MMG1004H S LEC0101 20261:A Practical Course in Programming for Biologists

Instructor: Brett Trost (brett.trost@sickkids.ca)

Course dates

- Monday, April 6, 2026 to Wednesday, May 13, 2026

Course times and locations

- Mondays: 10am - 11:15am in MS 2173 (Medical Sciences Building)
- Wednesdays: 10am - 11:15am in NL 6 (David Naylor Building)

Course overview

This course is designed to teach experimental biologists the basics and hands-on knowledge of bioinformatics programming. In today's world, most graduate students in Molecular Genetics will encounter situations where they have to make use of computational tools and deal with datasets that are too large to practically handle manually (or with Excel). The main objective of this class is to give students the power of automation via bioinformatics programming. The class teaches by example and makes students comfortable with doing basic programming in R and adapting existing programs to their needs, as well as how to interface with standard bioinformatics software. We focus on R because it is the most popular data science and bioinformatics language and is easier to learn compared to most other programming languages. However, we will also give a brief introduction to the more general language of Python.

Course format

We will be teaching using a "flipped classroom model", which will work as follows:

- All lectures (except lecture 1, which will be live in person) are presented as recorded videos, which are expected to be viewed prior to class. Links to each lecture are below.
- At the beginning of each in-person session (Mondays and Wednesdays 10am - 11:15am, as indicated above), we will gather in the classroom for a brief summary of the material, general discussion, and Q&A. After this, each student will work on the in-class exercises individually on their own laptop, with TAs and instructors present to provide immediate help if needed.
- You must bring your own laptop to class every day. Please make sure it is fully charged, as there are only a few power outlets in the classroom.

We follow this model to help you overcome micro-barriers to your progress (such barriers are often encountered when learning programming for the first time).

Assignments

One homework assignment will be handed out each week. Assignment 1 will be relatively easy, but they will **rapidly and severely get more difficult** because the concepts build up over time. **The only way to learn programming is through a lot of practical coding work, and the assignments are designed to give you that experience.** Completing these assignments is an integral component of the class and a large portion of the grade will be based on it. Assignment solutions will be presented in class a few days after they are due.

TA office hours

Each TA will hold weekly office hours. Consult the TA groups and join the TA office hours at the time noted for your TA group. **If you plan to visit your TA in-person, please e-mail them before you arrive.** If no Zoom password is indicated, then no password is required.

TA group	TA name	TA e-mail address	Day/time	Zoom link	Zoom password	In-person location
1	Xindi Zhang	xindi.zhang@mail.utoronto.ca	Fridays @ 11am	Link		OICR, room 6-16 WT boardroom (please use the 5th floor entrance)
2	Ira Horecka	ira.horecka@mail.utoronto.ca	Wednesdays @ 12pm	Link	officehour	CCBR, 6th floor
3	June Kim	yeekyung.kim@mail.utoronto.ca	Wednesdays @ 3pm	Link		OICR, 5th floor
4	Ali Fathi	ali.fathi@mail.utoronto.ca	Mondays @ 11:30am	Link	405928	The Donnelly building, 13th floor
5	Elena Peng	elena.peng@mail.utoronto.ca	Thursdays @ 10am	Link	062000	LTRI (Mount Sinai Hospital), 10th floor

Syllabus (may be updated during the course)

Note: if you are having difficulties downloading any of the files below and are using a browser other than Google Chrome, please switch to Google Chrome.

Week 1: Goal getting comfortable with basic tools

Lecture 1

- Introduction to programming.
- The Unix shell (command-line) environment.
- Getting comfortable with the shell, basic shell commands, some examples

Lecture 1 Notes: UofTMoGenPractProgBiol_Lecture1.pptx

Assignment 1: see Assignment section

Note: As a general policy in the course, packages are not allowed unless otherwise specified in the assignments. This is to help you learn the fundamentals.

Assignment 1 solution: Assignment1_solution.docx

ALL_SOURCES_ALL_FREQUENCIES_genes_to_phenotype.txt

Instructions for accessing the command line and R on your computer

Lecture 1 exercise solutions

Lecture 2

- Introduction to R, notebooks, RStudio
- R Statements, Basic Syntax and Variables

Lecture 2 part 1 (PPT) video (To be viewed before class)

Lecture 2 slides: UofTMoGenPractProgBiol_Lecture2.pptx

Lecture2 part 2 (R Notebook) (To be viewed before class)

R notebook for lectures 2-10

Data files for the in class exercises (unzip these into a data directory of your choosing) - these will be used over the course, so download it once and keep it in a convenient location. Only a few files will be used in the first lecture.

Exercise solutions to be posted

Week 2: Basic R

Lecture 3

- Numbers, operators, strings, vectors, matrices

Lecture 3 video: (To be viewed before class)

Assignment 2: see Assignment section

Exercise solutions: exerciseSolutions_module2.Rmd

Lecture 4

- lists, arrays, data frames

Lecture 4 video: (To be viewed before class)

Exercise solutions: exerciseSolutions_module3.Rmd

Week 3: Making more complex scripts

Lecture 5

- Flow control, files

Lecture 5 video: (To be viewed before class)

Assignment 3: see Assignment section

Exercise solutions: exerciseSolutions_module4.Rmd

Lecture 6

- Loops, vectorization

Lecture 6 video: (To be viewed before class)

Exercise solutions: exerciseSolutions_module5.Rmd

Week 4: Functions, external programs and recipes

Lecture 7

- Functions, regular expressions

Lecture 7 video: (To be viewed before class)

Assignment 4: see Assignment section

Exercise solutions: exerciseSolutions_module6.Rmd

Lecture 8

- Interfacing with external programs
- Using programming recipes

Lecture 8 video: (To be viewed before class)

Exercise solutions: exerciseSolutions_module7.Rmd

Week 5: Packages and Bioconductor

Lecture 9

- R packages, bioconductor

Lecture 9 video: (To be viewed before class)

Assignment 5: see Assignment section

Exercise solutions: exerciseSolutions_module8.Rmd

Lecture 10

- Plotting in R

Lecture 10 video: (To be viewed before class)

yeast_properties.txt

Exercise solutions: exerciseSolutions_module9.Rmd

Week 6: Intro to Python and GitHub

Lecture 11

- Intro to Python

-Install Anaconda Individual Edition at https://www.anaconda.com/products/individual and make sure you can run JupyterLab. Install for the Python 3.x series (currently 3.8). Note this requires up to 5GB of hard drive space to download the installer and install it. If you don't have the hard drive space, just use an online service for running python notebooks like https://colab.research.google.com/

Download the python notebook we will use for the lecture and lab: python.zip (there are two associated data files included)

Lecture 11 slides: UofTMoGenPractProgBiol_Lecture11.pptx

Lecture 11 extra slides on tidyverse: UofTMoGenPractProgBiol_Lecture11-tidyverse.pptx

Lecture 11 video: (To be viewed before class)

Assignment 6: see Assignment section

Exercise solutions: solutions.ipynb

Lecture 12

- Coding and data best practices
- Intro to version control with Git and Github
- Lab: Review of course concepts

Lecture 12 slides: UofTMoGenPractProgBiol_Lecture12.pptx

Lecture 12 extra slides on AI programming: UofTMoGenPractProgBiol_Lecture12-codingAI.pptx

Lecture 12 video: (To be viewed before class)

Solutions

- ***UofTMoGenPractProgBiol_Lecture1-TutorialSolutions.pptx***
- ***exerciseSolutions_module2.Rmd***
- ***exerciseSolutions_module3.Rmd***
- ***exerciseSolutions_module4.Rmd***
- ***exerciseSolutions_module5.Rmd***
- ***exerciseSolutions_module6.Rmd***
- ***exerciseSolutions_module7.Rmd***
- ***exerciseSolutions_module8.Rmd***
- ***exerciseSolutions_module9.Rmd***
- ***Assignment1_solution.docx***
- ***Assignment2_solution.zip***
- ***Assignment3_solution.zip***
- ***Assignment4_solution.zip***
- ***Assignment5_solution.Rmd***