Proposal for PhD Research: Quantifying the Uncertainty in both On and Off-target Activity of Prime Editing

Peiheng Lu

November 18, 2024

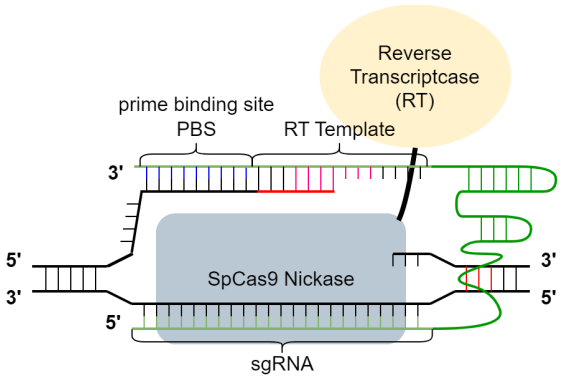
Prime editing is a novel gene editing technology that enables theoretically any base substitution, insertion, and deletion without the need for double-strand breaks or donor DNA templates[1, 2]. Shown in Figure 1a, the prime editors's versatility can mostly be attributed to the reverse transcription mechanism. After the SpCas9 nickase cleaves the target DNA strand, the PBS (prime binding site) portion of the pegRNA (prime editing guide RNA) hybridizes with the now floating 3' end of the target DNA strand. The reverse transcriptase (RT) then extends the 3' end using the RT template sequence in the pegRNA. The resulting DNA-RNA hybrid is then resolved by the endogenous DNA repair machinery, installing the desired edit into the genome.

Due to its superb versatility, prime editors have the potential to cure up to 90% of diseases resulted from genetic mutations[3]. However, its clinical application is still limited by the frequently low on target efficiency and the potential off-target activity[4].

My master's thesis focused on predicting the on-target activity of prime editing using ensemble learning techniques and the Transformer model to improve pegRNA guide design qualities and thus overall efficiency, and achieved state-of-the-art performance on several existing datasets. As a result, moving on to the prediction of off-target activity of prime editing is a natural next step.

A number of off-target prediction tools have been developed for CRISPR-Cas9, which makes up part of the prime editing system, while DeepPrime-Off developed via fine tuning DeepPrime on the off target sites remains as one of the few prime editing off-target prediction tools. Although the CRISPR-Cas9 system was an essential part of the prime editing system, as studies have pointed out, the reverse transcription process involving the PBS and RT sequences can also impact the off-target activity. Thus, novel tools are needed to produce more accurate predictions the off-target activity of prime editing[5, 6].

At the same time, with the recent development of a number of off-target site detection protocols supporting prime editors (such as Tracking-seq), it is now more and more possible to quantify their off target activity using big data and deep learning techniques[5, 7].



(a) Prime editing process

In addition to including off-target prediction, I also propose for a fundamental change in the model's architecture. Although majority of the existing solution (including my master's thesis) was aiming at producing a point estimation as the predicted editing efficiency[8, 6, 9], the energy-driven binding and cleavage processes are stochastic in nature[10]. Thus, it could be more informative to model the posterior distribution of the activity of prime editing given the target loci and pegRNA sequence. An example benefit would be allowing researchers to select the guide design that has smaller variance in editing outcome, if multiple guides showed similar mean editing efficiency.

Recently, a probabilistic deep learning model, crispAI, was proposed to fit a zero inflated negative binomial distribution to the off-target activity of CRISPR-Cas9 on each potential target site[11]. It was able to accuracy predict the off-target activity in both minimal activity and high activity sites, often outperforming existing off target prediction tools such as CRISPRoff and CRISOT.

To validate the probabilistic approach, I conducted some preliminary experiments on the on-target prediction of prime editing efficiency using the mathis et al's 90k PE2 HEK293T dataset(detail of which can be found here)[12].

On top of updating the output itself, additional methods including Monte Carlo dropout[13]/Deep ensemble can also be relatively easily applied to gauge the confidence of the model's prediction.

Overall, with the on and off target activity of prime editing quantified, we can provide a complete overview of the outcome of using a specific pegRNA sequence on a specific target loci in a specific cell line. This should noticeably improve the safety and efficiency of prime editing, and thus accelerate its clinical application, which would be a long-term goal of my PhD research.

Last year saw the first FDA approved CRISPR therapy - Casgevy - for the treatment of sickle cell disease, which took a remarkably short time of 11 years to go from the first CRISPR-Cas9 paper to the first FDA approval[14]. This is highly inspiring for more advanced CRISPR technologies including base editing and prime editing, both of which have been shown to have great potential in curing genetic diseases. I am planning to make contact with various wet labs to validate the model's prediction on

their prime editing experiments, understanding their needs and better incorporating this in silico guide design model into their workflow.

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).

[2] David R. Liu and Peter J. Chen. "Prime Editing for Precise and Highly Versatile Genome Manipulation". In: *Nature Reviews Genetics* 24.3 (Mar. 2023), pp. 161–177. ISSN: 1471-0056, 1471-0064. DOI: 10.1038/s41576-022-00541-1. URL: https://www.nature.com/articles/s41576-022-00541-1 (visited on 02/16/2024).

[3] Ariel Kantor, Michelle McClements, and Robert MacLaren. "CRISPR-Cas9 DNA Base-Editing and Prime-Editing". In: *International Journal of Molecular Sciences* 21.17 (Aug. 28, 2020), p. 6240. ISSN: 1422-0067. DOI: 10.3390/ijms21176240. URL: https://www.mdpi.com/1422-0067/21/17/6240 (visited on 02/08/2024).

[4] Zhihan Zhao et al. "Prime Editing: Advances and Therapeutic Applications". In: *Trends in Biotechnology* 41.8 (Aug. 1, 2023), pp. 1000–1012. ISSN: 0167-7799, 1879-3096. DOI: 10.1016/j.tibtech.2023.03.004. pmid: 37002157. URL: https://www.cell.com/trends/biotechnology/abstract/S0167-7799(23)00085-9 (visited on 10/16/2024).

[5] Shun-Qing Liang et al. "Genome-Wide Profiling of Prime Editor off-Target Sites in Vitro and in Vivo Using PE-tag". In: *Nature Methods* 20.6 (June 2023), pp. 898–907. ISSN: 1548-7105. DOI: 10.1038/s41592-023-01859-2. URL: https://www.nature.com/articles/s41592-023-01859-2 (visited on 09/30/2024).

[6] Goosang Yu et al. "Prediction of Efficiencies for Diverse Prime Editing Systems in Multiple Cell Types". In: *Cell* 186.10 (May 2023), 2256–2272.e23. ISSN: 00928674. DOI: 10.1016/j.cell.2023.03.034. URL: https://linkinghub.elsevier.com/retrieve/pii/S0092867423003318 (visited on 05/03/2024).

[7] Ming Zhu et al. "Tracking-Seq Reveals the Heterogeneity of off-Target Effects in CRISPR–Cas9-mediated Genome Editing". In: *Nature Biotechnology* (July 2, 2024), pp. 1–12. ISSN: 1546-1696. DOI: 10.1038/s41587-024-02307-y. URL: https://www.nature.com/articles/s41587-024-02307-y (visited on 09/28/2024).

[8] Nicolas Mathis et al. "Machine Learning Prediction of Prime Editing Efficiency across Diverse Chromatin Contexts". In: *Nature Biotechnology* (June 21, 2024), pp. 1–8. ISSN: 1546-1696. DOI: 10.1038/s41587-024-02268-2. URL: https://www.nature.com/articles/s41587-024-02268-2 (visited on 06/23/2024).

[9] Jonas Koeppel et al. "Prediction of Prime Editing Insertion Efficiencies Using Sequence Features and DNA Repair Determinants". In: *Nature Biotechnology* 41.10 (Oct. 2023), pp. 1446–1456. ISSN: 1087-0156, 1546-1696. DOI: 10.1038/s41587-023-01678-y. URL: https://www.nature.com/articles/s41587-023-01678-y (visited on 02/07/2024).

[10] Florian Störtz, Jeffrey K. Mak, and Peter Minary. "piCRISPR: Physically Informed Deep Learning Models for CRISPR/Cas9 off-Target Cleavage Prediction". In: *Artificial Intelligence in the Life Sciences* 3 (Dec. 2023), p. 100075. ISSN: 26673185. DOI: 10.1016/j.ailsci.2023.100075. URL: https://linkinghub.elsevier.com/retrieve/pii/S2667318523000193 (visited on 10/14/2024).

[11] Furkan Özden and Peter Minary. "Learning to Quantify Uncertainty in Off-Target Activity for CRISPR Guide RNAs". In: *Nucleic Acids Research* (Sept. 14, 2024), gkae759. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkae759. URL: https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkae759/7757940 (visited on 09/28/2024).

[12] Nicolas Mathis et al. "Predicting Prime Editing Efficiency and Product Purity by Deep Learning". In: *Nature Biotechnology* 41.8 (Aug. 2023), pp. 1151–1159. ISSN: 1087-0156, 1546-1696. DOI: 10.1038/s41587-022-01613-7. URL: https://www.nature.com/articles/s41587-022-01613-7 (visited on 04/24/2024).

[13] Yarin Gal and Zoubin Ghahramani. *Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning*. Oct. 4, 2016. DOI: 10.48550/arXiv.1506.02142. arXiv: 1506.02142. URL: http://arxiv.org/abs/1506.02142 (visited on 10/26/2024). Pre-published.

[14] *CRISPR Clinical Trials: A 2024 Update*. Innovative Genomics Institute (IGI). URL: https://innovativegenomics.org/news/crispr-clinical-trials-2024/ (visited on 11/15/2024).