



What can Bioinformatics do for you?

RI-MUHC Bioinformatics Platform Seminar

bioinformatics.rimuhc@mcgill.ca

Date: 2019-03-01

Seminar Outline

- **Overview of the Bioinformatics Platform**
 - Who we are
 - Mission
 - What we do
- **"Bioinformatics data skills 101" Workshops**
 - Why, What, When, Where, for Who?
 - Real-life examples of bash, python, R, Git
 - How to register
- **Questions & discussion**



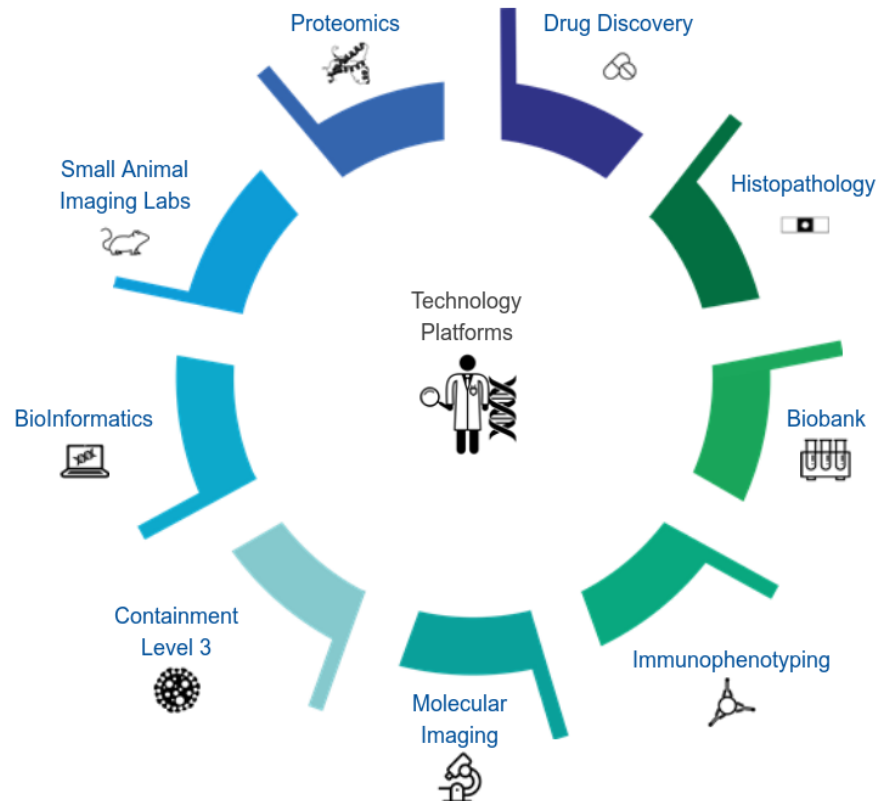
Platform Overview

Who we are & what we do



RI-MUHC Technology Platforms

- **Goal:** Accelerate research at the RI-MUHC
 - Access to technologies and instrumentation
 - Scientific expertise & training



Bioinformatics Platform

- Platform created in 2017
- **Website:** <http://rimuhc.ca/bioinformatics>
- **Team**
 - Manager & bioinformatician: **Joël Lafond-Lapalme**
 - Bioinformatician: **Daniel Jimenez**
 - Bioinformatician: **Rachade Hmamouchi** shared with the Proteomics Platform (0.5 FTE)
 - Small team, expected to grow



Mission

- **Goal:** Accelerate genomic research at RI-MUHC and its applications in clinical practice
- **Focus areas**
 1. Analytic services in genomics
 2. Support access to High-Performance Computing
 3. Management & integration of lab & clinical data
 4. Training in bioinformatics



1. Analytic Services

- **Goal:** Answer to project-specific needs in genomics
 - Project-specific data annotation & filtering
 - Focus on clinical interpretation of results
- **Analysis of genomic data**
 - **Research** (RI): WGS, WES, targeted DNA-Seq, RNA-Seq, targeted RNA-Seq
 - **Diagnostics** (MUHC): Targeted DNA/RNA-Seq
- **Data sources**
 - MUGQIC and local MiSeq instruments



2. HPC Support

- **Access to High Performance Computing (HPC)**
 - Support to create Calcul Québec and Compute Canada accounts
 - Compute Canada Rapid Access Service (RAS)
 - Support to apply to Compute Canada's annual Resource Allocation Competitions (RAC)



compute
canada



3. Management of Lab & Clinical Data

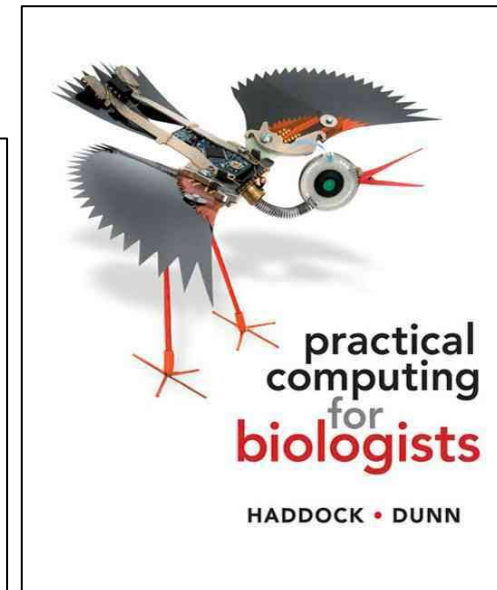
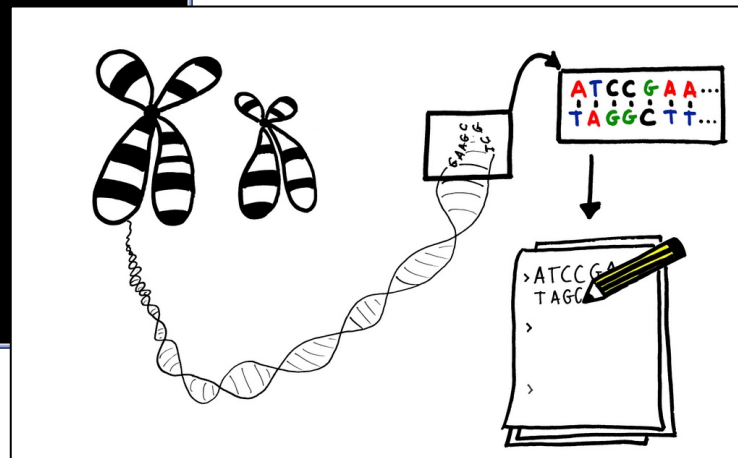
- The amount of scientific data being generated has greatly increased in recent years
- **Data accessibility** is essential but remains a major challenge
 - **Research data:** Minimize redundancy of efforts, and foster improve data sharing and pooling
 - **Clinical data:** To inspire new research questions
- **Goal:** To improve data accessibility at RI-MUHC
 - Support data an metadata collection
 - Help researchers to make their data accessible



4. Training in Bioinformatics

- **Bioinformatics workshops**
 - Training in bioinformatics data skills (Unix, Bash, R, Python, Git)
 - First series of workshops: for **beginners**
 - Target audience: **RI-MUHC trainees & staff**

```
Terminal
-rwxr-xr-x 1 sys 52850 Jun 8 1979 hptmunix
drwxrwxr-x 2 bin 320 Sep 22 05:33 lib
drwxrwxr-x 2 root 96 Sep 22 05:46 mdec
-rwxr-xr-x 1 root 50990 Jun 8 1979 rkunix
-rwxr-xr-x 1 root 51982 Jun 8 1979 r12unix
-rwxr-xr-x 1 sys 51790 Jun 8 1979 rphunix
-rwxr-xr-x 1 sys 51274 Jun 8 1979 rptmunix
drwxrwxrwx 2 root 48 Sep 22 05:50 tmp
drwxrwxr-x12 root 192 Sep 22 05:48 usr
# ls -l /usr
total 11
drwxrwxr-x 3 bin 128 Sep 22 05:45 dict
drwxrwxrwx 2 dmr 32 Sep 22 05:48 dmr
drwxrwxr-x 5 bin 416 Sep 22 05:46 games
drwxrwxr-x 3 sys 496 Sep 22 05:42 include
drwxrwxr-x10 bin 528 Sep 22 05:43 lib
drwxrwxr-x11 bin 176 Sep 22 05:45 man
drwxrwxr-x 3 bin 208 Sep 22 05:46 mdec
drwxrwxr-x 2 bin 80 Sep 22 05:46 pub
drwxrwxr-x 6 root 96 Sep 22 05:45 spool
drwxrwxr-x13 root 208 Sep 22 05:42 src
# ls -l /usr/dmr
total 0
#
```



5. Other Support

- Consulting in bioinformatics
- Involvement in grant applications
 - Methods
 - Experimental design
 - Budget
 - Support letter
- Contact: bioinformatics.rimuhc@mcgill.ca



Bioinformatics Workshops

Bioinformatics Data Skills 101



Why?

Ten simple rules for biologists learning to program

Maureen A. Carey¹, Jason A. Papin^{2*}

1 Department of Microbiology, Immunology, and Cancer Biology, University of Virginia School of Medicine, Charlottesville, Virginia, United States of America, **2** Department of Biomedical Engineering, University of Virginia, Charlottesville, Virginia, United States of America

* papin@virginia.edu

Introduction

As big data and multi-omics analyses are becoming mainstream, computational proficiency and literacy are essential skills in a biologist's tool kit. All "omics" studies require computational biology: the implementation of analyses requires programming skills, while experimental design and interpretation require a solid understanding of the analytical approach. While academic cores, commercial services, and collaborations can aid in the implementation of analyses, the computational literacy required to design and interpret omics studies cannot be replaced or supplemented. However, many biologists are only trained in experimental techniques. We write these 10 simple rules for traditionally trained biologists, particularly graduate students interested in acquiring a computational skill set.



What?

- Only **free & open-source** tools
- **4 half-day workshops**
- The **4 core components** to be up and running:
 1. Bash
 2. R
 3. Python
 4. Git



What?

- **Flipped Workshops**

- Online instructions prior to the workshops
- Software installation & reading
- A more effective pedagogical approach, but ...
- Students commitment required

- **In-class content**

- Almost exclusively practical computing (no lecture)
- Individual & small-group exercises
- Focused on problem-solving skills



When & Where?

Workshop	Date	From	To
Bash	2019-04-18 (Thursday)	12:30	16:30
R	2019-05-01 (Wednesday)	09:30	13:30
Python	2019-05-15 (Wednesday)	09:30	13:30
Git	2019-06-10 (Monday)	13:00	17:00



- **Where:** Glen, Bloc E - RI Auditorium (E S1.1129)

For Who?

- **For Beginners**
 - **Goal:** Help you being up and running with basic computing & programming tools
- Priority to **RI-MUHC members**
 - Trainees & staff
- Open to **non RI-MUHC attendees** depending on demand
 - MUHC trainees & staff (e.g. medical residents)



What will you need?

- **A Laptop:**
 - Mac or Linux (e.g. Ubuntu): Preferred
 - Windows 10: OK
 - Windows 7: Tolerated
- **A Wi-Fi connection**
 - CUSM-MUHC.MCGILL.CA
 - We can obtain a temporary login for you
- **Motivation**



Workshop #1 – The Command Line

Date: **2019-04-18**

Overview & illustrative examples



The Command Line Interface (Bash)

- **Command Line vs. Graphical user interface (GUI)**
 - A different way of interacting with your computer
 - Faster, more flexible and more powerful to navigate around your computer
- Many useful data science software work on the CLI
- Required for high-performance computing
- **Bash:** Default command-line interpreter for most Linux and macOS distributions (version also available for Windows 10)



Workshop #2 – R

Date: **2019-05-01**

Overview & illustrative examples



R

- Programming language and software environment for **statistical computing & graphics**
- Strong community
- Packages
- Application-focused development
- Easy to learn



Workshop #3 – Python

Date:

Overview & illustrative examples



Python

- **General-purpose** programming language
- Strong community
- Easy to learn
- Easy to read & understand
- **Widely used** – Organizations using Python:
<https://wiki.python.org/moin/OrganizationsUsingPython>



Workshop #4 – Git

Date:

Overview & illustrative examples



Git

- **A version control** tool
- Used to store different **versions of files**
- Used to **track changes** in the source code
- Allows multiple persons to work together
- Remote or local repository
- Excellent for **replicability** and **traceability**



Ten Simple Rules for Taking Advantage of Git and GitHub

Yasset Perez-Riverol^{1*}, Laurent Gatto², Rui Wang¹, Timo Sachsenberg³, Julian Uszkoreit⁴, Felipe da Veiga Leprevost⁵, Christian Fufezan⁶, Tobias Ternent¹, Stephen J. Eglen⁷, Daniel S. Katz⁸, Tom J. Pollard⁹, Alexander Konovalov¹⁰, Robert M. Flight¹¹, Kai Blin¹², Juan Antonio Vizcaíno^{1*}



How to Register

- **Event registration**
 - Eventbrite ([link to follow](#))
 - Link & slides on the RI-MUHC website (New & Events): <http://rimuhc.ca/bioinformatics>
- **Registration to 1 to 4 workshops**
 - Recommended to register to all four
- **Cost**
 - Nominal fee of \$20 / workshop
 - To reduce the number of no-shows



Other Training Opportunities (McGill)

- List of **data science classes** at McGill:

<http://micm.mcgill.ca/classes/>



The screenshot shows the website for the McGill initiative in Computational Medicine. The header includes the McGill logo and the text "McGill initiative in Computational Medicine". A navigation bar contains links for Home, About, Profiles, Events, Job Opportunities, Classes (which is highlighted), UNIWeb, and Contact. Below the navigation bar, the page title is "Classes (52)". There is a "Filter Classes" section with a search input field labeled "Search Classes:". At the bottom, a class is listed: "EXMD 634 Quantitative Research Methods (3 credits)".



Questions & Discussion

