# SYLLABUS: Foundational Computational Biology (FCB) Courses I and II (MMG1344H; MMG1345H)

**Spring 2025**

<u>FCB I</u>: March 21st – April 25th

<u>FCB II</u>: May 2nd – June 6th

## Course Description

The Foundational Computational Biology (FCB) courses, offered through the Molecular Genetics Graduate program, are taught as two 6-meeting topic courses that cover selected foundational concepts and current applications for computational biology and bioinformatics. The courses are targeted to 1$^{st}$ year graduate students, with preference given to students in the CBMG track of the Molecular Genetics Graduate Program.

Assignments will be both pen-and-paper and practical assignments requiring programming (e.g., Python) or statistical environments (e.g., R).

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.

- ❖ *When*:
    - ❖ FCB I will be from **2-4pm Fridays starting March 21st until April 25th except for Thursday April 17th 2-4pm**
    - ❖ FCB II will be on **Fridays 2-4pm starting May 2nd running until June 6th**
    - ❖ TA office hours and location will be TBC
- ❖ *Where*:
    - ➢ Lectures will be <u>in person</u> at the Donnelly Black Rm
    - ➢ Slack channel: https://foundationalc-kh01755.slack.com/
    - ➢ Current course content shared via DropBox: https://www.dropbox.com/scl/fo/xl0l7a4bs204xffvabnm5/ANIpCNZ3_WjOm6JFwubQuUk?rlkey=n6hxwkwlzout1hwez952iyjao&dl=0
    - ➢ Previous years course content shared: https://www.dropbox.com/scl/fo/v4z68c0w8yv04s2i6pd2m/h?dl=0&rlkey=xkmwg1uawc2n83p1yu4srrgtk https://www.dropbox.com/sh/5dvd87kssquhd5o/AAB0WVxidP1y8LtO4-3r_dNpa?dl=0

- ➤ Assignments will be uploaded to:
  https://www.dropbox.com/scl/fo/obrsljxtr4sqayg8ezmqu/h?rlkey=bvgaoa8qcz4sc95ot wy5ywzgb&dl=0
- ➤ Assignments will be submitted to DropBox file-request link:
  https://www.dropbox.com/request/iq5bLXbuROccPnfR7WGF
- ➤ Slides for lectures by Kieran Campbell will also be available here:
  https://www.camlab.ca/teaching/

- ❖ *Who:*
  - ➤ Co-instructors
    - ■ Juri Reimand (juri.reimand@utoronto.ca)
    - ■ Kieran Campbell
    - ■ Gary Bader

  - ➤ Teaching Assistant:
    - ■ TBA

The marking scheme will be **90%** assignments, **10%** participation.

The **lowest** participation mark will be dropped.  (Please bring a doctor's note for anything else!)

Late assignments may be handed in up to one week late with a penalty of 5% per business day until the maximum penalty of 25% is reached.

# FCB 1 Syllabus

DAY 1:  2 PM - 4 PM,  Fri March 21st, 2024

**Course overview - 10 minutes**

**Lecturers: Kieran Campbell, Gary Bader**

**Science intro (30 min) - Gary**

**The Practice of reproducible computational science (45 min) - Kieran**

- ➤ Why work reproducibly?
- ➤ Pseudorandom number generation and seeds
- ➤ An introduction to Snakemake
- ➤ Data and code management (Gary)
- ➤ Learn how to use HPC via scinet courses - https://education.scinet.utoronto.ca/

*ASSIGNMENT #1.1 MADE AVAILABLE*

**70 min Intro, Probability/Statistics** (Juri Reimand)

PDFs and PMFs, p-values, standard statistical tests

**Sequence Analysis (Gary Bader)**

Lecture: Substitution matrices, BLAST, Sequence similarity, Dynamic programming, Smith-Waterman sequence alignment

*ASSIGNMENT #1.2 MADE AVAILABLE*

DAY 3: 2 PM - 4 PM, Fri April 4th, 2024

2:05 - 2:55pm **Lecturer: Zhaolei Zhang (to confirm)**

➢ Lecture: Molecular evolution, Phylogenetics, Tree inference, paralogs, ancestral sequence inference
➢ Lecture: Conserved sequence elements, PHASTCONS, GERP, PHYLOP

3:05 - 3:55pm **Lecturer: Artem Babaian (to confirm)**

➢ Lecture: Finding distant homologs by structure/sequence similarity, sequence mining for pathogen discovery/understanding

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

DAY 4: 2 PM - 4 PM, Fri Apr 11, 2024

**Lecturer 1/2: Juri Reimand**

**Intro to unsupervised learning**
Lecture: Clustering -- distance and similarity measures, linkage methods, K-means, hierarchical clustering, agglomerative vs divisive clustering

**Lecturer 2/2: Kieran Campbell**

**Intro to supervised learning**

➢ 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

*ASSIGNMENT #1.3 made available*

DAY 5:  2 PM - 4 PM, Thu Apr 17, 2024

2:10 - 3:00pm **Lecturer 1/2: Jesse Gillis (to confirm)**

- ➢ Lecture: Fundamentals of counting-based 'omic methods (e.g.RNA-Seq, ChIP-Seq, ATAC-Seq) and mining to gain biological insight

3:05 - 4:00pm   **Lecturer 2/2: Hannes Rost (to confirm)**

- ➢ Lecture: Protein mass spectrometry analysis

*ASSIGNMENT #1.2 DUE Friday 11:59 pm*

DAY 6:  2 - 4  PM, Fri Apr 25, 2024

**Networks, Pathways and Function I**

- ➢ Lecture: Gene ontologies, gene function analysis, pathway enrichment analysis, network visualization and analysis, interaction networks

2:10 - 3:00pm  **Lecturer: Gary Bader**

Introduction to pathway enrichment analysis, pathway databases, Gene Ontology

3:05 - 4:00pm  **Lecturer: Juri Reimand**

Pathway enrichment analysis statistics, Interpreting pathway enrichment analysis using Enrichment Map and Cytoscape

*ASSIGNMENT #1.3 **DUE Friday May 2nd** at 11:59 pm*

**Thanks all!**

**[END FCB I]**

# FCB II

**Before or during the course, here are some opportunities to brush up on essential background:**

Linear Algebra:

Strang MIT course: https://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/index.htm

3blue1brown:

https://www.youtube.com/watch?v=kjBOesZCoqc&list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab

KhanAcademy: https://www.khanacademy.org/math/linear-algebra

KNOW: eigenvectors, singular value decomposition, Moore-Penrose pseudo-inverse

Gradients:

Khan Academy https://www.khanacademy.org/math/multivariable-calculus/multivariable-derivatives/gradient-and-directional-derivatives/v/gradient

DAY 1:  2 PM - 4 PM, Fri May 2nd, 2024

**Lecturer: Kieran Campbell**

➤ Lecture: **Bayesian inference**
   ○ Re-introduction to Bayes rule
   ○ Sampling methods: Gibbs sampling, Metropolis Hastings
   ○ Variational inference
   ○ An introduction to probabilistic programming languages & STAN
➤ Lecture: **Deep learning**
   ○ The perceptron, multi-layer perceptrons
   ○ Gradient descent and backpropagation
   ○ Deep learning for images: CNNs
   ○ Deep learning for sequence data: RNNs

*ASSIGNMENT #2.1 MADE AVAILABLE*

DAY 2:  2 PM - 4 PM, Fri May 9th, 2024

**Lecture 1/2: Gregory Schwartz — single-cell & spatial genomics**

**Lecturer 2/2: Sushant Kumar - variant interpretation in the context of protein structure**

DAY 2.3:  2 PM - 4 PM, Friday May 16th, 2024

**Lecturer: Kieran Campbell**

Lecture: Unsupervised learning:

➢ Continuous latent variable models: historical perspective, principal component analysis, probabilistic PCA, non-negative matrix factorization, tSNE, autoencoders
➢ Mixture models: probabilistic clustering, Gaussian mixture models, expectation maximization, model complexity

*ASSIGNMENT #2.2 MADE AVAILABLE*

*ASSIGNMENT #2.1 DUE Friday, May 16th at 11:59 pm*

DAY 2.4:  2 PM - 4 PM, Fri May 23rd 2024

**Lecturer 1/2: Michael Hoffman (to confirm)**

— Lecture: Evaluating supervised learning

**Lecturer 2/2: Philip Kim (to confirm)**

— Generative models for protein design

DAY 2.5: 2 PM - 4 PM, Fri May 30th, 2024

**Lecturer 1/2: Jared Simpson (to confirm)**

— Lecture: Read mapping algorithms for short and long reads, long read single molecule sequencing technology

**Lecturer 2/2: Philip Awadalla (to confirm)**

- computational and population genetics, genomics, cancer, etc

*ASSIGNMENT #2.2 DUE Friday, June 6th at 11:59 pm*

DAY 2.6: 2 PM - 4 PM, Fri June 6th 2024

**Lecturer 1/1 : Alan Moses (to confirm)**
Lecture: Deep learning for sequence analysis

**[END FCB II]**