



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 (6 hours of lectures, 18 hours of exercises) Other: Dr. Maria Fernanda Torres Jimenez (4 hours of lectures, 12 hours of exercises) Dr. Mingle Gabrielaitė (6 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	English/Lithuanian

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
<p>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.</p> <p>Specific competencies that students will acquire:</p> <ul style="list-style-type: none"> • 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 • Omics data integration <p>General skills:</p> <ul style="list-style-type: none"> • 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. Transcriptomics workflow.	4			12			16	12	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).
Phylogenomics	4			12			16	12	Example datasets analysis

Multi-omics data integration and analysis	4			12			16	22	Transcriptomics and epigenomics, transcriptomics and other omics data integration from published data.
Final project preparation and presentation	4			12			16	23	Integration of selected real life data and published data re-analysis (data pre-processing, biological analysis, data integration and results analysis).
In total:	16			48			64	69	

Assessment strategy	Weight, %	Timeline	Assessment criteria
Phylogenomics	10	During the semester	Scientific Research-type assignment. 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. Students must submit all homework assignments (empty forms do not count as submitted). Students who fail to submit all practical assignments will have to retake the course next year.
Project	70	Performed during the semester, reported 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 (to complete a project, students must get at least 50% of the maximum grade), the student will not be allowed to participate in the main exam session.
Exam	20	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 must get a passing grade (50% of the maximum grade) to pass the course. Students who fail to pass project assignment will not be able to participate in the reexamination session and will need to retake the course. During the re-examination session only exam may be retaken.

Author	Year of publication	Title	Issue of a periodical	Publishing place and house or web link
--------	---------------------	-------	-----------------------	--

			or volume of a publication	
Compulsory reading				
Editor Kang Ning	2023	Methodologies of Mutli-Omics Data Integration and Data Mining	-	Springer, https://link.springer.com/book/10.1007/978-981-19-8210-1
Basant K. Tiwary	2022	Bioinformatics and Computational Biology	-	Springer, https://link.springer.com/book/10.1007/978-981-16-4241-8
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.