspectos, o agente terapêutico é um agente profilático. Além disso, um agente terapêutico pode ser formulado farmacêuticamente. Um agente terapêutico também pode ser um isótopo radioativo ou agente ativado por alguma outra forma de energia, como energia luminosa ou ultrassônica, ou por outras moléculas circulantes que podem ser administradas sistemicamente. Um agente terapêutico também pode ser o processo de melhorar os efeitos não fisiológicos indiretos de uma doença, como planejamento familiar por meio de aconselhamento genético, ou informar um paciente ou um parente do paciente sobre o risco hereditário que uma variante representa para eles.

[00157] Uma quantidade "terapeuticamente eficaz", como utilizada neste documento, é uma quantidade de agente terapêutico que fornece alguma melhora ou benefício a um sujeito que tem uma doença ou distúrbio. Assim, uma quantidade "terapeuticamente eficaz" é uma quantidade que fornece algum alívio, mitigação e/ou diminuição em pelo menos um sintoma clínico da doença ou distúrbio. Aqueles versados na técnica compreenderão que os efeitos terapêuticos não precisam ser completos ou curativos, desde que seja fornecido algum benefício ao sujeito.

[00158] Como utilizado neste documento, uma "quantidade sufici-

ente" ou "uma quantidade suficiente para" alcançar um resultado específico em um paciente com uma doença ou distúrbio refere-se a uma quantidade de um agente terapêutico que é eficaz para produzir um efeito desejado, que é opcionalmente um terapêutico efeito (isto é, pela administração de uma quantidade terapeuticamente eficaz).

[00159] O termo "amostra", como utilizado neste documento, inclui qualquer fluido ou amostra biológica, como sangue total, soro, músculo, saliva obtido de um sujeito. As amostras incluem qualquer fluido ou tecido biológico, como sangue total, soro, músculo, saliva, urina, líquido sinovial, medula óssea, líquido cefalorraquidiano, secreções nasais, escarro, líquido amniótico, líquido de lavagem broncoalveolar, tecido pulmonar, células mononucleares do sangue periférico, total de glóbulos brancos, linfonodos, baço, amígdalas ou pele. Em alguns aspectos específicos, essa amostra é sangue ou uma fração deste, músculo, pele ou uma combinação destes. As amostras podem ser obtidas por qualquer meio conhecido na técnica. Em alguns aspectos, uma amostra pode ser obtida colhendo amostras biológicas de vários sujeitos e as agrupando ou agrupando uma alíquota da amostra biológica de cada sujeito. A amostra agrupada pode ser tratada como uma amostra de um único sujeito. O termo amostra também inclui frações separadas experimentalmente de todas as anteriores. Por exemplo, uma amostra de sangue pode ser fracionada em soro ou em frações contendo tipos específicos de células. Em alguns aspectos, uma amostra pode ser uma combinação de amostras de um indivíduo, como uma combinação de uma amostra de tecido e fluido.

[00160] Para aplicar os métodos e sistemas da divulgação, amostras de um paciente podem ser obtidas antes ou após a administração de uma terapia para tratar uma doença ou distúrbio. Em alguns casos, amostras sucessivas podem ser obtidas do paciente após o início da terapia ou após o término da terapia. As amostras podem, por exem-

plo, ser solicitadas por um profissional de saúde (por exemplo, um médico) ou por um profissional de assistência médica, obtidas e/ou processadas pelo mesmo ou por outro profissional de saúde (por exemplo, uma enfermeira, um hospital) ou um laboratório clínico, e após o processamento, os resultados podem ser encaminhados para o profissional de saúde original ou para outro profissional de saúde, prestadores de benefícios da saúde ou ao paciente. Da mesma forma, a medição/determinação de uma ou mais pontuações, comparações entre pontuações, avaliação das pontuações e decisões de tratamento podem ser realizadas por um ou mais profissionais de saúde, prestadores de benefícios de saúde e/ou laboratórios clínicos.

[00161] Como utilizado neste documento, o termo "profissional de saúde" refere-se a indivíduos ou instituições que interagem e administram diretamente a sujeitos vivos, por exemplo, pacientes humanos. Exemplos não limitativos de profissionais de saúde incluem médicos, enfermeiros, técnicos, terapeutas, farmacêuticos, conselheiros, médicos de medicina alternativa, instalações médicas, consultórios médicos, hospitais, prontos-socorros, clínicas, centros de atendimento de urgência, clínicas/instalações de medicina alternativa e qualquer outra entidade que fornece tratamento geral e/ou especializado, avaliação, manutenção, terapia, medicação e/ou aconselhamento relacionado a todo ou qualquer parte do estado de saúde de um paciente, incluindo, entre outros, cuidados médicos gerais, médicos especializados, cirúrgicos e/ou qualquer outro tipo de tratamento, avaliação, manutenção, terapia, medicação e/ou aconselhamento. Um profissional de saúde também pode se referir ao indivíduo ou a um associado do indivíduo que procura uma interpretação de variantes para o indivíduo, como na busca de entender um fenótipo familiar.

[00162] Como utilizado neste documento, o termo "laboratório clínico" refere-se a uma instalação para o exame ou processamento de

materiais derivados de um sujeito vivo, por exemplo, um ser humano. Exemplos não limitativos de processamento incluem exames biológicos, bioquímicos, sorológicos, químicos, imuno-hematológicos, hematológicos, biofísicos, citológicos, patológicos, genéticos ou outros exames de materiais derivados do corpo humano com o objetivo de fornecer informações, por exemplo, para o diagnóstico, prevenção ou tratamento de qualquer doença ou comprometimento ou avaliação da saúde de sujeitos vivos, por exemplo, seres humanos. Esses exames também podem incluir procedimentos para coletar ou obter uma amostra, preparar, determinar, medir ou descrever a presença ou ausência de várias substâncias no corpo de um sujeito vivo, por exemplo, um ser humano ou uma amostra obtida do corpo de um sujeito vivo, por exemplo, um ser humano.

[00163] Conforme usado neste documento, o termo "prestador de benefícios de saúde" abrange partes individuais, organizações ou grupos que fornecem, apresentam, oferecem, pagam no todo ou em parte ou estão associados de outra forma a oferecer ao paciente acesso a um ou mais benefícios de saúde, benefício de planos de saúde, e/ou programas de contas de despesas com assistência médica.

[00164] Em alguns aspectos, um profissional de saúde pode administrar ou instruir outro profissional de saúde a administrar uma terapia para tratar uma doença ou distúrbio. Um profissional de saúde pode implementar ou instruir outro profissional de saúde ou paciente a executar as seguintes ações: obter uma amostra, processar uma amostra, enviar uma amostra, receber uma amostra, transferir uma amostra, analisar ou medir uma amostra, quantificar uma amostra, fornecer os resultados obtidos após analisar/medir/quantificar uma amostra, receber os resultados obtidos após analisar/medir/quantificar uma amostra, comparar/pontuar os resultados obtidos após analisar/medir/quantificar uma ou mais amostras, fornecer a comparação/pontuação

de uma ou mais amostras, obter a comparação/pontuação de uma ou mais amostras, administrar uma terapia, iniciar a administração de uma terapia, interromper a administração de uma terapia, continuar a administração de uma terapia, interromper temporariamente a administração de uma terapia, aumentar a quantidade de um agente terapêutico administrado, diminuir a quantidade de um agente terapêutico administrado, continuar a administração de uma quantidade de um agente terapêutico, aumentar a frequência de administração de um agente terapêutico, diminuir a frequência de administração de um agente terapêutico, manter a mesma frequência de dosagem em um agente terapêutico, substituir uma terapia ou agente terapêutico por pelo menos outra terapia ou agente terapêutico, combinar uma terapia ou agente terapêutico com pelo menos outra terapia ou agente terapêutico adicional.

[00165] Em alguns aspectos, um prestador de benefícios de saúde pode autorizar ou negar, por exemplo, a coleta de uma amostra, processamento de uma amostra, envio de uma amostra, recebimento de uma amostra, transferência de uma amostra, análise ou medição de uma amostra, quantificação de uma amostra, fornecimento dos resultados obtidos após análise/medição/quantificação de uma amostra, transferência de resultados obtidos após análise/medição/quantificação de uma amostra, comparação/pontuação dos resultados obtidos após análise/medição/quantificação de uma ou mais amostras, transferência da comparação/pontuação a partir de uma ou mais amostras, administração de uma terapia ou agente terapêutico, início da administração de uma terapia ou agente terapêutico, interrupção da administração de uma terapia ou agente terapêutico, continuação da administração de uma terapia ou agente terapêutico, interrupção temporária de administração de uma terapia ou agente terapêutico, aumento da quantidade de agente terapêutico administrado, diminuição da quanti-

dade de administração de agente terapêutico, continuação da administração de uma quantidade de um agente terapêutico, aumento na frequência de administração de um agente terapêutico, diminuição na frequência de administração de um agente terapêutico, manutenção da mesma frequência de dosagem de um agente terapêutico, substituição de uma terapia ou agente terapêutico por pelo menos outra terapia ou agente terapêutico ou combinar uma terapia ou agente terapêutico com pelo menos outra terapia ou agente terapêutico adicional.

[00166] Além disso, os benefícios de assistência médica podem, por exemplo, autorizar ou negar a prescrição de uma terapia, autorizar ou negar cobertura para a terapia, autorizar ou negar reembolso pelo custo da terapia, determinar ou negar a elegibilidade para a terapia, etc.

[00167] Em alguns aspectos, um laboratório clínico pode, por exemplo, coletar ou obter uma amostra, processar uma amostra, enviar uma amostra, receber uma amostra, transferir uma amostra, analisar ou medir uma amostra, quantificar uma amostra, fornecer os resultados obtidos após analisar/medir/quantificar uma amostra, receber os resultados obtidos após analisar/medir/quantificar uma amostra, comparar/pontuar os resultados obtidos após analisar/medir/quantificar uma ou mais amostras, fornecer a comparação/pontuação de uma ou mais amostras, obter a comparação/pontuação de uma ou mais amostras ou outras atividades relacionadas.

[00168] Em aspectos particulares, os métodos divulgados neste documento incluem informar o sujeito de um resultado, por exemplo, o impacto fenotípico de uma variante molecular, obtido de acordo com os métodos divulgados neste documento. O paciente pode ser informado verbalmente, por escrito e/ou eletronicamente. Essas informações também podem ser registradas no prontuário do paciente. Por exemplo, em vários aspectos, o diagnóstico de uma doença ou distúr-

bio tratável com um agente terapêutico específico é registrado em um prontuário médico. O termo "prontuário médico" ou "prontuário médico do paciente" refere-se a uma conta do exame e/ou tratamento de um paciente que normalmente inclui um ou mais dos seguintes itens: histórico médico e queixas do paciente, constatações físicas do médico, resultados de diagnóstico testes e procedimentos e medicamentos para pacientes e procedimentos terapêuticos. Um prontuário médico é normalmente feito por um ou mais médicos e/ou assistentes médicos e é um prontuário e/ou histórico escrito, transcrito ou registrado de outro modo de várias doenças ou lesões que requerem cuidados médicos e/ou inoculações e/ou alergias e/ou tratamentos e/ou prognóstico e/ou frequentemente informações sobre saúde sobre pais, irmãos e/ou ocupação. O prontuário pode ser revisado por um médico, por exemplo, ao diagnosticar uma condição ou tomar uma decisão de tratamento.

[00169] O prontuário médico pode estar em formato de papel e/ou pode ser mantido em um meio legível por computador. O prontuário médico pode ser mantido por um laboratório, consultório médico, um hospital, uma organização de manutenção da saúde, uma companhia de seguros e/ou um site pessoal de prontuário médico. Em alguns aspectos, um diagnóstico, baseado pelo menos em parte nos métodos divulgados neste documento, é registrado em um artigo de alerta médico, como um cartão, um artigo de vestimenta e/ou uma etiqueta de identificação por radiofrequência (RFID). Como utilizado neste documento, o termo "artigo de vestimenta" refere-se a qualquer artigo que possa ser usado no corpo de um sujeito, incluindo, mas não limitado a uma etiqueta, pulseira, colar, faixa para braço ou faixa para cabeça.

[00170] Os métodos divulgados neste documento também incluem prescrever, iniciar e/ou alterar a profilaxia e/ou terapia para uma doença ou distúrbio. Em certos aspectos, os métodos podem envolver a

encomenda e/ou a realização de um ou mais ensaios adicionais. Por exemplo, um teste genético pode ser repetido para descartar um resultado falso negativo e/ou um ou mais testes adicionais podem ser realizados para monitorar o status do sujeito.

[00171] Uma pessoa versada na técnica entenderia que os métodos divulgados neste documento podem ser utilizados, por exemplo, em métodos de tratamento, diagnóstico e monitoramento, como (i) seletores positivos, ou seja, uma ação específica seria tomada (por exemplo, tratar um paciente com doença ou distúrbio) após uma determinação do potencial efeito clínico de um genótipo; ou (ii) seletores negativos, ou seja, uma ação específica seria tomada (por exemplo, não tratar um paciente com uma doença ou distúrbio) após a determinação do potencial efeito clínico de um genótipo; ou (iii) seletores positivos e negativos, por exemplo, um tratamento específico pode cessar e um tratamento diferente pode começar após a determinação do potencial efeito clínico de um genótipo.

[00172] Esta divulgação fornece um método de tratamento de um paciente suspeito de ter uma doença, distúrbio ou fenótipo, compreendendo a administração de um agente terapêutico ao paciente se uma determinação do potencial efeito clínico de um genótipo de acordo com os métodos divulgados neste documento indica que o paciente pode se beneficiar do tratamento com o agente terapêutico.

[00173] Esta divulgação também fornece métodos e sistemas para facilitar uma determinação por um profissional de saúde, um prestador de benefícios de saúde ou um laboratório clínico para saber se um paciente se beneficiará do tratamento com um antagonista do agente terapêutico se uma determinação do potencial efeito clínico de um genótipo de acordo com o método divulgado neste documento indicar que o paciente pode se beneficiar do tratamento com o agente terapêutico.

[00174] Os métodos fornecidos neste documento também facilita-

rão a determinação por um profissional de saúde, prestador de benefícios de saúde ou laboratório clínico para saber se um paciente se beneficiará do tratamento com quaisquer outros agentes terapêuticos.

[00175] A presente divulgação também fornece um método de tratamento de um paciente com ou suspeito de ter uma doença ou distúrbio, compreendendo a administração de um agente terapêutico ao paciente se o impacto fenotípico de uma variante molecular identificada de acordo com os métodos divulgados neste documento indicar que o paciente se beneficiaria de esse tratamento. Em alguns aspectos, uma amostra é obtida do paciente e submetida a testes genéticos, por exemplo, a um laboratório clínico.

[00176] Também é fornecido um método de tratamento de um paciente com ou com suspeita de ter uma doença ou distúrbio compreendendo (a) enviar uma amostra retirada do paciente para teste genético; e (b) administrar um agente terapêutico ao paciente se o impacto fenotípico de uma variante molecular identificada a partir do referido teste genético de acordo com os métodos divulgados neste documento indicar que o paciente pode se beneficiar do tratamento com o agente terapêutico.

[00177] A divulgação também fornece um método de tratamento de um paciente com ou suspeito de ter uma doença ou distúrbio compreendendo (a) medir o impacto fenotípico de uma variante molecular identificada de acordo com os métodos divulgados neste documento em uma amostra obtida de um paciente com ou suspeito de ter uma doença ou distúrbio; (b) determinar se o paciente pode se beneficiar do tratamento com um agente terapêutico com base na presença/ausência de uma variante alélica; e (c) aconselhar um profissional de saúde a administrar o agente terapêutico ao paciente se a variante alélica estiver presente/ausente.

[00178] Em certos aspectos, um laboratório clínico (por exemplo,

um laboratório de testes genéticos) determinando o impacto fenotípico de uma variante molecular identificada de acordo com os métodos da presente divulgação aconselhará o profissional de saúde se o paciente pode se beneficiar do tratamento com um determinado agente terapêutico. Em alguns aspectos, o laboratório clínico pode aconselhar o profissional de saúde se o paciente pode se beneficiar do início, cessação ou modificação do tratamento com um determinado agente terapêutico.

[00179] Em alguns aspectos, os resultados de uma determinação do impacto fenotípico de uma variante molecular conduzida de acordo com os métodos da presente divulgação podem ser submetidos a um profissional de saúde ou um prestador de benefícios de saúde para determinar se o seguro do paciente cobrirá o tratamento com um certo agente terapêutico.

[00180] Em certos aspectos, esta divulgação fornece um método de tratamento de um paciente com ou suspeito de ter uma doença ou distúrbio compreendendo: determinar, por exemplo, em um laboratório de testes genéticos, o impacto fenotípico de uma variante molecular identificada de acordo com os métodos da presente divulgação; e aconselhar um profissional de saúde a administrar um certo agente terapêutico ao paciente se o impacto fenotípico da variante molecular identificado de acordo com os métodos divulgados neste documento indica que o paciente pode se beneficiar do tratamento com o agente terapêutico.

[00181] Em certos aspectos, o método de tratamento pode compreender: determinar, por exemplo, em um laboratório de testes genéticos, o impacto fenotípico de uma variante molecular identificada de acordo com os métodos da presente divulgação; determinar se o impacto fenotípico da variante molecular indica que o paciente pode se beneficiar do tratamento com um agente terapêutico; e aconselhar um

profissional de saúde a ajustar a dosagem do agente terapêutico, se indicado, por exemplo, para aumentar ou manter a quantidade ou a frequência do agente terapêutico administrado ao paciente, interromper a terapia ou manter ou reduzir a quantidade ou frequência do agente terapêutico.

[00182] Em alguns aspectos, além da determinação do impacto fenotípico de uma variante molecular identificada de acordo com os métodos divulgados neste documento, os métodos divulgados neste documento podem compreender determinar, enviar uma amostra retirada do paciente para determinação ou instruir um laboratório clínico a realizar testes adicionais, por exemplo, para determinar a ausência ou presença e/ou nível de expressão e/ou atividade de um determinado biomarcador ou biomarcadores.

[00183] A determinação do impacto fenotípico de uma variante molecular identificada de acordo com os métodos divulgados neste documento pode ser usada, como discutido acima, como parte do tratamento de uma doença ou condição. Além disso, a determinação do impacto fenotípico de uma variante molecular identificada de acordo com os métodos divulgados neste documento pode ser usada, por exemplo, para selecionar um paciente para tratamento com um agente terapêutico, para selecionar um agente terapêutico entre várias opções potenciais para tratamento, para selecionar ou excluir um paciente para um ensaio clínico ou determinar o prognóstico do paciente. Em resposta ao potencial impacto fenotípico de uma variante molecular identificada de acordo com os métodos divulgados neste documento, um profissional de saúde, um prestador de benefícios de saúde ou um terapeuta pode fornecer orientação sobre o estilo de vida. Por exemplo, em resposta à identificação de uma variante molecular ligada à obesidade, um sujeito pode ser aconselhado a ajustar sua dieta; em resposta à identificação de uma variante molecular ligada ao câncer de pul-

mão, um sujeito pode ser aconselhado a parar de fumar, etc.

[00184] Em alguns aspectos, os resultados de uma determinação do impacto fenotípico de uma variante molecular podem ser utilizados em aplicações de manipulação biomolecular, biomanipulação molecular, manipulação genética ou biomanipulação, informando os efeitos de variantes em uma biomolécula, sugerindo alterações na biomolécula para atingir uma determinada propriedade, comportamento ou finalidade da biomolécula, sistema biológico ou tecnologia biomédica.

[00185] Como usado neste documento, o termo "biomolécula" inclui todas as moléculas, derivadas biologicamente e fabricadas pelo homem, como proteínas humanas e não humanas, proteínas sintéticas, peptídeos, ácidos nucleicos ou bioprodutos destes, como analitos, metabólitos ou moléculas que interagem com eles, como ligantes, moléculas pequenas e outros peptídeos. Por exemplo, a proteína humana "butirilcolinesterase" é uma biomolécula de proteína.

[00186] Como utilizado neste documento, "manipulação biomolecular", "biomanipulação molecular", "manipulação genética" ou "biomanipulação" é usada para significar a aplicação dos princípios da biologia e das ferramentas de manipulação para produzir produtos com propriedades específicas. Por exemplo, a proteína humana "butirilcolinesterase humana" foi manipulada novamente para produzir uma hidrolase de cocaína 1390 vezes mais eficaz do que na sua forma original (Xue et al., Design, preparation, and characterization of high-activity mutants of human butyrylcholinesterase specific for detoxification of cocaine. Molecular pharmacology. 2011).

[00187] Como utilizado neste documento, "sistema biológico" é usado para significar uma entidade ou grupo biológico de entidades, como um grupo de micro-organismos, um órgão humano ou um grupo de órgãos. Por exemplo, a epiderme é um sistema biológico.

[00188] Como utilizado neste documento, "tecnologia biomédica" é

usada para significar uma tecnologia roteada, parcial ou totalmente baseada ou inspirada na biologia. Por exemplo, o PacBio Sequencing consegue o sequenciamento em tempo real de molécula única usando DNA polimerases manipuladas.

REIVINDICAÇÕES

1. Método implementado por computador para prever o impacto fenotípico da variante molecular, caracterizado pelo fato de que compreende:

(a) registrar um modelo de evidência compreendendo dados de evidência;

(b) avaliar dados de desempenho de validação para o modelo de evidência com base em dados de produção;

(c) gerar um valor hash de dados de suporte para o modelo de evidência;

(d) avaliar dados de desempenho de teste para o modelo de evidência com base nos dados de evidência e nos dados de teste em resposta ao recebimento dos dados de teste para o modelo de evidência;

(e) classificar o modelo de evidência em um conjunto de modelos de evidência para a entidade alvo com base nos dados de desempenho da validação ou nos dados de desempenho do teste; e

(f) fornecer o impacto fenotípico previsto usando um modelo de evidência com melhor desempenho para a entidade alvo com base na classificação em resposta a uma consulta para o impacto fenotípico previsto da variante molecular para a entidade alvo a partir de um terminal de interpretação de variantes;

em que,

(i) os dados de evidência descrevem um impacto fenotípico previsto de uma variante molecular para uma entidade alvo;

(ii) a geração do valor hash permite avaliação prospectiva dos dados de evidência em resposta ao recebimento de dados de teste para o modelo de evidência; e

(iii) o valor hash de dados de suporte para o modelo de evidência é armazenado em um banco de dados distribuído, em que o

banco de dados associa ao valor hash aos dados de suporte.

2. Método, de acordo com a reivindicação 1, caracterizado pelo fato de que a entidade alvo compreende um elemento funcional, uma molécula ou uma variante molecular e um fenótipo de interesse.

3. Método, de acordo com a reivindicação 1, caracterizado pelo fato de que o registro do modelo de evidência compreende gerar o modelo de evidência com base nos dados de produção.

4. Método, de acordo com a reivindicação 1, caracterizado pelo fato de que o registro do modelo de evidência compreende importar o modelo de evidência ou os dados de evidência.

5. Método, de acordo com a reivindicação 1, caracterizado pelo fato de que os dados de suporte são gerados a partir dos dados de evidência, dados de produção, dados de teste, dados de desempenho de validação, dados de desempenho de teste, ou de uma combinação destes.

6. Método, de acordo com a reivindicação 1, caracterizado pelo fato de que o valor hash dos dados de suporte para o modelo de evidência é armazenado no banco de dados distribuído com uma marca de tempo.

7. Método, de acordo com a reivindicação 3, caracterizado pelo fato de que os dados de produção são recebidos a partir de uma base de conhecimento clínico.

8. Método, de acordo com a reivindicação 1, caracterizado pelo fato de que a avaliação dos dados de desempenho de validação compreende:

(1) calcular uma pontuação de impacto fenotípico para uma ou mais variantes moleculares da entidade alvo nos dados de produção usando o modelo de evidência e uma técnica de validação de modelo; e

(2) gerar os dados de desempenho de validação com

base nas pontuações de impacto fenotípico usando uma métrica de desempenho de interesse.

9. Método, de acordo com a reivindicação 1, caracterizado pelo fato de que a avaliação dos dados de desempenho de teste compreende:

(1) calcular uma pontuação de impacto fenotípico para uma ou mais variantes moleculares da entidade alvo nos dados de teste usando o modelo de evidência e uma técnica de validação de modelo; e

(2) gerar os dados de desempenho de teste com base nas pontuações de impacto fenotípico usando uma métrica de desempenho de interesse.

10. Método, de acordo com a reivindicação 1, caracterizado pelo fato de que o banco de dados distribuído é imutável.

11. Método, de acordo com a reivindicação 1, caracterizado pelo fato de que o banco de dados distribuído é uma estrutura de dados em "blockchain".

12. Método, de acordo com a reivindicação 1, caracterizado pelo fato de que o banco de dados distribuído é um "feed" distribuído.

13. Sistema de terminal de interpretação de variante, caracterizado pelo fato de que compreende:

(i) uma memória; e

(ii) pelo menos um processador acoplado à memória;

em que pelo menos um processador é configurado para:

(a) enviar uma consulta de suporte a um sistema de interpretação de variante para dados de suporte para um modelo de evidência que atenda a um conjunto de métricas de desempenho para uma entidade alvo;

(b) receber os dados de suporte e um registro de auditoria associado para os dados de suporte do sistema de interpretação de

variante;

(c) enviar uma consulta de auditoria para uma estrutura de dados distribuídos, em que a consulta de auditoria compreende o registro de auditoria para os dados de suporte;

(d) receber um certificado de validação para o registro de auditoria do banco de dados distribuído em resposta ao envio da consulta de auditoria; e

(e) determinar um estado de dados dos dados de suporte em um momento específico com base no registro de auditoria;

e em que pelo menos um processador é configurado para:

(1) calcular um valor hash dos dados de suporte para o modelo de evidência e armazenar o dito valor hash em um banco de dados distribuído; e

(2) determinar que o valor hash corresponda a um valor hash no registro de auditoria para os dados de suporte para o modelo de evidência.

14. Sistema, de acordo com a reivindicação 13, caracterizado pelo fato de que a entidade alvo compreende um elemento funcional, uma molécula ou uma variante molecular e um fenótipo de interesse.

15. Método, de acordo com a reivindicação 3, caracterizado pelo fato de que o modelo de evidência baseado nos dados de produção é gerado usando:

(i) uma técnica de aprendizagem de máquina;

(ii) um ensaio funcional;

(iii) um estímulo biofísico; ou,

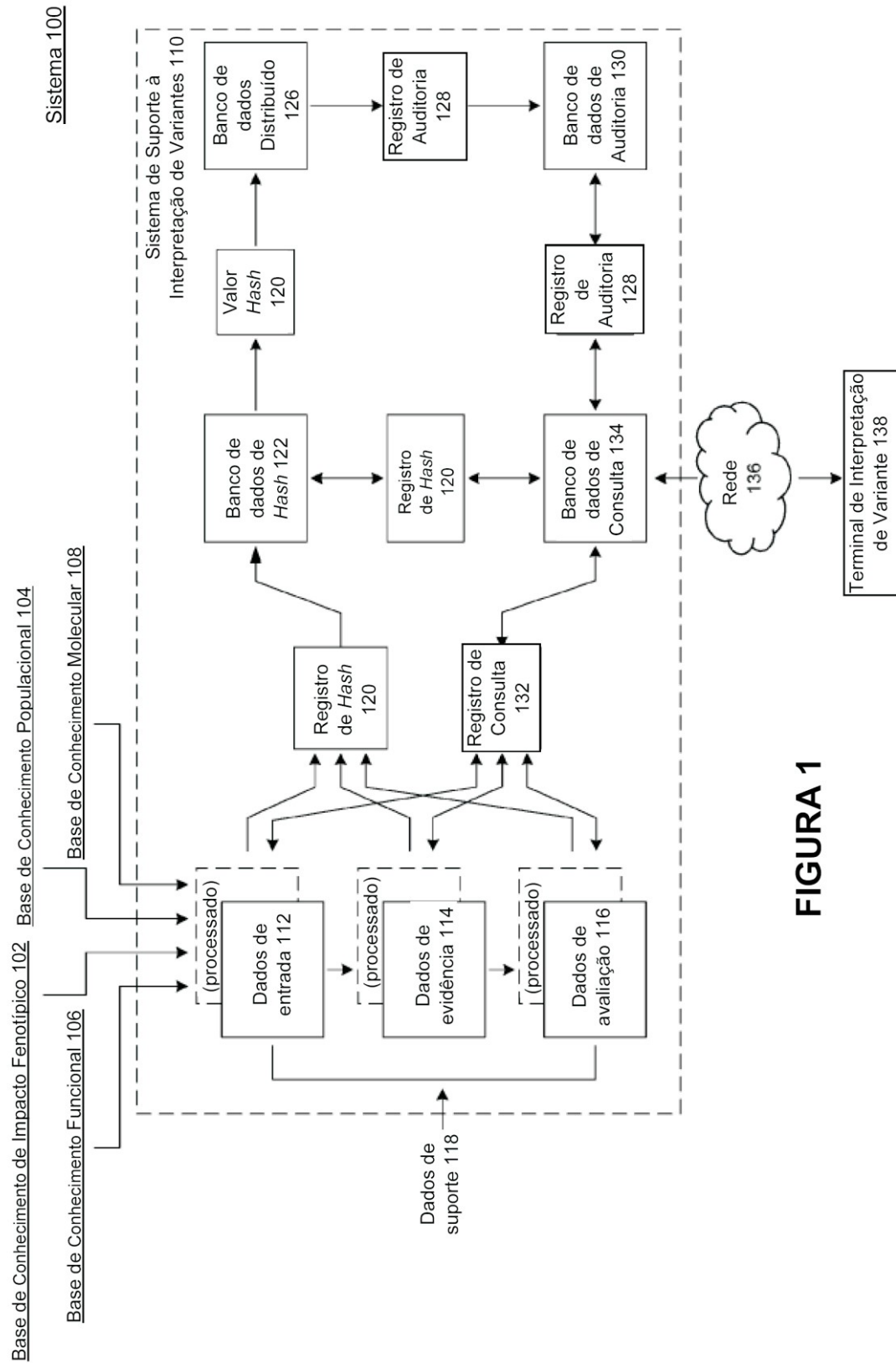
(iv) uma combinação destes.

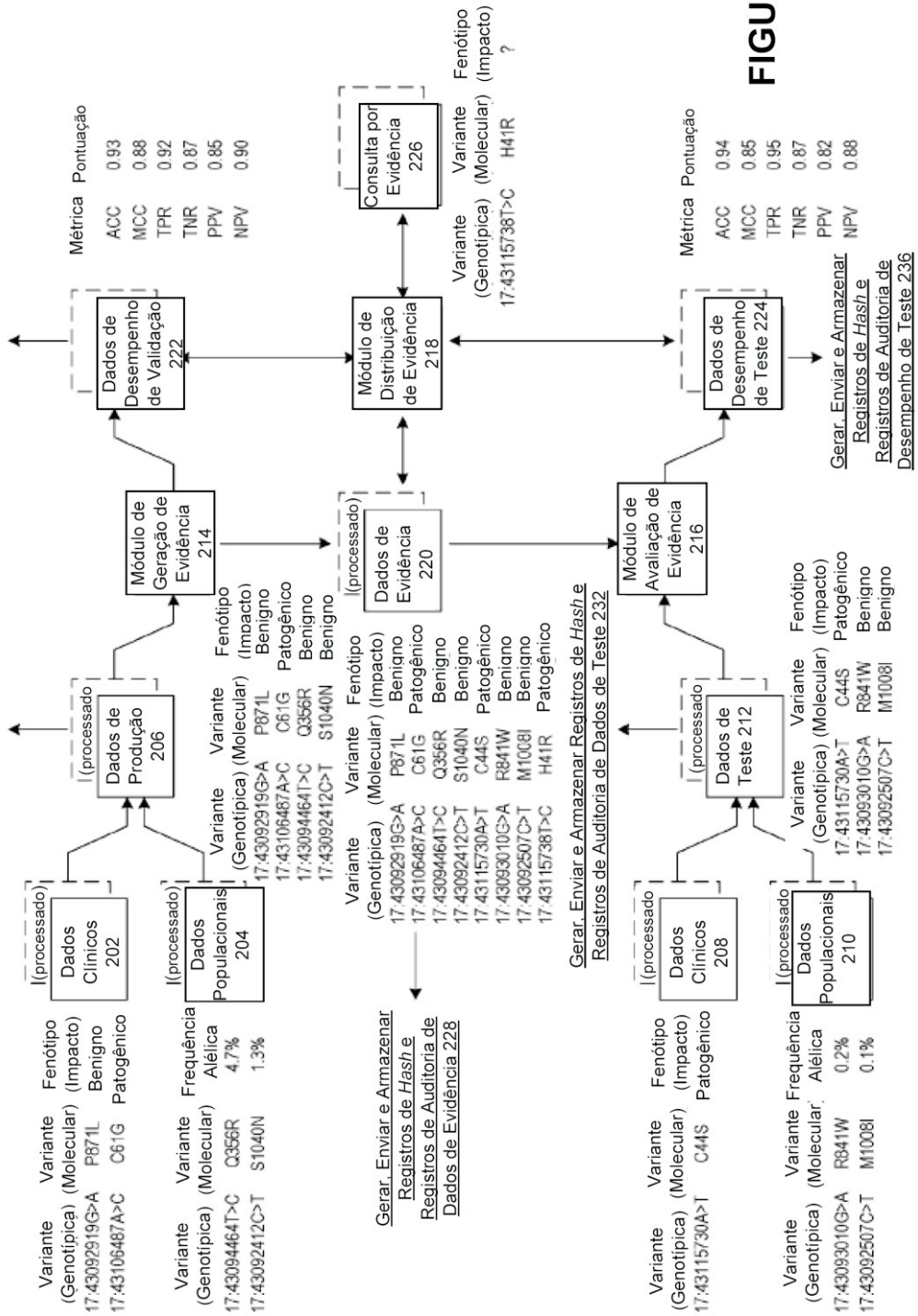
16. Método, de acordo com a reivindicação 15, caracterizado pelo fato de que a técnica de aprendizagem de máquina é não supervisionada, supervisionada ou semisupervisionada.

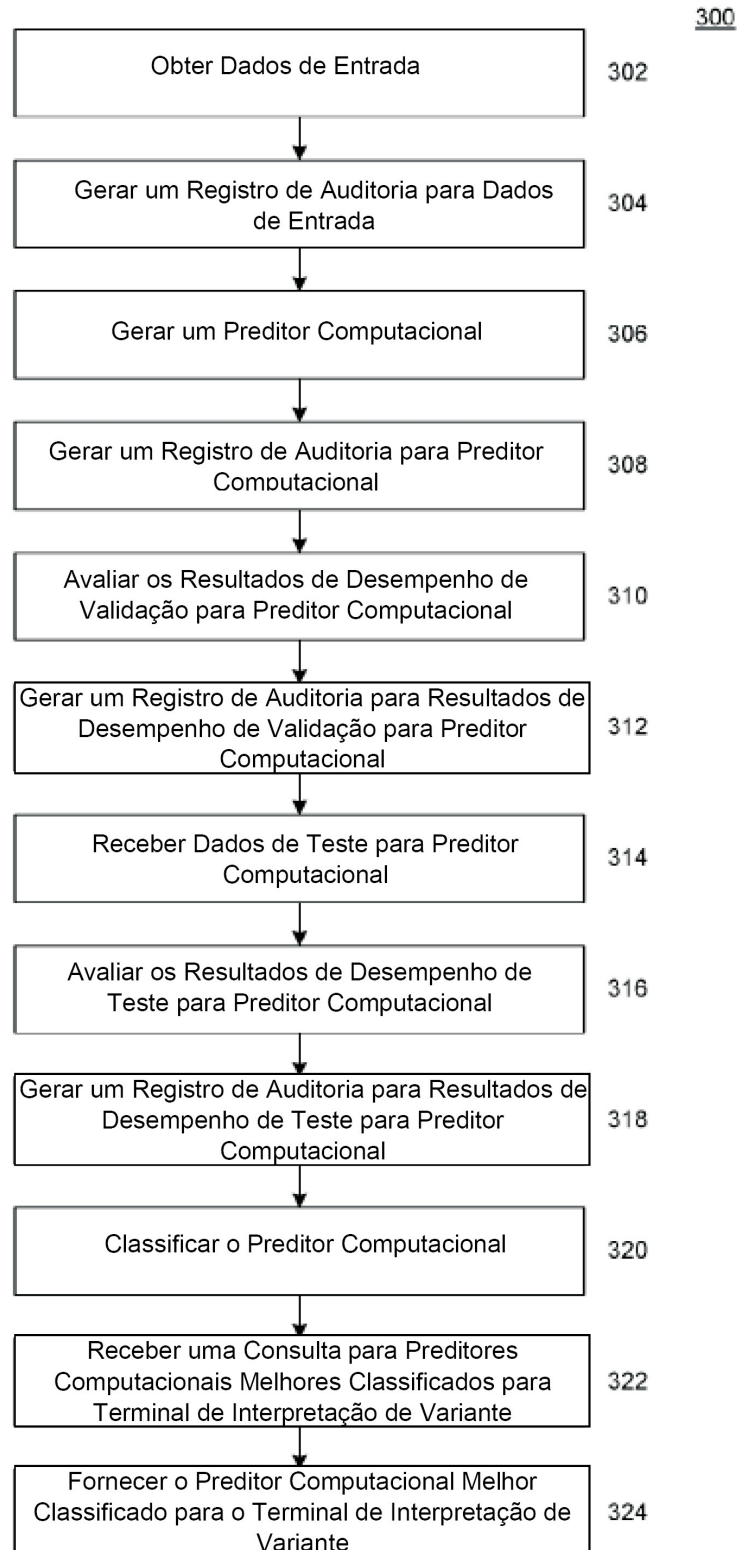
17. Método, de acordo com a reivindicação 1, caracterizado pelo fato de que compreende ainda fornecer um registro de auditoria a um terminal de interpretação de variante, como definido na reivindicação 13, em que

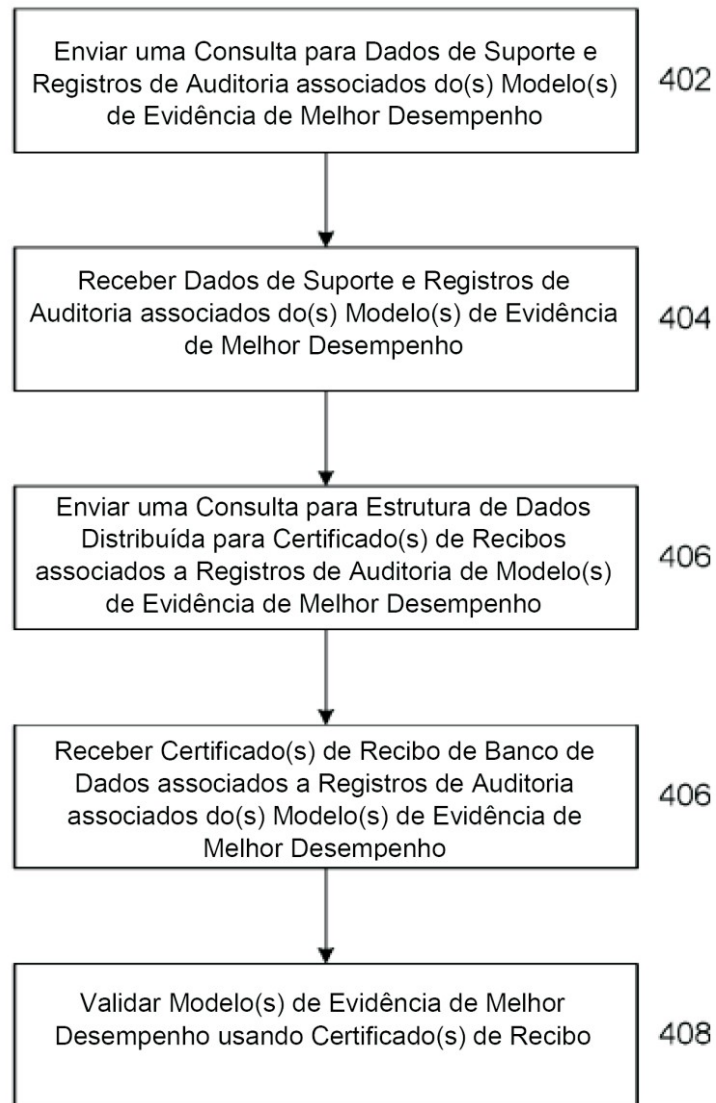
(i) o registro de auditoria refere-se a uma entrada para os dados de suporte no banco de dados distribuído, e

(ii) o registro de auditoria permite que o terminal de interpretação de variante examine o conteúdo dos dados de suporte e um tempo de criação dos dados de suporte.





**FIGURA 3**

400**FIGURA 4**

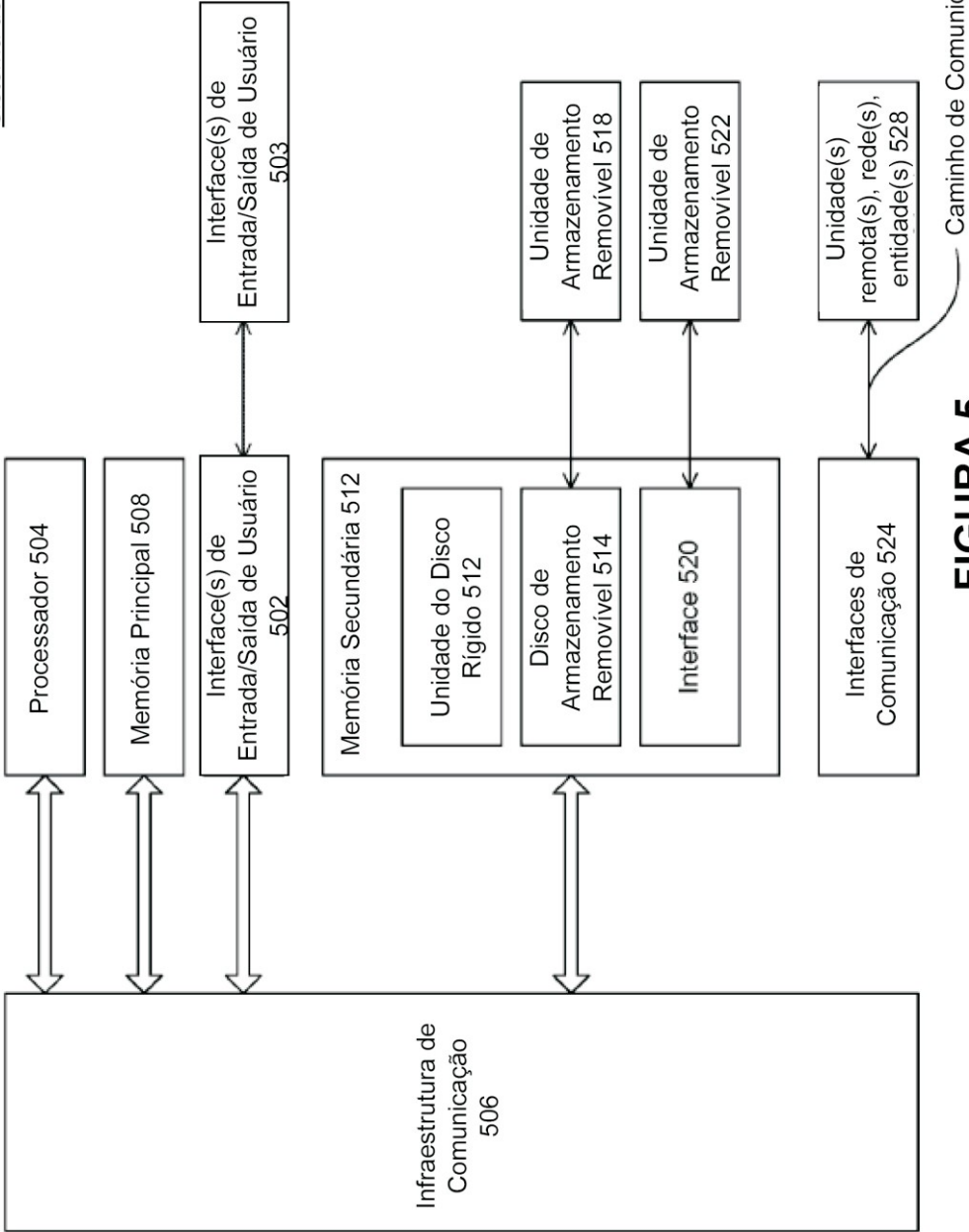
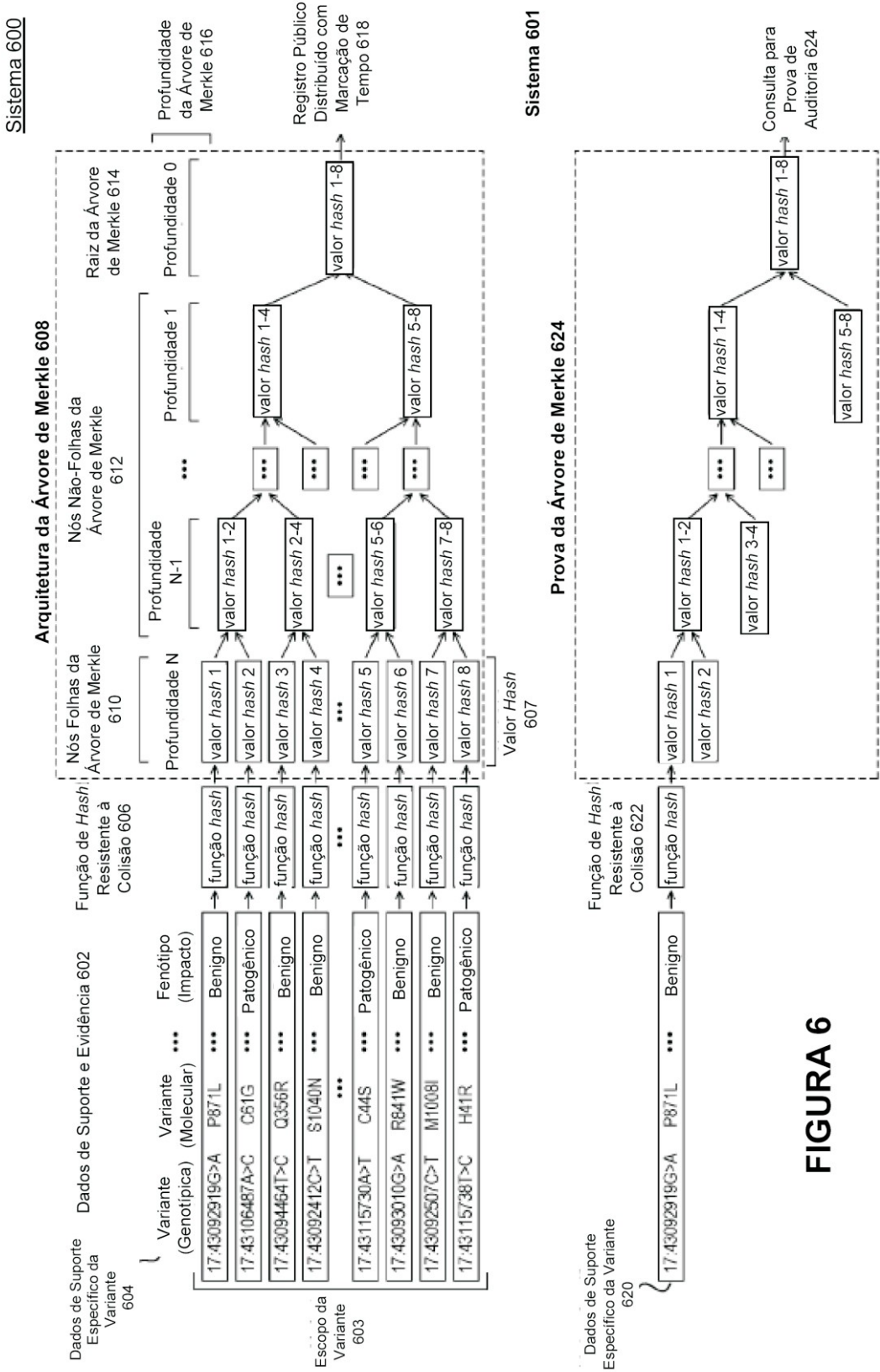


FIGURA 5



RESUMO

Patente de Invenção: **"PLATAFORMA DE EVIDÊNCIA MOLECULAR PARA OTIMIZAÇÃO AUDITÁVEL E CONTÍNUA DE INTERPRETAÇÃO DE VARIANTES EM TESTES E ANÁLISES GENÉTICAS E GENÔMICAS"**.

A presente invenção refere-se a modalidades de sistema, método e produto de programa de computador para otimizar a determinação de um impacto fenotípico de uma variante molecular identificada em testes, amostras moleculares ou relatórios de sujeitos por meio de incorporação, atualização, monitoramento, validação, seleção e auditoria, executados regularmente, dos modelos de evidência com melhor desempenho para a interpretação de variantes moleculares através de uma pluralidade de classes de evidência.