Work program of the course (training module)

course:	Transcriptomics, Epigenomics and OMICS Data Analysis/Транскриптомика, эпигеномика и анализ ОМИКСных данных
major:	Applied Mathematics and Physics
specialization:	Applied Bioinformatics/Прикладная биоинформатика
	Phystech School of Biological and Medical Physics
	Chair of Bioinformatics and Systems Biology
term:	1
qualification:	Master

Semester, form of interim assessment: 2 (spring) - Grading test

Academic hours: 30 AH in total, including:
      lectures: 30 AH.
      seminars: 0 AH.
      laboratory practical: 0 AH.

Independent work: 60 AH.

In total: 90 AH, credits in total: 2

Authors of the program:

    I.V. Antonov, phd (candidate of biological sciences)
    Y.A. Medvedeva, candidate of biological sciences

The program was discussed at the Chair of Bioinformatics and Systems Biology 04.06.2020

<div align="center">**Annotation**</div>

The purpose of this discipline is give an idea of the mechanisms of the implementation of genetic information, transcription regulation and practical skills in applying bioinformatics methods for analyzing and interpreting biological data. After completing the course, the student will understand physical and chemical properties of nucleic acids and chromatin proteins, processes involving nucleic acids and chromatin proteins, research methods of the genome and evolution of living organisms, the main mechanisms of regulatory genomics, transcriptome and epigenetics, basic bioinformatics algorithms, databases, file formats used in transcriptome and epigenomics, modern molecular biological methods for studying cell processes.

1. Study objective

Purpose of the course

Give an idea of the mechanisms of the implementation of genetic information, transcription regulation and practical skills in applying bioinformatics methods for analyzing and interpreting biological data.

Tasks of the course

• give an idea of the basic mechanisms of gene transcription regulation;
• make students familiar with modern understanding of regulatory transcriptome and epigenomics;
• teach how to use the main databases in the field;
• introduce basic algorithms and data formats for epigenomics and transcriptomics.

2. List of the planned results of the course (training module), correlated with the planned results of the mastering the educational program

Mastering the discipline is aimed at the formation of the following competencies:

Code and the name of the competence	Competency indicators
UC-1 Use a systematic approach to critically analyze a problem,and develop an action plan	UC-1.1 Systematically analyze the problem situation, identify its components and the relations between them
	UC-1.2 Search for solutions by using available sources
	UC-1.3 Develop a step-by-step strategy for achieving a goal, foresee the result of each step, evaluate the overall impact on the planned activity and its participants
Gen.Pro.C-1 Gain fundamental scientific knowledge in the field of physical and mathematical sciences	Gen.Pro.C-1.1 Apply fundamental scientific knowledge in the field of physical and mathematical sciences
	Gen.Pro.C-1.2 Consolidate and critically assess professional experience and research findings
	Gen.Pro.C-1.3 Understand interdisciplinary relations in applied mathematics and computer science and apply them in professional settings
Gen.Pro.C-3 Select and/or develop approaches to professional problem-solving with consideration to the limitations and specifics of different solution methods	Gen.Pro.C-3.1 Analyze problems, plan research strategy to achieve solution(s), propose, and combine solution approaches
	Gen.Pro.C-3.2 Employ research methods to solve new problems and apply knowledge from various fields of science (technology)
	Gen.Pro.C-3.3 Gain knowledge of analytical and computational methods of problem-solving, understand the limitations of the implementation of the obtained solutions in practice

3. List of the planned results of the course (training module)
As a result of studying the course the student should:
know:

- physical and chemical properties of nucleic acids and chromatin proteins;
- processes involving nucleic acids and chromatin proteins;
- research methods of the genome and evolution of living organisms;
- the main mechanisms of regulatory genomics, transcriptome and epigenetics;
- basic bioinformatics algorithms, databases, file formats used in transcriptome and epigenomics;
- modern molecular biological methods for studying cell processes;

be able to:
- use the main databases of regulatory genomics;
- reproduce the algorithms for analyzing transcriptome and epigenomics data

master:
- of working with large volumes of biological data;
- of planning and implementation of multi-stage bioinformatic analysis.

4. Content of the course (training module), structured by topics (sections), indicating the number of allocated academic hours and types of training sessions

4.1. The sections of the course (training module) and the complexity of the types of training sessions

№	Topic (section) of the course	Types of training sessions, including independent work			
		Lectures	Seminars	Laboratory practical	Independent work
1	Introduction to regulatory genomics	1			4
2	Basic consepts of transcription and epigenetic regulation	1			4
3	Transcriptomics: basic steps of the NGS protocols (RNA-seq)	2			8
4	Algorithms for analyzing RNA sequencing data	2			4
5	Epigenetics: basic molecular mechanisms of epigenetic regulation	2			4
6	Important examples of biological prosesses based on epigenetic regulation	6			8
7	Epigenetics: detection of DNA methylaiton	4			4
8	Epigenetics: detection of histon modifications	4			8
9	Epigenetics: determination of the chromatic structure via DNA:DNA interactions and DNA:RNA interactions. LncRNA in epiganetic regulation	4			8
10	Epigenetics: single cell NGS data analysis. Multiomics data analysis.	4			8
	AH in total	30			60
	Exam preparation	0 AH.			
	Total complexity	90 AH., credits in total 2			

4.2. Content of the course (training module), structured by topics (sections)

Semester: 2 (Spring)

1. Introduction to regulatory genomics

The central dogma of molecular biology. Differences in phenotypes of cells of a multicellular organism. Stages of transcription.

2. Basic consepts of transcription and epigenetic regulation

Elements of the regulatory structure of a gene: promoters, enhancers, insulators.

3. Transcriptomics: basic steps of the NGS protocols (RNA-seq)

RNA sequencing protocol. Stages, modifications. RNA-seq and CAGE-seq. Quality control, primary data analysis

4. Algorithms for analyzing RNA sequencing data

Alignment of sequencing reads to intron-containing eukaryotic genes, HiSAT / STAR alignment algorithms, search for exon-intron boundaries, determination of gene expression levels, normalization. Differential expression (edgeR, DESeq, etc).

5. Epigenetics: basic molecular mechanisms of epigenetic regulation

Methylation of cytosisines in DNA, other modifications of DNA and RNA Modifications of histone proteins: methylation, acetylation. Combinations of histone marks. Histone code. Database epigenetic regulators.

6. Important examples of biological prosesses based on epigenetic regulation

Genomic imprinting, X chromosome inactivation, cancer and other diseases, environmental effects, transgenerational inheritance.

7. Epigenetics: detection of DNA methylaiton

Bisulfite sequencing, protocol modifications (WGBS, RRBS), algorithm for determining the methylation degree of individual Bismark cytosines, differential methylation analysis (Methpipe), search for variable methylation regions

8. Epigenetics: detection of histon modifications

Protocol for determining modifications of histone proteins (ChIP-seq), analysis of open chromatin areas (ATAC-seq), "good practices" of building an experiment, searching for peaks (MACS, etc.), visualizing the distribution of readings by metagen, genome segmentation (chromHMM).

9. Epigenetics: determination of the chromatic structure via DNA:DNA interactions and DNA:RNA interactions. LncRNA in epiganetic regulation

Protocols for determining chromatin structure through DNA-DNA interaction (Hi-C), analysis of Hi-C data, determination of DNA-RNA interactions, long ncRNAs in epigenetic regulation, targeting mechanisms of epigenetic regulators using ncRNA

10. Epigenetics: single cell NGS data analysis. Multiomics data analysis.

ATAC-seq for single cells, NOM-seq, protocols for simultaneous analysis of RNA and chromatin structure. Computational data integration

5. Description of the material and technical facilities that are necessary for the implementation of the educational process of the course (training module)

Equipment needed for lectures and seminars: computer and multimedia eqipment (projector, sound system)

6. List of the main and additional literature, that is necessary for the course (training module) mastering

Main literature
Provided at the department:
1. Razin, Bystritskiy. Chromatin: Packed genome. Binom. Laboratory of knowledge, ISBN: 978-5-9963-1611-3

Additional literature
Provided at the department:
Рэндал Л. Шварц, Том Феникс, Брайан Д Фой. Изучаем Perl. Символ-Плюс, 2009
Р. Дурбин, Ш. Эдди, А. Крог, Г. Митчисон. Анализ биологических последовательностей. Регулярная и хаотическая динамика, 2006. Марк Бородовский, Светлана Екишева. Задачи и решения по анализу биологических последовательностей. Регулярная и хаотическая динамика, 2008

7. List of web resources that are necessary for the course (training module) mastering

Scientific bibliographic and patent databases in the field of physico-chemical biology, available on the Internet in free mode - Science Citation Index (Web of Science), Medline (PubMed), Scientific Electronic Library (NEB), Russian Patent DB of FGU FIPS and American USPAFULL patent database; email addresses of major scientific publishers who provide access to the full text of current and archival issues of these journals.

8. List of information technologies used for implementation of the educational process, including a list of software and information reference systems (if necessary)

Internet access, Unix server. For some of the lessons, you need Zoom. Google Drive to access course materials. The presence of smartphones / laptops during classes is encouraged to participate in interactive exercises.

9. Guidelines for students to master the course

A student who studies discipline must, on the one hand, master a general conceptual apparatus, and on the other hand, must learn to apply theoretical knowledge in practice.
As a result of studying the discipline, the student should know the basic definitions of the discipline, be able to apply this knowledge to solve various problems.

Successful learning requires:
- visits to all classes provided by the curriculum for the discipline;
- conducting the abstract of occupations;
- intense independent work of the student.

Independent work includes:
- reading recommended literature;
- study of educational material, preparation of answers to questions intended for self-study;
- solving problems offered to students in the classroom;
- preparation for performance of tasks of the current and intermediate certification.

An indicator of possession of the material is the ability to answer questions on discipline topics without an outline.

It is important to achieve an understanding of the material being studied, and not its mechanical memorization. If it is difficult to study individual

topics, questions, you should seek advice from the teacher.

Intermediate control of students' knowledge in the form of problem solving in accordance with the subject of classes is possible

Assessment funds for course (training module)

major:	Applied Mathematics and Physics
specialization:	Applied Bioinformatics/Прикладная биоинформатика
	Phystech School of Biological and Medical Physics
	Chair of Bioinformatics and Systems Biology
term:	1
qualification:	Master

Semester, form of interim assessment: 2 (spring) - Grading test

Authors:

I.V. Antonov, phd (candidate of biological sciences)

Y.A. Medvedeva, candidate of biological sciences

1. Competencies formed during the process of studying the course

Code and the name of the competence	Competency indicators
UC-1 Use a systematic approach to critically analyze a problem,and develop an action plan	UC-1.1 Systematically analyze the problem situation, identify its components and the relations between them
	UC-1.2 Search for solutions by using available sources
	UC-1.3 Develop a step-by-step strategy for achieving a goal, foresee the result of each step, evaluate the overall impact on the planned activity and its participants
Gen.Pro.C-1 Gain fundamental scientific knowledge in the field of physical and mathematical sciences	Gen.Pro.C-1.1 Apply fundamental scientific knowledge in the field of physical and mathematical sciences
	Gen.Pro.C-1.2 Consolidate and critically assess professional experience and research findings
	Gen.Pro.C-1.3 Understand interdisciplinary relations in applied mathematics and computer science and apply them in professional settings
Gen.Pro.C-3 Select and/or develop approaches to professional problem-solving with consideration to the limitations and specifics of different solution methods	Gen.Pro.C-3.1 Analyze problems, plan research strategy to achieve solution(s), propose, and combine solution approaches
	Gen.Pro.C-3.2 Employ research methods to solve new problems and apply knowledge from various fields of science (technology)
	Gen.Pro.C-3.3 Gain knowledge of analytical and computational methods of problem-solving, understand the limitations of the implementation of the obtained solutions in practice

2. Competency assessment indicators

As a result of studying the course the student should:

know:
- physical and chemical properties of nucleic acids and chromatin proteins;
- processes involving nucleic acids and chromatin proteins;
- research methods of the genome and evolution of living organisms;
- the main mechanisms of regulatory genomics, transcriptome and epigenetics;
- basic bioinformatics algorithms, databases, file formats used in transcriptome and epigenomics;
- modern molecular biological methods for studying cell processes;

be able to:
- use the main databases of regulatory genomics;
- reproduce the algorithms for analyzing transcriptome and epigenomics data

master:
- of working with large volumes of biological data;
- of planning and implementation of multi-stage bioinformatic analysis.

3. List of typical control tasks used to evaluate knowledge and skills

During the current control, the student should be able to answer the following questions:
1. Central dogma of molecular biology
2 Structure of eukaryotic genes (promoter, enhancer, insulator)
3. Mechanisms that provide phenotypic differences of genetically identical cells of a multicellular organism
4. Stages of transcription and processing of RNA, levels of regulation of RNA expression
5. The main mechanisms of regulation of transcription
6. Main epigenetic mechanisms
7. DNA methylation and its regulatory role

8. DNA packaging in the cell, the structure of the nucleosome

9. Modifications of histone proteins and their regulatory role, histone code

10. Transcription initiation factors

11. Basic principles of sequencing technology

12. Types of sequencing technologies, biological mechanisms that can be studied on their basis

13. Technology RNA-seq, issues that can be solved with its help

14. The difficulties of implementing RNA-seq

15. Stages of RNA-seq data analysis (bowtie, DESeq programs, RPKM / FPKM measure, DESeq / edgeR normalization, splice-junction search and isoform search, differentially expressed genes search)

16. Bisulfite sequencing

17. Analysis of bisulfite sequencing data (Bismark algorithm, differential methylated regions search by MethPipe)

18. Chromatin immunoprecipitation method, ChIP-seq

19. ChIP-seq data analysis (peak search (MACS2), functional analysis, motive search, visualization using sequence-LOGO)

20. Analysis of ChIP-seq data for histone modifications (MACS2 broadpeak, NGS-plot program, chromHMM segmentation),

21. Search for regulatory areas (DNase-seq, ATAC-seq), specific readings and data processing

22. "Good practices" of the design of the experiment based on sequencing (replica, Input, spike-in) and analysis of the obtained data

23. Methods of analysis of the epigenetic profile at the level of unit cells (scATAC-seq, scBS-seq)

24. Multi-mix protocols

During the class, interactive discussions can take place in the course chats, which will be homework. It is possible to perform patent search as an independent task. Successful completion of all tasks in the course and the completion of control slices of knowledge gives an advantage in the differential credit.

4. Evaluation criteria

1. Central dogma of molecular biology

2 Structure of eukaryotic genes (promoter, enhancer, insulator)

3. Mechanisms that provide phenotypic differences of genetically identical cells of a multicellular organism

4. Stages of transcription and processing of RNA, levels of regulation of RNA expression

5. The main mechanisms of regulation of transcription

6. Main epigenetic mechanisms

7. DNA methylation and its regulatory role

8. DNA packaging in the cell, the structure of the nucleosome

9. Modifications of histone proteins and their regulatory role, histone code

10. Transcription initiation factors

11. Basic principles of sequencing technology

12. Types of sequencing technologies, biological mechanisms that can be studied on their basis

13. Technology RNA-seq, issues that can be solved with its help

14. The difficulties of implementing RNA-seq

15. Stages of RNA-seq data analysis (bowtie, DESeq programs, RPKM / FPKM measure, DESeq / edgeR normalization, splice-junction search and isoform search, differentially expressed genes search)

16. Bisulfite sequencing

17. Analysis of bisulfite sequencing data (Bismark algorithm, differential methylated regions search by MethPipe)

18. Chromatin immunoprecipitation method, ChIP-seq

19. ChIP-seq data analysis (peak search (MACS2), functional analysis, motive search, visualization using sequence-LOGO)

20. Analysis of ChIP-seq data for histone modifications (MACS2 broadpeak, NGS-plot program, chromHMM segmentation),

21. Search for regulatory areas (DNase-seq, ATAC-seq), specific readings and data processing

22. "Good practices" of the design of the experiment based on sequencing (replica, Input, spike-in) and analysis of the obtained data
23. Methods of analysis of the epigenetic profile at the level of unit cells (scATAC-seq, scBS-seq)
24. Multi-mix protocols

Can be left unchanged

The mark is excellent (10 points) - it is given to a student who has shown comprehensive, systematic, deep knowledge of the curriculum of the discipline, who has an interest in this subject area, has demonstrated the ability to confidently and creatively put them into practice in solving specific problems, and a free and proper substantiation of decisions.

The mark is excellent (9 points) - it is given to a student who has shown comprehensive, systematic, in-depth knowledge of the curriculum of the discipline and the ability to confidently put them into practice in solving specific problems, free and proper substantiation of the decisions made.

The mark is excellent (8 points) - given to a student who has shown comprehensive, systematic, in-depth knowledge of the curriculum of the discipline and the ability to confidently apply them in practice in solving specific problems, correct justification of decisions made, with some shortcomings.

A mark is good (7 points) - it is put up for a student, if he knows the material firmly, sets it up competently and in essence, knows how to apply the knowledge gained in practice, but does not competently substantiate the results obtained.

Evaluation is good (6 points) - it is put up to a student, if he knows the material firmly, sets it up correctly and in essence, knows how to apply this knowledge in practice, but admits some inaccuracies in the answer or in solving problems.

A mark is good (5 points) - it is given to a student, if he basically knows the material, correctly and essentially sets it out, knows how to apply this knowledge in practice, but allows a sufficiently large number of inaccuracies to answer or solve problems.

Grade satisfactorily (4 points) is given to a student who has shown the fragmented, fragmented nature of knowledge, insufficiently correct formulations of basic concepts, violations of the logical sequence in the presentation of program material, but at the same time he has mastered the main sections of the curriculum necessary for further education and can apply knowledge is modeled in a standard situation.

Grade satisfactorily (3 points) - given to a student who showed the fragmented, scattered nature of knowledge, making mistakes in formulating basic concepts, disrupting the logical sequence in presenting program material, poorly masters the main sections of the curriculum required for further education and even applies the knowledge gained in a standard situation.

The rating is unsatisfactory (2 points) - is given to a student who does not know most of the main content of the curriculum of the discipline, makes gross mistakes in the wording of the basic principles and does not know how to use this knowledge when solving typical tasks.

Unsatisfactory mark (1 point) - is given to a student who does not know the main content of the discipline's curriculum, makes gross errors in the wording of the basic concepts of the discipline and does not have any skills to solve typical practical problems.

5. Methodological materials defining the procedures for the assessment of knowledge, skills, abilities and/or experience

When conducting an oral differential test, the student is given 60 minutes to prepare. Interrogation of a student on a ticket on an oral differential test should not exceed one astronomical hour.