Bioinformatics training

Gabriella Rustici, PhD - School of Biological Sciences  
gr231@cam.ac.uk
Why is bioinformatics training important?

- Data analysis is now the major bottleneck to research in the molecular life sciences.
- Many biomedical professionals feel under-qualified to make the most of biological data.
- There are an estimated 3 million life scientists in Europe alone; >20 million healthcare professionals.
- Potentially all of them are producers or consumers of data managed by Europe’s biomedical research infrastructures.
ELIXIR: a distributed infrastructure for life science

Building a **sustainable European infrastructure for biological information**, supporting life science research and its translation to medicine, agriculture, bioindustries and society

ELIXIR unites Europe’s leading life science organisations in **managing and safeguarding the massive amounts of data** being generated every day by publicly funded research

ELIXIR will **provide the facilities necessary for life science researchers to make the most of our rapidly growing store of information** about living systems, which is the foundation on which our understanding of life is built
ELIXIR “platforms” organization

• Data
  Sustain core data resources

• Tools
  Services & connectors to drive access and exploitation

• Compute
  Access, Exchange & Compute on sensitive data

• Standards
  Integration and interoperability of data and services.

• Training
  Professional skills for managing and exploiting data
ELIXIR-UK: Training remit

To facilitate training of research scientists and infrastructure technologists in bioinformatics, computing, statistics and biology, in partnership with UK centres, industry and other ELIXIR Nodes.

- Computational Genomics Analysis and Training (CGAT)
- Queen Mary University London
- The Genome Analysis Centre (TGAC)
- The Oxford e-Research Centre
- The Wellcome Trust Sanger Institute
- University College London
- University of Birmingham
- University of Cambridge
- University of Cardiff & NERC EOS Centre
- University of Edinburgh
- University of Liverpool Centre for Genomic Medicine
- University of Manchester
ELIXIR training strategy

- Facilitate accessibility to Europe’s bioinformatics resources by up-skilling researchers who can more effectively exploit the data, tools, standards and compute services provided by ELIXIR.
- Support and train users through e-learning, face-to-face courses and programs held across Europe.
- Develop a coordinated pan-European training program of high quality and impact.
- Partnership with global efforts such as GOBLET.
GOBLET - http://mygoblet.org/

• Global Organisation for Bioinformatics Learning, Education & Training

• Provide a global, sustainable support and networking structure for bioinformatics trainers and trainees, including (i) a training portal for sharing materials, tools, tips and techniques; (ii) guidelines and best practice documents; (iii) facilities to help “Train the Trainers”; and (iv) offering different learning pathways for different types of learner
Bioinformatics postgraduate training @ UoC

**Total number of courses**

- Downing Site
- CRUK
- EMBL-EBI
- External

**Total number of trainees**

- 2011
- 2012
- 2013
- 2014
- 2015
Bioinformatics postgraduate training @ UoC

Core skills

- Interpreting the clinical genome with Decipher
- Biological data analysis using Intermine
- Mouse Genome Informatics (MGI) workshop
- EMBL-EBI courses:
  - Introduction to EMBL-EBI resources
  - An introduction to sequence searching
  - Exploring protein sequence and functional information with UniProt
  - GWAS catalog
  - Interactions & pathways – IntAct
  - Interactions & pathways – Reactome
  - Introduction to ontologies
  - Metabolomics databases and tools
  - Network analysis - Cytoscape and PSICQUIC
  - Small molecules resources
  - Transcriptomics data and tools
  - Ensembl API workshop
  - Using the Ensembl genome browser

Databases and services

- Introduction to genome variation analysis using NGS
- Introduction to RNA-seq and ChIP-seq data analysis
- Introduction to Scientific Figure Design
- Mathematical and computational modelling in biology
- Network visualization and analysis
- Molecular phylogenetics
- R object-oriented programming and package development
- Variant analysis with GATK

Specialized training

- Analysis of DNA methylation using sequencing
- Analysis of HTS data with Bioconductor
- Analysis of mapped HTS data with SeqMonk
- Analysis of single cell RNA-seq
- How to get started with sequencing analysis: the metagenomics example
- Image analysis for biologists
- Introduction to Galaxy: data manipulation and visualisation
- Introduction to Galaxy: RNA-seq and ChIP-seq data analysis
- Analysis of mapped HTS data with SeqMonk
- Analysis of single cell RNA-seq
- How to get started with sequencing analysis: the metagenomics example
- Image analysis for biologists
- Introduction to Galaxy: data manipulation and visualisation
- Introduction to Galaxy: RNA-seq and ChIP-seq data analysis

Data Carpentry
- Introductory statistics and experimental design for genomics
- Software Carpentry
- Statistical analysis using R

Data Carpentry
- Introductory statistics and experimental design for genomics
- Software Carpentry
- Statistical analysis using R

http://bioinfotraining.bio.cam.ac.uk/
http://training.csx.cam.ac.uk/bioinformatics/event-timetable
What do we aim for?

- **Course content.** Courses consist of a well balance mixture of lectures and hands on sessions

- **Software choice.** Focus on the use of open source, stable, actively developed and well-maintained software tools (i.e. Bioconductor, Galaxy,…)

- **Objectives.** Trainees should learn:
  - how to interpret biological data;
  - what a specific data analysis pipeline entails; and
  - how to critically evaluate the data analysis tools available.

- **Objectives.** We want to enable you to establish a partnership with your statistician and/or bioinformatician collaborators, based on mutual understanding
Data Carpentry develops and teaches workshops on the fundamental data skills needed to conduct research

Focus is on the introductory computational skills needed for data management and analysis in all domains of research

Target audience is learners who have little to no prior computational experience

Topics:

1. Data organization in spreadsheets
2. Data cleaning with OpenRefine
3. Introduction to R
4. Data analysis and visualization in R
5. SQL for data management
Upcoming training

• 1st summer school – will include training on basic programming, including UNIX scripting and command-line tools, in the context of HTS data analysis and interpretation

  NGS Training Biology Genetics Software Approach
  Data Analysis Introductory RNA-seq
  Bioinformatics Sequencing Protein Genome

• Scientific databases, including relevant resource from EMBL-EBI such as nucleotide databases, literature services, gene expression databases, etc.

• Data publishing
Sources of training materials/information

- ELIXIR training portal: https://tess.elixir-uk.org/
- GOBLET training portal: http://mygoblet.org/training-portal
- See individual course pages on: http://bioinfotraining.bio.cam.ac.uk/
- Online resources: EMBL-EBI Train online, EdX (Data for life sciences from Harvard-Irizarry), Coursera, etc.
- Training course catalogue: https://www.on-course.eu/
- And many more…….