COURSE DESCRIPTION CARD - SYLLABUS

Course name
NGS Data Processing [S2Bioinf2>NGS]

Course

Field of study
Bioinformatics

Year/Semester
1/2

Area of study (specialization)
–

Profile of study
general academic

Level of study
second-cycle

Course offered in
Polish

Form of study
full-time

Requirements
compulsory

Number of hours

Lecture
15

Laboratory classes
15

Other
0

Tutorials
0

Projects/seminars
0

Number of credit points

2,00

Coordinators

dr hab. inż. Aleksandra Świercz
aleksandra.swiercz@put.poznan.pl

Lecturers

Prerequisites

A person undertaking a second degree in Bioinformatics should have achieved the results of education from the 1st degree of this field of study, as defined in the Resolution of the Senate PUT - these effects are presented on the website of the Faculty of www.cat.put.poznan.pl. The student starting this module should have a basic knowledge of statistics, algorithms and genomics. The student must present attitudes such as honesty, responsibility, perseverance, cognitive curiosity, creativity, personal culture, respect for other people.

Course objective

1. Providing students with basic knowledge of new high-pass sequencing technologies. 2. Familiarize students with various problems and issues that can be solved with next-generation sequencing. Familiarise students with alternative splicing, determining differences between genomes of individuals of the same species, de novo assemby and resequencing. 3. Develop students' ability to apply the knowledge they have learned to solve the above-mentioned problems.

Course-related learning outcomes

Knowledge:

1. Is familiar with the methods and tools used in the process of solving complex bioinformatics tasks, mainly of an engineering nature
2. Knows and can use specialized IT and bioinformatics tools
3. Has knowledge of bioinformatic analysis on the genome scale based on a statistical background
4. Is familiar with basic sequencing technologies, has knowledge and is able to use basic genomic sequence analysis methods and tools in the context of these issues
5. Knows the development trends of bioinformatics, understands and can draw conclusions from scientific publications

Skills:
1. Is able to use the methods and IT tools learned to solve biological problems, is able to assess their usefulness and, if necessary, propose an alternative solution
2. Is able to draw conclusions from conducted experiments,
3. Uses various sources of data, including scientific publications and is able to use them to perform current tasks
4. Formulates and tests hypotheses related to the bioinformatics issues discussed, e.g. studying differences in gene expression levels, studying alternative splicing, etc.

Social competences:
1. Understands the need to systematically search for new solutions, familiarize oneself with scientific journals, also in English, in order to deepen bioinformatics knowledge
2. Is willing to share knowledge with others and to seek expert opinions.

Methods for verifying learning outcomes and assessment criteria

Learning outcomes presented above are verified as follows:

Formative assesment:
(a) Lectures, verification of the intended learning outcomes shall be carried out:
• based on your activity in discussing the material in question;
(b) laboratories: verification of the intended learning outcomes shall be carried out:
• on the basis of the current progress of the tasks;
• assessment for activity in class or lack of preparation for classes
• evaluation of reports prepared partly during the classes and partly after their completion
• assessment of the student's laboratory exercises
Summary assessment:
(a) Lectures, verification of the intended learning outcomes shall be carried out by:
• assessment of knowledge and skills demonstrated in a written, problem-based final test. The test consists of six questions; the total number of points is 5. To get a positive mark, you must get 50% of the points.
• those who prepare a presentation related to a scientific publication concerning the subject are exempt from the test. In such a case, the final grade will be the same as the grade from the laboratory
b) laboratories: verification of the intended learning outcomes shall be carried out:
• assessment of reports from classes prepared independently by students. A maximum of 5 points can be awarded for each report
• additionally, a presentation of one of the scientific publications agreed with the instructor. A maximum of 10 points can be awarded for this task
• passing the subject requires obtaining a total of 50% of points
Activity during classes is rewarded with additional points (which can increase the final grade), in particular for:
• discussion of additional aspects of the issue,
• effectiveness of applying the knowledge acquired when solving a given problem,
• comments leading to the improvement of teaching materials or the teaching process.

Programme content

The lecture programme covers the following topics: Familiarize students with new high-throughput sequencing technologies (NGS, TGS). Different methods of generating libraries for the Illumina sequencer. Sequencing of single reads and paired reads. DNA sequencing; de novo sequencing approaches and methods; resequencing, i.e. mapping reads to a reference genome; Smith-Waterman algorithm; Burrows-Wheeler transformation. Graphs in the context of assembly algorithms; overlap

graphs and DNA graphs. RNA sequencing; searching for new splice sites for transcriptomes, examining gene expression levels for several samples, analyzing short RNAs (e.g. miRNA, piRNA). Single cell sequencing. Different approaches and algorithms for solving these problems. Methods of analyzing data for genomic differences (CNV, SNP).
Laboratory exercises are conducted in the form of seven/eight two-hour classes held in a computer laboratory. The first classes are designed to familiarize students with the rules of laboratory use and the reckoning of exercises. The exercises are carried out independently by each student. The laboratory program covers the following topics: familiarize students with the available tools for analyzing data from next-generation sequencers, browsing databases for experimental data. Use of available tools for problems discussed in lectures and laboratory activities: pre-filtering data by quality, finding and cutting off adapters in sequencer readings, DNA and RNA mapping to reference genome, gene expression level testing, alternative splicing gene search, short RNA analysis, de novo transcriptome assembly, visualization of mapping results in IGV.

Course topics

DNA sequencing using various sequencers (NGS, TGS), whole-genome sequencing, short molecule sequencing, or selected genome fragments.
Gene expression level studies - discussion of activities from RNA library preparation, quality control, normalization to expression arrays.
Single-cell sequencing experiments - advantages and problems.
De novo genome sequencing.
Study of genomic differences.

Teaching methods

A lecture illustrated with a multimedia presentation containing the programme content in question, enriched with examples;
Laboratories: practical exercises in data analysis, presentations, discussion, group work

Bibliography

Basic:
1. N. Rodriguez-Ezpelta, M. Hackenberg, A.M. Aransay eds. „Bioinformatics for high throughput sequencing”, Springer, 2012
2. TA Brown, „Genomy”, PWN

Additional:
1. M.Zhao, Q. Wang, Q. Wang, P. Jia, Z. Zhao “Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives”, BMC Bioinformatics, 2013, 14:S1
2. Briefings in Bioinformatics vol. 11 issue 5, 2010, Special Issue: Second generation sequencing.

Breakdown of average student's workload

	Hours	ECTS
Total workload	50	2,00
Classes requiring direct contact with the teacher	30	1,00
Student's own work (literature studies, preparation for laboratory classes/ tutorials, preparation for tests/exam, project preparation)	20	1,00