

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ

Проректор по учебной работе
и образовательным инновациям

О. Н. Здрок

2020 г.

Регистрационный № УД-8274/уч.



Deep Analysis of Transcriptomics Data

**Учебная программа учреждения высшего образования
по учебной дисциплине для специальности:**

1-31 80 01 Biology
профилизация Molecular Cell Biology

2020 г.

Учебная программа составлена на основе ОСВО 1-31 80 01-2019 и учебного плана УВО № G 31а-092/уч., утвержденного 11.04.2019 г.

СОСТАВИТЕЛЬ:

В.В. Гринев, доцент кафедры генетики Белорусского государственного университета, кандидат биологических наук

РЕЦЕНЗЕНТЫ:

Л.Н. Валентович, заведующий лабораторией «Центр аналитических и генно-инженерных исследований» ГНУ «Институт микробиологии Национальной академии наук Беларуси», кандидат биологических наук;

Е.А. Николайчик, доцент кафедры молекулярной биологии Белорусского государственного университета, кандидат биологических наук, доцент

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой генетики
(протокол № 20 от 2 июня 2020 г.)

Научно-методическим Советом БГУ
(протокол № 5 от 17 июня 2020 г.)

Зав. кафедрой генетики,
профессор



Н. П. Максимова

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цель и задачи учебной дисциплины

Целью учебной дисциплины является формирование у обучающихся навыков обработки и анализа транскриптомных данных.

В рамках поставленной цели **задачами учебной дисциплины** является изучение ключевых форматов данных, применяемых в транскриптомике, базовых принципов картирования прочтений, методов анализа экспрессии генов, а также основных методов функционального анализа транскриптомных данных.

Место учебной дисциплины в системе подготовки магистра

Учебная дисциплина относится к компоненту учреждения образования учебного плана и входит в учебный модуль «Analysis of Genomic and Transcriptomic Data».

Связи с другими учебными дисциплинами, включая учебные дисциплины компонента учреждения высшего образования, дисциплины специализации и др.

Учебная программа составлена с учетом межпредметных связей с учебными дисциплинами «Structural and Functional Organization of Genomes», «Fundamentals of Bioinformatics», «Introduction to R Programming» и «Molecular Diagnostics».

Требования к компетенциям:

Освоение учебной дисциплины «Deep Analysis of Transcriptomics Data» совместно с учебной дисциплиной «Molecular Taxonomy» должно обеспечить формирование специализированной компетенции СК-5 «Быть способным применять знания алгоритмов и подходов, используемых в анализе геномных и транскриптомных данных, для решения молекулярно-генетических задач в фундаментальных и прикладных исследованиях, владеть методами молекулярной систематики».

В результате освоения учебной дисциплины обучающийся должен:

знать:

- базовые форматы данных, используемые в транскриптомике;
- ключевые подходы и программные решения в картировании прочтений;
- основные подходы и программные решения в анализе экспрессии генов;
- методы функционального анализа транскриптомных данных

уметь:

- делать предварительную обработку транскриптомных данных;
- картировать прочтения, получаемые с помощью ОМИКС-методов транскриптомики;
- проводить анализ дифференциальной экспрессии генов;

– делать функциональную аннотацию транскриптомных данных;
– творчески и эффективно использовать полученные знания в профессиональной деятельности

владеть:

– базовым набором методов анализа транскриптомных данных;
– навыками применения языка программирования R для решения задач транскриптомики.

Структура учебной дисциплины

Дисциплина изучается в 3 семестре. Всего на изучение учебной дисциплины «Deep Analysis of Transcriptomics Data» отведено:

– для очной формы получения высшего образования – 198 часов, в том числе 54 аудиторных часа, из них: лекции – 14 часов, практические занятия – 10 часа, управляемой самостоятельной работы – 30 часов, в т.ч. контроль управляемой самостоятельной работы (ДО) – 8 часов.

Трудоемкость учебной дисциплины составляет 6 зачетных единиц.

Форма текущей аттестации – экзамен.

CONTENT OF COURSE SYLLABUS

Section 1. Introduction

Topic 1.1. Transcriptomics as an integral part of modern molecular biology: from objectives to fundamental and practical applications. Objects of transcriptomics. Overview of experimental technologies in transcriptomics. The relationship between object oriented and technology oriented approaches in transcriptomics. Main data formats in transcriptomics.

Section 2. Pre-processing and alignment of short reads

generated by transcriptome-oriented high-throughput sequencing

Topic 2.1. Raw data formats. FASTQ file format. Base and reads quality control. Alignment of reads. Alignment algorithms: merge sorting, spaced-seed indexing with hash table, indexing with suffix array/Burrows-Wheeler transformation. Local versus global/spliced alignment. Overview of alignment software. Quality control of aligned reads. Storing aligned reads.

Section 3. Analysis of gene expression

Topic 3.1. Brief introduction into analysis of gene expression with RNA-Seq data. Principal workflow of gene-level analysis. Read summarization and count matrix. Filtering of data. Estimation of normalization, or scaling, factors. Overview of current solutions. Variability of RNA-Seq data. Stabilization of data variability. Heteroscedasticity of data and elimination of the mean-variance relationships. Assessment and correction of batch effects.

Topic 3.2. Differential gene expression. Design matrix and matrix of contrasts. Linear modeling: simple linear model versus generalized linear model. Fitting of the linear models to experimental data. Calculation of the gene-wise differences between conditions. Calculation of the concomitant statistics. Adjustment of p-values. Parsing, consolidation, graphical representation and saving the final results.

Section 4. Analysis of RNA splicing

Topic 4.1. Assembling of experimental transcriptomes. Formalization of the problem. Types of transcriptome assembly: de novo assembly versus assembly supported by annotation. De novo assembly: main algorithms, software overview, advantages and limitations. Assembly supported by annotation: sources of annotation, main algorithms, software overview, advantages and limitations. Assessment of assembly quality. Annotation of assembled transcriptomes. Storing of assembled transcriptome.

Topic 4.2. RNA splicing and differential RNA splicing analysis. Introduction into RNA splicing phenomena. Alternative and differential splicing. Splicing analysis at exon, exon-exon junction and full-length transcript levels. Differential exon usage. Exon-exon junctions: definition, identification, classification, differential usage during splicing. Splicing analysis at full-length

transcript level: main approaches, advantages and disadvantages. Comprehensive annotation of splicing events.

Section 5. Functional annotation of transcriptomics data

Topic 5.1. Identification and annotation of the coding and non-coding RNA transcripts. Gene co-expression networks and gene regulatory networks: definition, types, algorithms of inferring and quality assessment. Analysis of gene ontologies in transcriptomics. Enrichment analysis versus gene set enrichment analysis. Pathway analysis in transcriptomics.

EDUCATIONAL AND METHODOLOGICAL MAP OF COURSE
Full-time higher education scheme using distance learning technologies

Section or topic #	Name of section or topic	Number of classroom hours					Control of managed independent work	Knowledge control approaches
		Lectures	Practical classes	Seminar classes	Laboratory classes	Managed independent work		
1.	Introduction							
1.1	Introduction to transcriptomics	2					2 (ДО)	Test by LMS Moodle
2.	Pre-processing and alignment of short reads							
2.1	Pre-processing of short reads. Alignment of short reads.	2	2				4 2 (ДО)	Practical activity report. Test by LMS Moodle
3.	Analysis of gene expression							
3.1	Analysis of gene expression	2					2 1 (ДО)	Test by LMS Moodle
3.2	Analysis of differential gene expression	2	2				2 1 (ДО)	Practical activity report. Test by LMS Moodle
4.	Analysis of RNA splicing							
4.1	Assembling of experimental transcriptomes	2	2				2 1 (ДО)	Practical activity report. Test by LMS Moodle
4.2	RNA splicing and differential RNA splicing analysis	2	2				4 1 (ДО)	Practical activity report. Test by LMS Moodle
5.	Functional annotation of transcriptomics data							
5.1	Functional annotation of transcriptomics data	2	2				6	Practical activity report
	Total	14	10				30 (8 ДО)	

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Перечень основной литературы

- 1) *Колчанов Н. А.* Введение в информационную биологию и биоинформатику. Учебное пособие для студентов вузов. В 5-и томах. Том 3 / Н. А. Колчанов, О. В. Вишневский, Д. П. Фурман. – Новосибирск: РИЦ НГУ, 2015. – 298 с.
- 2) *Мастуцкий С. Э.* Статистический анализ и визуализация данных с помощью R / С. Э. Мастуцкий, В. К. Шитиков. Хайдельберг – Лондон – Тольятти, 2014. – 401 с.
- 3) *Mardis E. R.* DNA sequencing technologies: 2006-2016 / E. R. Mardis. – Nature Protocols. – 2017. – Vol. 12. – P. 213-218.
- 4) *Reinert K.* Alignment of next-generation sequencing reads / K. Reinert, B. Langmead, D. Weese, D. J. Evers. – Annual Review of Genomics and Human Genetics. – 2015. – Vol. 16. – P. 133-151.

Перечень дополнительной литературы

- 1) *Carbon S.* The Gene Ontology resource: 20 years and still GOing strong / The Gene Ontology Consortium. – Nucleic Acids Research. – 2019. – Vol. 47. – P. D330-D338.
- 2) *Cellerino A.* Transcriptome analysis / A. Cellerino, M. Sanguanini. – Scuola Normale Superiore, 2018. – 185 p.
- 3) *Delgado F. M.* Computational methods for gene regulatory networks reconstruction and analysis: A review / F. M. Delgado, F. Gómez-Vela. – Artificial Intelligence in Medicine. – 2019. – Vol. 95. – P. 133-145.
- 4) *Hölzer M.* De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers / M. Hölzer M, M. Marz. – Gigascience. – 2019. – Vol. 8. – P. giz039.
- 5) *Song Q. A.* Computational analysis of alternative splicing in plant genomes / Q. A. Song, N. S. Catlin, W. Brad Barbazuk, S. Li. – Gene. – 2019. – Vol. 685. – P. 186-195.
- 6) *Wang Y.* Transcriptome data analysis. Methods and protocols / Y. Wang, M.-an Sun. – Humana Press, 2018. – 238 p.

Интернет-ресурсы

- 1) Гринев В. В. Авторские видеолекции на канале Transcriptomics (www.youtube.com/channel/UCawO0gkL7GDS7cSsRJpQqDg)
- 2) Гринев В. В. Анализ транскриптома (s-transcriptomics.blogspot.com)
- 3) Мастуцкий С. Э. R: Анализ и визуализация данных (r-analytics.blogspot.com)
- 4) BioStar (www.biostars.org)
- 5) RNA-Seq Blog (www.rna-seqblog.com)
- 6) SEQanswers (seqanswers.com)

Перечень рекомендуемых средств диагностики и методика формирования итоговой отметки

Для оценки профессиональных компетенций обучающихся используется следующий диагностический инструментарий:

- отчет по заданиям практического занятия;
- тестирование.

При формировании итоговой оценки используется рейтинговая оценка знаний студента, дающая возможность проследить и оценить динамику процесса достижения целей обучения. Формирование отметки за текущую успеваемость:

- отчеты по заданиям практических занятий – 20 %;
- тестирование – 80 %.

В качестве формы текущей аттестации по учебной дисциплине используется экзамен.

Рейтинговая оценка по дисциплине рассчитывается на основе оценки текущей успеваемости и экзаменационной оценки с учетом их весовых коэффициентов. Оценка по текущей успеваемости составляет 40 %, экзаменационная оценка – 60 %. Рейтинговая оценка выставляется только в случае успешной сдачи экзамена (при получении 4-х баллов и выше).

Примерный перечень заданий для управляемой самостоятельной работы студентов

Раздел 1. Введение в дисциплину

- 1) ОМИКС-методы в транскриptomике.
- 2) Использование методов молекулярной биологии в транскриptomике.
- 3) Основные форматы данных, используемые в транскриptomике.

Форма контроля – тестирование в системе LMS Moodle.

Раздел 2. Базовые принципы картирования прочтений, получаемых в транскриptomных исследованиях

- 4) Оценка качества высокопроизводительного секвенирования.
- 5) Алгоритмы картирования прочтений.
- 6) Контроль качества картирования прочтений.

Форма контроля – тестирование в системе LMS Moodle.

Раздел 3. Анализ экспрессии генов в транскриptomике

- 7) Суммаризация прочтений для целей анализа экспрессии прочтений.
- 8) Различные методы нормализации и трансформации суммаризованных прочтений.
- 9) Решение проблем высокой вариабельности и гетероскедастичности транскриptomных данных.

- 10) Методы сборки транскриptomов.
 - 11) Методы количественной оценки экспрессии транскриptomов.
 - 12) Линейное моделирование в оценке экспрессии генов.
- Форма контроля – тестирование в системе LMS Moodle.

Раздел 4. Методы функционального анализа в транскриптомике.

- 13) Методы оценки кодирующего потенциала РНК.
 - 14) Подходы, используемые для оценки альтернативного и дифференциального сплайсинга РНК.
 - 15) Методы реконструкции сетей генных регуляций.
 - 16) Методы реконструкции молекулярных путей.
 - 17) Генные онтологии и их использование в транскриптомике.
- Форма контроля – тестирование в системе LMS Moodle.

Примерная тематика практических занятий

- 1) Анализ дифференциальной экспрессии генов (2 часа).
- 2) Функциональная аннотация транскриптомных данных (2 часа).

Описание инновационных подходов и методов к преподаванию учебной дисциплины

При организации образовательного процесса используется *эвристический подход*, который предполагает:

- осуществление обучающимися личностно-значимых открытий окружающего мира;
- демонстрацию многообразия решений большинства профессиональных задач и жизненных проблем;
- творческую самореализацию обучающихся в процессе создания образовательных продуктов;
- индивидуализацию обучения через возможность самостоятельно ставить цели, осуществлять рефлексию собственной образовательной деятельности.

Кроме этого, в образовательном процессе используется также *метод учебной дискуссии*, который предполагает участие обучающихся в целенаправленном обмене мнениями, идеями для предъявления и/или согласования существующих позиций по определенной проблеме. Использование этого метода обеспечивает появление нового уровня понимания изучаемой темы, применение знаний при решении проблем, определение способов их решения. Наконец, образовательный процесс *включает методы и приемы развития критического мышления*, которые представляют собой систему, формирующую навыки работы с информацией в процессе чтения и письма; понимания информации как отправного, а не конечного пункта критического мышления.

Методические рекомендации по организации самостоятельной работы обучающихся

Для организации самостоятельной работы обучающихся разработан примерный перечень тем и заданий, составлен список рекомендуемых

источников информации (включая Интернет-ресурсы), а также организована работа на образовательном портале биологического факультета БГУ.

List of questions for the exam

- 1) Transcriptomics as an integral part of modern molecular biology. Objects of transcriptomics.
- 2) Overview of experimental technologies in transcriptomics.
- 3) Main data formats in transcriptomics.
- 4) Alignment of short reads. Alignment algorithms.
- 5) Overview of alignment software. Quality control of aligned short reads.
- 6) Storing aligned short reads.
- 7) Analysis of gene expression with RNA-Seq.
- 8) RNA-Seq short reads summarization. Filtering of data.
- 9) Estimation of normalization factors for RNA-Seq libraries. Overview of current solutions.
- 10) Variability and heteroscedasticity of RNA-Seq data.
- 11) Linear modeling in transcriptomics.
- 12) Differential gene expression.
- 13) Assembling of experimental transcriptomes.
- 14) Annotation of assembled transcriptomes.
- 15) Analysis of RNA splicing analysis at exon, exon-exon junction and full-length transcript levels.
- 16) Differential RNA splicing analysis.
- 17) Comprehensive annotation of splicing events.
- 18) Identification and annotation of the coding and non-coding RNA transcripts.
- 19) Gene co-expression networks and gene regulatory networks.
- 20) Analysis of gene ontologies in transcriptomics.
- 21) Enrichment analysis and gene set enrichment analysis.
- 22) Pathway analysis in transcriptomics.

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УВО

Название учебной дисциплины, с которой требуется согласование	Название кафедры	Предложения об изменениях в содержании учебной программы учреждения высшего образования по учебной дисциплине	Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола)
Fundamentals of Bioinformatics	Молекулярной биологии	Отсутствуют	Утвердить согласование, протокол № 20 от 2 июня 2020 г.
Introduction to R Programming	Генетики	Отсутствуют	Утвердить согласование, протокол № 20 от 2 июня 2020 г.
Molecular Diagnostics	Молекулярной биологии	Отсутствуют	Утвердить согласование, протокол № 20 от 2 июня 2020 г.
Structural and Functional Organization of Genomes	Молекулярной биологии	Отсутствуют	Утвердить согласование, протокол № 20 от 2 июня 2020 г.

ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ УВО

на ____/____ учебный год

№№ ПП	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры
_____ (название кафедры) (протокол № ____ от _____ 20 г.)

Заведующий кафедрой

_____ (ученая степень, ученое звание)

_____ (подпись)

_____ (И. О. Фамилия)

УТВЕРЖДАЮ
Декан факультета

_____ (ученая степень, ученое звание)

_____ (подпись)

_____ (И. О. Фамилия)