Implementation of Galaxy to expand the use of command-line-based bioinformatics analysis of genome editing experiments

Alexis L. Norris[1], Mike Mikailov[2], Fu-Jyh Luo[2], Nadia Guimont[2], Heather Lombardi[1], Daniel A. Tadesse[3], Nicholas Petrick[2]

[1]Center for Veterinary Medicine (CVM), Office of New Animal Drug Evaluation (ONADE), Division of Animal Bioengineering and Cellular Therapies (DABCT)
[2]Center for Devices and Radiological Health (CDRH), Office of Science and Engineering Laboratories (OSEL), Division of Imaging Diagnostics and Software Reliability (DIDSR)
[3]Center for Veterinary Medicine (CVM), Office of Research (OR), Division of Animal and Food Microbiology (DAFM)

Background

HPC: Standard NGS tools are command-line based

The evaluation of next generation sequencing (NGS) data requires rigorous, reproducible bioinformatics analysis pipelines, as well as the flexibility to rapidly deploy new pipelines with technology changes. NGS data continue to grow in both size and complexity, overwhelming the ability of most computer systems to store, analyze, and transfer the data, necessitating the use of high performance computing (HPC) environments.

Galaxy: Command-line tools in a web browser

One drawback to command-line-based HPC is that it requires some programming/coding experience, and thus requires some bioinformatics training. For those without a programming background, command-line based bioinformatics pipelines can be converted into graphical workflows on the web-based Galaxy[1] platform.

Test Dataset: Screening for genome editing errors

In genome editing experiments, unintended structural alterations may occur and their detection is computationally intensive, thus requiring an HPC. Here we describe our work implementing our command-line-based pipeline[2] into a Galaxy workflow using publicly available data[3] from germline genome editing in cattle for hornlessness.
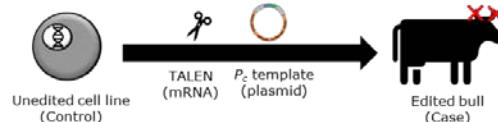
Method

1. **Create** a Galaxy workflow of our command-line pipeline that detects structural alterations.
2. **Validate** the Galaxy workflow using a test dataset.

Test Dataset[3]:
Whole genome sequencing (WGS) data from genome-edited cattle.

Unedited cell line (Control) → TALEN (mRNA) / P_c template (plasmid) → Edited bull (Case)

No coding required.

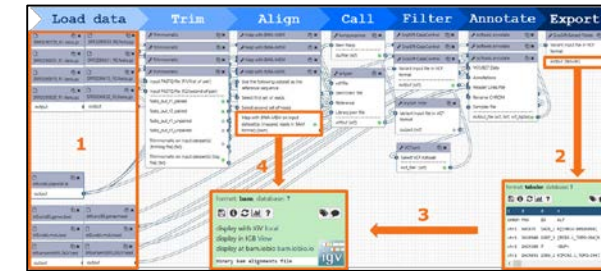The HPC power and tools to analyze omics data in your web browser.

Insert QR code here
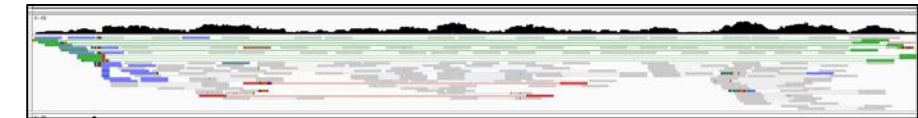
Scan QR code to access additional information online

Results

Galaxy workflow identified the intended genome edit and an unintended plasmid integration in two steps:

Step 1: Identify alterations using the Galaxy workflow.

Step 2: Confirm alterations using Galaxy's IGV extension.

Conclusions & Next Steps

- ☑ Our command-line pipeline to detect structural alterations is now available as Galaxy workflow
- ☑ Galaxy produced expected test dataset's results, without any coding
- ☑ Galaxy meets our regulatory needs (cost, speed, documentation)
- ☐ Develop additional workflows
- ☐ Deploy beta and production versions of Galaxy

References
1. www.usegalaxy.org
2. Norris and Lee, et al. bioRxiv 2019
3. Carlson, et al. Nature Biotech 2016

Acknowledgments