Bioinformatics Training Opportunities

Mark Dunning, Bioinformatics Training Coordinator
Bioinformatics Core, Cancer Research UK, Cambridge Institute

http://bioinformatics-core-shared-training.github.io/

Introduction

Modern technologies are able to generate vast amounts of data on an unprecedented scale. Therefore numerical and computational competency are increasingly being seen as core skills for bench scientists.

University-run Courses

All members of The University of Cambridge are eligible for courses hosted in the Department of Genetics. A list of currently-scheduled courses can be obtained from http://tiny.cc/biocourses. Several courses that involve C.I. staff as demonstrators, or might be of particular interested, are listed below

- Beginner
  - An Introduction to Solving Biological Problems with Python
  - An Introduction to Solving Biological Problems with PERL
  - An Introduction to Solving Biological Problems with R
  - An Introduction to MATLAB for biologists
  - Analysing mapped NGS data with SeqMonk
  - Galaxy: Data Manipulation and Visualisation
  - Galaxy: RNA-seq and ChIP-seq data analysis
  - Whole Exome Sequencing and RNA-sequence data analysis
  - Introduction to genome variation analysis using NGS
  - Using the Ensembl Genome Browser
- Intermediate
  - Analysis of High-throughput sequencing data with Bioconductor
  - Microarray Analysis with Bioconductor
- Advanced
  - A Practical Introduction to Good Programming Practices
  - R object-oriented programming and package development

Interested in becoming a trainer?

We are always keen to recruit Bioinformaticians to the training roster. If you think you can contribute to one of our existing courses, or can spot a gap in our schedule, please let us know!

Acknowledgements

The current roster of course developers and presenters is as follows;

- Chandu Chilamakuri
- Ines de Santiago
- Henry Farmery
- Peter MacCallum
- Suraj Menon
- Marc O'Brien
- Anne Pajon
- Bernard Pereira
- Oscar Rueda
- Rory Stark
- Jing Su
- Sarah Vowler

References

Vance Ashlee. Data analysts captivated by R's power. *The New York Times*, New York Edition:B6, 09.

James Cheshire. Great maps with ggplot2, February 2012. URL `http://spatial.ly/2012/02/great-maps-ggplot2/`.

Aims

- Ensure that researchers leave the institute with competency in numerical and computational analysis. We focus on core skills in statistics, experimental design, R, data visualisation and unix. These are discussed briefly in more detail below.
- Increase researcher independence by promoting interfaces such as IGV, Galaxy and UCSC for routine Bioinformatics analysis and data exploration
- Monitor and advertise Bioinformatics training opportunities that arise in the University

Bioinformatics Courses at CI - 2015 Schedule

Seminar room 215 in Cancer Research Uk Cambridge Institute is block-booked for training two days a month. The courses in this room are run in collaboration with the Bioinformatics Training Programme at the Department of Genetics. However, priority to attend the course and first notification of courses will be given to CCC members. Participants are expected to bring their own laptops and instructions are provided on how to install the relevant software.

- Monday 19th October - Further Staistical Analysis using R
- Monday 16th November - Experimental Design
- Tuesday 17th November - Introduction to Statistical Analysis
- Monday 30th November - Basic Unix
- Friday 4th December - Scientific Figure Design (live webcast from Department of Genetics)
- Monday 14th - Tuesday 15th December - Introduction to Solving Biological Problems With Python

Unix

The Unix shell has existed since the early days of computers, and yet is still the preferred way to run many popular Bioinformatics tools. The concept of typing commands to run programs is often alien to those that are used-to a GUI interface, so we aim to take the novice and turn them into a beginning Linux user. We describe the Linux environment so they can start to utilize command-line tools and feel comfortable using a text-based way of interacting with a computer. Teaching is done via a 'virtual machine' so that participants can experiment without making permanent changes to their own machine.

R

Figure 1 : R is gaining press attention (Ashlee [09])

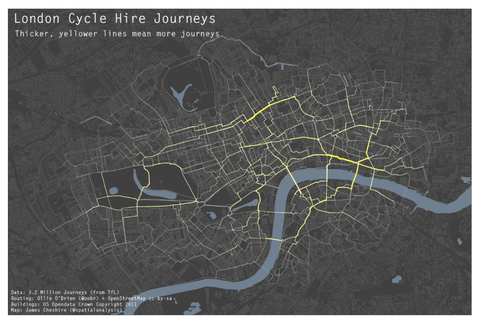

Figure 2 : R can produce high-quality graphics (Cheshire [2012])

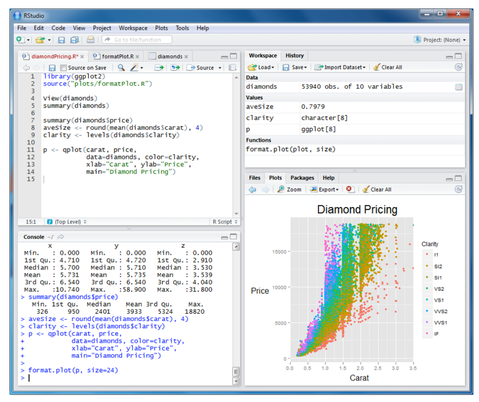

Figure 3 : Rtsudio promotes reproducible research

The R programming language is our language of choice for data analysis. Not only is it open-source and freely-available, but it is capable of generating publication-quality images and has a large user-base amongst the scientific community. However, it comes with a steep learning-curve especially for those without prior exposure to command-line tools. Although it requires an investment of time, it is ultimately rewarding and a valuable skill to acquire.
R, and especially the RStudio interface, also promotes the practice of Reproducible Research. We are actively-engaged in restructuring the University's long-standing R course so that a key component is the generation of reproducible analyses and reports.

Experimental Design & Statistics

"To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of." - R.A. Fisher
This popular quote by Fisher sums up why experimental design is a crucial part of the experimental process. For this reason, our training course on Experimental Design is highly-recommended for all new PhD students.
For statistics training, we have developed a series of online tools using Shiny (http://shiny.rstudio.com/); a framework for developing web applications using R. Therefore the focus of this training is on the statistical concepts, rather than wrestling with the R language. In this course we introduce the basics of statistics and hypothesis testing so that participants can conduct their own simple statistical testing, and be able to interpret published results. We are also developing follow-on courses in Statistics for those familiar with R.

Data Visualisation and Exploration

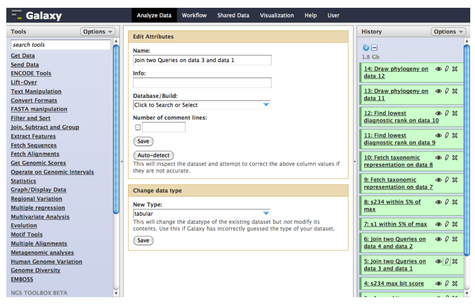

Figure 4 : Galaxy
https://usegalaxy.org/

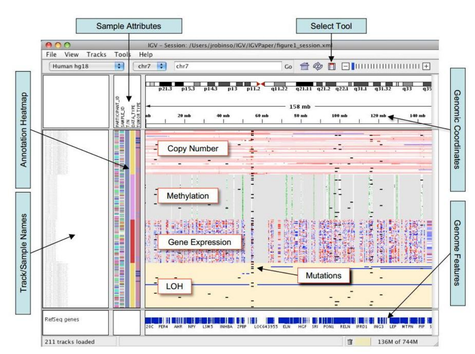

Figure 5 : Integrative Genomics Viewer (IGV)
http://www.broadinstitute.org/igv/

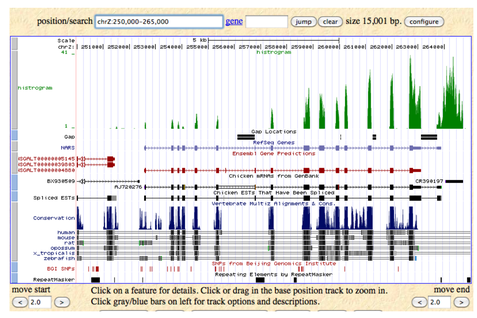

Figure 6 : U.C.S.C Genome Browser
https://genome.ucsc.edu/

Arguably the best way to develop an understanding of high-throughput data is to get into the habit of visualising the data. There are many tools that provide the opportunity to do this. We are currently developing introductory and advanced courses to introduce such tools. Researchers will be show how to load their own data into these tools and overlay with other genomic 'tracks' and perform simple operations in the most-popular tools.

Course Materials

Our courses are based-on open-source and freely-available software. Materials from our courses are developed and shared via Github: https://github.com/bioinformatics-core-shared-training. We also have collaborations and share material with the Bioinformatics Training Programme (Department of Genetics) and MRC Clinical Sciences Centre (London).

GitHub