Breaking the Bottleneck in Genetic Variant Interpretation for Precision Medicine

Xavier de la Cruz[1]

[1]Vall d'Hebron University Hospital, Barcelona

Personalized medicine, a promising branch of modern healthcare, has been made possible by the rapid development of next-generation sequencing (NGS), which has revolutionized genetic diagnostics and provided unprecedented opportunities for tailored treatments. However, the clinical utility of NGS remains constrained by the challenge of interpreting the impact of the genetic variants it uncovers. A significant portion of these variants remains classified as Variants of Uncertain Significance (VUS), undermining their clinical utility and creating anxiety for patients and their families. This situation has driven the development of computational pathogenicity predictors, machine learning tools trained to produce binary classifications—benign or pathogenic—of variants. While these methods have been integrated into clinical workflows, their accuracy and interpretability still fall short of meeting the stringent requirements of medical applications.

In this context, recent years have witnessed a paradigm shift toward continuous prediction models, which aim to provide more precise quantitative assessments of variant impacts on protein function. These approaches leverage a combination of technologies that include data from deep mutational scanning experiments and machine learning techniques. By moving beyond binary labels, continuous predictors hold the promise of elucidating critical aspects of variant effects, such as disease severity and therapeutic response, thereby enhancing their relevance for clinical decision-making in precision medicine.

This talk will explore the current state of methodologies to estimate the impact of protein variants, focusing on an original approach developed in our group to address the problem of using a small amount of protein-specific datasets to generate predictions for any protein, combining regression models and an ensemble-based approach. I will discuss, among other things, the results obtained both in rigorous validation experiments as well as in our participation in the CAGI5 and CAGI6 challenges, comparing our performance with that of other methods in the field.

Keywords: