Usage Report for Fiscal Year 2016

Project Title:

FANTOM6 CRISPR target design

Name: Jessica Severin

Laboratory at RIKEN:
RIKEN Center for Life Science Technologies, Division of Genomic Technology

---

The FANTOM6 project is aimed at doing large scale knockdowns of long-non-coding-RNA (lncRNA) in a human cell line to help discover new and novel roles for lncRNA and their possible function in both normal and
disease cells. Next phase of the project will focus on using CRISPR technology to perform the knockdowns.
To help with the picking of potential lncRNA targets and CRISPR sgRNA oligos for knockdown, we planned to
perform a whole genome scan for potential CRISPR sites using the CasFinder.pl program. Initial trial runs on our local cluster give us an estimated run time of 19,000 CPU-hours based on 3100 jobs each taking around 6 CPU-hours with total run time taking 1 to 2months. These jobs also require a large memory footprint (sometimes growing to >64GB each). Our goal and hope was that by using the RICC, potentially with 100-200 jobs in parallel, that we could complete the whole genome search within a week so that we would have time to use the results for planning additional experiments before the end of the fiscal year.

2: System usage
The software package we utilized is called CasFinder and is a set of perl scripts using the bowtie alignment program to search for CRISPR target site and score them for off-target hits. Setup of the software on the HIKUSAI Greatwave ACSL (Application Computing Server with Large Memory) was completed easily. We also tried to run the package on acsg, but the programs used too much memory and the jobs failed to run. We did not try to run the programs on the MPC. Because we could only run on the ACSL we were not able to get the 100-200 parallel jobs that we had hoped for.

[Q16349] FANTOM6 CRISPR target design

Resource unit	Limit	Used	Used(%)	Expiry date
gwmpc	1,011,916.8	-	-	2017-03-31
gwacsg	21,081.6	3.6	0.0%	2017-03-31
gwacsl	3,513.6	532.7	15.2%	2017-03-31
severin	-	536.2	-	-

3: Results
We had hoped to be able to run 100-200 jobs in parallel on this cluster to complete the search quickly, but we only seemed to be able to run around 30 jobs in parallel. In addition, the software package was not very stable and would seem to randomly fail about 10% of the time, which would require jobs to be resubmitted. To perform a whole genome scan for all CRISPR sites we changed our jobs to be smaller but more. Initially we divided the genome into 3100 1-megbase regions, but after initial trials we decided to divided the genome into 31,099 100-kilobase segments. Because of problems with the software and only being able to run on the ACSL and only with ~30 jobs in parallel, it did not appear that we would complete the jobs faster on the HOKUSAI GreatWave than on our own local computers. We were able to submit 986 of the 31,099 jobs (3%) and utilized 15.2% of our CPU limit within the first week of using the HOKUSAI GreatWave. Also since each job took around 45minutes to run, our estimated total run time was approximately 5 weeks and we

would exceed our CPU limit. So we made the decision to run the remaining jobs on our own large memory server even though they would run slower. We also decided to change our timeline expectations about using the resulting data for experiments before the end of the fiscal year.

4: Conclusion

We were glad for the experience of using the HOKUSAI GreatWave servers and learned many things. The servers are very good and easy to use. But we also learned that large-memory jobs, like whole genome scanning for features like CRISPR target sites, are maybe not the best use for the GreatWave servers. The GreatWave servers seem to be better optimized for lower memory, large CPU demand jobs were many jobs can run in parallel. But the ACSL large memory server can still be of use to our group for other less demanding large-memory searches in the future.

5: Schedule

We completed the search using our own large memory CPU server machine in February 2017. We are currently working on post processing the results for loading into the ZENBU system and for use by our team for our next phase of experiments. We hope to publication these results either in a small bioinformations paper, or include this result in a large publication within the near future. When we publish the results we will cite our use of the HOKUSAI GreatWave server in addition to using our own servers.