

TENDÊNCIA DE DISCREPÂNCIA CLASSIFICATÓRIA DE VARIANTES DE SEQUÊNCIA QUANTO À SUA NATUREZA POR MODELOS DE LINGUAGEM DE GRANDE ESCALA: UM ESTUDO DE ANÁLISE DE RESÍDUOS PADRONIZADOS

Felício de Freitas Netto¹, Taís Rosa Calisto¹ e Mara Sanches Guaragna¹

¹Universidade Estadual de Campinas (UNICAMP)

INTRODUÇÃO

A análise de variantes genéticas constitui um desafio central da genética clínica e laboratorial, especialmente diante do volume crescente de dados oriundos do sequenciamento genômico. Com o surgimento das tecnologias de inteligência artificial (IA), surgem novas perspectivas para o uso dessas ferramentas no suporte ao processo de interpretação.

OBJETIVO

Avaliar a distribuição das classificações de variantes de sequência (SNVs) atribuídas por três ferramentas de IA (ChatGPT, Gemini e Claude), segundo seus tipos, e identificar os principais padrões de discrepância em relação aos critérios do ACMG.

METODOLOGIA

Foram analisadas 56 variantes SNVs curadas por especialistas no ClinVar e classificadas de acordo com os critérios do ACMG, abrangendo variantes dos tipos *frameshift*, *missense*, *sinônima*, *intrônica*, *in-frame* e *splicing*.

As mesmas variantes foram submetidas à classificação por três modelos distintos de IA. Para cada IA, foram construídas tabelas de contingência entre o tipo de variante e a classificação atribuída por convenção (benigna, VUS ou patogênica), sendo aplicados o teste qui-quadrado e a análise de resíduos padronizados para identificação dos desvios significativos em relação ao esperado.

RESULTADOS E DISCUSSÃO

Os resultados evidenciaram padrões característicos de cada ferramenta de IA (Fig. 1). O ChatGPT apresentou excesso de classificações de variantes intrônicas como VUS (resíduo padronizado = +2,49) e déficit na classificação destas como patogênicas (resíduo = -2,20), além de excesso na atribuição de benignidade às variantes sinônimas (resíduo = +2,47).

O Gemini mostrou excesso expressivo na classificação de variantes intrônicas como VUS (resíduo = +4,24) e sinônimas como benignas (resíduo = +6,55), ao passo que apresentou déficit na classificação das intrônicas e sinônimas como patogênicas (resíduos = -2,34 e -2,06, respectivamente).

O Claude, por sua vez, evidenciou excesso na classificação de variantes intrônicas como benignas (resíduo = +2,68) e de *missense* como VUS (resíduo = +2,18), além de um déficit na classificação de intrônicas como patogênicas (resíduo = -2,20).

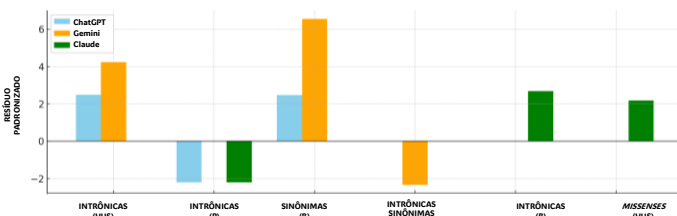


Fig. 1. Resíduos padronizados por ferramenta de IA e categoria das variantes. VUS (variante de significado incerto); P (patogênica); B (benigna). FONTE: elaborado pelos autores, 2025.

CONCLUSÃO

Este estudo contribui para o entendimento das características e restrições das abordagens baseadas em IA ao sinalizar caminhos para o aprimoramento futuro dessas tecnologias na prática da genética molecular.

REFERÊNCIAS

Richards S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015 May;17(5):405-24.