bioinformatics.ca

CANADIAN BIOINFORMATICS WORKSHOP

Using Clouds for Big Cancer Data Analysis

Saturday, November 2, 2019
8:30am-12:30pm

Registration Fee: $100
Registration for the workshop is required and will be available through CCRC registration.

Target Audience: Graduates, postgraduates, staff bioinformaticians, and PIs who want to know how to access and work with cloud compute infrastructure to analyze their big cancer data sets (e.g. cancer data from the International Cancer Genome Consortium (ICGC) or PanCancer Analysis of Whole Genomes (PCAWG)). Familiarity with the Unix command line interfaces is recommended but not required. A laptop is required.

Course Description: The Cancer Genome Collaboratory is a compute cloud environment that was set up to facilitate analysis on big cancer genome data projects, including the ICGC and PCAWG. The Collaboratory provides access to configurable virtual machines (VM) with which to compute on this data and the Dockstore provides container packages of common genomic analysis tools and workflows. The CBW has developed a half-day course providing a hands-on introduction to launching and configuring your own virtual machine (VM), logging into your VM, copying unaligned sequence files into your VM, and performing sequence analysis on these files and saving the results.

Participants will gain practical experience and skills to be able to:

- Understand the difference between HPC and Cloud
- Understand containers and find workflows in Dockstore
- Launch a VM on Collaboratory (VM)
- Run BWA-Mem alignment
- Run a PCAWG variant calling workflow

Schedule:

8:30	Welcome (Michelle Brazas) Module 1: Introduction to the Collaboratory and Dockstore (Jared Baker, Denis Yuen) • Introduction to Cloud computing vs. HPC • The Cancer Genome Collaboratory Cloud • Introduction to containers and Dockstore
9:00	Module 1: Lab Practical • Connect to the Cancer Genome Collaboratory console • Upload your cryptographic key • Launch a personal Virtual Machine (VM) • Log into your VM • Copy unaligned genomic sequence into your VM

10:30	Coffee Break
10:45	Module 1: Lab Practical Continued • Run the BWA-Mem sequence aligner • Save your results • Run the PCAWG Sanger variant calling workflow • Shut down your VM
12:15	Module 2: Resources and Research on Big Cancer Datasets • Large-scale research projects that generate datasets used in the Cloud • Searching for ICGC data stored on Collaboratory • How to access data in the Cloud and accessing non-protected data
12:30	Survey and Closing Remarks

We reserve the right to cancel this workshop if we experience poor attendance. Notification and refunds would be provided by October 31, 2019.