POSTER - PROTEINS AND PROTEOMICS

FUNCTIONAL AND STRUCTURAL INSIGHTS OF ALPHAMISSENSE, SIFT, AND POLYPHEN-2 IN CLASSIFYING CFTR MISSENSE VARIANTS

Ana Katarina Campos Nunes (nunesmusic12@gmail.com)

Arthur Felipe Vasconcelos Ferreira Reis (arthurreis.ufpa@gmail.com)

Camila Forte (camila.s.forte@gmail.com)

Gustavo Barra Matos (gustavobarra16@gmail.com)

Gilderlanio Santana De Araujo (gilderlanio@ufpa.br)

Accurate classification of missense variants remains a significant challenge in medical genomics, especially in distinguishing between pathogenic mutations and benign variants in monogenic disorders such as Cystic Fibrosis (CF). This limitation highlights the need for robust computational and functional approaches to improve diagnostic accuracy. Widely used predictors, such as SIFT, PolyPhen-2, and now on AlphaMissense (AM), have shown varying usefulness. AlphaMissense, for example, is a deep learning-based tool that integrates sequence, structural and evolutionary information, showing superior performance compared to sequence-based methods. This study comprehensively evaluated the performance of AlphaMissense, SIFT, and PolyPhen-2 in classifying clinically significant CFTR missense variants, using 164 variants from the CFTR2 database as ground truth. Our results demonstrate that SIFT achieved the highest accuracy (0.99) for predicting CF-causing variants, while AlphaMissense outperformed other tools (accuracy: 0.78) for non-CF-causing variants. Notably, AlphaMissense showed the

strongest correlation with FoldX-derived ΔΔG values ($r = 0.5$; $p \leq 2.2e{-}16$), suggesting its predictions are more closely tied to protein destabilization energetics than SIFT ($r = 0.21$) or PolyPhen-2 ($r = 0.39$). Further regression analyses revealed that each unit increase in the AlphaMissense score corresponded to a 4.3-fold rise in ΔΔG ($p \leq 2.69e{-}08$), and logistic regression confirmed that ΔΔG itself is a significant predictor of pathogenicity (OR = 1.376 per unit; 95% CI: 1.103–1.792; $p \leq 0.0095$), increasing the odds of a variant being CF-causing by 37.6%. Collectively, these findings suggest that AlphaMissense's pathogenicity predictions extend beyond simple destabilization effects, potentially capturing broader functional impacts of missense variants. The integration of AlphaMissense with FoldX free energy calculations emerges as a robust complementary strategy for interpreting CFTR variants, especially the ones of uncertain significance in CF patients, offering more accuracy and a structural overview of pathogenicity significance.

Palavras-chave: pathogenicity predictors; cftr; gibbs free energy; missense variants; databases.