**RNASeq Analysis Workshop**

**Nov. 5-7, 2020 Florida (Florida International University)**

**Nov. 12-14, 2020 Puerto Rico (University of Puerto Rico)**

**Nov. 19-21, 2020 Puerto Rico (University of Puerto Rico)**

**Workshop will be held remotely**

**Sponsors of the Workshop**

* Puerto Rico Science, Technology and Research Trust
* Department of Biology, University of Puerto Rico
* Institute of Environment, NSF CREST Center for Aquatic Chemistry and Environment, Florida International University
* National Science Foundation

**Instructor:** Dr. Ravi Kiran Donthu

**Organizing Team**

* Tugrul Giray, University of Puerto Rico, San Juan, Puerto Rico, USA
* Rosanna Giordano, Puerto Rico Science Technology and Research Trust, San Juan, Puerto Rico, USA
* Shu-Ching Chen, Florida International University, Florida, USA
* Ravi Kiran Donthu, Puerto Rico Science Technology and Research Trust, San Juan, Puerto Rico, USA
* Maria Eugenia Presa Reyes, Florida International University, Florida, USA
* Tianyi Wang, Florida International University, Florida, USA

**Workshop schedule outline**

**Day 1: (9 AM to 12 noon)**

**Goals for the day**

* Brief introduction
* Overview- From DNA to protein
* How we can use RNASeq analysis?
* Examples from published literature
* *10-10:10 AM: 10 mins break*

**Brief introduction**

* Hypothesis testing. Considerations for experimental design with RNASeq
* *11-11:10 AM: 10 mins break*
* Sequencing, annotation and alignment data formats –
* Fastq, Fasta, Bam, GTF, GFF, SAM
* *12 noon to 1 PM – Lunch*

**Day 1: (1 PM to 5 PM)**

**Data preprocessing**

* FastQC – quality check of reads – commands to run FastQC, how to interpret the output – (Introduction and running FastQC by creating and submitting a job using SLURM script).
* *2-2:10 PM: 10 mins break*
* MultiQC (Introduction and running the program).
* Trimmomatic (Introduction).
* *3:15-3:30 PM: 15 mins break*
* Trimmomatic (Running the program).
* Summary of accomplishments of Day 1

**Day 2: (9 AM to 12 noon)**

* **Goals for the day**
* Reference genome – Introduction; Gene annotations; Functional annotations; GFF file format; Public sources of genome data (e.g. honeybee reference genome).
* *10-10:10 AM: 10 mins break*
* Genome sequence alignment:
* Two basic type of aligners:
* Prokaryotes (splice-unaware aligner)
* Eukaryotes (splice aware aligner)
* *11-11:10 AM: 10 mins break*
* Genome sequence alignment parameters.
* *12 noon to 1 PM – Lunch*

**Day 2: (1 PM to 5 PM)**

* Writing script to run alignments.
* *2-2:10 PM: 10 mins break*
* Abundance estimation of genes/transcripts.
* 3:15-3:30 PM: 15 mins break
* Normalization methods for read counting.
* Summary of accomplishment for Day 2.

**Day 3: (9 AM to 12 noon)**

**Goals for the day** –

Demonstration of steps involved in differential gene expression analysis using R. (Students will observe instructor and take notes during presentation and run the analysis on their own after the workshop. Workshop will be recorded and accessible to the participants for consultation when analyzing their own data).

* *10-10:10 AM: 10 mins break*
* Continuing with the running of differential gene expression analysis pipeline.
* 1*1-11:10 AM: 10 mins break*
* Visualization of gene expression data. Heat maps to illustrate patterns of gene expression in different treatments.
* *12 noon to 1 PM – Lunch*

**Day 3: (1 PM to 5 PM)**

* **Test Exercise:**

1. Import read counts file into R.

2. Normalization of data.

3. Output of list of differentially expressed genes as a csv file.

* *2-2:10 PM: 10 mins break*
* Continuation of Test Exercise and opportunity to ask questions.
* *3:15-3:30 PM: 15 mins break*
* Continuation of Test Exercise and opportunity to ask questions.
* Summary of accomplishments of Day 3 and the workshop.