Analysis of the relationship between missense variants and diseases.

Hinami Kadono

MITA International School, Tokyo, Japan

Missense variants have the potential to result in disease development attributable to protein variants. Current genetic tests evaluate the potential for disease development by querying various databases for information on known missense variants. Therefore, it is impossible to address unknown amino acid variants not included in these databases. In this study, we aimed to evaluate the impact of missense variants on human health and predict the potential of unknown missense variants to cause diseases. In particular, we considered the subcellular localization nature of the protein.From previous year's research, localization information was downloaded from UniProt, while data on missense variants and diseases were obtained from ClinVar. For specific variant patterns across different subcellular localizations, heatmaps were created by scoring the degree of association with the disease. This score was normalized by gene length and frequency of missense variants.The results revealed that certain subcellular localizations are more deeply linked to disease than others. By using these findings, and an index for evaluating missense variants as training data, a machine learning model was developed to predict disease associations from unknown missense variants. It correctly predicted disease-related missense variants with an accuracy of approximately 81%. From these analyses, subcellular localizations may serve as a valuable indicator for assessing the disease potential of missense variants. This research can be expanded in the future for application to basic medical research, drug discovery research, and the development of software for clinical use.