Title: RNA-seq data of invasive ductal carcinoma and adjacent normal tissues from a Korean patient with breast cancer

Group 4:

Nikhita Gogate, Nuerye Ainiwan, Christian Zevallos Delgado, Wei Shao

The publication used for building the BCO utilized 7 tools: Trim Galore v 0.4.2 with a prerequisite of Cutadapt v 1.1.2 in Step 1, STAR v 2.5.2.b in Step 2, Sambamba v 0.6.5 in Step 3, RSeQC v 2.6.4 in Step 4, Cufflinks v 2.2.1 with cuffnorm function in Step 5, Cufflinks v 2.2.1 with cuffdiff function in Step 6, and Metascape v 3.0 in Step 7.

Trim Galore is a wrapper tool used to apply quality and adapter trimming to Fastq files. This tool was found in CGC but was not found in the HIVE Platform. Equivalent tools such as Adapters FIlter, Quality and Length Filter, Sequence Filter, SequenceTrimmer are found on HIVE.

STAR is a tool used to map sequencing reads to the reference genome (hg38). This tool was found in CGC but was not found in the HIVE Platform. Equivalent tools such as HISAT2 and Tophat are available on HIVE.

Sambamba is a tool used to remove PCR-duplicates of mapped reads. This tool was found in CGC but was not found in the HIVE Platform. Equivalent tool such as Sequence Filter are found on HIVE.

RSeQC is a tool that evaluates the quality of the sequence data based on sequence saturation, mapped read distribution, converge uniformity, strand specificity and transcript level RNA integrity. It outputs a transcript integrity number score. This tool was found in CGC but was not found in the HIVE Platform. Equivalent tool such as MultiQC are found on HIVE.

Cufflinks tool utilizes a cuffnorm functionality that generates abundances of expression levels and normalized fragment counts based on library size. The cuffdiff functionality of Cufflinks is used to identify transcript expression levels and differential expressed genes between cancer and normal samples. This tool was found in CGC but was not found in the HIVE Platform. Equivalent tool such as Gene Expression Analysis portal is found on HIVE. The portal is under implementation and cannot be accessed at the moment.

Metascape is a web-based tool used to perform gene ontology analysis to identify key pathways with differentially expressed genes. This tool was not found in CGC as well as the HIVE Platform. A comparable workflow such as KnowEnG is available on CGC platform. Gene ontology tools are not implemented in HIVE.

An email has been sent to [gwu.hive.team@gmail.com](mailto:gwu.hive.team@gmail.com) (HIVE Platform) and [david.roberson@sbgenomics.com](mailto:david.roberson@sbgenomics.com) (CGC) sharing feedback on the respective tools not found on the platform.