

Algorithms for Functional Genomics
COMP90014
Subject Outline
2012

Credit Points	12.5
Coordinator	Dr L Stern
Prerequisites	One semester computer programming would be desirable
Semester	2
Contact	Twenty-four hours of lectures Eleven hours of laboratory

In 2012 the timetable for this subject is:

- Lecture: Thursday 12:00–2:000, Chemical and Biomolecular Engineering Theatre 2
- Laboratory: Thursday 2:15–3:15, Alice Hoy 109

1 Students and Prerequisites

This subject is an offering in the Master of Biomedical Engineering, the Master of Engineering (Biomedical), and the MSc(Bioinformatics). It is also open to students in the MSSE (Master of Software Systems Engineering), the MIT (Master of Information Technology), the MEDC (Master of Distributed Computing), Computer Science Honours, to NICTA PhD students, and to other students, including undergraduate students, by permission of the coordinator.

No prior biology background is assumed.

The ability to read scientific literature is required and some facility with using and exploring new software is expected. No computer programming will be required.

2 Subject Description

This subject is an in-depth look at the computational side of functional genomics, an active area of bioinformatics in which gene expression is studied. While every cell in the body has the same set of genes, these genes are expressed differently in different tissues. Brain cells, for example do not make specialized kidney proteins, or *vice versa*. Importantly, expression levels are aberrant in some medically important diseases, like cancer.

In this subject we look at current methods for studying functional genomics, with an emphasis on computation. We concentrate on two current approaches to functional genomics: microarrays and next

generation sequencing (RNA-seq). The content includes theoretical background and low-level and high-level analysis of data. Students will read and discuss selected research papers in the current literature and will use a software application associated with next generation sequencing assembly.

3 Aims

The aim of this subject is to give students an in-depth view of a currently active area in bioinformatics, with a view to preparing them for working in this or related areas or for further study. Students will be exposed to the role of computation in this area, the biomedical context of expression analysis, and current issues in the leading edge technology in the field.

4 Objectives

On completion of this subject, students should be able to:

- explain key concepts in functional genomics.
- describe current research issues in functional genomics.
- read current articles in the functional genomics literature.
- describe algorithms for analysis of functional genomics data, and evaluate their strengths and weaknesses.
- effectively use application software for assembling next generation sequencing data.

5 Generic Skills

On completion of this subject, students should be able to:

- apply electronic resources effectively in solving new problems.
- relate experimental findings to theory.
- read and critically analyze scientific literature.

6 Lecture schedule (approximate)

Week 1: 27 July

Introduction to Functional Genomics

Weeks 2–4: 2, 9, 16 August

Microarray analysis, including:

Microarray background

Data Normalization

Clustering, classification, heatmaps

Week 5: 23 August

Gene signatures for disease prognosis (Guest lecturer)

Week 6: 30 August

Feature Selection

Workshop on Gene Signature papers

Weeks 7–8: 6, 13 September

Next generation sequencing and RNA-seq, including:

Sequencing background

Base calling

Contig assembly

Non-teaching period: 20, 27 September

Week 9: 4 October

Sequence assembly (Guest lecturer)

Week 10: 11 October

Workshop on assembly paper

Week 11: 18 October

Advanced topics in NGS for functional genomics, including:

RNA-Seq

Error Correction (Guest lecturer)

Week 12: 25 October

Review

7 Laboratories and Tutorials

Laboratory and tutorial sessions will not be held every week. There will be tutorial sessions in weeks 6 and 7 to support assignment 1, and laboratory sessions in weeks 8, 10 and 11 to support assignment 2. The session in week 8 will be for students without previous experience with UNIX. Additional tutorial sessions will be held as needed. Tutorial questions will be generated by students.

8 Assessment

The weighting of assessment components is:

- Practical work (30%), spread over two assignments.
- End of semester examination (70%).

In order to pass the subject, students must achieve at least 15/30 on the practical work and at least 35/70 on the final examination.

Practical Work

Practical work stresses the application of the theoretical work covered in lectures. In the first assignment, students will apply their knowledge of microarray work to the development of gene signatures for clinical work. In the second assignment, students will get hands-on experience with software used for next generation sequence assembly.

- **Assignment 1: Critical analysis of gene signatures as clinical tools**

- Available 24 August
- Due 12 September
- 15 marks

In this assignment, students write a critical analysis of the use of gene signatures as tools in the clinical setting. The report will be based on reading of seminal papers on developing gene signatures for clinical use (provided) and later papers discussing the issues involved in developing stable, useful signatures (also provided). Reading beyond the provided papers will help deepend students' understanding and the quality of analysis.

- **Assignment 2: VELVET algorithm for Next Generation Sequencing Assembly**

- Available 5 October
- Due 24 October
- 15 marks

In this assignment, students will use a computer program for analyzing next generation sequence data, and explore the assembly process. The proposed software is the **Velvet** suite of programs. Real data will be used, and students will compare their results with published results.

A report with a critical analysis of results and of the software will be written, approximate length: 3–5 pages.

9 Resources

There is no prescribed text for the subject. Written resources available to students include:

- Lecture notes and recordings of lectures.
- Subject handouts.

References to relevant articles will be given in the lectures and lecture notes. Students are expected to read these articles as relevant, and are also encouraged to browse the current bioinformatics literature and read relevant articles. A particularly relevant journal is *Bioinformatics*, available electronically through the university library.

Lecture notes

Lecture notes and a link to the recordings will be available on the LMS.