

# БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Биологический факультет

Кафедра молекулярной биологии

СОГЛАСОВАНО

Председатель учебно-методической  
комиссии биологического факультета  
Поликсенова В.Д.



« 10 » ноября 2015 г.

СОГЛАСОВАНО

Декан  
биологического факультета  
Лысак В.В.



« 10 » ноября 2015 г.

Регистрационный номер № УД- 416

## УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС ПО УЧЕБНОЙ ДИСЦИПЛИНЕ

### Геномика

для специальностей

1-31 01 01 Биология (по направлениям);

1-31 01 02 Биохимия;

1-31 01 03 Микробиология

Составитель: канд. биол. наук, доцент Е.А. Николайчик

Рассмотрено и утверждено

на заседании

Научно-методического совета БГУ

« 11 » ноября 2015 г.

протокол № 2

**РЕЦЕНЗЕНТЫ:**

кафедра биотехнологии и биоэкологии Белорусского государственного технологического университета;

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

## СОДЕРЖАНИЕ

<b>ПОЯСНИТЕЛЬНАЯ ЗАПИСКА</b>	4
<b>1. ТЕОРЕТИЧЕСКИЙ РАЗДЕЛ</b>	5
<b>2. ПРАКТИЧЕСКИЙ РАЗДЕЛ</b>	20
<b>3. КОНТРОЛЬ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ</b>	20
Вопросы для самоконтроля	20
Вопросы для подготовки к зачету	22
<b>4. ВСПОМОГАТЕЛЬНЫЙ РАЗДЕЛ</b>	24
Учебно-программные материалы	24
Список рекомендуемой литературы и Интернет-ресурсов	24

## ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Учебно-методический комплекс (УМК) по учебной дисциплине «Геномика» создан в соответствии с требованиями Положения об учебно-методическом комплексе на уровне высшего образования и предназначен для студентов специальностей 1-31 01 01 Биология (по направлениям), 1-31 01 02 Биохимия, 1-31 01 03 Микробиология. Содержание разделов УМК соответствует образовательным стандартам высшего образования по данным специальностям. Главная цель УМК – оказание методической помощи студентам в систематизации учебного материала в процессе подготовки к итоговой аттестации по курсу «Геномика».

Структура УМК включает:

1. Учебно-методическое обеспечение дисциплины.

1.1. Теоретический раздел (учебное издание для теоретического изучения дисциплины в объеме, установленном типовым учебным планом по специальности).

1.2. Практический раздел (материалы для проведения лабораторных занятий по дисциплине в соответствии с учебным планом).

2. Контроль самостоятельной работы студентов (материалы текущей и итоговой аттестации, позволяющие определить соответствие учебной деятельности обучающихся требованиям образовательных стандартов высшего образования и учебно-программной документации, в т.ч. вопросы для подготовки к экзамену, задания, тесты, вопросы для самоконтроля, тематика рефератов и др.).

3. Вспомогательный раздел.

3.1. Учебно-программные материалы (типовая учебная программа, учебные программы (рабочий вариант) для студентов дневной формы получения образования).

3.2. Информационно-аналитические материалы (список рекомендуемой литературы, перечень электронных образовательных ресурсов и их адреса и др.).

Работа с УМК должна включать на первом этапе ознакомление с тематическим планом дисциплины, представленным в типовой учебной программе. С помощью рабочего варианта учебной программы по дисциплине можно получить информацию о тематике лекций и лабораторных занятий, перечнях рассматриваемых вопросов и рекомендуемой для их изучения литературы. Для подготовки к лабораторным занятиям и промежуточным зачетам необходимо, в первую очередь, использовать материалы, представленные в разделе учебно-методическое обеспечение дисциплины, а также материалы для текущего контроля самостоятельной работы. В ходе подготовки к итоговой аттестации рекомендуется ознакомиться с требованиями к компетенциям по дисциплине, изложенными в типовой учебной программе, структурой рейтинговой системы, а также перечнем вопросов к экзамену. Для написания рефератов могут быть использованы информационно-аналитические материалы, указанные в соответствующем разделе УМК.

## 1. ТЕОРЕТИЧЕСКИЙ РАЗДЕЛ

### Краткий конспект лекций по курсу «Молекулярная биология»

#### Лекция №1. История геномики

- 1866 - Законы Менделя
- 1869 - Открытие ДНК
- 1900 - Генетика
- 1910 - Хромосомная теория
- 1913 - Генетические карты
- 1944 - Роль ДНК
- 1953 - Структура ДНК
- 1961 - Генетический код
- 1965 - База данных RDB
- 1970 - Рестриктазы
- 1972 - Клонирование ДНК
- 1977 - Секвенирование
- 1982 - GenBank
- 1982 - фаг  $\lambda$
- 1983 - ПЦР
- 1986 - Автоматический секвенатор
- 1986 - Проект "Геном человека"
- 1988 - NCBI
- 1988 - Пиросеквенирование
- 1991 - EST
- 1995 - Геном бактерии
- 1996 - Геном дрожжей
- 1998 - Геном нематоды
- 1999 - Геном мухи
- 2000 - Геном растения
- 2003 - Геном человека
- 2004 - Геномный секвенатор
- 2007 - Синтез генома
- 2010 – Синтетический организм
- 2014 - "Нанопоровое" секвенирование

#### Лекции №2-3. Как устроены геномы и как они работают?

Геномы прокариот чаще всего имеют одну (реже 2) кольцевую или линейную хромосому размером **0.5-13 м.н.п.**; могут иметь **несколько** кольцевых или линейных **плазмид**. **Плазмида** отличается от хромосомы только тем, что может быть утрачена без потери клеткой жизнеспособности. Плазмиды, как правило, имеют размер от нескольких тысяч до десятков тысяч

нуклеотидных пар, но могут достигать размера более миллиона н. п. (симбиотические мегаплазмиды ризобий).

Геномы эукариот состоят из различного числа расположенных в ядре линейных хромосом размером до нескольких сот м.н.п. и множества копий небольших (6-300 т.н.п.) геномов органелл.

Геномная молекула ДНК любой клетки обязательно должна иметь следующие функциональные участки:

- **точки инициации репликации** (одну у прокариот и много у эукариот);
- **центромеру** для прикрепления веретена деления (или аналогичной структуры);
- (только для линейных хромосом) **теломеры** для завершения репликации линейных концов и их стабилизации

Основной смысловой единицей ДНК является ген – часть молекулы ДНК, определяющая при транскрипции синтез одной или нескольких (но имеющих общие участки) молекул функциональных РНК. Такими молекулами РНК могут быть:

- информационные (иРНК), кодирующие белки,
- рибосомные (рРНК),
- транспортные (тРНК),
- малые ядерные (мяРНК; несколько классов с разными функциями),
- различные регуляторные РНК.

Различия в механизмах контроля транскрипции определяют разный размер регуляторных областей генов. У прокариот на один ген приходится обычно не более 100 н.п. регуляторной ДНК (<10% генома). У высших эукариот регуляторные области занимают несколько тысяч или даже десятков тысяч н.п. – их размер сопоставим с размером гена.

Геномы прокариот компактны, их размер пропорционален числу генов. Геномы прокариот обычно содержат не более 5-10% некодирующей ДНК. Средний размер гена – около 1 т. н. п. (чуть меньше у архей). Количество генов – от 500 до 11 тыс. (обычно 3-5 тыс.). Размер генома – от 0,5 до 14 млн. н. п. (обычно 3-5 млн. н. п.).

Размеры геномов эукариот варьируют в широких пределах и не соответствуют сложности организма (и числу генов), однако минимальный размер генома внутри каждой морфологической группы возрастает у высших организмов. Размер генома конкретного организма определяется несколькими факторами:

- количеством сателлитной ДНК (40% генома риса – один минисателлитный повтор);
- числом мобильных генетических элементов (около 50% генома человека – активные и реликтовые МГЭ);
- частотой событий дупликации генома или его частей и последующей стабильностью полиплоидного генома;

- эффективностью процессов экспрессии прерывистых генов, в первую очередь сплайсинга пре-мРНК.

Только у прокариот количество генов в геноме оказывает определяющее влияние на его размер. Геном человека на 98.5% состоит из некодирующей ДНК.

Сложность организма нельзя однозначно оценивать по числу генов в его геноме. У многих организмов, особенно у высших эукариот, один ген может кодировать более одного белкового продукта за счет альтернативных путей экспрессии генов, срабатывающих преимущественно на стадии процессинга первичных транскриптов:

- альтернативного сплайсинга,
- транс-сплайсинга,
- альтернативного полиаденилирования,
- использования альтернативных промоторов и старт-кодонов.

У млекопитающих эти процессы ( в основном альтернативный сплайсинг) увеличивают число белковых продуктов по сравнению с числом генов в несколько раз.

Геном - это больше, чем полная последовательность ДНК одного организма. Геномы различных особей одного вида существенно различаются (от 0,1% различий между двумя особями у человека до 2-3% у асцидий). Многие свойства вида, важные для его выживания (например, устойчивость к патогенам у растений), реализуются на надорганизменном уровне и зависят от геномного полиморфизма и внутривидовой рекомбинации. Полное понимание биологии вида возможно только с учетом внутривидового полиморфизма генома, механизмов его поддержания и использования.

#### **Лекции №4-6. Методы секвенирования ДНК**

Разработка технологий определения последовательностей (секвенирования) ДНК началась в 70-е годы прошлого века. С тех пор сменилось три поколения таких технологий, принципиально отличающихся друг от друга как по методическим подходам, так и по своим возможностям.

В основе методик секвенирования ДНК первого поколения лежат следующие принципы:

- Создается набор коротких фрагментов ДНК, один из концов которых идентичен для всех фрагментов, а второй варьирует и соответствует всем возможным позициям одного из 4 нуклеотидов.
- Электрофорезом в полиакриламидном геле с точностью до 1 нуклеотида определяется длина полученных фрагментов (размер фрагмента соответствует позиции конкретного нуклеотида).
- Считывание последовательности производится путем сопоставления размеров фрагментов, соответствующим позициям каждого из 4 нуклеотидов.
- Оригинальные методики для детекции продуктов секвенирующих реакций использовали радиоактивные метки, радиоавтографию и “ручное”

прочтение последовательности ДНК, более поздние - флуоресцентные метки и автоматическое считывание последовательности ДНК.

Основной методикой секвенирования первого поколения является ферментативный метод дидезокситерминаторов (метод Сэнгера). Секвенирование крупных геномов по методу Сэнгера происходит в специализированных центрах с большим количеством автоматических капиллярных секвенаторов. Статистика современных автоматических капиллярных секвенаторов:

- Длина читаемой последовательности – до 1000 н.п. (в среднем около 750)
- Количество одновременных реакций: 96(384)
- Продолжительность рабочего цикла: около 2 часов
- Количество оснований, производимых одним прибором за сутки: ~1(4) Млн. н.п.

Применение метода Сэнгера в настоящий момент ограничено высокой трудоемкостью (и, следовательно, ценой). Основные причины высокой трудоемкости и стоимости методики Сэнгера:

использование электрофореза для регистрации продуктов реакции, необходимость создания библиотеки клонов для секвенирования, необходимость выделения и очистки ДНК из большого числа клонов.

Все более современные методы секвенирования не используют электрофорез, что позволяет существенно удешевить каждую реакцию, но требует значительно более сложного (и более дорогого) оборудования.

- Основные принципы методик секвенирования ДНК II поколения:
- Исключено клонирование *in vivo*
- Полный отказ от электрофореза
- Только флуоресцентные метки
- Миниатюризация и параллелизация реакций
- Секвенируемые фрагменты короче (от пары десятков до пары сотен п.о.), но их гораздо больше (сотни тысяч)
- Максимальная автоматизация
- Возможность однократно секвенировать геном человека за неделю.

Первый метод второго поколения (пиросеквенирование) был предложен в 2004 г., следующий ("секвенирование путем синтеза") – в 2008, и с начала 2010-х до сегодняшнего дня второй метод является наиболее производительным и наименее дорогим. Принцип метода: детекция флуоресценции присоединяемых ДНК-полимеразой нуклеотидов. Статистика наиболее производительной модели секвенатора II поколения на март 2014 г.:

Длина читаемой последовательности: 125 н.п.

Количество одновременных реакций: 8 млрд.

Продолжительность рабочего цикла: 6 дней

Количество оснований в сутки: до 1000 млрд. н.п. (300 геномов человека)

Стоимость секвенирования генома человека: ~3 000\$

Основные характеристики методик третьего поколения:



Сиквенс одиночных молекул ДНК  
 Не используется клонирование ни *in vivo*, ни *in vitro*  
 Максимальная миниатюризация  
 Значительно повышена скорость реакций  
 Увеличена длина читаемой последовательности  
 Существенное удешевление секвенирования  
 Разработка нескольких методов на стадии технологической демонстрации, начало продаж одной модели в 2011 г., второй – 2015 г.

Наиболее совершенной на сегодня методикой секвенирования III поколения является SMRT-технология, основанная на детекции присоединения флуоресцентных нуклеотидов к одной молекуле ДНК непосредственно по ходу ее синтеза. Возможности наиболее совершенного на середину 2015 г. прибора, использующего SMRT-технологии:

- 150 тыс. ячеек на одном чипе
- 10-15 т.н.п. (макс. >40 т.н.п.!) с одной ячейки
- До 1 млрд н.п. с чипа
- Точность однократного прочтения – 89%
- Число повторностей для получения точности 99,99% – 13, 99,999% – 20.

### **Лекция №7. Молекулярные базы данных и аннотация геномных последовательностей**

В ходе аннотации геномных последовательностей решаются следующие основные задачи:

- идентификация кодирующих последовательностей
- идентификация регуляторных последовательностей
- идентификация функций генов.

Их решение возможно как экспериментально, так и методами биоинформатики. Для идентификации кодирующих последовательностей применяются сложные алгоритмы. Наиболее надежными при идентификации генов являются программы, использующие все доступные критерии кодирующих и регуляторных последовательностей. Для интегральной оценки данных большинство таких программ используют скрытые марковские модели генов или искусственные нейронные сети (а некоторые программы – комбинацию двух подходов). Наиболее известными являются программы GeneMark, Grail, GenScan, GeneBuilder.

Основу функционального анализа последовательностей ДНК заложили алгоритмы сравнения последовательностей. Превые работы в этом направлении были выполнены М. Дайхофф и ее коллегами в 1960-70-х гг: создание первой базы данных белковых последовательностей (PIR 1965 г.) и предложение использовать аминокислотные матрицы замещения для сравнения белковых последовательностей (1972 г.). Первый достаточно быстрый алгоритм (FASTA) поиска гомологичных нуклеотидных последовательностей в базах данных был предложен в 1985 г. Сегодня основным является сходный алгоритм BLAST. Эти

алгоритмы используют эвристический подход, пытаясь найти не оптимальное, а "достаточно хорошее" решение, и поэтому работают значительно быстрее.

Анализ нуклеотидных и белковых последовательностей сегодня сложно представить без молекулярных баз данных. Их незаменимость обусловлена: экспоненциальным ростом объема информации; тем, что многие экспериментальные данные (последовательности, трехмерные структуры, данные масс-спектрометрии и т.д.) больше не публикуются в литературе и присутствуют только в базах данных; возможностью использовать мощные аналитические инструменты (не только компьютеры) в биологических исследованиях.

Основные молекулярные базы данных.

Нуклеотидные:

EMBL/GenBank/DBJ.

Белковые:

Аннотированные последовательности:

SWISS-PROT.

PIR.

Трехмерные структуры:

Protein Data Bank (PDB)

## **Лекция №8. Эволюция геномов**

Можно выделить различные по своим масштабам и молекулярным механизмам уровни изменений геномов, в конечном итоге определяющие их эволюцию:

геномный – аутополиплоидия (дупликация всего генома) и аллополиплоидия);

хромосомный – хромосомные перестройки (транслокации, делеции, дупликации и инверсии);

генный – дупликации генов с последующей дивергенцией, делеции, транспозиции, горизонтальный перенос;

доменный – дупликация и перетасовка доменов;

нуклеотидный – точечные мутации, короткие делеции и дупликации.

Эволюция геномов во многом зависит от рекомбинационных процессов. Две группы механизмов рекомбинации (гомологичная и сайт-специфическая) оказываются значительно эффективнее в случае присутствия в геноме мобильных генетических элементов. Мобильные генетические элементы (МГЭ или транспозоны) – небольшие (несколько тысяч пар нуклеотидов) фрагменты ДНК, способные перемещаться из одного места генома в другое. Исходный элемент может сохраняться в "старом" месте, а в "новое" может перемещаться его точная или измененная копия. Один или несколько ферментов, необходимых для перемещения, как правило, кодируют гены самого МГЭ, т.е. транспозоны сами обеспечивают свое перемещение и распространение по геному, занимая у высших эукариот значительную часть генома. Многие МГЭ не приносят пользы непосредственно индивидуальному организму, в геноме которого они

содержатся, из-за чего их иногда называют "эгоистичной ДНК". Несмотря на отсутствие очевидной пользы для индивидуума, МГЭ способны резко ускорить эволюцию вида за счет существенного облегчения рекомбинационных процессов.

МГЭ прокариот включают два основных класса. Геномы эукариот имеют подобные прокариотическим ДНК-транспозоны и три типа ретротранспозонов, перемещающихся с обязательным снятием РНК-копии с элемента с ее последующей обратной транскрипцией: ретровирусоподобные и неретровирусоподобные (элементы LINE и SINE).

### **Лекции №9-12 Организация геномов различных групп организмов.**

Геном – совокупность всех молекул ДНК организма. Кроме ДНК хромосом в него входит ДНК плазмид и молекулы ДНК эукариотических органелл. Размер генома прокариот пропорционален количеству генов и варьирует в пределах от 500 тысяч до 14 миллионов пар нуклеотидов. Для прокариот характерны компактные геномы с очень небольшой долей некодирующей ДНК (не более 5-10%), представленной в основном регуляторными последовательностями. Средний размер гена прокариот – 1000 пар нуклеотидов (чуть меньше для некоторых архей, чуть больше для самых больших геномов). Повторы (в том числе мобильные генетические элементы) присутствуют, однако занимают не более пары процентов генома. В связи с такой организацией и размер генома, и количество генов являются хорошими показателями сложности прокариотического организма. Так, геном облигатного паразита *Mycoplasma genitalium* содержит всего около 500 генов, самый компактный для свободноживущей бактерии геном (1,5 млн н.п. и 1512 генов) имеет узкоспециализированный хемолитоавтотроф *Aquifex aeolicus*, энтеробактерии имеют геномы размером около 5 млн н.п. с 4,5-5 тыс. генов, а миксобактерии могут иметь геном размером более 13 млн н.п.

Для большинства эукариот нет жестких эволюционных ограничений на увеличение размера генома, поэтому эукариотические геномы обычно существенно больше бактериальных, а также значительно разнообразнее. Размер генома эукариот определяется в основном некодирующей белки ДНК, количество которой может существенно различаться даже у близких представителей одной таксономической группы. Для эукариот не прослеживается прямая корреляция между сложностью организма и размером его генома или количеством содержащихся в нем генов. Размер генома эукариотического организма определяется в первую очередь различными повторяющимися последовательностями, которые можно разделить на следующие классы:

семейства гомологичных генов, среди которых выделяют ортологи и паралоги, а также псевдогены.

тандемные повторы (прежде всего рРНК)

сателлиты (около 80% генома кукурузы)

мобильные генетические элементы

Определенная корреляция структуры генома и сложности организма прослеживается, если рассматривать минимальные по размеру и количеству генов геномы в каждой таксономической группы. Наиболее изученными из типичных для эукариот разной сложности организации являются геномы:

хлебных дрожжей *Saccharomyces cerevisiae* (12 млн н.п., 5 тыс. генов),  
 нематоды *Caenorhabditis elegans* (100 млн н.п., 19 тыс. генов) и мухи *Drosophila melanogaster* (180 млн н.п., 13 тыс. генов) как представителей многоклеточных беспозвоночных животных,  
 человека (3 млрд н.п., 20 тыс. генов) как представителя многоклеточных позвоночных животных,  
 крестоцветного сорняка *Arabidopsis thaliana* (100 млн н.п., 19 тыс. генов) как представителя растений.

Следует также отметить существенные различия среднего размера генов эукариот – в пределах от 2 т.н.п. у дрожжей до 50 т.н.п. у человека, что связано как с разными размерами кодируемых этими генами белков, так и с возможностями аппарата сплайсинга интронов.

### **Лекции №13-17 Специализированные разделы геномики**

#### **Две задачи функциональной геномики:**

1. выяснить функции (биохимические, физиологические, клеточные, онтогенетические и т. д.) каждого структурного компонента генома (в первую очередь генов);
2. исследовать, каким образом геном организма обеспечивает его функционирование как **целостной системы**.

Для функциональной геномики характерен набор высокопроизводительных методов, применяемых в масштабах всего генома.

В функциональных исследованиях геномов применяют два типа высокопроизводительных методических подходов:

Индивидуальный: инсерционная инактивация, сайленсинг или сверхэкспрессия отдельных генов (но в масштабах всего генома) с последующим исследованием фенотипа организма

Тотальный: одновременные количественные измерения всех (или значительного числа) транскриптов или белков или метаболитов организма, чем занимаются соответственно **транскриптомика, протеомика и метаболомика**.

Индивидуальный подход использует высокопроизводительные методы прямой и обратной генетики. Прямая генетика подразумевает получение набора мутантов, анализ их фенотипов, а затем идентификацию мутировавшего гена. Обратная генетика подразумевает получение мутации (или сверхэкспрессию) конкретного гена с последующим поиском фенотипа.

Основной подход прямой генетики – насыщающий транспозоновый мутагенез. Транспозоновый мутагенез удобнее химического, поскольку:

- Транспозон обычно несет селективный маркер, по присутствию которого можно гарантированно отобрать мутантные организмы;
- Для многих транспозонов после мутагенеза в геноме присутствует только одна новая копия транспозона;
- Используя клонирование по маркеру транспозона, обратную ПЦР или прямое секвенирование, мутантный ген легко идентифицировать.

Насыщающий транспозоновый мутагенез *M. genitalium* показывает необходимость 382 из 482 кодирующих белок генов. Использован транспозон Tn4001tet. Получено 3,321 инсерций. Обратная ПЦР и сиквенс позволили картировать 2,462 инсерций. 84% мутаций – в кодирующих белки генах. Ни одной инсерции в генах рРНК, тРНК, прочих структурных РНК. Многие мутанты росли медленнее, но два - быстрее.

Недостатки транспозонового мутагенеза:

- Инсерция транспозона не дает 100%-ной гарантии инактивации гена (особенно инсерции в самом начале или в самом конце гена)
- У бактерий инсерция транспозона часто имеет полярный эффект на последующие гены оперона
- Методы прямой генетики неэффективны для большинства эукариот из-за низкой доли кодирующих последовательностей, поэтому для них чаще применяются методы обратной генетики.

Для основных модельных организмов разработаны эффективные методы направленной инактивации большого числа генов. Простейший метод инактивации гена за счет одиночного кроссинговера с его фрагментом, расположенным на плазмиде, возможен у бактерий.

Направленная инактивация генов дрожжей использует двойной кроссинговер. Дрожжи *S. cerevisiae* имеют эффективный аппарат гомологичной рекомбинации, для работы которого достаточно очень коротких областей гомологии. Это позволяет легко заменить произвольный ген на кассету антибиотикорезистентности. Инактивация большинства генов дрожжей не имеет видимых эффектов. Менее 20% генов дрожжей им действительно необходимы для роста в лабораторных условиях.

РНК-интерференция (РНКи) – экспериментальный подход, при котором введение двухцепочечной РНК (дцРНК) в клетки используется для снижения активности (или полного выключения) гомологичного гена за счет индукции деградации его мРНК. РНК-сайленсинг – природная системная супрессия при помощи дцРНК соответствующего гена (наиболее распространен у растений). РНКи и природный РНК-сайленсинг используют тот же молекулярный механизм, что и гены микроРНК – комплексы DICER и RISC.

РНК-интерференция была открыта у нематод и технически достаточно проста у этих организмов. Для того, чтобы осуществить масштабный сайленсинг генов у нематод, нужно:

- создать библиотеку фрагментов генов нематод в специальном бактериальном векторе (который обеспечит транскрипцию обеих цепей клонированной ДНК);
- ввести плазмиды с клонированными фрагментами в клетки *E. coli*;
- выращивать нематод на газоне клеток *E. coli*, экспрессирующих двухцепочечные РНК с клонированных фрагментов ДНК.

Небольшая часть дцРНК попадает в клетки нематод через пищеварительный тракт, и этого оказывается достаточно, чтобы вызвать РНК-интерференцию

**Транскриптомика** исследует, как меняются количества всех транскриптов в клетке при изменении внешних или внутренних условий.

Это помогает выяснить функции генов, поскольку:

- Экспрессия генов отражает физиологическое состояние клетки. Набор генов, экспрессируемых в клетке, определяет, как устроена клетка, как она регулируется, а также что она может и чего не может делать.
- Экспрессия генов коррелирует с их функциями. Естественный отбор определяет, что гены, экспрессируемые в определенных клетках при определенных условиях, обеспечивают функции, необходимые для надлежащей приспособленности организма.
- Гены, необходимые для выполнения одной или сходных функций, имеют одинаковую или сходную регуляцию и, соответственно, похожий характер экспрессии. Поэтому для генов со сходной экспрессией можно предполагать сходные функции.

Экспериментальные подходы к оценке количеств транскриптов основаны на трех разных принципах:

Количественная полимеразная цепная реакция (кПЦР)

Гибридизация РНК с короткими ДНК-зондами (ДНК-чипы)

Количественное секвенирование РНК (RNA-Seq).

**Протеомика** исследует набор клеточных белков и более непосредственно отражает состояние клетки, поскольку позволяет:

- Исключить вариации стабильности мРНК;
- Учесть посттрансляционные модификации белков;
- Учесть различную скорость деградации белков.

Недостатки протеомных подходов:

- Значительно большая трудоемкость и себестоимость;
- Меньшая чувствительность
- Невозможность одновременной детекции всех белков клетки.

Транскриптомные и протеомные подходы позволяют:

- Выявить гены (белки), экспрессирующиеся на разных стадиях онтогенеза,
- Выявить гены (белки), экспрессирующиеся в разных органах/тканях
- Выявить гены (белки), контролируемые одним регулятором
- Выявить гены (белки), активируемые в стрессовых условиях (и, вероятно, вовлеченные в адаптацию к стрессу...

Функциональные гипотезы, предложенные на основе "омных" подходов, как правило, проверяются для конкретных генов путем их инактивации/сверхэкспрессии с последующим изучением фенотипа, экспрессией и очисткой исследуемых белков с последующим исследованием их свойств и т.д.

Конечная цель **синтетической геномики** – создание синтетических (полностью искусственных) геномов (и, соответственно, организмов) с заданными свойствами.

Основные задачи синтетической геномики:

- Выявление минимального набора генов, необходимого для функционирования клетки
- Синтез *in vitro* полных последовательностей геномов
- Замещение ДНК клетки на синтетическую (трансплантация генома)
- Реконструкция метаболизма клетки по геномной последовательности
- Создание геномов с заданными свойствами.

Как выявить минимальный геном? □ Самый простой способ – тщательный анализ многих геномных последовательностей для выявления универсально консервативных генов, которые скорее всего являются обязательными для функционирования клетки. Проблемы такого подхода:

- упускается слишком много генов из-за того, что многие функции могут выполняться аналогичными, а не гомологичными генами;
- оценка минимального числа генов может быть занижена в 2 раза;
- как либой компьютерный анализ, такой подход просто предлагает гипотезу, которую все равно надо проверять.

Генетический подход к выявлению минимального генома заключается в инактивации всех генов по одному. Подход экспериментальный, поэтому более надежный (но значительно более трудоемкий). Основные методы: транспозоновый мутагенез (бактерии), рекомбинация инактивирующих кассет (дрожжи), РНКи (многие эукариоты). Недостатки: даже в самом маленьком геноме многие жизненно важные гены дублируются, поэтому их одиночная инактивация не имеет эффекта, но совместная – будет; большинство методов инактивации не дает 100%-ной гарантии выключения гена, поэтому "мутанты" по жизненно важным генам могут оставаться жизнеспособными. Следовательно, инактивация также дает заниженное значение для минимального числа генов.

Биохимический подход к выявлению минимального генома заключается в реконструкции базового метаболизма *in vitro*. Клеточные субсистемы, относительно охарактеризованные *in vitro*:

- Репликация ДНК
- Транскрипция
- Процессинг РНК (не до конца!)
- Трансляция
- Посттрансляционные модификации.

Для их кодирования требуется всего 151 ген (113 кодирующих белок и 38 – РНК) и геном размером 113 т.н.п.

И еще один подход к выявлению минимального генома заключается в его синтезе *in vitro*. Если клетка с синтетическим геномом живет и делится – все жизненно важные гены есть. Проблемы такого подхода:

- Чтобы такой геном создать, нужно точно знать, какие гены в него должны быть включены

В первых синтетических геномах наверняка останется много "лишних" генов

Технически сложно, пока был создан только один функциональный синтетический геном клетки (реконструирован природный геном с минимальными модификациями).

Зачем нужно создавать синтетические геномы:

- Понимание биологии клетки: если нам удастся создать функциональный геном (и "включить" его), это покажет, что мы наконец поняли, как работает клетка
- Биотехнология: из синтетического генома можно убрать все ненужные метаболические пути и, наоборот, добавить необходимые, получив клетку, синтезирующую нужные продукты без лишних затрат энергии.
- Безопасность: синтетический геном дает возможность создать клетку, гарантированно неспособную существовать вне лаборатории. Более того, использование альтернативного генетического кода позволит полностью исключить перенос синтетических генов в геномы природных организмов.

В догеномную эру уже были разработаны способы синтеза генов и целых плазмид. В 2002 г. осуществлена пошаговая сборка синтетического генома полиовируса. В 2003 г. осуществлен синтез генома бактериофага ФХ174.

Последовательность синтеза:

- химический синтез 44-мерных олигонуклеотидов и их очистка в геле
- Таq-лигирование олигонуклеотидов в строгих условиях отжига (55°C)
- полимеразная циклическая сборка полноразмерного генома
- электропорация в клетки *E. coli*, получение бляшек, выделение ДНК и ее секвенс.

Оценка числа ошибок: до 1 летальной замены на каждые 500 п.о., однако один секвенированный геном не имел ни одной ошибки (что удивительно, поскольку только каждый второй исходный олигонуклеотид не содержал ошибок)

В 2008 г. осуществлен химический синтез и сборка генома *M. genitalium* □. Геном был разбит на 101 фрагмент, каждый из которых был химически синтезирован и собран *in vitro*, после чего правильность сборки была подтверждена секвенированием. Исходный геном был изменен добавлением пяти коротких "водяных знаков" и вставкой одной кассеты антибиотикорезистентности в ген вирулентности (*msrA*). Однако в этом году



получить живую клетку *M. genitalium* с синтетическим геномом не удалось.

Возможные причины:

- Трансформации микоплазм мешают системы рестрикции-модификации. В этой работе использованы штаммы с интактной рестрикцией. Проблема была решена через год
- Инсерция рTARVAC3 в синтетический геном в неправильном месте: разорван ген РНК-компонента РНКазы Р.

Первая успешная реконструкция полностью синтетического генома осуществлена в 2010 г. Почему на этот раз получилось:

- Использован быстрорастущий вид микоплазмы
- Отработана (для этого вида) методика трансплантации
- Учтены ошибки первой попытки (метилование ДНК, инактивация жизненно важных генов)
- Сборка всего генома полностью в клетках дрожжей.

**Метагеномика** исследует природные сообщества микроорганизмов путем секвенирования их тотальной ДНК. Синонимы: геномика сообществ, популяционная геномика.

Если геном – полная генетическая информация организма, то **Метагеном** – полная генетическая информация сообщества организмов

Наибольший вклад метагеномика вносит в микробиологические исследования, поскольку:

- большая часть организмов на планете имеет микроскопический размер
- микроорганизмы есть практически везде, в том числе в сообществах с макроорганизмами (как обязательный компонент)
- в наших телах клеток бактерий как минимум в 10 раз больше, чем клеток человека
- классические методы микробиологии рассчитаны на чистые культуры организмов, которых можно культивировать в лабораторных условиях
- подавляющее большинство микроорганизмов культивировать в лаборатории не удастся. Культивируются не более 0,1-1% почвенных микроорганизмов и гораздо меньше водных. Основная причина – большинство организмов зависят от других членов сообщества.
- в результате **мы не знаем даже о самом факте существования большинства видов на этой планете**, и тем более ничего не знаем о биологии представителей этих видов.

Чистые культуры не дают реального представления о природных сообществах микроорганизмов. Аномалия количественных посевов заключается в том, что во многих случаях прямой подсчет под микроскопом дает значительно более высокую (в сотни и даже тысячи раз) оценку числа бактерий в популяции, чем подсчет числа колоний после посева. Несмотря на невозможность культивировать подавляющее большинство микроорганизмов из почвенных, водных и др. образцов, их жизнеспособность можно легко доказать

целым рядом методов (окрашиванием витальными красителями, детекцией продуктов метаболизма и т.д.). Введение в практику генов рРНК в качестве эволюционных хронометров и изобретение ПЦР позволили предложить первые надежные методы детекции и идентификации некультивируемых организмов.

Большинство таксономических групп бактерий представлены некультивируемыми организмами, однако геномика дает возможность изучать практически любые некультивируемые организмы. Даже если культивировать микроорганизмы невозможно, из них все равно можно выделить ДНК. Если ДНК есть, можно определить ее последовательность. Зная последовательность значительной части генома, можно получить представление о таксономии организма и его биологических свойствах. Проблема такого подхода – в естественной среде обитания микроорганизмы практически всегда существуют в сложных сообществах, состоящих из особей многих видов, часто в значительной степени сходных, из-за чего полные последовательности геномов отдельных видов собрать чаще всего нереально.

Стадии метагеномного исследования:

- отбор образцов из окружающей среды
- фильтрация частиц, обычно по размеру
- лизис клеток и экстракция ДНК
- создание библиотеки
- секвенирование
- анализ геномной информации.

Основная проблема метегеномики – сборка полных геномов при анализе сложных сообществ. При стандартном геномном подходе вся ДНК происходит от одного клона, что делает полную сборку возможной (с использованием либо иерархического подхода либо шотгана в комплекте с прицельным закрытием пробелов).

При метагеномном подходе ДНК выделяется из гетерогенных сообществ (от 10 до 1000 и более видов), что определяет неполноту, неравную представленность, а часто и более низкое качество данных. Компьютерная сборка метагенома требует специализированных (и значительно более сложных) алгоритмов, занимает значительно больше времени (до полугода на одном быстром современном компьютере).

Основной объект метегеномики – гены, а не геномы. Метагеномика по определению может использовать только шотган-подход к секвенированию. Вероятностный характер шотган-подхода приводит к тому, что с сегодняшними технологиями сборка полных геномов конкретных организмов при анализе метагенома невозможна! В лучшем случае для простых сообществ (с малым числом видов) удастся получить относительно протяженные контиги, с высокой долей вероятности собранные из ДНК одного вида. По этой причине метагеномика делает акцент на генах, а не на геномах, что все равно дает возможность понять, как такие сообщества функционируют, а также позволяет при необходимости клонировать наиболее интересные гены. Такой подход

полностью оправдан с учетом горизонтального переноса (вспомните также концепцию пангенома): гены большинства организмов не ограничены рамками одной клетки (одного генома), а достаточно легко перемещаются в пределах сообщества.

Основные вопросы метагеномики:

- Какие организмы присутствуют в сообществе?
- Как эти организмы взаимосвязаны и как функционирует сообщество в целом?
- Какую пользу может получить от этих организмов человек?

Области применения метагеномики:

- Получение информации о всех формах жизни на планете
- Обнаружение новых (практически полезных) ферментов и метаболитов
- Исследование влияния загрязнений на экосистемы и выявление сообществ, способных к очистке загрязненных экосистем
- Исследование влияния микробиома человека на его здоровье
- Палеогеномика.

**Палеогеномика** – исследование геномов вымерших организмов и реконструкция их свойств путем анализа геномной информации.

Палеогеномные исследования возможны, поскольку после гибели организмов их ДНК миллионы лет может сохраняться в ископаемых останках. Большинство ископаемых останков – кости. Из них удается в небольших количествах выделить короткие фрагменты ДНК (как правило, для костей не старше 100 млн лет). Гораздо больше ДНК и лучшего качества можно выделить из мягких тканей, которые могут сохраняться в условиях вечной мерзлоты для относительно недавно вымерших видов (напр., мамонт). Волосы – идеальный источник ископаемой ДНК, поскольку для них минимален риск контаминации.

Сложности секвенирования «ископаемой» ДНК:

- Сильная фрагментация ДНК (фрагменты часто короче 100 н.п.) делает очень сложным физическое картирование палеогеномов и фактически позволяет использовать только шотган-подход
- ДНК из ископаемых образцов выделяется в очень малых количествах, поэтому требуются специальные методы для ее амплификации
- Как правило, в образцах присутствует примесь гораздо больших количеств ДНК бактерий и грибов
- С ДНК за миллионы лет могут происходить химические реакции, меняющие ее последовательность, поэтому по возможности следует секвенировать обе цепи
- Низкое качество исходной ДНК требует большей степени перекрытия генома (15-20 раз и более).

Все эти особенности определяют использование для секвенирования древней ДНК только методик II-III поколений

Первый палеогеномный проект – геном мамонта (2008 г.)

Оценка размера генома	4,7 млрд н.п.
Всего секвенировано	4,168 млрд н.п.
(из них можно совместить с геномом <i>Loxodonta africana</i> )	3,3 млрд н.п.)
Средняя длина одного сиквенса	128 н.п.
Митохондриальный геном мамонта М4 прочтен	4430 раз
Частота ошибок (оценка по митохондриальному геному) – 0,35%, в т.ч. :	
замены оснований из-за повреждения ДНК	6 на 10,000 н.п.
замены из-за ошибок секвенирования	8 на 10,000 н.п.
Инсерции при секвенировании (пиросеквенатор!)	21 на 10000 н.п.
Сходство последовательностей мамонта и африканского слона:	
на уровне ДНК	99.41%
на уровне белка	99.78%

## 2. ПРАКТИЧЕСКИЙ РАЗДЕЛ

В соответствии с учебным планом практических занятий по курсу не предусмотрено.

## 3. КОНТРОЛЬ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

### Вопросы для самоконтроля

#### Лекция №1. История геномики

1. История развития геномных исследований.
2. Геномная революция конца XX века.
3. Иерархический и шотган-подход к расшифровке геномных последовательностей.
4. Эволюция подходов к расшифровке геномных последовательностей

#### Лекции №2-3. Как устроены геномы и как они работают?

1. Синтез ДНК *in vivo* и *in vitro*: компоненты и продукты реакции, свойства ДНК-полимераз.
2. Способы использования реакции полимеризации ДНК для определения нуклеотидных последовательностей.
3. Хромосомы про- и эукариот: форма, количество, структурные элементы, обеспечивающие стабильность и репликацию.
4. Структура гена у различных организмов: прерывистые и непрерывные кодирующие последовательности, размеры и расположение регуляторных элементов.
5. Организация оперонов у про- и эукариот.

#### Лекции №4-6. Методы секвенирования ДНК

1. Секвенирование ДНК по методу Сэнгера: возможности и ограничения.
2. Принцип действия, достоинства и недостатки геномных секвенаторов второго поколения, использующих реакцию пиросеквенирования

3. Принцип действия, достоинства и недостатки геномных секвенаторов второго поколения, использующих ДНК-полимеразную реакцию (секвенирование путем синтеза, Illumina)

4. Принцип действия, достоинства и недостатки геномных секвенаторов второго поколения, использующих детекцию протонов (Ion Torrent)

5. Геномные секвенаторы третьего поколения, использующие технологию SMRT (Pacific Biosciences): принцип действия, преимущества и недостатки

6. Нанопоровые геномные секвенаторы третьего поколения (Oxford Nanopore): принцип действия, преимущества и недостатки

### **Лекция №7. Молекулярные базы данных и аннотация геномных последовательностей**

1. Аннотация геномных последовательностей: основные задачи и подходы к их решению.
2. Как можно найти кодирующие последовательности в неаннотированном геноме?
3. Как можно определить функции кодирующих последовательностей?
4. Молекулярные базы данных. Специализация, структура и методы поиска информации.
5. Основные алгоритмы сравнения нуклеотидных и белковых последовательностей между собой и программные пакеты, их реализующие.
6. Какая из молекулярных баз данных содержит больше всего информации?
7. Какая из основных молекулярных баз данных содержит наиболее достоверную информацию и почему?

### **Лекция №8. Эволюция геномов**

1. Ранние этапы эволюции геномов. Мир РНК
2. Уровни изменений генома.
3. Относительный вклад мутационных и рекомбинационных процессов в эволюцию генома.
4. Мобильные генетические элементы как горячие точки рекомбинации.
5. Классификация, строение и основные свойства мобильных генетических элементов эукариот.
6. Классификация, строение и основные свойства мобильных генетических элементов прокариот.
7. Вклад горизонтального переноса генов в эволюцию геномов про- и эукариот. Острова патогенности.
8. Дайте определение термина "вид" для прокариот.
9. Что такое пангеном?

### **Лекции №9-12 Организация геномов различных групп организмов.**

1. Характерные черты геномов прокариот.
2. Характерные черты геномов факультативных и облигатных патогенов.

3. Взаимная адаптация геномов патогена и его хозяина.
4. Разнообразие и характерные особенности геномов одноклеточных эукариот.
5. Основные характеристики геномов грибов.
6. Организация геномов нематод.
7. Организация генома *Drosophila melanogaster*.
8. Особенности организации геномов позвоночных животных.
9. Сравнительная характеристика геномов *Homo sapiens* и *Pan troglodytes*.
10. Отличительные черты геномов растений.

### Лекции №13-17

1. Функциональная геномика. Подходы к идентификации генов в геномных последовательностях и определению их функций.
2. Концепция минимального генома. Природные минимальные геномы бактерий, архей, эукариот – их размер, число генов, особенности организации.
3. Синтетическая геномика. Методы синтеза, клонирования и трансплантации полных геномных последовательностей. Достижения и перспективы синтетической геномики.
4. Метагеномика. Геномные подходы к исследованию сообществ некультивируемых микроорганизмов.
5. Палеогеномика. Технические трудности, достижения и перспективы палеогеномных проектов.

### Вопросы для подготовки к зачету

1. История развития геномных исследований. Геномная революция конца XX века.
2. Геномные проекты. Иерархический и шотган-подход. Фазы геномного проекта.
3. Современные методы картирования геномов.
4. Сложности расшифровки генома высших эукариот и пути их преодоления.
5. Синтез ДНК *in vitro*: компоненты и продукты реакции, свойства ДНК-полимераз. Способы использования реакции полимеризации ДНК для определения нуклеотидных последовательностей.
6. Секвенирование ДНК по методу Сэнгера: возможности и ограничения.
7. Принцип действия, достоинства и недостатки геномных секвенаторов второго поколения, использующих реакцию пиросеквенирования
8. Принцип действия, достоинства и недостатки геномных секвенаторов второго поколения, использующих ДНК-полимеразную реакцию (секвенирование путем синтеза, Illumina)
9. Принцип действия, достоинства и недостатки геномных секвенаторов второго поколения, использующих детекцию протонов (Ion Torrent)
10. Геномные секвенаторы третьего поколения, использующие технологию SMRT (Pacific Biosciences): принцип действия, преимущества и недостатки
11. Нанопоровые геномные секвенаторы третьего поколения (Oxford Nanopore): принцип действия, преимущества и недостатки

12. Аннотация геномных последовательностей: основные задачи и подходы к их решению.
13. Молекулярные базы данных. Специализация, структура и методы поиска информации.
14. Функциональная геномика. Подходы к идентификации генов в геномных последовательностях и определению их функций.
15. Возможности и ограничения компьютерного анализа при идентификации кодирующих и регуляторных последовательностей, а также для предсказания их функций.
16. Транскриптомные и протеомные подходы к идентификации функций структурных элементов генома.
17. Классификация, строение и основные свойства мобильных генетических элементов эукариот.
18. Классификация, строение и основные свойства мобильных генетических элементов прокариот.
19. Вклад горизонтального переноса генов в эволюцию геномов про- и эукариот. Острова патогенности. Концепция пангенома.
20. Хромосомы про- и эукариот: форма, количество, структурные элементы, обеспечивающие стабильность и репликацию.
21. Структура гена у различных организмов: прерывистые и непрерывные кодирующие последовательности, размеры и расположение регуляторных элементов.
22. Организация оперонов у про- и эукариот. Проблема экспрессии внутренних генов оперонов эукариот и молекулярный механизм ее решения.
23. Концепция минимального генома. Природные минимальные геномы бактерий, архей, эукариот – их размер, число генов, особенности организации.
24. Характерные черты геномов прокариот.
25. Характерные черты геномов факультативных и облигатных патогенов. Взаимная адаптация геномов патогена и его хозяина.
26. Разнообразие и характерные особенности геномов одноклеточных эукариот.
27. Основные характеристики геномов грибов.
28. Организация геномов нематод.
29. Организация генома *Drosophila melanogaster*.
30. Особенности организации геномов позвоночных животных.
31. Сравнительная характеристика геномов *Homo sapiens* и *Pan troglodytes*.
32. Отличительные черты геномов растений.
33. Синтетическая геномика. Методы синтеза, клонирования и трансплантации полных геномных последовательностей. Достижения и перспективы синтетической геномики.
34. Метагеномика. Геномные подходы к исследованию сообществ некультивируемых микроорганизмов.
35. Палеогеномика. Технические трудности, достижения и перспективы палеогеномных проектов.

## 4. ВСПОМОГАТЕЛЬНЫЙ РАЗДЕЛ

### Учебно-программные материалы

Мультимедийные презентации по всем лекциям курса доступны через систему e-University по адресу <http://euniversity.bsu.by>.

Учебная программа по дисциплине «Геномика» для специальностей 1-31 01 01 Биология (по направлениям), 1-31 01 02 Биохимия, 1-31 01 03 Микробиология доступна по адресу:

<http://elib.bsu.by/handle/123456789/105411>

Учебная программа (рабочий вариант) по дисциплине «Геномика» для специальностей 1-31 01 01 Биология (по направлениям), 1-31 01 02 Биохимия, 1-31 01 03 Микробиология доступна по адресу:

<http://elib.bsu.by/handle/123456789/105412>

### Список рекомендуемой литературы и Интернет-ресурсов

Список рекомендуемой литературы и Интернет-ресурсов приведен в учебной программе по дисциплине «Геномика» по специальностям 1-31 01 01 Биология (по направлениям), 1-31 01 02 Биохимия, 1-31 01 03 Микробиология доступна по адресу:

<http://elib.bsu.by/handle/123456789/105411>