Medical care, Drug development, Bioinformatics

ACIDES: Technological innovation in protein screening analysis algorithms

Premium Research Institute for Human Metaverse Medicine (WPI-PRIMe)

Specially Appointed Associate Professor Takahiro Nemoto

Researchmap https://researchmap.jp/takahiro_nemoto?lang=en

Abstract

Protein screening is an experimental technique used for creating new proteins and investigating their functions. It is frequently employed in medical research, such as drug developments and understanding disease mechanisms. To conduct highly accurate screenings, it is necessary to analyze experimental results with greater precision. However, the evaluation of statistical errors when combined with next-generation sequencing (NGS) had not been fully understood. In collaboration with a research group from the Institut de la Vision in Paris, we developed a new method that combines a statistical model to describe highly dispersed NGS noise with a mathematical model for protein screening. This approach, called ACIDES, enables an unprecedented level of precision in determining statistical errors in protein screening experiments.

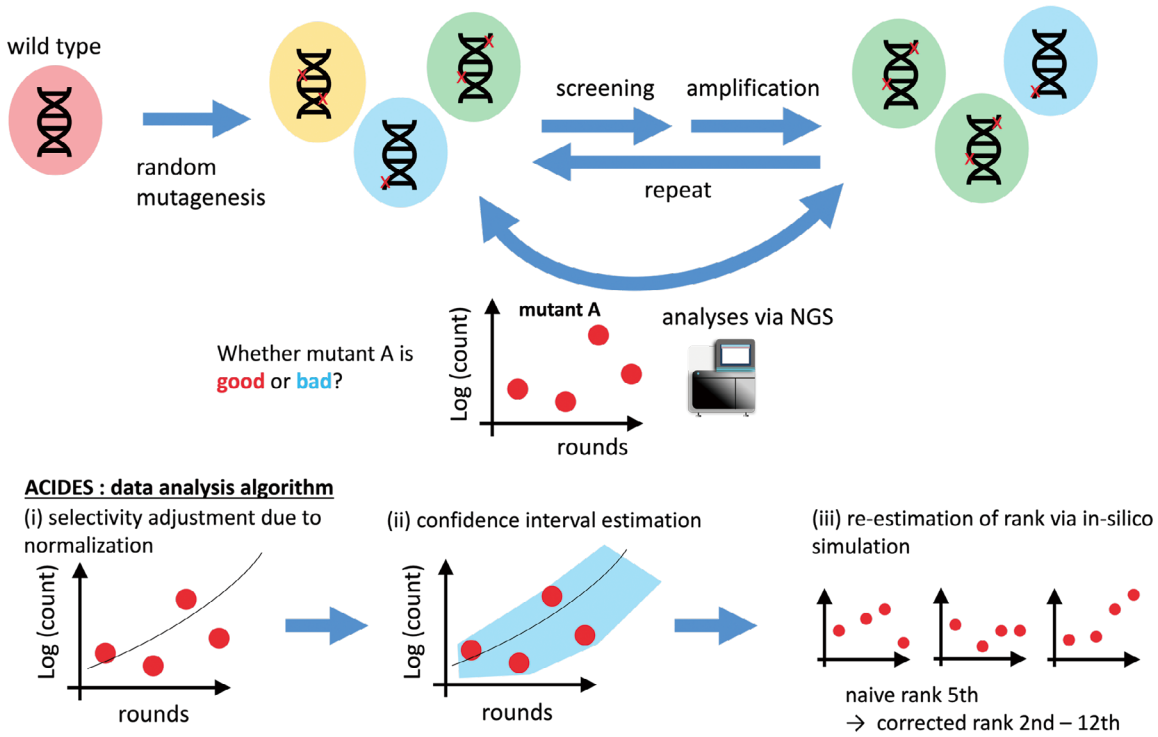
Background & Results

Protein screening involves examining proteins that have undergone random genetic mutations to understand the effects of these mutations on the proteins or to design new proteins. It includes techniques like Deep Mutational Scanning, used to study the receptor-binding domain of SARS-CoV-2, and Directed Evolution, which led to the Nobel Prize in Chemistry in 2018. Especially when combined with NGS, which can simultaneously determine millions of DNA sequences, researchers can obtain large datasets on proteins being screened, and this has recently attracted attention for applications using machine learning. However, statistical errors from NGS are generally high, and their impact on protein screening has not been well understood.

In this study, in collaboration with Deniz Dalkara and Ulisse Ferrari from the Institute de la Vision, we investigated how statistical errors from NGS propagate in protein screening experiments. Highly dispersed NGS noise, amplified through processes such as PCR amplification, is known to be statistically described by a negative binomial distribution. We integrated this statistical model with a mathematical model of protein screening experiments using maximum likelihood approach. By systematically comparing it with existing algorithms, we found that our approach allows for unprecedented accuracy in evaluating protein screening experiments. The proposed algorithm, publicly available under the name ACIDES (Accurate Confidence Intervals for Directed Evolution Scores), offers improved precision in protein screening analysis.

Significance of the research and Future perspective

By applying ACIDES, the experimental pipeline for directed evolution could be designed with greater precision, leading to the development of viral vectors that will advance the field of gene therapy. The research team at the Institute de la Vision has a proven track record of developing an optimized adeno-associated virus variant (AAV7m8) for gene delivery to retinal cells. Given the active research community focused on improving adeno-associated viruses through directed evolution and protein screening, ACIDES could play a significant role in the future for gene therapy development.



Patent

Treatise Nemoto, Takahiro; Dalkara, Deniz; Ferrari, Ulisse et al. ACIDES: on-line monitoring of forward genetic screens for protein engineering. Nature Communications. 2023, 14, 8504. doi: 10.1038/s41467-023-43967-9

URL

Keyword gene therapy, directed evolution, next-generation sequencing, algorithm developments