

## Syllabus for MMG1344H: Foundational Computation Biology I (FCB I)

**Course Instructor:** Fritz Roth and Kieran Campbell

**Course Time and Dates:** Apr 7 - May 12 (Thursdays 1-3pm).

**Course Location:** Donnelly Centre Black Room or Zoom (<https://utoronto.zoom.us/j/6808789917>) if necessary

**Total Student Contact Hours (0.25 FCE):** 12 hours of lectures plus 12 hours of (optional) TA office hours.

**Prerequisite:** This course is targeted to 1<sup>st</sup> year graduate students with substantial computational biology experience. For example, students with computational biology experience which is too advanced for MMG1004 (A Practical Course in Programming for Biologists) will be encouraged to take this course. Priority will be given to first-year Molecular Genetics Department admitted graduate students. On a case-by-case basis, the instructors will also admit Molecular Genetics Department students in year 2 or later as well as students from other UofT Departments.

Enrollment is subject to Instructor approval, and **will require:**

- 1) evidence of comfort with computer programming and
- 2) excellence in two or more quantitative subjects, which may include: calculus, linear algebra, probability/statistics or other mathematics courses.

**Course Objectives and Learning Outcomes:** The Foundational Computational Biology (FCB) courses are two 6-meeting topic courses, offered through the Molecular Genetics Graduate program, covering foundational concepts and current applications for computational biology and bioinformatics.

**Course Description and Format:** The course will consist of lectures as well as hands on assignments. Assignments will be both pen-and-paper and practical assignments requiring programming (e.g., Python) or statistical environments (e.g., R).

### Delivery of the course:

#### Lecture 1: Intro, Probability and Statistics, and the Practice of Reproducible Computational Science

**Lecturer:** Fritz Roth (90 min)

- Course overview
- Basic probability & statistics (Bayesian and frequentist)

- Error propagation / delta method

**Lecturer:** Kieran Campbell (30 min)

- Why work reproducibly?
- Pseudorandom number generation and seeds
- An introduction to Snakemake

*ASSIGNMENT #1.1 MADE AVAILABLE*

## Lecture 2: Sequence Analysis

Lecturer: Fritz Roth

- Dynamic programming and Smith-Waterman sequence alignment
- Sequence similarity and substitution matrices
- BLAST

Paper Discussion: Pitfalls of irreproducibility

Corrigendum: [https://www.cell.com/fulltext/S0092-8674\(12\)01427-4](https://www.cell.com/fulltext/S0092-8674(12)01427-4)

Original: [https://www.cell.com/cell/fulltext/S0092-8674\(08\)00886-6](https://www.cell.com/cell/fulltext/S0092-8674(08)00886-6)

Corrected: [https://www.cell.com/fulltext/S0092-8674\(12\)01427-4#supplementaryMaterial](https://www.cell.com/fulltext/S0092-8674(12)01427-4#supplementaryMaterial)

*ASSIGNMENT #1.2 MADE AVAILABLE*

*ASSIGNMENT #1.1 DUE Friday at 11:59 pm*

## Lecture 3: Unsupervised and Supervised Learning

**Lecturer:** Fritz Roth (45 minutes)

- Overview of unsupervised learning
- distance and similarity measures
- agglomerative vs divisive clustering
- hierarchical vs partitional clustering
- linkage methods
- K-means clustering

**Lecturer:** Kieran Campbell (75 minutes)

- An overview of supervised learning
- Linear regression models
- Loss optimization via gradient descent
- Classification with logistic regression
- More complex models
- Train/test splits
- Model complexity: overfitting and underfitting
- Penalized regression

#### **Lecture 4: Networks, Pathways and Function I**

**Lecturers:** Gary Bader and Jüri Reimand

- Gene ontologies
- gene function analysis
- pathway enrichment analysis
- network visualization and analysis
- interaction networks

**Paper Discussion:** HRDetect paper Davies, Helen, et al. "HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures." *Nature medicine* 23.4 (2017): 517. <https://www.nature.com/articles/nm.4292>

**ASSIGNMENT #1.3 made available**

**ASSIGNMENT #1.2 DUE Friday at 11:59 pm**

#### **Lecture 5: Phylogenetics and Tumour Genome Evolution**

**Lecturer:** Zhaolei Zhang (45 min)

- Phylogenetics, Tree inference (+ UPGMA), ancestral sequence inference
- Conserved sequence elements, PHASTCONS, PHYLOP

**Lecturer:** Sagi Abelson (45 min)

- Tumour genome analysis/evolution

#### **Paper discussion:**

- Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and Enrichment Map

Jüri Reimand,, Gary D. Bader, *Nature Protocols* (2019)

<https://www.nature.com/articles/s41596-018-0103-9>

## Lecture 6: Proteomics and Imaging Technologies

**Lecturer:** Hannes Rost (45 min)

- Protein mass spectrometry and analysis

**Lecturer:** Hartland Jackson (45 min)

- Highly multiplexed imaging technologies and data analysis

### **Paper Discussion:**

Analysis of protein-coding genetic variation in 60,706 humans

M Lek, KJ Karczewski,, D MacArthur, Exome Aggregation Consortium *Nature* 2016

<https://www.nature.com/articles/nature19057>

**ASSIGNMENT #1.3 DUE Friday at 11:59 pm**

### **Method of Evaluation:**

The marking scheme will be **75%** assignments, **25%** participation. The **lowest two** participation marks will be dropped. Bring a doctor's note for anything else! The maximum mark on late assignments will be reduced by 5% per day until the maximum penalty of 25% is reached.