BIOL436/FNH436: Integrated Functional Genomics (3 credits)

Term 1, 2018W  Tues and Thurs 11:00-12:30 pm  (West Mall Swing Space 309)
Prerequisite: BIOL 335 or 338

Instructor: Dr. Jae-Hyeok Lee (jae-hyeok.lee@botany.ubc.ca)
Office: Biological Sciences Building, Rm. 1326C (on the first floor in South Wing)
TA: Nolan Shelley, Rm BIOSCI 2327 (nolan.shelley@botany.ubc.ca)

Course Website: https://canvas.ubc.ca/courses/9297

Course design and main objectives:
Lectures and in-class tutorials will provide practical exercises suited for those who prepare or just begin graduate-level study in genomics-facilitated research environment. This course intends to introduce students to

1. Modern functional genomics research utilizing genome-wide mutant collections,
2. High-throughput technology used in genomics research, and
3. Hands-on practices on big-data analysis and bioinformatics tools.

Major learning outcomes: By the end of this course, you should be able
1. To design a gene-discovery study using genomics approaches for a given biological question.
2. To collect candidate genes of interests by analyzing gene expression or other quantitative data using R.

Tutorial sessions to provide:
1. Hands-on practices including the following examples.
   - RNA-seq data analysis pipeline.
   - How to use differential expression analysis/hierarchical clustering methods.
   - Statistical evaluation of the genome-wide analysis using the false-discovery-rate criterion.
2. Group activities that help students to elaborate what they have learned in class and during homework by peer teaching/evaluation.

Evaluation:  Students will be assessed on
1) their knowledge about the materials covered in the lectures/tutorials,
2) course participation ranging from tutorials worksheets and pre-reading quizzes.
3) their ability to analyze genomics data to produce biologically meaningful conclusions.
Two midterms will be a written test (two-stage exam) in class on the knowledge-base, and the final will consist of take-home questions similar to tutorial exercises/term-paper to design your own project.

The mark breakdown is shown below:

Two midterms (two-stage exam): 20 X 2 = 40%  
Final take-home exam in R-markdown: 20%  
R-Tutorial worksheets in R-markdown: 10 X 2 = 20%  
Homework: 20%  
   - Pre-reading quizzes: 5 X 1.5 = 7.5%  
   - 5-min mini-lecture on one technique: 2.5% (>80% correct, or 2%)  
   - Term paper in R-markdown: 10% (group project)
What you need in class (during tutorial):
A laptop or an equivalent device that can run various on-line and off-line software, including EXCEL and R-studio.

What you will expect to do outside class:
Average 3 hours preparation per each class for pre-readings and assignments.

<table>
<thead>
<tr>
<th>Tentative Class Schedule</th>
<th>Date</th>
<th>Tuesdays</th>
<th>Thursdays</th>
<th>Reading/HW</th>
<th>Submission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 0 (9.6)</td>
<td>Imagine day</td>
<td></td>
<td>Introduction to functional genomics</td>
<td>R-installation/Dent et al. (2004) (wk1-2)</td>
<td></td>
</tr>
<tr>
<td>Week 1 (9.11-13)</td>
<td>Mutants to discover biological functions 1) Model system 2) Genetics 3) High-throughput technologies: NGS</td>
<td>Mutagenesis 1) Induced mutations: Ins/UV 2) Existing variations: RI lines 3) Targeted mutations Tutorial (T) 0 (What is R?) Peer teaching (PT) 1 (NGS- technology, CRISPR)</td>
<td>On-line tutorial parts 1/2</td>
<td>Dent et al quiz by 9.13</td>
<td></td>
</tr>
<tr>
<td>Week 2 (9.18-20)</td>
<td>Phenotypes/Probes (Guest: Jacob Munz)</td>
<td>Mutation identification 1) FSTs for tagged mutations 2) Chr.Walk for random SNP T1 (Data types, simple functions) PT 2 (FST-fishing)</td>
<td>Matrix, Dataframe, Dataimport/ Fang et al. (2012) (wk3-)</td>
<td>T1-WS</td>
<td></td>
</tr>
<tr>
<td>Week 3 (9.25-27)</td>
<td>Gene expression study: How to quantify (Guest: PhD. Meysam Abbasi)</td>
<td>RNA-seq workflow: DE/HC/co-expression network T2 (Dataimport, Q-PCR analysis) PT 3 (SEQ-tech)</td>
<td>Plotting by ggplot2/Mackinder et al. (2017) (wk4)</td>
<td>Fang et al quiz by 9.27 T2-WS</td>
<td></td>
</tr>
<tr>
<td>Week 4 (10.2-4)</td>
<td>CCM project workout (Guest: Yuan or Thamali) 1) Genetic screening project 2) Use of probes</td>
<td>Omics approach (Guest: Last) T3 (basic plotting) PT 4 (Protein-protein interaction detection)</td>
<td>Advanced visualization</td>
<td>Mackinder et al quiz by 10.4 T3-WS</td>
<td></td>
</tr>
<tr>
<td>Week 5 (10.9-11)</td>
<td>Mid-term 1 (wk1-4)</td>
<td>Statistics for genomics R-review and survey T4 (advanced plotting)</td>
<td>None</td>
<td>T4-WS</td>
<td></td>
</tr>
<tr>
<td>Week 6 (10.16-18)</td>
<td>RNA-seq 1: Data Quality control - Source of error - TUT + CCM data</td>
<td>RNA-seq 2: Experiment Quality control using visualization tools 1) PCA, 2) Pearson correlation</td>
<td>CCM markdown draft 1 (T5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 7 (10.23-25)</td>
<td>RNA-seq 3: DATA normalization</td>
<td>RNA-seq 4: Differential expression analysis 1) DE, 2) T-test, FDR</td>
<td>Duanmu et al. (2013)</td>
<td>CCM markdown final (T6)</td>
<td></td>
</tr>
<tr>
<td>Week 8 (10.30-11.1)</td>
<td>Ex1. Billin-signaling</td>
<td>Use of Venn diagram Group project for Bilin</td>
<td>Zones et al. (2015)</td>
<td>Duanmu et al. quiz by 10.30 Bilin markdown (T7)</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------</td>
<td>---------------------------------------------</td>
<td>---------------------</td>
<td>-----------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Week 9 (11.7-9)</td>
<td>Ex2. Diurnal cycle Data normalization Algorithms to study patterns (e.g. PCA)</td>
<td>Multi-way comparison Group project for Diurnal: From data collection to testing candidate genes.</td>
<td>Zones et al. quiz by 11.7 Diurnal markdown (T8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 10 (11.13-16)</td>
<td>Mid-term 2 (wk6-9) Introducing term projects unpublished N-transcriptome</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 11 (11.21-23)</td>
<td>Analysis by annotation GO/KEGG/ Enrichment test Revisiting the diurnal data</td>
<td>Hierarchical Clustering method for co-expressed gene networks</td>
<td>Annotation markdown (T9), HC markdown (T10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 12 (11.28-11.30)</td>
<td>RNA-seq Data visualization <strong>Group project presentation</strong></td>
<td>Final group presentation on N-transcriptome</td>
<td>Term paper submission (markdown)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final Exam</td>
<td>Take home exam as a review of tutorial exercises.</td>
<td></td>
<td>Final exam submission</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Breakdown activities:**

**Main reading material:**
1. Approaches of biological research in genomics era
   **Mackinder et al. (2017)** A spatial interactome reveals the protein organization of the algal CO₂-concentrating mechanism. Cell

2. RNA-seq study examples
   **Duanmu et al. (2013)** Retrograde bilin signaling enables Chlamydomonas greening and phototrophic survival. PNAS.
   **Zones et al. (2015)** High resolution profiling of a synchronized diurnal transcriptome from Chlamydomonas reinhardtii reveals continuous cell and metabolic differentiation. Plant Cell,

3. Suggested readings for statistics
   **Eddy (1996)** Hidden Markov models
4. Miscellaneous readings:

**Model organisms**: https://en.wikipedia.org/wiki/Model_organism

**Student-run Mini-lecture topics**

**Module 1**
- a. Illumina sequencing technology: how does it collect millions of sequencing reads at once?
- b. CRISPR technology: how does it work? (specific gene-targeting)
- c. CRISPR applications: what can be done with CRISPR?

**Module 2**
- a. TAIL PCR
- b. FST by Mmel
- c. ID mutation by WGS

**Module 3 (sequencing-saavy)**
- a. CHIP-seq
- b. BiSulfite-seq
- c. Ribo-seq

**Module 4 (protein-X interaction)**
- a. Calorimetry (quant)
- b. BiFC (localize)
- c. Immunoprecipitation (id complex)

**Tutorial**: will be held for 30 min in Thursday classes
(four groups, each of 2 or 3 students, working together)

T1. Installation check-up, Self-learning module for data types
T2. Big-DATA handling with Excel VS R.
T3. Q-PCR analysis and graphing
T4. Big-DAT visualization

T5-6. Differential expression analysis with R for the cia5/CCM study.
T7. Venn diagram analysis with R for the billin study.
T8. Correlation-based analysis with R for the diurnal study.
T9. Working with annotation, revisiting the diurnal study.
T10. Hierarchical clustering by R

**On-line resources adopted for class materials**
https://www.bioconductor.org/help/workflows/rnaseqGene/
http://combine-australia.github.io/RNAseq-R/06-rnaseq-day1.html
https://dwheelerau.com/2014/02/17/how-to-use-deseq2-to-analyse-rnaseq-data/

**Two-tier exam structure:**
Two midterm exams will contain one or two questions that can be revisited for group discussion. The second-stage exam answers will be used to revise the individual exams only for improvement.
Pre-required knowledge check-up

BIOL234 - Relevant Learning Objectives:

● Explain the role of mutagenesis and mutant isolation in biological research.
● Describe the process of mutagenesis including the reasons why mutants are normally identified in the generation following mutagenesis.
● List the genetic analyses that are done on newly isolated mutants.
● Explain how molecular markers can be used to map the location of genes (and how this is dependent on linkage analysis).
● If necessary do a complementation test (or explain how to do one) or analyze the results of a complementation test to determine how many genes are involved to produce one type of phenotype.
● Given the functions of two or more gene products that control the same phenotype (genetic interactions) predict the phenotype if there are changes to one or more of the genes involved.