

FantasticLamp: a genome graph based pipeline for calculating the efficacy of genomic edits

Casper Schutte^{1*}, Ian Fiddes^{1*}, and Erik Garrison^{1*¶}

¹ The University of Tennessee Health Science Center, United States ¶ Corresponding author * These authors contributed equally.

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Open Journals](#) ↗

Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

Accurately calculating the efficacy of genomic edits is crucial to understanding the performance of the editing techniques, in order to optimize methods and improve the success rate of the edits. Additionally, by understanding the success rate of genomic edits, researchers can identify any potential problems or limitations of the techniques and work to overcome them. This can help to improve the accuracy and precision of the editing methods, which is essential for many applications, such as creating genetically engineered cells for therapeutic purposes ([Doudna & Charpentier, 2014](#)), understanding gene functions ([Liu et al., 2019](#)), and studying genetic diversity in a population of cells ([Hsu et al., 2014](#)). FantasticLamp is an open source pipeline for calculating the efficacy of genomic edits performed on multiple populations of cells. It constructs a genome graph from the reference (unedited) genome with nodes representing sequences and edges representing the relationship between them. It then maps reads from the edited populations onto this graph in order to calculate the coverage of the edited sequences compared to the unedited sequences. This pipeline aims to provide a quantitative measure of the success of each genomic edit.

Statement of need

In the field of genome engineering, researchers often perform genomic edits in cells, such as CRISPR/Cas9, TALEN, and ZNF-based systems, to study gene functions and to create new cell lines ([Gaj et al., 2013](#)). However, these edits may not always be successful, and it may be challenging to identify and quantify the success rate of these edits in large-scale data sets ([Güell et al., 2014](#)), ([Haasteren et al., 2020](#)). Compared to linear alignment methods, genome graphs can provide a more accurate and comprehensive view ([Garrison et al., 2018](#)) of the relationships between sequences ([Paten et al., 2017](#)), and by using this method, one can identify and quantify the success rate of genomic edits. FantasticLamp is a pipeline that consists of a bash script that can be initiated from the command line. The bash script calls two Python scripts and (the bioinformatics tools) vg, odgi, and minimap2 (refs. and very short descriptions?).

FantasticLamp was designed to calculate the coverage of genomic edits in multiple populations of cells, simultaneously. Given a reference genome and reads sequenced from multiple edited populations, the pipeline uses a design library CSV file, which contains the intended edits and reference sequences at the intended edit sites, to construct a genome graph. The genome graph is made up of the reference genome, the intended edit sequences ("homology arms"), and the reference sequences at the edit sites ("reference homology arms"). The pipeline then maps reads from the edited populations to this graph using vg, a graph-based alignment tool, creating a GAF (ref) file representing the alignment. The pipeline can handle both paired-end and single-end reads.

A Python script is called to parse the alignment file and calculate the coverage of the homology arms compared to the reference homology arms. The output of this pipeline is a coverage table, which displays the name of each intended edit sequence along with the homology arm coverage and reference homology arm coverage. This allows FantasticLamp to simultaneously quantify the efficacy of edits in multiple edited populations, which can be used to test novel editing methods as well as to verify current methods in experiments that involve genomic edits. By using reads aligned to a graph instead of linear alignment, FantasticLamp can handle the complexity of genomic data, making it a useful tool for researchers in the field of genome engineering and other related fields.

Motivate importance. Compare to linear methods. Discuss shortcomings. Stay between 250 and 1000 words (can be more, but keep it as short as possible). Repo needs to adhere to specific criteria mentioned under "Submissions". A simple figure would be a good idea as well (awaiting feedback) Add a way for the reviewer to test the software themselves. Can use test data set? Does it matter that it is so small? Otherwise need to make one of the real sets smaller, and somehow make sure to include numerous reads that actually map to both the homology arms and ref homology arms. Also need to remember that this paper needs to be in the repo in the form of a markdown doc.

See example paper and references on JOSS website. There are mandatory metadata that need to be included (created markdown doc, will push this doc when I make sure the fields are correct).

Acknowledgements

We acknowledge ...

References

Citations to entries in paper.bib should be in [rMarkdown](#) format.

Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346(6213), 1258096.

Gaj, T., Gersbach, C. A., & Barbas III, C. F. (2013). ZFN, TALEN, and CRISPR/cas-based methods for genome engineering. *Trends in Biotechnology*, 31(7), 397–405. <https://doi.org/10.1016/j.tibtech.2013.04.004>

Garrison, E., Sirén, J., Novak, A. M., Hickey, G., Eizenga, J. M., Dawson, E. T., Jones, W., Garg, S., Markello, C., Lin, M. F., & others. (2018). Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nature Biotechnology*, 36(9), 875–879.

Güell, M., Yang, L., & Church, G. M. (2014). Genome editing assessment using CRISPR genome analyzer (CRISPR-GA). *Bioinformatics*, 30(20), 2968–2970.

Haasteren, J. van, Li, J., Scheideler, O. J., Murthy, N., & Schaffer, D. V. (2020). The delivery challenge: Fulfilling the promise of therapeutic genome editing. *Nature Biotechnology*, 38(7), 845–855.

Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 157(6), 1262–1278.

Liu, J.-J., Orlova, N., Oakes, B. L., Ma, E., Spinner, H. B., Baney, K. L., Chuck, J., Tan, D., Knott, G. J., Harrington, L. B., & others. (2019). CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature*, 566(7743), 218–223.

Paten, B., Novak, A. M., Eizenga, J. M., & Garrison, E. (2017). Genome graphs and the evolution of genome inference. *Genome Research*, 27(5), 665–676.