**Designing a CURE that engages students in RNA-seq analysis - Instructions for Instructors**

Our resource **BIOME/RNA-Seq Lessons with GEA** (Genomics Education Alliance) will enable instructors to build a CURE in which students will analyze RNA sequence data. The resource will provide instructors with the tools necessary to understand how to analyze RNA sequence data and how to design a CURE based on these data. The resource will also help instructors understand the relationships between GEA, CyVerse, CyberDuck, Jupyter, and the DNA Subway Green Line.

Goals for students include

* Deepening understanding of the process of science
* Developing bioinformatics skills

Throughout the CURE, students' understanding will be guided through formative assessment.

The premise of RNA-Seq based CUREs will include a few lectures introducing the research to be performed and specific research articles addressing the research question of interest. Then students complete the DNA Subway Green Line with the data provided in the tutorial. Finally, students look into data provided by the instructor (student-directed) or find data according to their subject of interest (student-centered).

Unlike many genomics CUREs which focus on structural genomics (validating gene structure), RNA-Seq CUREs provide opportunities to study both structural information (splicing) and functional information (gene expression levels), for eukaryotic or prokaryotic transcriptome analysis. In a RNA-Seq CURE, students engage in hypothesis generation based on functional information by examining differential gene expression and gene ontology/pathway analysis, and reviewing literature on the topic of interest.

**Learning Objectives**

*Process of Science*

During this course, the student will:

1. Analyze papers to understand the potential of the techniques used in datasets to formulate and answer relevant biological questions (examples from the group include: microbes, stem cells, Piwi, cardiogenesis, a differentially expressed gene or protein, etc.).
2. Search specific databases for datasets of interest to be analyzed on a known pipeline/workflow for the student's own independent research project.
3. Formulate a research question that can be answered with RNA sequence data (new or published).
4. Use RNA sequencing data to **generate new hypotheses** and design experiments to test their hypothesis.
5. Decide whether a particular research question can be answered with previously published RNA seq data or requires acquisition of new RNA seq data.
6. Describe and present (oral, written, or with Jupyter) how they performed their analysis, and conclude on their hypothesis.
7. How the identified pathways fit into the broader context of expression changes and their research question.
8. Collaborate with one another and with the teaching staff to problem-solve issues arising with experimental design and techniques.

*RNA-sequencing*

1. Describe the steps involved in an RNA Seq experiment (from sample processing, data generation, clean up, analysis, and interpretation).
2. Determine the necessary experimental design according to the research question, and preliminary data.
3. Discuss for each of the steps involved in RNA-seq experiments the choices (number of technical and biological replicates, depth of sequencing, coverage, choice of RNA extraction, isolation and processing, etc)

*Genomics*

1. Molecular biological processes (DNA replication, RNA transcription, gene regulation (up or down), adaptation to the environment, gene structure - promoter, UTR, cis/trans regulating sequences, phase variation, RNA splicing, etc)
2. Molecular evolution (changes in DNA sequence, changes in RNA sequences, impact on gene expression/regulation
3. Chromatin structure

*Bioinformatics*

1. Learn how coding is used in RNA analysis by using Jupyter Notebook
2. Understand the steps of DNA Subway Green Line - why are you doing all of this? - use this as a standalone or else as a tutorial for more in-depth work
3. Understand and be able to complete the processes below:

|  |  |  |
| --- | --- | --- |
| **Process for Eukaryotes** | **Cyverse Learning Center Tutorials** | **Software** |
| Store Data (user data or published data) to Cyberduck | [Jupyter Notebook](https://cyverse-leptin-rna-seq-lesson-dev.readthedocs-hosted.com/en/latest/section-5.html#downloading-and-converting-data-from-sra) or  [DNA Subway Green Line](https://learning.cyverse.org/projects/dnasubway_guide/en/latest/step6.html) | CyberDuck |
| Perform Quality Analysis of Data | [Jupyter Notebook](https://cyverse-leptin-rna-seq-lesson-dev.readthedocs-hosted.com/en/latest/section-5.html#downloading-and-converting-data-from-sra) or [DNA Subway Green Line](https://learning.cyverse.org/projects/dnasubway_guide/en/latest/step6.html) | FastQC |
| Trim Data | [Jupyter Notebook](https://cyverse-leptin-rna-seq-lesson-dev.readthedocs-hosted.com/en/latest/section-5.html#downloading-and-converting-data-from-sra) or  [DNA Subway Green Line](https://learning.cyverse.org/projects/dnasubway_guide/en/latest/step6.html) | FastQC |
| Pseudoalign reads to a transcriptome | [Jupyter Notebook](https://cyverse-leptin-rna-seq-lesson-dev.readthedocs-hosted.com/en/latest/section-5.html#downloading-and-converting-data-from-sra) or  [DNA Subway Green Line](https://learning.cyverse.org/projects/dnasubway_guide/en/latest/step6.html) | Kalisto |
| Identify genes that are differentially expressed | [DNA Subway Green Line](https://learning.cyverse.org/projects/dnasubway_guide/en/latest/step6.html) | Sleuth |
| Analyze differential gene analysis with GO Terms and Pathways | [Rshiny integrated in CyVerse](https://learning.cyverse.org/projects/vice/en/latest/user_guide/quick-rshiny.html) | Sleuth |

**Example list of tools and workflow for prokaryotic RNA-Seq:** [**https://github.com/snandiDS/prokseq**](https://github.com/snandiDS/prokseq)

**Rockhopper:** [**https://cs.wellesley.edu/~btjaden/Rockhopper/**](https://cs.wellesley.edu/~btjaden/Rockhopper/)

|  |  |  |
| --- | --- | --- |
| **Process for Prokaryotes** | **Cyverse Learning Center Tutorials** | **Software** |
| Store Data (user data or published data) to Cyberduck | [Jupyter Notebook](https://cyverse-leptin-rna-seq-lesson-dev.readthedocs-hosted.com/en/latest/section-5.html#downloading-and-converting-data-from-sra) or [DNA Subway Green Line](https://learning.cyverse.org/projects/dnasubway_guide/en/latest/step6.html) | CyberDuck |
| Perform Quality Analysis of Data | [Jupyter Notebook](https://cyverse-leptin-rna-seq-lesson-dev.readthedocs-hosted.com/en/latest/section-5.html#downloading-and-converting-data-from-sra) or [DNA Subway Green Line](https://learning.cyverse.org/projects/dnasubway_guide/en/latest/step6.html) | FastQC |
| Trim Data | [Jupyter Notebook](https://cyverse-leptin-rna-seq-lesson-dev.readthedocs-hosted.com/en/latest/section-5.html#downloading-and-converting-data-from-sra) or  [DNA Subway Green Line](https://learning.cyverse.org/projects/dnasubway_guide/en/latest/step6.html) | FastQC |
| Align reads to an indexed genome | TBA | BWA |
| Identify genes that are differentially expressed | [Rshiny integrated in CyVerse](https://learning.cyverse.org/projects/vice/en/latest/user_guide/quick-rshiny.html) | DESeq2 or EdgeR |
| Analyze differential gene analysis with GO Terms and Pathways | [Rshiny integrated in CyVerse](https://learning.cyverse.org/projects/vice/en/latest/user_guide/quick-rshiny.html) | {options: cytoscape; IGV; STRING; KEGG Mapper} |

1. Use sites that support the analysis of the data, whether using the command line or point & click processes **maps to**
2. Understand different ways of analyzing datasets from SRA (NCBI)
   1. [DrosSRA Workflow](https://github.com/jfear/ncbi_remap)
   2. Kallisto etc. (uses Illumina data for read coverage)
   3. Older software that may still be useful:  Tuxedo pipeline BowTie2 and TopHat2 good for short reads
3. Use a Jupyter Notebook to create an independent RNA analysis
4. Design a new RNA analysis using Jupyter Notebooks
5. Create a Jupyter Notebook OR GitHub page/repository explaining the process and presented as a research study (including Introduction, Hypothesis/Predictions, Methods, Results with relevant figures/tables, Conclusions, in text references/citations, and troubleshooting/FAQ page).
6. Move beyond the RNA-Seq data
   1. Pathways & Networks : Cytoscape; STRING; KEGG Mapper
   2. Protein structure
   3. Protein-protein interactions

**Learning Outcomes**

Students will be able to

* Apply the principles of data analysis to a new data set and biological question.
* Identify the underlying rationale for the logic of a particular step in the process.
* Develop skills in experimental design.