

HSPH

Quantitative Biomedical Research Center(qBRC)

CNAP System Single-End RNASeq Analysis

Walkthrough

Upload your files



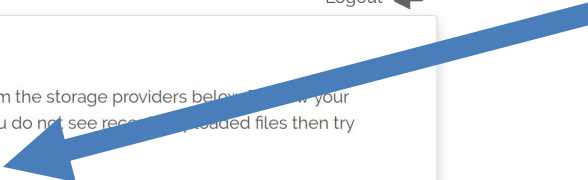
Logout

Uploads

Upload files to our system for analysis by choosing from the storage providers below. To view your currently uploaded files, see the file manager tab. If you do not see recently uploaded files then try refreshing the page.

Dropbox **Google Drive**

Direct to Dropbox/Google Drive login page for file upload



Choose the analysis to run:





Logout 



Analyses

Choose an analysis to run.



Analysis name	Description
Paired-end RNA-Seq basic differential expression 	For determining differential expression from a paired-end RNA-seq experiment
Paired-end RNA-Seq basic differential expression 	For determining differential expression from a basic paired-end RNA-seq experiment



- List of available analysis created for current user
- Opens the analysis in a new tab

A sample analysis page for Single-End RNASeq Analysis



Logout

Single-end RNA-Seq basic differential expression

Use this workflow for aligning with STAR, quantifying, and testing differential expression with DESeq2 for single-end RNA-seq experiment.

Input files:
Choose input fastq-format files to analyze. Files should end with "_R1.fastq.gz".

Filter files...

Select highlighted:

Uploads

Sample annotations:
Choose a sample annotation file. This should have two columns, separated by a "tab" (tab-delimited). The first has the sample names, which need to match the "prefix" of the fastq files you will be using. For example, if the fastq file is named "SampleA_R1.fastq.gz", then the sample name that goes in this file is "SampleA". The second column tells us which group the sample belongs to. The names of the groups NEED to match the values you enter below for the base and experimental groups. There should NOT be a column header. An example may be found [here](#)

Filter files...

Select highlighted:

Uploads

Single-end RNA-Seq basic differential expression (Completed April 16, 2019 (18:18:51))

Reference genome
Choose the reference genome to use for this process.

Select Fastq files for analysis

Reference genome
Choose the reference genome to use for this process.

Ensembl Homo sapiens GRCh38.95

Differential expression contrasts
Choose which groups to compare when performing simple differential expression. Add as many contrasts as you like. Note that the name of the contrast groups must exactly match those provided in the sample annotation file.

Base condition	Experimental condition
<input type="text"/>	<input type="text"/>

Provide a Contrast

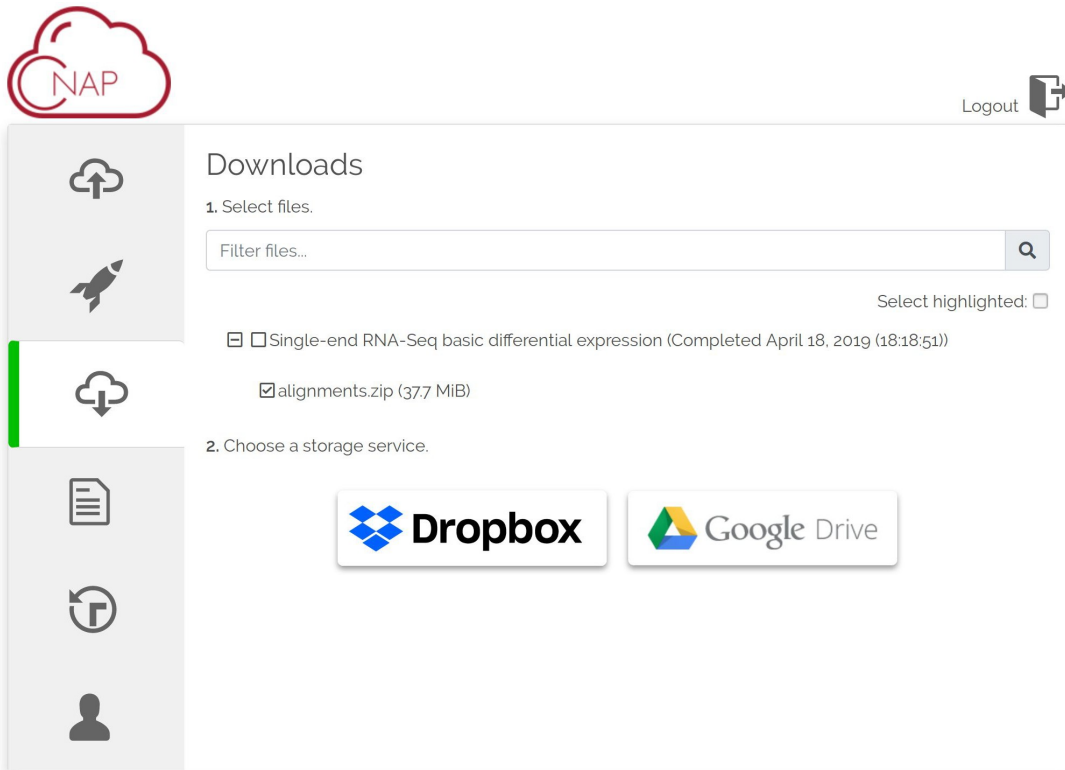
Output zip-archive name:
Name the output file. Result will be a "zip" archive. If you enter "my_output", then the results will be "my_output.zip".

Analyze

Provide an output

Sample Annotation file

Wait for analysis to finish, then download results



The screenshot shows the NAP web interface. At the top left is the NAP logo (a red cloud with 'NAP' inside). At the top right is a 'Logout' button with a document icon. The main content area is titled 'Downloads' and is divided into two steps:


- 1. Select files.** This section includes a search bar labeled 'Filter files...' with a magnifying glass icon. Below the search bar is a checkbox labeled 'Select highlighted:' followed by a smaller checkbox. A file entry is listed: Single-end RNA-Seq basic differential expression (Completed April 18, 2019 (18:18:51)). Underneath this entry is a checked checkbox next to 'alignments.zip (37.7 MiB)'. A green vertical bar is visible on the left side of the interface.
- 2. Choose a storage service.** This section features two buttons: 'Dropbox' with its logo and 'Google Drive' with its logo.


A vertical sidebar on the left contains several icons: a cloud with an upward arrow, a rocket, a cloud with a downward arrow, a document, a circular arrow, and a person silhouette.

Once a project is submitted for analysis, an email will be sent once the analysis is completed.

Result packages will show up here for download to Dropbox/Google Drive


See all your uploaded files (and delete if necessary)



Logout 

Files

Below are the currently active files for use with analyses and for download. Click on a file for details, or select files for deletion. If you have recently uploaded a file that you do not see, try a [refresh](#).

Select highlighted:

Delete selected

- Uploads
 - HBR_Rep1_chr22_R1.fastq.gz (6.3 MiB)
 - HBR_Rep2_chr22_R1.fastq.gz (7.6 MiB)
 - HBR_Rep3_chr22_R1.fastq.gz (6.9 MiB)
 - UHR_Rep1_chr22_R1.fastq.gz (12.8 MiB)
 - UHR_Rep2_chr22_R1.fastq.gz (9.7 MiB)
 - UHR_Rep3_chr22_R1.fastq.gz (10.6 MiB)
 - samples.tsv (113 Bytes)
- Single-end RNA-Seq basic differential expression (Completed April 18, 2019 (18:18:51))

Example Results

Example Output - MultiQC

MultiQC
v1.7

- General Stats
- featureCounts
- Picard
- STAR
- FastQC
- Sequence Counts
- Sequence Quality Histograms
- Per Sequence Quality Scores
- Per Base Sequence Content
- Per Sequence GC Content
- Per Base N Content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Adapter Content



A modular tool to aggregate results from bioinformatics analyses across many samples into a single report.

Report generated on 2019-04-12, 15:57 based on data in: /cromwell_root

Welcome! Not sure where to start?

[Watch a tutorial video](#) (8:00)

[don't show again](#) ✕

General Statistics

[Copy table](#)

[Configure Columns](#)

[Plot](#)

Showing 10 rows and 9 columns.

Sample Name	% Assigned	M Assigned	% Dups	% Aligned	M Aligned	% Dups	% GC	M Seqs
HBR_Rep1_chr22				51.4%	0.1			
HBR_Rep1_chr22.primary_filtered	58.9%	0.0	10.2%					
HBR_Rep1_chr22_R1						42.7%	50%	0.1
HBR_Rep2_chr22				51.1%	0.1			
HBR_Rep2_chr22.primary_filtered	59.7%	0.0	12.2%					
HBR_Rep2_chr22_R1						44.3%	50%	0.1
HBR_Rep3_chr22				51.3%	0.1			
HBR_Rep3_chr22.primary_filtered	58.8%	0.0	10.6%					

Example Output – Analysis Report

Report for alignment and differential expression analysis

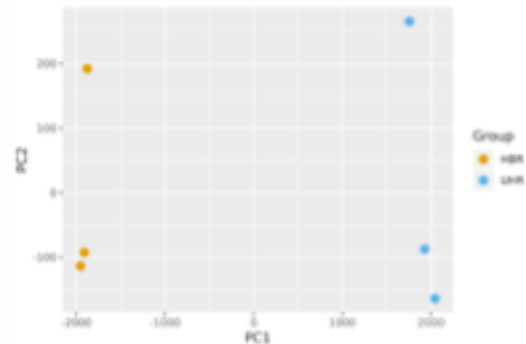
This document discusses the steps that were performed in the analysis pipeline. It also describes the format of the output files and some brief interpretation. For more detailed questions about interpretation of results, consult the documentation of the various tools.

Results:

We summarize some brief results in this section. Full results can be found in the files, as described in the Outputs section.

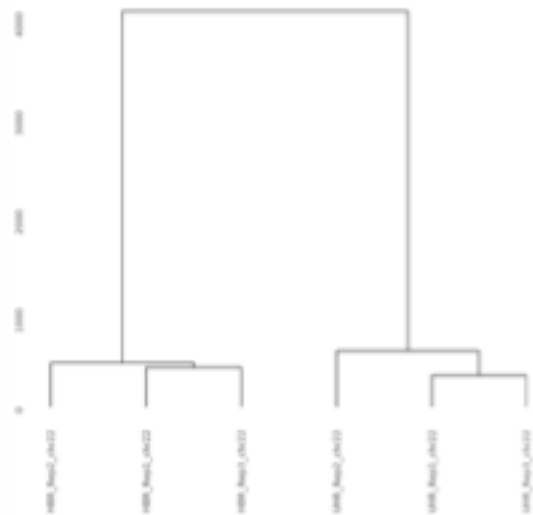
PCA

Principle component analysis (PCA) of the full normalized count matrix was performed. The first two components, PC1 and PC2, are shown. Each sample group are represented with a different color.



Hierarchical clustering of samples

Hierarchical clustering tree of all samples computed with euclidean distance considering all genes.



The following contrasts were performed, yielding the differentially expressed gene counts shown below. The threshold for significance was set such that the adjusted p -value is less than 0.05. For the heatmap figures, the plotted genes were further limited to those with log-fold change magnitudes of 1.5 or greater. When referenced, the "top" genes refers to the 40 genes with lowest p -value.

Experimental condition	Base condition	Upregulated	Downregulated	Result table	Heatmap of significant genes	Heatmap of top DE genes	Volcano
CHIK	HBR	126	128	Table	Figure	Figure	Figure

Outputs:

This section describes the contents of the delivered results.

Alignments

Individual alignment files (in compressed BAM format, ending with "bam") are available for download, but are provided separately due to their typically large size. If you download the BAM files, ensure that you also have the corresponding "index" files, which end with "bai". Index files allow programs like IGV to use the BAM file in an efficient manner.

Main results

The main results are contained in a zip-archive and should be downloaded as "unzipped" on your local computer. It contains several sub-directories which contain files produced in each step of the pipeline.

- **QC**
 - This directory contains an interactive HTML-based QC report which summarizes read quality, alignment quality, and other metrics. It was produced by MultiQC, and information can be found at <https://multiqc.info/>.
 - Other QC plots are provided, produced by the RSeQC tool. See documentation at <http://www.aaref.org/rseqc/> for details on each plot.
- **Quantifications**
 - Quantification tables, which give the number of reads aligned to each gene. Files are tab-delimited. These may be opened with your software of choice, including spreadsheet software such as Excel (note: <https://doi.org/10.1080/10717339-016-1044->

For more information:

<https://www.hsph.harvard.edu/qbrc/services/cloud-services/cnap/>

Quantitative Biomedical Research Center

[Home](#) > [Quantitative Biomedical Research Center](#) > [Services](#) > [Cloud Services](#) > [CNAP](#)

QUANTITATIVE BIOMEDICAL RESEARCH CENTER

Search this section 

[Home](#)

[Services](#) 

[Analysis Service](#)

[Cloud Services](#) 

[CNAP](#)

[WebMeV](#)

[Events](#)

[Publications](#)

CNAP



CNAP (Cloud Native Application Platform) takes advantage of the elasticity, reproducibility, and global access nature of cloud computing technologies to disseminate data analysis pipelines for large data sets such as Next Generation Sequencing (NGS) data. CNAP optimizes for large data file transfer from Dropbox and Google Drive for their ease of use and permeated availability.

CNAP currently offers following services for NGS data analysis:

- Bulk RNA-Seq alignment, quantification, and differential expression (example [output](#))
- DNA Variant calling for both somatic and germline mutations by [GATK](#), with QC by [MultiQC](#)
- The extra-cellular RNA processing toolkit ([exceRpt](#)) for small RNA-Seq

CNAP is designed specifically for experimental focused laboratories with limited computing resources and bioinformatics support. CNAP provides on-demand access to primary NGS analysis with extensive QC such that researchers can immediately evaluate the results of their NGS experiments.

Please email qbrc@hsph.harvard.edu for any inquiries.

For example RNASeq output:

<https://www.dropbox.com/sh/0eoik7bgoe5jcz8/ABdi14CEuamdYEADKze9ISca?dl=0>