An Interactive Review of In-silico Prime Editing Guide Design Tools

Abstract

Prime editing is a novel genome editing technology that enables precise base editing without the need for double-strand breaks. The design of prime editing guides is a critical step in the prime editing workflow. In this review, we evaluate the performance of several in-silico prime editing guide design tools. We compare the quality of the guide designed by these tools, and to improve the usage of these state of art tools, they were reimplemented and integrated into a single web base application. Additionally, we provided the ability to aggregate the results from multiple tools using ensemble learning to improve the overall guide design quality.

Keywords: Prime editing, Machine Learning, in-silico tools, Ensemble Learning

1 Background

Prime editing is a versatile and precise genome editing technology that enables the introduction of all 12 possible base-to-base conversions as well as insertions and deletions without the need for double-strand breaks[1].

The versatility of prime editors comes from the fusion of a reverse transcriptase (RT) and a Cas9 nickase (nCas9) to a prime editing guide RNA (pegRNA) (Add figure here on prime editing process)

In-silico Prime Editing Guide Design Tools

2 Methods

Data Acquisition and Preprocessing

The dataset used in this study was obtained from the DeepPrime and PRIDICT study[**mathisMachineLea**
yuPredictionEfficienciesDiverse2023], which contains 220,000 and 20,000 prime editing guides, respectively.

Ensemble Learning

Three ensemble learning approaches were investigated in this study: weighted average, bagging and AdaBoost. The algorithms were implemented in Python, but without the use of Scikit-learn ensemble library, as it does not support having different types of base learners in the ensemble.

Details of the implementation can be found in subsection A.1.

However, no significant difference in performance was observed among the three ensemble learning methods, possibly due to the high correlation in error between the base models (Add figure here). The weighted average method was chosen for the final implementation due to its simplicity and ease of interpretation.

3 Results

4 Discussion

A Appendix

A.1 Ensemble Learning Methods

References

[1] Liu David R. et al. "Search-and-Replace Genome Editing without Double-Strand Breaks or Donor DNA". In: *Nature* 576.7785 (Dec. 5, 2019), pp. 149–157. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-019-1711-4. URL: https://www.nature.com/articles/s41586-019-1711-4 (visited on 02/08/2024).