RNA-Seq Course-based Undergraduate Research Experiences (CUREs) at James Madison University

Ray A Enke

Available at: https://works.bepress.com/raymond_enke/90/
RNA-Seq Course-based Undergraduate Research Experiences (CURES) at James Madison University

Ray Enke Ph.D.
James Madison University
Department of Biology
2017 CGEMS Summer Workshop

contact: enkera@jmu.edu
Twitter: @Enke_Lab
Talk Overview

1. Intro to the *Infrastructure & Training to Bring Next-Generation Sequence Analysis into Undergraduate Education* project

2. Examples of my classroom-tested RNA-Seq CURE modules
The Impact of Next Generation Sequencing (NGS)

**Cost per Genome**

- **advent of “Next Gen” sequencing (2007)**
- **10K fold reduction in sequencing cost over 4 yrs (2007-2011)**

Source: NHGRI
Next Generation Sequencing (and beyond)

- Roche 454 (2007)
- Illumina (2009)
- Ion Torrent (2010)
- Nanopore MinION (2015)

**NGS Platforms**

- Reduced cost of NGS platforms have made genome-wide analysis available to the masses

“3rd Generation!”

**RNA-Seq for the Next Generation**
Infusing Next-Generation Sequence Analysis into Undergraduate Education

Barriers to integrating genomics analysis into undergraduate curriculum:

• Cost prohibitive to generate data sets
• Complex bioinformatics analysis & advanced computing capabilities
• Limited instructor resources for training
Main objective:
• Assist PUI faculty to integrate authentic NGS analysis into undergraduate course work
  - RNA-sequencing (RNA-seq) analysis as vehicle

Infrastructure: provided by the CyVerse Collaborative
• Data storage space
• High performance computing power
• Intuitive bioinformatics web interfaces designed for non-experts

Faculty Training:
• Summer workshops for bioinformatics training
• Development & dissemination of teaching materials
  - Summer 2014 Workshop: 11 PUI faculty trained
  - Summer 2015 Workshops: 33 PUI faculty trained
  - Summer 2016 Virtual Workshop: 40 PUI faculty trained
  ❖ total of 84 PUI faculty trained over 3 years
DNA Subway Green Line Bioinformatics Analysis

- User friendly interface for RNA-seq bioinformatics analysis
- Software packages: FASTQC, FastX, TopHat, Cufflinks, Cuffdiff
- Workshop members charged with creating CURE curriculum using data sets

http://dnasubway.iplantcollaborative.org/
Program Website
www.rnaseqforthegenextgeneration.org/

- Teaching Resources, Virtual Training, RNA-Seq Data Sets (in development)
Selected Works of Ray Enke Ph.D.
Assistant Professor of Biology

My lab studies epigenetic regulation of gene expression in the vertebrate retina. Eukaryotic genomes acquire heritable and reversible chemical modifications that play a large role in influencing expression patterns of...

Classroom Exercises & Activities (14)

Classroom ready protocols and web-based resources for active learning and applied research in the areas of molecular biology, molecular ecology, genomics, and bioinformatics.

<table>
<thead>
<tr>
<th>Handout</th>
<th>Handout</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Heat map analysis of RNA-seq data</strong></td>
<td><strong>Intro to RStudio</strong></td>
</tr>
<tr>
<td>Advanced Molecular Biology; Genomics (2016)</td>
<td>Advanced Molecular Biology; Genomics (2016)</td>
</tr>
<tr>
<td>Raymond A Enke and Ashton Holub</td>
<td>Raymond A Enke and Ashton Holub</td>
</tr>
<tr>
<td>This in class exercise focuses on using the CummeRbund package in RStudio to create heat maps for analyzing differential gene...</td>
<td>This in class exercise is designed to teach novices about the basic features of R and RStudio using a non-biological...</td>
</tr>
</tbody>
</table>

- Publically available classroom tested RNA-seq teaching resources
BIO480 Advanced Molecular Biology

BIO481 Genomics

- Junior/senior level courses
- 18 students/section
- 4 credit lecture/lab course
  - 4 hours/week
- 2 sections/year
- ~50% have lab research experience other than classes

- Developed & incorporated RNA-seq analysis modules over the past 3 academic years
**Research Interest:** Developmental & tissue-specific patterns of gene expression in the vertebrate retina

What genes are differentially expressed between:
- E8 retina vs E18 retina (developmental changes in gene expression)
- E18 retina vs E18 cornea (tissue-specific changes in gene expression)
BIO480 RNA-Seq CURE Modules (4 hour lab/week)

(in class student activities)

- week #1: Embryo dissection & RNA extraction for qRT-PCR
- week #2: RNA quantity & quality
- week #3: cDNA synthesis
- week #4: qPCR primer standard curves
- week #5-6: qPCR setup
- week #7: Student Poster Session

RNA extraction
- For RNA-Seq

Illumina RNA-Seq
- Analysis of TopHat Mapping

*Bioinformatics analysis of raw sequence data

*using DNA Subway Green Line

RNA-Seq for the Next Generation

Key:
- Wet lab module
- Computational module
1. Analysis of TopHat Alignment Mapping

- Students create a DNA Subway account and navigate to a pre run public project
1. Analysis of TopHat Alignment Mapping

- Students navigate to & view individual TopHat alignment files to gather info for each sample
1. Analysis of TopHat Alignment Mapping

- Activity gives students an introductory experience with Green Line
- Also demonstrates the huge amount of sequence data in analysis pipeline
- Students gain experience creating a meaningful data visualization
1. Analysis of TopHat Alignment Mapping

https://www.youtube.com/watch?v=n77EAk8ClEs
2. Visualizing TopHat Alignments in a Genome Browser

- TopHat alignments (BAM files) can be visualized in a genome browser (IGV, UCSC, Ensembl, etc) as BAM density plot.

**Example: IGV Browser**
The UCSC Genome Browser

homepage: http://genome.ucsc.edu/

- Curated by the University of California, Santa Cruz
- Research grade bioinformatics tools
- Intuitive user friendly interface - excellent tutorials available
- Many preloaded data sets
- Custom data builds available
- Link to training modules
Open Helix training modules

http://www.openhelix.com/ucsc

- 6 training module videos
- ~75 minutes to complete
- excellent overview of basic browser features
- You will view these as an out of class assignment
2. Visualizing TopHat Alignments in the UCSC Genome Browser

- View TopHat alignment results in DNA Subway
- Navigate to TopHat BAM files in CyVerse Discovery Environment
2. Visualizing TopHat Alignments in the UCSC Genome Browser

- Navigate to TopHat alignment files in Discovery Environment
  -(2.06 GB file!)
- Copy BAM URL location from Discovery Environment
2. Visualizing TopHat Alignments in the UCSC Genome Browser

- Copy BAM URL location from Discovery Environment
- Paste BAM URL into UCSC Genome Browser custom tracks window
2. Visualizing TopHat Alignments in the UCSC Genome Browser

- Aligned RNA-seq reads for the chicken PERIPHERIN gene & downstream
  - Activity gives students experience with Green Line, DE, UCSC GB
  - Opportunity for original discovery and creating data visualizations

*Aligned RNA-seq reads for the chicken PERIPHERIN gene & downstream*
3. RNA-Seq Analysis Programming in RStudio

**completed prior to course**

- RNA extraction
  - For RNA-Seq
- **Illumina** RNA-Seq
  - Embryo dissection & RNA extraction for qRT-PCR
  - Analysis of TopHat Mapping

**in class student activities**

- **week #1**
  - RNA extraction
  - RNA quantity & quality
  - Intro to R Studio
  - Visualizing TopHat Alignments
- **week #2**
  - cDNA synthesis
  - Gene ontology analysis
  - Sequence annotation
- **week #3**
  - qPCR primer standard curves
  - RNA-Seq Data Analysis in R Studio
  - qPCR primer design
- **week #4**
  - qPCR setup
  - qPCR data analysis
- **week #5-6**
  - Student Poster Session

**Key:**
- Wet lab module
- Computational module

*using DNA Subway Green Line*
**Introduction to R and RStudio**

**Week #2**

1. **Console:** Where the processing and computation occurs.
2. **Source:** Where you type your code.
3. **Environment:** Lists the variables, tables, etc.
4. **Files, Plots, Packages, help, View:** Lists all the files, plots, packages, and help.

---

**Key:**

- **Wet lab module**
- **Computational module**

---

**RNA extraction**

For RNA-Seq

**RNA quantity & quality**

**Intro to R Studio**

**Visualizing TopHat Alignments**

*Bioinformatics analysis of raw sequence data*
Introduction to R and RStudio

Gapminder data set (National GDPs)

- Students learn fundamentals of coding language using non-biological data sets
- Focus on creating data visualization using large data sets (ggplots package)

Baseball stats

RNA-Seq for the Next Generation
3. RNA-Seq data visualizations using RStudio

- **CuffDiff list of differentially expressed genes (DEGs)**
  - 12,939 differentially regulated transcripts between E8 retina & E18 retina

- **Analysis of differential expression using CummeRbund RStudio package**
  - What’s the quality and content of the data set?
3. RNA-Seq data visualizations using RStudio

- Students use RStudio to create data visualizations for robust QC analysis of RNA-seq data sets (CummeRbund package)
3. RNA-Seq data visualizations using RStudio

- Students use RStudio to create data visualizations for robust gene expression analysis of RNA-seq data sets
Course-embedded RNA-Seq Research Modules
(4 hour lab/week)

(in class student activities)

week #1
Embryo dissection & RNA extraction for qRT-PCR
RNA extraction For RNA-Seq
completed prior to course

RNA quantity & quality
Intro to R Studio
RNA-Seq Data Analysis in R Studio

week #2
RNA-Seq Data Analysis
Sequence annotation
Gene ontology analysis
qPCR primer design

week #3
cDNA synthesis

week #4
qPCR primer standard curves

week #5-6
qPCR setup

week #7
Student Poster Session

*using DNA Subway Green Line

Key: [Wet lab module] [Computational module]

*Bioinformatics analysis of raw sequence data
4. RNA-seq Student Poster Session

The JMU Advanced Molecular Biology Class invites you to

The Donuts & Transcriptomics Poster Session

In partnership with:

RNA-Seq for the Next Generation
4. RNA-seq Student Poster Session

The JMU Advanced Molecular Biology Class invites you to

The Donuts & Transcriptomics Poster Session

RNA-Seq for the Next Generation
Reflections & Future Directions

• (If possible) mixing wet bench & computational activities is key
  - no budget for wet lab? Find collaborators!
  - computational modules are completely free
  - do not need your own data (NCBI SRA database has 1,000s of datasets!)
Reflections & Future Directions

• (If possible) mixing wet bench & computational activities is key
  - no budget for wet lab? Find collaborators!
  - computational modules are completely free
  - do not need your own data (NCBI SRA database has 1,000s of datasets!)

• Make your course activities dovetail with your research program
  - maintains research focus during the academic year
  - course budgets can be used to propel your research
  *planning & troubleshooting efforts for class will benefit your lab research
Reflections & Future Directions

• (If possible) mixing wet bench & computational activities is key
  - no budget for wet lab? Find collaborators!
  - computational modules are completely free
  - do not need your own data (NCBI SRA database has 1,000s of datasets!)

• Make your course activities dovetail with your research program
  - maintains research focus during the academic year
  - course budgets can be used to propel your research
  * planning & troubleshooting efforts for class will benefit your lab research

• Start small, build up over several iterations/semesters
  - start with activities you feel most comfortable with & build out
  - visit the websites to download posted classroom tested protocols
Thanks!

- CSHL DNA Learning Center
- CyVerse support staff
- JMU Biology & GCEMS
- Ashton Holub
- JMU BIO 480/481 students

contact: enkera@jmu.edu
Twitter: @Enke_Lab