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

Spring 2026

FCB I*:* March 20th – April 24th

FCB II: May 1st – June 5th

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 1st 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 20th until April 24th except for Thursday April 3rd 2-4pm**
  - FCB II will be on **Fridays 2-4pm starting May 1st running until June 5th**
  - TA office hours will be **held on Zoom (details TBA)**
- *Where*:
  - Lectures will be in person at the Donnelly Black Rm
  - TA office hours will be on Zoom
  - Discussion channel in Quercus
  - Current course content shared in Quercus
  - Assignments will be uploaded in Quercus
  - 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 (kierancampbell@lunenfeld.ca)
- Gary Bader (gary.bader@utoronto.ca)

➢ Teaching Assistant (to be confirmed)

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

The **lowest** participation mark will be dropped.

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. Assignments handed in later than one week after the deadline will receive a score of 0%.

If you are late on an assignment due to a medical reason, please tell us before the assignment deadline.

AI policy: strictly no AI (e.g. Chat GPT) allowed for any assignment. You need to do the assignment fully by yourself so you can learn the material. Assignments done with AI may be penalized by at least 50%.

FCB 1 Syllabus

DAY 1: 2 PM - 4 PM, Fri March 20st, 2026 (Kieran & Gary hosting)

Course overview - 10 minutes

Lecturer: Gary Bader - Science intro (30 min)

Lecturer: Kieran Campbell - The Practice of reproducible computational science (45 min)

- ➢ 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

DAY 2: 2 PM - 4 PM, Fri March 27th, 2026 (Gary & Juri hosting)

Lecturer: Juri Reimand - 70 min Intro, Probability/Statistics ()

- ➢ probability
- ➢ p-values
- ➢ standard statistical tests

Lecturer: Gary Bader - Sequence Analysis

- ➢ Substitution matrices
- ➢ BLAST
- ➢ Sequence similarity
- ➢ Dynamic programming
- ➢ Smith-Waterman sequence alignment

ASSIGNMENT #1.2 MADE AVAILABLE

DAY 3: 2 PM - 4 PM, Thu April 2nd, 2026 (Kieran hosting) - NB - lecture moved due to Good Friday

Lecturer 1: Zhaolei Zhang (to confirm)

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

Lecturer 2: Artem Babaian (to confirm)

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

ASSIGNMENT #1.1 ***DUE Thur Apr 2nd*** *at 11:59 pm*

DAY 4: 2 PM - 4 PM, Fri Apr 10, 2024 (Gary & Juri hosting)

Networks, Pathways and Function I

Lecturer 1: Gary Bader

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

Lecturer 2: Juri Reimand

- ➢ Pathway enrichment analysis statistics
- ➢ Interpreting pathway enrichment analysis using Enrichment Map and Cytoscape

ASSIGNMENT #1.3 made available

DAY 5: 2 PM - 4 PM, Fri Apr 17, 2026 (Juri & Kieran hosting)

Lecturer 1: Juri Reimand

Intro to unsupervised learning

- ➢ An overview of supervised learning
- ➢ Clustering
- ➢ distance and similarity measures, linkage methods
- ➢ K-means, hierarchical clustering
- ➢ agglomerative vs divisive clustering

Lecturer 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 #1.2 ***DUE Fri Apr 17*** *11:59 pm*

DAY 6: 2 - 4 PM, Fri Apr 24, 2026 (Kieran hosting)

Lecturer 1: Jesse Gillis (to confirm)

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

Lecturer 2: Hannes Rost (to confirm)

- ➢ Protein mass spectrometry analysis

ASSIGNMENT #1.3 ***DUE Fri May 1st*** *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 1, 2026 (Kieran hosting)

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 8, 2026 (Kieran hosting)

Lecture 1: Gregory Schwartz (to confirm)

- single-cell & spatial genomics

Lecturer 2: Sushant Kumar (to confirm)

- ➢ variant interpretation in the context of protein structure

DAY 2.3: 2 PM - 4 PM, Friday May 15th, 2026 (Kieran hosting)

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 Fri, May 15th*** *at 11:59 pm*

DAY 2.4: 2 PM - 4 PM, Fri May 22 2026 (Gary hosting)

Lecturer 1: Michael Hoffman (to confirm)

- ➢ Lecture: Evaluating supervised learning

Lecturer 2: Philip Kim (to confirm)

- ➢ Generative models for protein design

DAY 2.5: 2 PM - 4 PM, Fri May 29, 2026 (Juri hosting)

Lecturer 1: Jared Simpson (to confirm)

- ➢ Read mapping algorithms for short and long reads, long read single molecule sequencing

Lecturer 2: Aleks Goeva (to confirm)

- ➢ Dimensionality reduction and prediction methods for single cell and spatial transcriptomics data

DAY 2.6: 2 PM - 4 PM, Fri June 5th 2026 (Gary hosting)

Lecturer 1 : Alan Moses (to confirm)

- ➢ Deep learning for sequence analysis

ASSIGNMENT #2.2 ***DUE Fri June 5th*** *at 11:59 pm*

[END FCB II]