COURSE OUTLINE: BS4017

Course Title	**High-Throughput Bioinformatics**		
Course Code	**BS4017**		
Offered	Study Year 3, Semester 1		
Course Coordinator	Jarkko Tapani Salojarvi (Asst Prof)	Jarkko@ntu.edu.sg	6904 7231
Pre-requisites	BS1009		
AU	3		
Contact hours	Lectures: 26, Laboratories: 15, Tutorials: 7		
Approved for delivery from			
Last revised	9 Apr 2021, 13:58		

Course Aims

High throughput sequencing has made it possible to quantify the genomes and transcriptomes of any species. Transcriptome sequencing can produce evidence about molecular responses to environmental stimuli, whereas genome sequencing is used to study the evolution and adaptation in the particular species. In this course you will develop skills to process and analyze both genome and transcriptome sequencing data from most common sequencing platforms and to carry out data analysis in species with or without a reference genome. Practical analysis is done using free and open-source software tools with an emphasis on the command-line Linux computing environment.

Intended Learning Outcomes

Upon successfully completing this course, you should be able to:

1. Preprocess next generation sequencing data such that it can be used in subsequent analyses
2. Align genomic sequencing data to a reference genome or carry out a de novo assembly to obtain a reference
3. Perform RNA-sequencing data analysis in a model organism
4. Carry out variant calling in population data and perform general data analyses

Course Content

How next-generation sequencing data is produced and preprocessed

How to use a Linux operating system and command line interface

General tools for processing different data types and the file format standards for each data type.

How to analyze transcriptomics and genomics sequencing data

The common research questions that can be addressed with genome sequencing and transcriptomics data

Assessment

Component	Course ILOs tested	SBS Graduate Attributes tested	Weighting	Team / Individual	Assessment Rubrics
Continuous Assessment					
Lectures					
Multiple Choice Questions	1, 2, 3, 4	**1.** a, b, d **2.** a, b, c, d, e, f **3.** c, g **5.** c, e **7.** d	5	individual	
Tutorials					
Project	1, 2, 3, 4	**1.** a, b, d **2.** a, b, c, d, e, f, g, h **3.** a, b, c, e **4.** a, b, c **5.** a, b, c, d, e **6.** c, d **7.** a, b, c	30	individual	See Appendix for rubric
Mid-semester Quiz					
Multiple Choice Questions	1, 2	**1.** a, b, d **2.** a, b, c, d, e **3.** c, g **5.** c, d, e **7.** a, b, c, d	15	individual	
Examination (3 hours)					
Short Answer Questions	3, 4	**1.** a, b, d **2.** a, b, c, d, e, f **3.** a, b, c, e, g **5.** c, d, e **6.** c, d **7.** a, b, c	50	individual	See Appendix for rubric
		Total	**100%**		

These are the relevant SBS Graduate Attributes.

1. **Recognize the relationship and complexity between structure and function of all forms of life, resulting from an academically rigorous in-depth understanding of biological concepts**

   a. Possess a conceptual framework that identifies the relationships between the major domains in the field of biology.

   b. Explain the relationship between structure and function of all forms of life at the molecular level

   d. Explain the relationship between structure and function of all forms of life at the organism level

2. **Critically evaluate and analyze biological information by applying the knowledge, scientific methods and technical skills associated with the discipline**

   a. Identify the assumptions behind scientific problems and issues

   b. Create and evaluate hypotheses

   c. Create abstract models of data

   d. Design experiments relevant to authentic problems and their models

   e. Analyze the validity of qualitative and quantitative scientific data

   f. Evaluate results in primary biological literature

   g. Evaluate the results of their own experiments and decide on the next step

   h. Identify unintended results as opportunities for discovery

3. Develop and communicate biological ideas and concepts relevant in everyday life for the benefit of society

a. Simplify and explain scientific concepts and results of experiments to a non-biologist (avoiding jargon)

b. Display and explain scientific results clearly and persuasively to peers both verbally and in writing (includes the ability to graph data appropriately and accurately).

c. Demonstrate an understanding of the recursive nature of science, where new results continually modify previous knowledge

e. Discuss current critical questions in the field of biology

g. Demonstrate an understanding of the history of ideas and development of the major fields of biology

4. Acquire transferable and entrepreneurial skills for career development

a. Demonstrate innovative approaches to solving problems in biological science, leading to new approaches or techniques

b. Demonstrate a flair for developing new technologies, attracting funding, marketing products and respecting IP rights

c. Demonstrate a flair for conducting research

5. Develop communication, creative and critical thinking skills for life-long learning

a. Learn independently and then share that knowledge with others

b. Learn collaboratively and be willing to share expertise with peers

c. Demonstrate critical thinking skills such as analysis, discrimination, logical reasoning, prediction and transforming knowledge

d. Question the assumptions, sources, and contexts of scientific investigation

e. Demonstrate good observation skills and a curiosity about the world

6. Develop codes of social responsibility and scientific ethics, particularly in relation to biological advancement and applications

c. Respect regulations involving plagiarism and copyright

d. Respect requirements regarding confidentiality, data protection, conflict of interest, and falsification of data

7. Demonstrate information literacy and technological fluency

a. Locate and evaluate information needed to make decisions, solve problems, design experiments, and understand scientific data

b. Work effectively with common technologies in biology

c. Evaluate and use biological databases (literature and public datasets)

d. Complete online learning independently

Formative Feedback

You will carry out an individual project on genome or transcriptome sequencing data during this course.
The project will include data preprocessing (LO1), sequence alignment (LO2) and data analysis of transcriptome (LO3) or genome data (LO4), for which you will obtain feedback during the tutorials and computer classroom exercises.
In addition to the project work, your performance will be assessed by MCQs for the lectures (LOs 1-4) and a mid-term exam (LOs 1-2) for which there will be feedback on the overall performance of the class. Feedback on the final exam (assessing LOs 3-4) is given in examiner's report after the exam period.

Learning and Teaching Approach

Lectures (26 hours)	During the lectures we will go through the procedures of next generation sequencing data analysis (LOs 1-4) with emphasis on the critical thinking of the experimental design and possible pitfalls in the analysis. You may encounter these issues in the individual project work carried out during the course.
Laboratories (15 hours)	In the computer classes we will go through the necessary software tools for carrying out the project work (LOs1-4). Half of the time is used for working with the individual project data and for troubleshooting.
Tutorials (7 hours)	Each of you will have an individual project work where the person analyses either genomics or transcriptomics sequencing data. In the project, you will carry out all data analysis tasks for the sequencing data, such as preprocessing of the data (LO1), alignment of the data (LO2) and, depending on the project type, either analysis of the transcriptome data (LO3) or population data (LO4).

Reading and References

The lecture slides will give provide the necessary information for passing the course. Recommended further reading of each lecture topic will be given in the end of each lecture.

Course Policies and Student Responsibilities

In preparation for the tutorials you are expected to complete your project work up to the required milestone.

Academic Integrity

Good academic work depends on honesty and ethical behaviour. The quality of your work as a student relies on adhering to the principles of academic integrity and to the NTU Honour Code, a set of values shared by the whole university community. Truth, Trust and Justice are at the core of NTU's shared values.

As a student, it is important that you recognize your responsibilities in understanding and applying the principles of academic integrity in all the work you do at NTU. Not knowing what is involved in maintaining academic integrity does not excuse academic dishonesty. You need to actively equip yourself with strategies to avoid all forms of academic dishonesty, including plagiarism, academic fraud, collusion and cheating. If you are uncertain of the definitions of any of these terms, you should go to the Academic Integrity website for more information. Consult your instructor(s) if you need any clarification about the requirements of academic integrity in the course.

Course Instructors

Instructor	Office Location	Phone	Email
Jarkko Tapani Salojarvi (Asst Prof)	SBS-02s-88f	6904 7231	Jarkko@ntu.edu.sg

Planned Weekly Schedule

Week	Topic	Course ILO	Readings/ Activities
1	Introduction to Linux and command line	1	
2	Sequencing platforms, raw data types, data repositories	1	
3	Data preprocessing and quality control	1	
4	De novo assembly - transcriptomes and genomes	2	
5	De novo assembly - annotation	2	
6	Transcriptomes: experiment design, data processing, differential expression	3	
7	Transcriptomes: multiple testing, enrichment analysis, annotation	3	
8	Transcriptomes: network analysis	3	
9	Genome re-sequencing: SNPs, structural variation, pan-genomes	4	
10	SNP calling	4	
11	Population data analysis	4	
12	Comparative genomics	4	
13	Wrap up - summary	1, 2, 3, 4	

Appendix 1: Assessment Rubrics

Rubric for Tutorials: Project (30%)

This is an individual project work where transcriptomics or genomic sequencing data is analyzed.

In order to pass, you will need to complete all the project work steps. Each step will be explained in the tutorial lessons, and you have to complete that step for the next tutorial in two weeks time in order to obtain a pass mark.

In addition to passing the analysis steps, you will summarize the work in a project report, which will be graded. A high mark results if you are able to provide biological insight from the data analysis. Medium mark will result if you demonstrate good understanding of the project work and show some ability to interpret the results. Weak mark results from superficial understanding of the results.

Rubric for Examination: Short Answer Questions (50%)

The final examination will consist of short answer questions testing learning outcomes.

A high mark will result if you can apply your knowledge of LOs 1-4 in practice, for example by showing ability to design experiments for answering new research questions or by showing ability to express criticism to possible problems in existing experiments.

A medium mark will result from showing some insight and demonstrating good understanding of learning outcomes.

A low mark will result from erroneous and superficial demonstration of the learning outcomes.

Appendix 2: Intended Affective Outcomes

As a result of this course, it is expected you will develop the following "big picture" attributes:

Aware of the differences between exploratory, data-driven and hypothesis driven research

Aware of the analysis problems resulting from high dimensional data

Aware of the genomic variation between individuals and different species and its consequences.