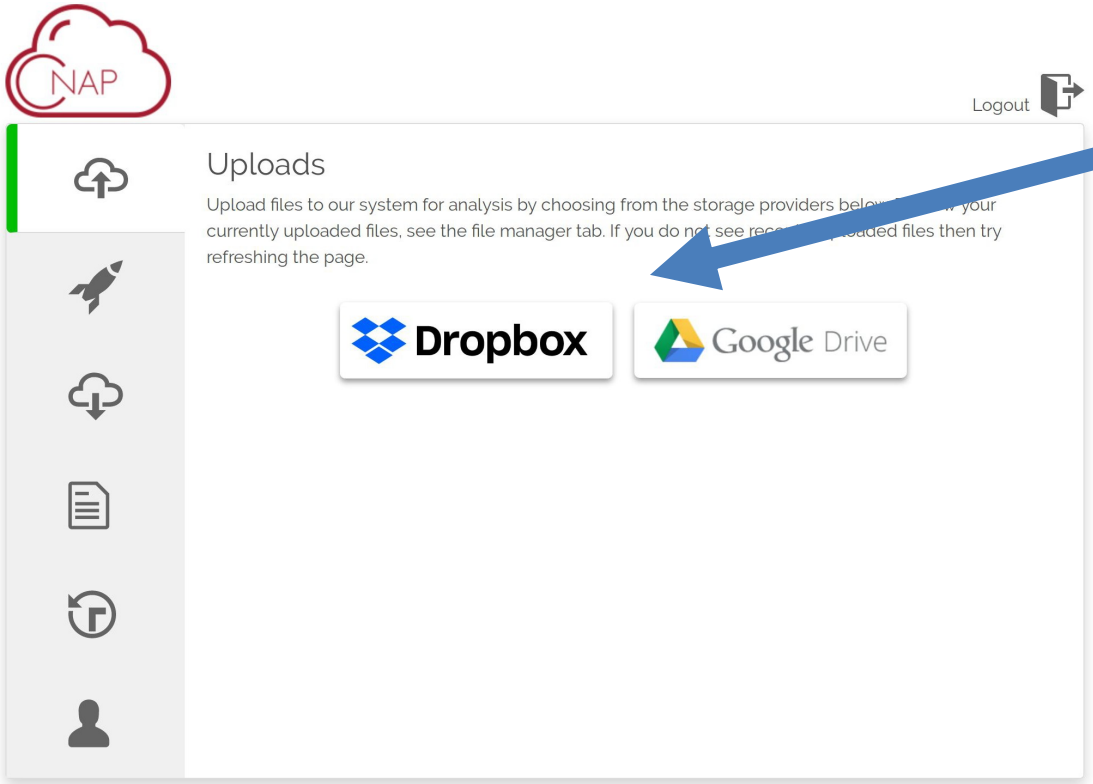
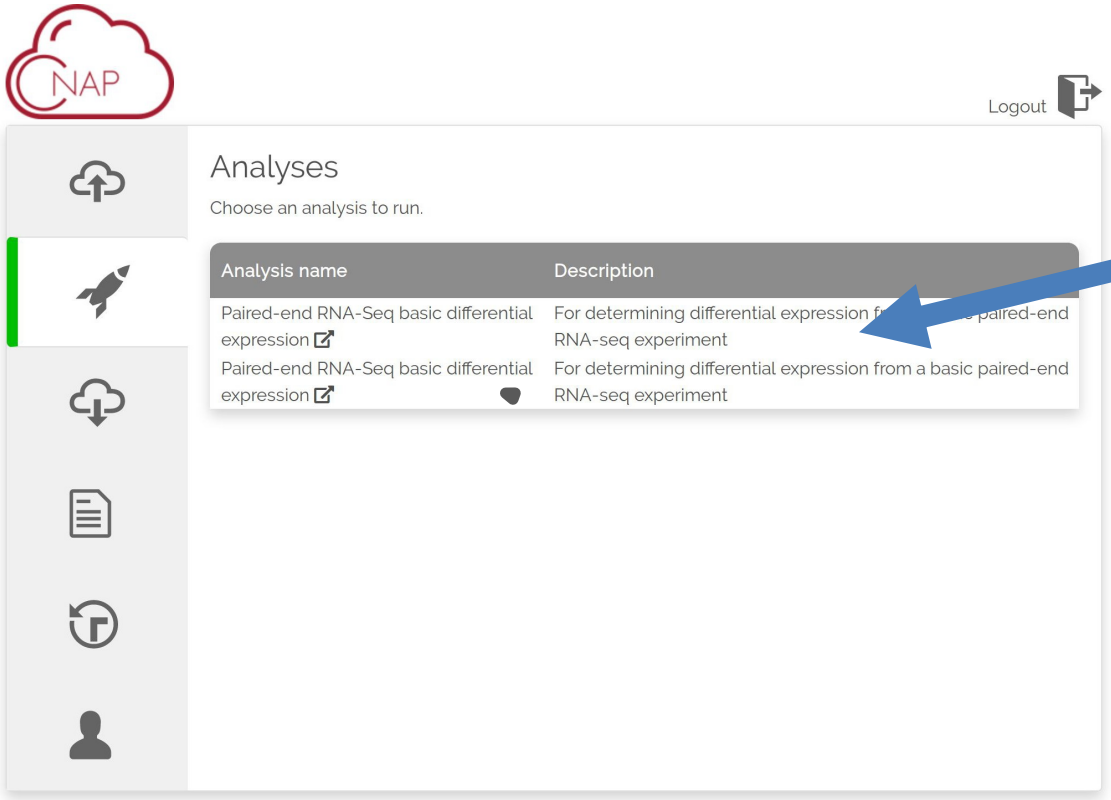
HSPH
Quantitative Biomedical Research Center(qBRC)
CNAP System Single-End RNASeq Analysis Walkthrough

Upload your files



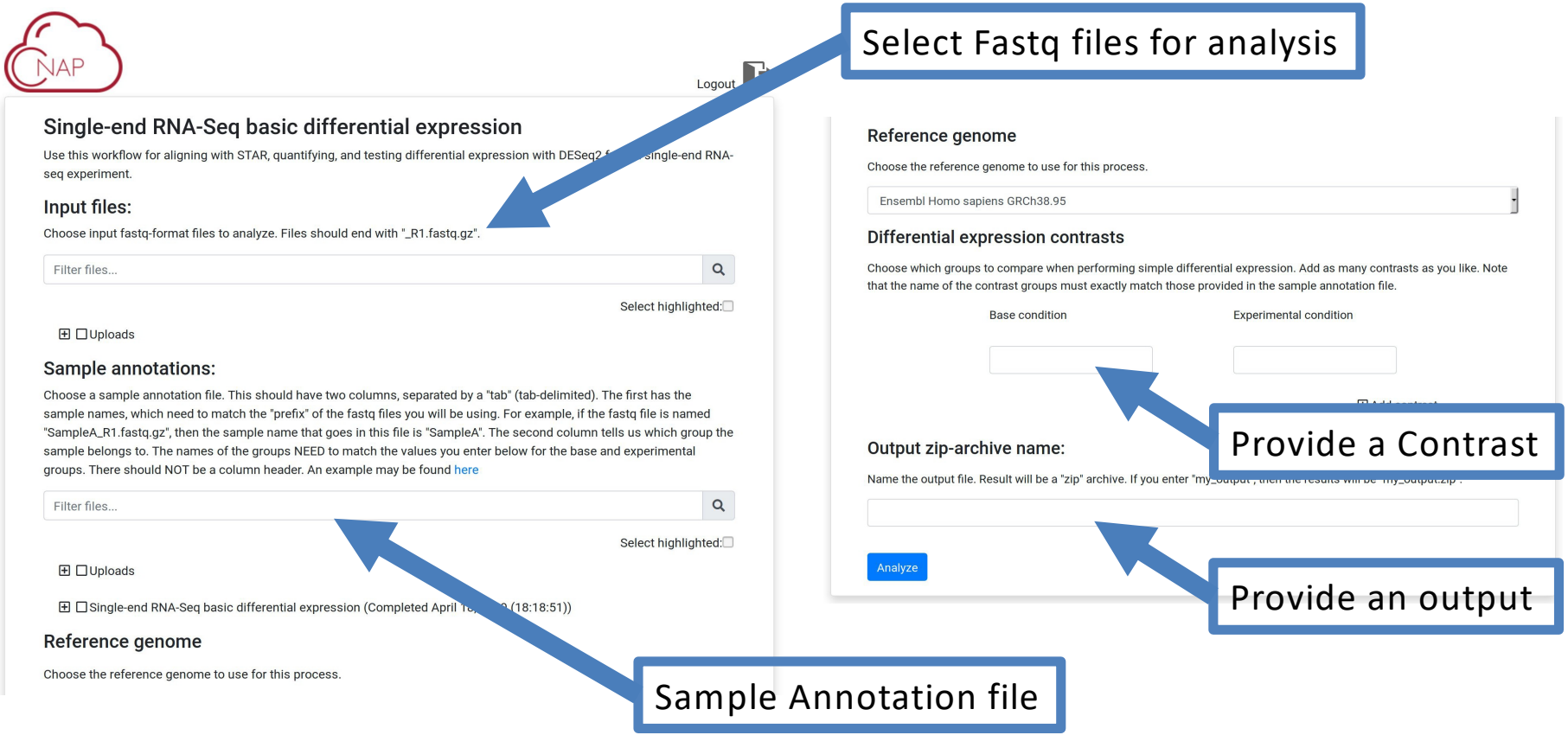
Direct to Dropbox/Google Drive login page for file upload

Choose the analysis to run:

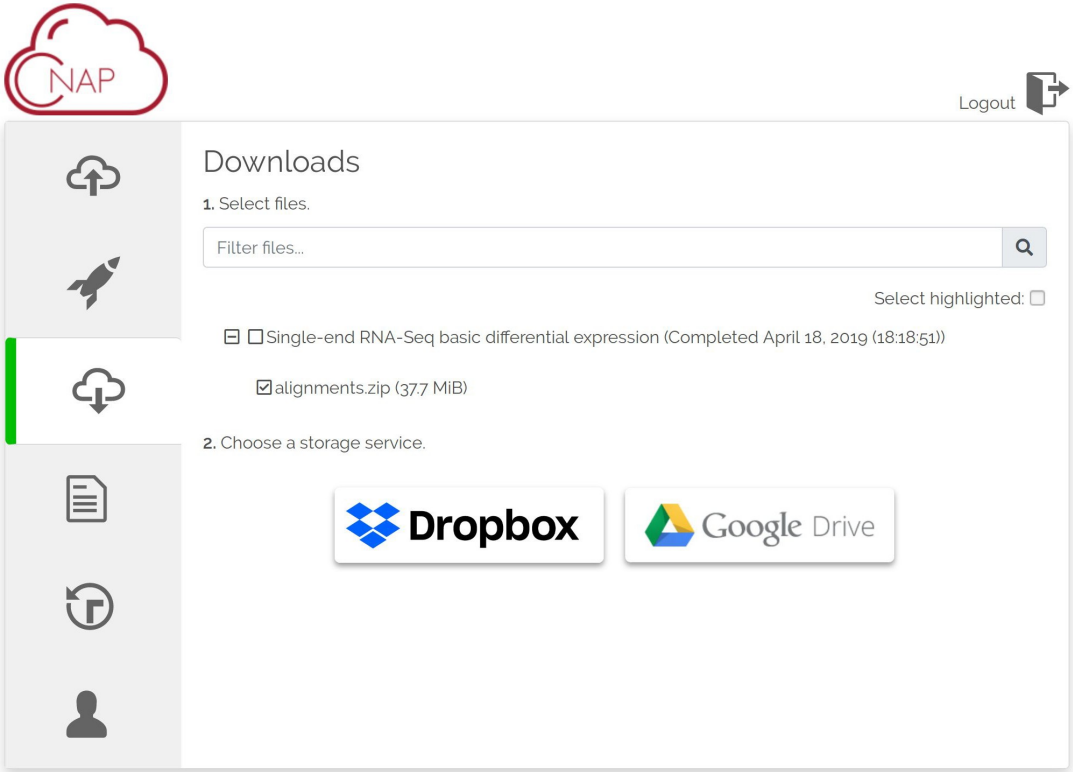


- List of available analysis created for current user
- Opens the analysis in a new tab

A sample analysis page for Single-End RNASeq Analysis



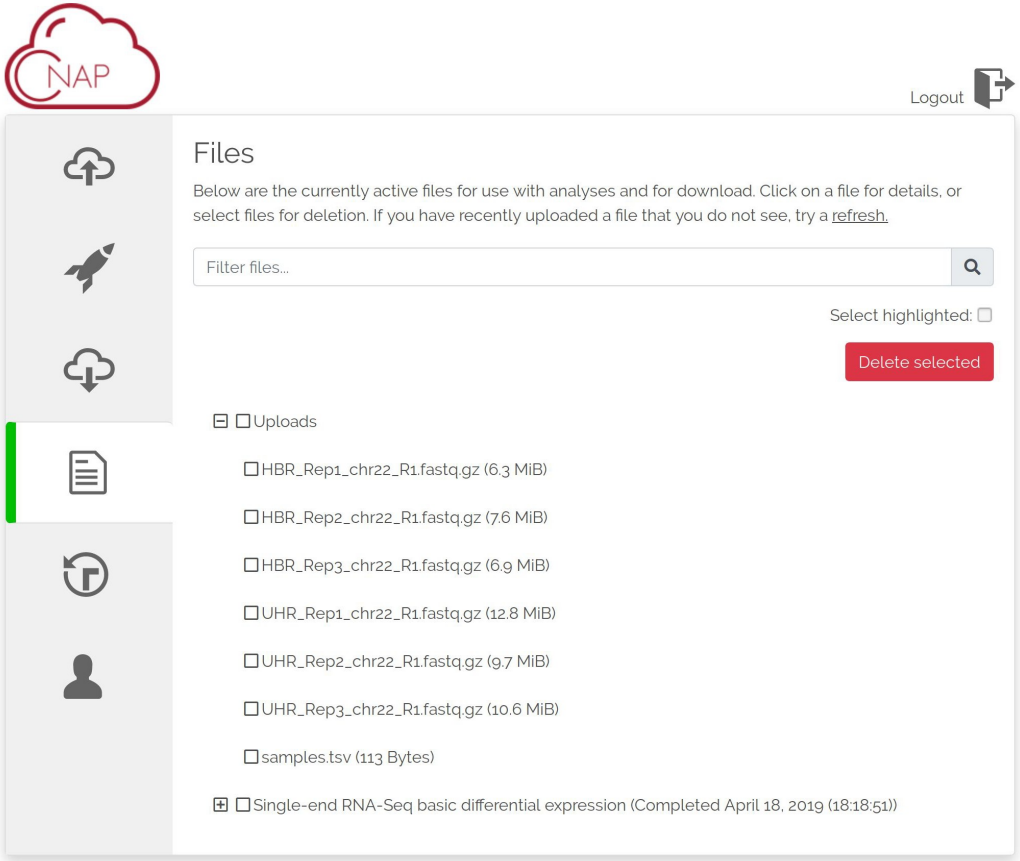
Wait for analysis to finish, then download results



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

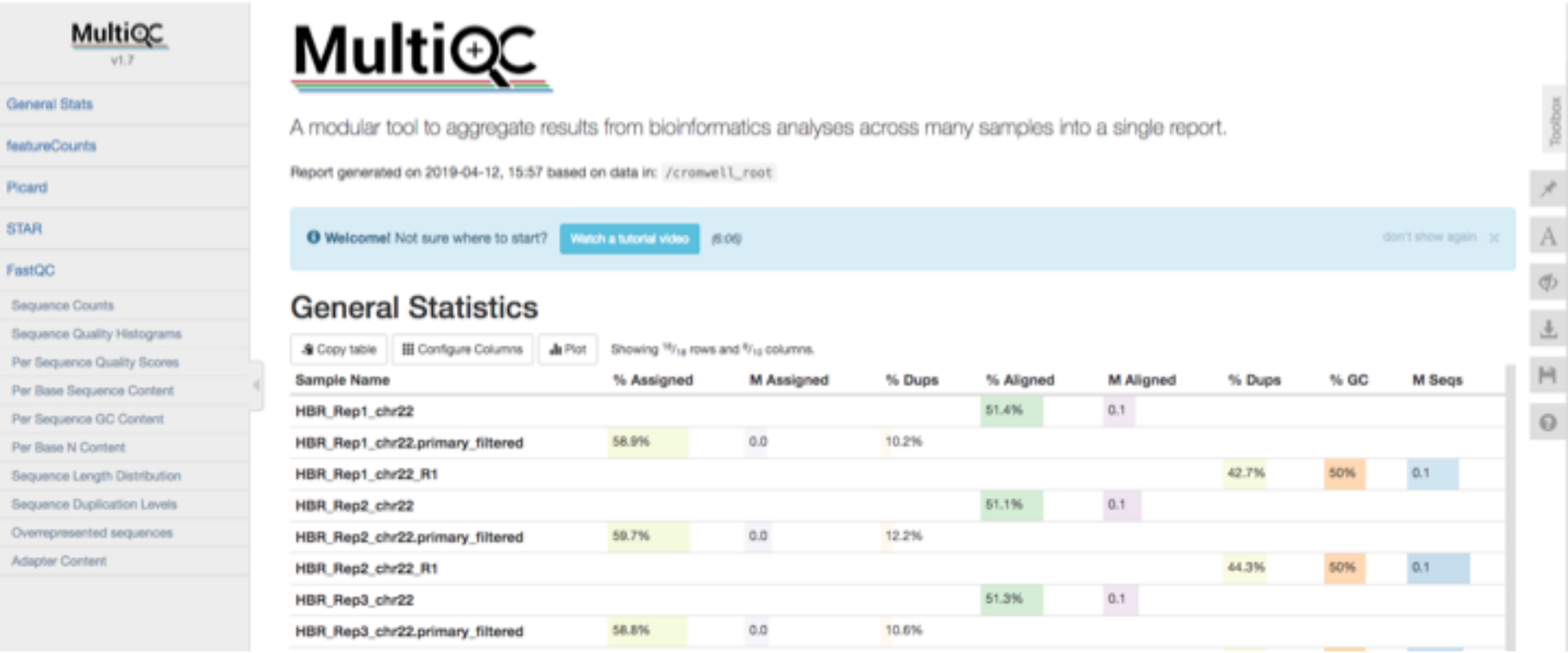
Result packages will show up here for download to Dropbox/Google Drive

See all your uploaded files (and delete if necessary)



Example Results

Example Output - MultiQC



Example Output – Analysis Report

#Report for alignment and differential expression analysis

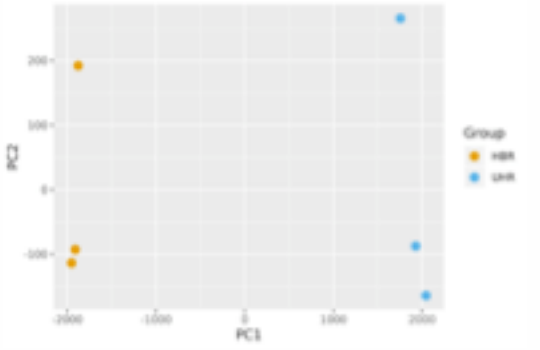
This document discusses the steps that were performed in the analysis pipeline. It also describes the format of the output files and some brief interpretation. For more detailed questions about interpretation of results, consult the documentation of the various tools.

Results:

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

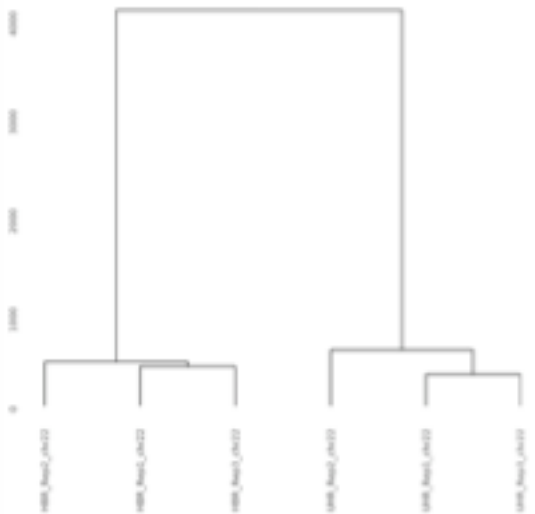
PCA

Principle component analysis (PCA) of the full normalised count matrix was performed. The first two components, PC1 and PC2, are shown. Each sample groups 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 0.01. 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
UHR	HBR	126	128	Table	Figure	Figure	Figure

Outputs:

This section describes the contents of the delivered results.

Alignments

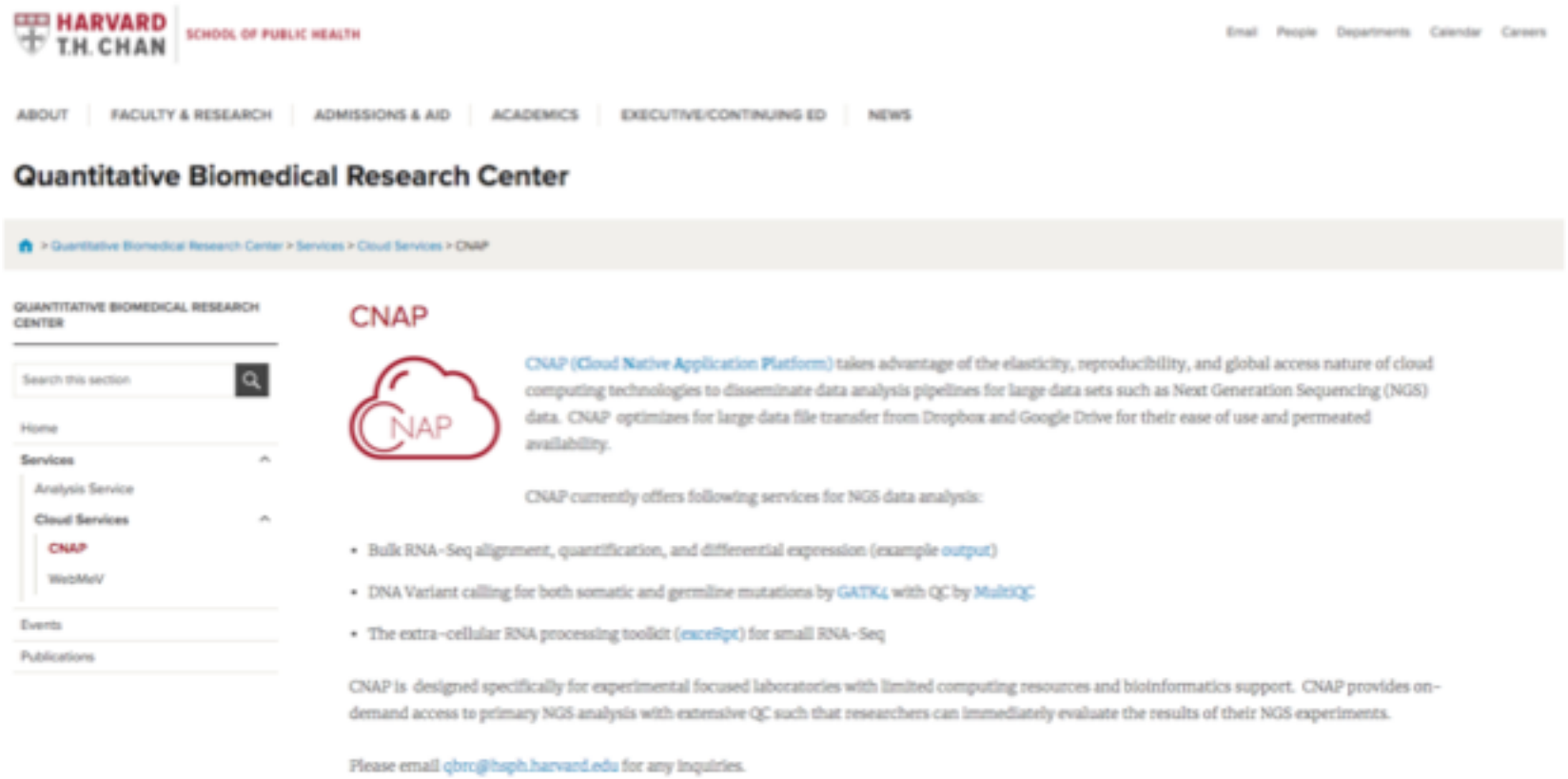
Individual alignment files (in compressed BAM format, ending with "bam") are available for download, but are provided separately due to their typically large size. If you download the BAM files, ensure that you also have the corresponding "index" files, which end with "bai". Index files allow programs 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 an "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://rseqc.sourceforge.net/ 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/10.1186/s13059-016-1044-

For more information:
https://www.hsph.harvard.edu/qbrc/services/cloud-services/cnap/



For example RNASeq output:
https://www.dropbox.com/sh/0eoik7bgoe5jcz8/AABdi14CEuamdYEADKze9ISca?dl=0