**RNA-SEQ ANALYSIS WORKSHOP**

**Schedule**

* Introduction to High Throughput Sequencing and RNA-seq
  + High throughput sequencing
  + RNA-seq Experimental Design
  + RNA Sequencing and Sample Preparation
* Overview of RNA-seq Data Analysis
  + Pipeline Workflow/Stages
  + Software
  + Metadata input
  + Files and Directory Structure
* Workshop Dataset
* Introduction to the Command Line
* Read Preprocessing
  + Description of the expHTS preprocess pipeline
  + Parameters and what they mean
  + Preprocessing the Workshop Data
  + QA/QC
* Read Mapping
  + Description of the expHTS mapping pipeline
  + Parameters and what they mean
  + Mapping the Workshop Data
  + QA/QC
* Estimate known genes and transcripts expression – Counting
  + Description of the expHTS counting pipeline
  + Parameters and what they mean
  + Counting the Workshop Data
  + QA/QC
* Differential Expression Analysis using edgeR
  + Overview of Differential Expression Analysis
  + Models and model formulation
  + QA/QC
  + Perform Differential Expression the Workshop Data
* Summarization and Visualization of Output

**Software List:**

Python 2.7

expHTS – experimental High Throughput Sequencing

<https://github.com/msettles/expHTS>

to install module, download and within the directory

python setup.py install

**Preprocessing**:

Python 2.7

Modules: argparse, optparse, distutils

bowtie2 - contaminant screening

http://bowtie-bio.sourceforge.net/bowtie2/index.shtml

Super-Deduper – Identify and remove PCR duplicates

https://github.com/dstreett/Super-Deduper

Sickle – Trim low quality regions

https://github.com/dstreett/sickle

Scythe – Identify and remove adapters in SE reads

https://github.com/ucdavis-bioinformatics/scythe

FLASH2 – Join overlapping reads, identify and remove adapter in PE reads

<https://github.com/dstreett/FLASH2>

**Mapping**:

Python 2.7

bwa mem – map reads to a reference

http://sourceforge.net/projects/bio-bwa/files/

samtools – processing of sam/bam files

http://www.htslib.org/

**Read Counting:**

Python 2.7

HTSeq-0.6.1 htseq\_count – Count reads occurrences within genes

http://www-huber.embl.de/users/anders/HTSeq/

**Analysis of differential expression:**

R 3.2.0 or greater

<http://www.r-project.org/>

R Package: optparse, parallel, tools from cran

R Package: ShortRead, EdgeR from bioconductor

<http://bioconductor.org/packages/release/bioc/html/edgeR.html> <http://bioconductor.org/packages/release/bioc/html/ShortRead.html>

RStudio

https://www.rstudio.com/