COURSE UNIT (MODULE) DESCRIPTION

Course unit (module) title	Code
Big data analysis in genomics	

Lecturer(s)	Department(s) where the course unit (module) is delivered
Coordinator: lect. K. Kvederavičiūtė (12 hours of lectures, 36 hours of exercises) Other: Dr. Maria Fernanda Torres Jimenez (4 hours of lectures, 12 hours of exercises)	Life Sciences Centre, Vilnius University

Study cycle	Type of the course unit (module)
Full-time studies (2^{nd} stage)	Obligatory

Mode of delivery	Period when the course unit (module) is delivered	Language(s) of instruction
Face to face	Spring semester	Lithuanian/English

Requirements for students	
Prerequisites: Fundamentals of bioinformatics, statistics and R (or programming). Fundamentals of genetics, biology, molecular biology and biochemistry.	**Additional requirements (if any):**

Course (module) volume in credits	Total student's workload	Contact hours	Self-study hours
5	133	64	69

Purpose of the course unit (module): programme competences to be developed

The aim of this course is to introduce master's students of the genetics study program to the origin of Next Generation Sequencing (NGS) data, principles of analysis, and provide basic skills for solving various NGS tasks. These skills will enable students to independently deepen their knowledge in specific NGS areas.

Specific competencies that students will acquire:

- Origin of NGS data and related limitations
- Quality control and assurance of NGS data
- Selection and practical application of NGS tools for standardized data analysis
- Interpretation and presentation of results from NGS analysis tools

General skills:

- Critical and analytical thinking
- Ability to search for data information sources, analyze, visualize, and organize acquired data
- Ability to search for and select appropriate tools/methods/resources for problem-solving
- Ability to properly describe and present scientific results

After successfully completing the course, students will be able to appropriately select NGS specifications based on biological problems: suitable tools, analysis algorithms, and perform standardized data analyses. Additionally, having successfully completed the course, students will be able to apply chosen tools to real-life standardized NGS data and, if necessary, find and adapt new tools. By the end of the course, students will be able to create small workflows (automate tasks) using freely available tools. Ultimately, having successfully completed the course, students will be able to answer posed scientific questions and present results properly using NGS data and their analysis outcomes.

Learning outcomes of the course unit (module)	**Teaching and learning methods**	**Assessment methods**
Will be able to assess the quality of both new and publicly available NGS data types, and if necessary, clean the data considering identified issues.	Lectures, problem-based teaching, analysis of model real life cases, discussions, literature review	Tests (closed-ended, open-ended questions), practical assignments.
Will be able to analyze/collect various types of NGS data (e.g. RNA-Seq, genome assembly, WGBS, etc.) and evaluate the quality of analysis and obtained results.		
Based on individual experiment parameters and ideas, will be able to search for and find suitable software tools for solving specific questions.		
Based on the obtained individual results, will be able to select appropriate and optimal parameters for analysis, as well as properly visualize and interpret the obtained results.		

Content: breakdown of the topics	**Contact hours**						**Total contact hours**	**Self-study hours**	**Self-study work: time and assignments**
	Lectures	Consultations	Seminars	Exercises	Laboratory work	Internship/work placement			**Assignments**
NGS Data Sources, NGS Data Quality Inspection and Assurance	4			5			9	4	Application of different tools for data quality assessment. Comparison of tool results, identification of issues, and problem-solving using various tools. Practical tasks aimed at learning to retrieve information from public online databases and correctly download data, both using graphical interfaces and specialized tools (e.g., SRA-toolkit).
Transcriptome-level NGS Data Analysis: Experimental Section Methodology, Data Preparation, Classical Tools, Standard Results, and Result Interpretation.	4			10			14	16	Collection and analysis of reference based transcriptome-level data in prokaryotes. Quality control and visualization of collected data. Identification of differentially

									expressed genes and analysis of the identified genes.
Genome-level NGS Data Analysis: Methodology of the Experimental Section, Data Preparation, Classical Tools, Standard Results, and Interpretation of Results.	4			13			17	15	Assembly of representative genomes with/without a reference sequence. Quality control and visualization of assembled data. Assessment of the assembled genome, statistical analysis. Utilization of long reads in genome assembly.
Epigenetic/ or Methylation-Level NGS Data Analysis: Methodology of the Experimental Section, Data Preparation, Classical Tools, Standard Results, and Interpretation of Results	4			13			17	12	Analysis of epigenetic-level data, application of specific programs, work with the GALAXY system
Final project preparation and presentation				7			7	16	Independent preparation and presentation of a scientific research project.
Exam								6	-
In total:	16			48			64	69	

Assessment strategy	Weight,%	Timeline	Assessment criteria
Class/home-work	30	During the semester	Scientific Research-type assignment (3 planned assignments). During the assignment, students will need to examine and analyze real data and present their obtained results. The assignment will be assessed based on the portion of tasks students completed and the accuracy of their task completion. Failing to complete at least one practical assignment (a passed assignment is considered when at least 40% of the maximum grade is achieved) will prohibit the student from participating in the main exam session. Late submissions of assignment will result in a grade reduction of 5% of the maximum score for each delayed day (delayed days are rounded up, meaning even a delay of 1 second incurs a 5% reduction). If a student fails to pass more than one assignment during the semester, the student will need to retake the course. If a student fails to pass one assignment during the semester, they will have the opportunity to complete the failed assignments till the reexamination session.
Weekly tests	10	During the semester, weekly	Small weekly assignments from the covered material. Most assignments will have optional or fill-in-the-blank answers. There will be fewer weekly quizzes than there are weeks in the semester, and the quizzes will be available on the E-Learning system. Assessment will be based on the number of correct answers provided by the student. The maximum score can be achieved by correctly solving N-1 (where N is the number of quizzes per semester) quizzes without errors. Weekly quizzes must be completed strictly within the specified timeframe, and delays are not allowed (unless for reasons complying with VU regulations).

Project	20	At the end of the semester	The Independent (performed in groups) Scientific Research Project will involve the analysis of a specific type of NGS data based on published data (supported by scientific articles). Students, working in groups, will independently comprehend the issues addressed in the article, the experimental research section, replicate part of the author's analysis, and attempt to enhance/optimize the analysis. Assessment will be based on the student's executed work (whether all requirements and formatting are met), the quality of the presentation, and the answers to questions. The project is prepared/presented in English. Without completing this project, the student will not be allowed to participate in the main exam session.
Exam	40	During the exam session	During the exam, the assessment will be based on how many of the provided tasks the student solved correctly and how many questions they answered correctly, taking into account the maximum possible score. The exam questions/tasks will cover the entire course. Students who have accumulated less than 30% of the total course points (i.e., have less than 3.0 points in a 10-point system) by the exam session will not be allowed to take the exam.. Students who do not meet the minimum course requirements by the beginning of the re-examination session will not be able to participate in the reexamination session and will need to retake the course

Author	**Year of publication**	**Title**	**Issue of a periodical or volume of a publication**	**Publishing place and house or web link**
Compulsory reading				
Edited by: Urszula Demkow and Rafal Ploski	2016	Clinical Applications for Next-Generation Sequencing	-	Academic Press
Wing-Kin Sung	2017	Algorithms for Next-Generation Sequencing	-	CRC Press
Eija Korpelainen, Jarno Tuimala, Panu Somervuo, Mikael Huss, Garry Wong	2015	RNA-seq Data Analysis: A Practical Approach	-	CRC Press
Optional reading				
-	-	Online lectures	-	http://www.biocenter.helsinki.fi/bi/loytynoja/EBI_workshop/Practicals/RNASeqPractical/doc/RNASeqPractical.pdf and etc.
-	-	Scientific articles	-	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4728800/ and etc.