Signature Analysis Pipeline

Cancer Genomics Cloud (CGC) version
QUICKSTART GUIDE

knoweng

Updated: 01/29/2018

This Quickstart guide describes how to run the Signature Analysis (SA) pipeline/workflow on the Cancer Genomics Cloud (CGC).

This KnowEnG workflow matches genomic profiles from a user-supplied query spreadsheet to a library of genomic signatures determined in previous studies. The query genomic profiles will come from different biological samples under investigation and will often be measurement of gene expression. The library of signatures must be provided by the user and should contain similarly derived measurement for specific conditions or treatments. In this Signature Analysis Pipeline, we find the list of common genomic features between the query and library and then rank the library signatures in order of the selected Similarity Measure for each query.

Initial Steps

1. Login to the CGC at https://cgc.sbgenomics.com/.

   If you don't have a CGC account, create one using the "Create a free account" link on this page just below the login box.
2. Click on the project's link to go to the project dashboard.
   If you don't have a project to use, create one using the "Create a project" button.
3. Click on the "Apps" tab on the project dashboard.
   If you don't have the KnowEnG Signature Analysis Workflow app in your project, add it using the "Add app" button:
   a. *Click on the green "Add app" button.*
   b. *Search for the app, e.g., enter “signature analysis” in the search box.*
   c. *Click on the “Copy” button in the app box.*
   d. *Click on the green “Copy” button.*
   e. *Click on the “X” in the top right of the window to close it.*
4. Click on the KnowEnG Signature Analysis Workflow link to go to the app page and view information about the app.

Running Apps

There are actually several ways to run apps on the CGC; for example, on the Apps tab, each app has some Actions, one of which is a green arrow run button; in addition, if you have run the app before, you can go to the Tasks tab, and each task includes an Action to re-run the task; and if you are on a task page, there is a green "Edit and rerun" button.

Run the App

5. Click on the Run button to run the app.

Data Files

6. These instructions use demo input files already available in this project (demo_SA.samples.txt and demo_SA.signatures.txt). For information on uploading your own data to the CGC, see the section "Uploading Files to the CGC" further down in this document.
   a. ***Samples File***:
      i. *Click on the “Select file(s)” button.*
      ii. *Enter "demo" in the search box.*
      iii. *Find the input file in the list of files returned (the file's name is "demo_SA.samples.txt").*
      iv. *Click on the checkbox to the left of the file name.*
      v. *Click on the “Save selction” button.*
   b. ***Signatures File***:
      i. *Click on the “Select file(s)” button.*
      ii. *Enter "demo" in the search box.*
      iii. *Find the input file in the list of files returned (the file's name is "demo_SA.signatures.txt").*
      iv. *Click on the checkbox to the left of the file name.*
      v. *Click on the “Save selction” button.*

About the Demo File: Samples file

The demo samples file used here, demo_SA.samples.txt, contains gene expression data derived from the TCGA GRCh38 data set, as outlined in the description of the Signature Analysis workflow on the CGC.

About the Demo File: Signatures file

The demo signatures file used here, demo_SA.signatures.txt, was taken from the publication [PMID: 20643781], and is available here. (The original name of the file was predictor.centroids.csv.)

Parameters

7. Click on the "Next: Define App Settings" button (or the "Define App Settings" tab).
   The parameters are listed below in their suggested order of entry, although they may be listed in a different order on this page.
   For each parameter, the "?" link contains a description of what the parameter is used for and what type of value is expected.
   a. ***Similarity Measure****: choose a value from the pull-down menu; the possible values are* ***cosine****,* ***pearson****, and* ***spearman****.*
   b. ***Dont Map Samples Flag****: set this value to True if you do not want to map the names in the samples file. (Mapping will be done if this value is not specified, or if it is set to False.)*

c. ***Dont Map Signatures Flag****: set this value to True if you do not want to map the names in the signatures file. (Mapping will be done if this value is not specified, or if it is set to False.)*

d. ***Species Taxon ID****: this value will be used in the gene name mapping process; sample values are listed at KN Contents by Species (the values are listed within parentheses); e.g., "9606" (for human) is a sample value.*
(*Species Taxon ID is listed three times on this page, but the input boxes are linked together so that when you enter the value in one text box it appears in the other****.)***
(*This value is only relevant if mapping is done on at least one of the samples or signatures files.)*

Links

KnowEnG Main Website
KnowEnG Analytics
Knowledge Network Overview
Knowledge-Guided Pipelines
SA Pipeline
Pipeline Quickstart Guides
KnowEnG YouTube Channel

Launch the App

8. When the parameters are all entered, click on the green Run button to start the workflow running.

   The app generally takes a few minutes to run (e.g., 4 to 9 minutes).

Results and Download

9. When the task is finished, the task page will show a green "COMPLETED" image to the left of the task name. The page shows the input files, the parameter settings, and the output files. There are several output files that give information about the run, but the main output file is the **Similarity Matrix File**. The **README** file describes the output files.

   To view an output file, click on the file name, and scroll down to the bottom of the page to view the file's contents (or just a portion of the contents, for large files). (You may need to click on the “Display raw data” button to view the contents.) From that page, you can also Download the file (the "Download" link is available under the "..." "More actions" button).

   You can also download a file, or multiple files, one at a time, from the task page by clicking the folder image "Browse files" button next to a particular file, or the "Outputs" header.

Uploading Files to the CGC

There are several ways to upload files to the CGC, as described on this page:

https://docs.cancergenomicscloud.org/docs/upload-to-the-cgc

The best way will depend on where the files are, how you access them.

If they are on your personal computer, you can use the CGC Uploader GUI tool.

If they are on a server, you can use the Command Line Uploader.

If they are accessible via FTP or the web, you can use the FTP/HTTP(S) import tool.

And you can also upload from a cloud volume, such as Amazon Web Services (AWS) or Google Cloud Storage (GCS).