УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ
УНИВЕРСИТЕТ ИМЕНИ МАКСИМА ТАНКА»

Факультет естествознания
Кафедра биологии и методики преподавания биологии
(рег. № УМК 25-1-285-2024)

СОГЛАСОВАНО
Заведующий кафедрой
_____ И. И. Жукова
16.01. 2024 г.

СОГЛАСОВАНО
Декан факультета
_____ Г. В. Скриган
17.01 2024 г.

ЭЛЕКТРОННЫЙ УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС ПО УЧЕБНОЙ ДИСЦИПЛИНЕ «БИОИНФОРМАТИКА»

для специальности
7-06-0113-03 Природоведческое образование
Профилизация: Биология, география, химия

Составитель:
Деревинский А. В., доцент кафедры биологии и методики преподавания биологии, кандидат сельскохозяйственных наук, доцент

Рассмотрено и утверждено
на заседании Совета БГПУ «25» января 2024 г., протокол № 5

ОГЛАВЛЕНИЕ

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА
1. ТЕОРЕТИЧЕСКИЙ РАЗДЕЛ
1.1. Краткое содержание лекционных занятий
2. ПРАКТИЧЕСКИЙ РАЗДЕЛ
2.1. Тематика и задания для практических занятий
3. РАЗДЕЛ КОНТРОЛЯ ЗНАНИЙ
3.1. Вопросы к экзамену
3.2. Примерная тематика рефератов
3.3. Тематика докладов для самостоятельной работы
3.4. Итоговый тест
4. ВСПОМОГАТЕЛЬНЫЙ РАЗДЕЛ
4.1. Учебно-программная документация
4.2. Список рекомендуемой литературы

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Электронный учебно-методический комплекс (ЭУМК) по учебной дисциплине «Биоинформатика» создан в соответствии с образовательным стандартом углубленного высшего образования: ОСВО 7-06-1113-03-2023 по специальности 7-06-0113-03 Природоведческое образование (18.05.2023, № 160); учебным планом по специальности (23.02.2023, № 076-2023/у) и требованиями Положения об учебно-методическом комплексе на уровне высшего образования.

Цель ЭУМК – оказание методической помощи студентам в систематизации учебного материала в процессе подготовки к промежуточной аттестации по учебной дисциплине «Биоинформатика».

Материалы ЭУМК дают возможность студентам планировать и осуществлять самостоятельную работу, обеспечивают рациональное распределение учебного времени по темам учебных занятий.

К основным задачам ЭУМК относятся:

– создать условия для поэтапного и эффективного овладения знаниями, формирования умений и навыков по учебной дисциплине «Биоинформатика», специализированных компетенций, необходимых для решения профессиональных задач;

– предоставить учебно-методические материалы и рекомендации для эффективного выполнения заданий к практическим занятиям в соответствии с учебным планом и учебно-программной документацией;

– разработать комплекс материалов промежуточной аттестации, позволяющий определить соответствие результатов учебной деятельности обучающихся требованиям образовательного стандарта высшего образования;

– предоставить учебно-программную документацию, перечень учебных изданий и информационно-аналитических материалов, рекомендуемых для изучения учебной дисциплины.

Структура учебно-методического комплекса учебной дисциплины состоит из четырех разделов: теоретического, практического, раздела контроля знаний и вспомогательного.

Теоретический раздел включает конспективно изложенный лекционный материал, соответствующий требованиям учебной программы учебной дисциплины, в объеме, установленном учебным планом по специальности.

В практическом разделе размещены учебно-методические материалы, необходимые для проведения практических занятий по учебной дисциплине в соответствии с учебным планом по специальности.

Раздел контроля знаний включает материалы промежуточной аттестации, позволяющие осуществлять мониторинг результатов учебно-познавательной деятельности студентов.

Вспомогательный раздел содержит учебно-программные материалы (учебную программу для студентов дневной и заочной форм получения образования), список рекомендуемой литературы, алгоритмы выполнения

индивидуальных творческих заданий. С помощью учебной программы по учебной дисциплине можно получить информацию о тематике лекций и практических занятий, перечнях рассматриваемых вопросов и рекомендуемой для их изучения литературы.

Для подготовки к практическим занятиям и экзамену рекомендуется использовать материалы, представленные в теоретическом, практическом разделе, а также материалы для промежуточного контроля знаний и самостоятельной работы.

1. ТЕОРЕТИЧЕСКИЙ РАЗДЕЛ

1.1. Краткое содержание лекционных занятий

Тема 1. 1. Предметная область биоинформатики

Вопросы для рассмотрения:

1. Биоинформатика как наука.
2. Понятие информации Свойства информации Методологические аспекты моделирования.
3. Понятие об аналогии. Ассоциативная аналогия.
4. Определение модели. Аналогия и моделирование. Типы моделирования.
5. Материальные модели. Мысленные модели. Аналитические, конструктивные и имитационные модели. Материальное моделирование и модельный эксперимент.

Краткое содержание лекции

Биоинформатика – это наука о хранении, извлечении, организации, анализе, интерпретации и использовании биологической информации.

Современная биоинформатика возникла в конце семидесятых годов двадцатого века с появлением эффективных методов расшифровки нуклеотидных последовательностей ДНК. Датой выделения биоинформатики в отдельную научную область можно считать 1980 год, когда началось издание журнала Nucleic Acids Research, целиком посвященного компьютерным методам анализа последовательностей.

Важной вехой в становлении и развитии биоинформатики стал проект по секвенированию генома человека. Именно с этого времени биоинформатика перестала быть только вспомогательным инструментом. Переход к обработке, анализу и сравнению полных геномов организмов был невозможен без использования компьютерных методов информационного анализа, в результате эти исследования вылились в самостоятельное научное направление. Геномы содержат огромное количество генов, многие из которых до настоящего времени не идентифицированы экспериментально.

Предмет биоинформатики включает в себя три компонента:

1) создание баз данных, позволяющих осуществлять хранение крупных наборов биологических данных и управление ими;

2) разработка алгоритмов и методов статистического анализа для определения отношений между элементами крупных наборов данных;

3) использование этих средств для анализа и интерпретации биологических данных различного типа – в частности, последовательностей ДНК, РНК и белков, белковых структур, профилей экспрессии генов и биохимических путей.

Цели биоинформатики следующие:

– организовывать данные таким образом, чтобы исследователи имели доступ к текущей информации, хранящейся в базах данных, и могли вносить в нее новые записи по мере получения новых сведений;

– развивать программные средства и информационные ресурсы, которые помогают в управлении данными и в их анализе;

– применять эти средства для анализа данных и интерпретации полученных результатов таким образом, чтобы они имели биологический смысл.

Задачи биоинформатики состоят в анализе информации, закодированной в биологических последовательностях, в частности:

• обнаруживать гены в последовательностях ДНК различных организмов;

• развивать методы изучения структуры и (или) функции новых расшифрованных последовательностей и соответствующих структурных областей РНК;

• определять семейства родственных последовательностей и строить модели;

• выравнивать подобные последовательности и восстанавливать филогенетические деревья с целью выявления эволюционных связей.

Помимо перечисленных выше задач, следует упомянуть *еще один важнейший* вопрос биоинформатики – *обнаружение мишеней* для медикаментозного воздействия и отыскание перспективных опытных соединений.

Информация - *запомненный выбор одного варианта из нескольких возможных и равноправных*.

Запомненный выбор еще называют *макроинформацией*. Если информация создается, но не запоминается, то ее называют *микроинформацией*.

Цитата Н. Винера: «Информация – это обозначение содержания, полученного из внешнего мира в процессе нашего приспособления к нему и приспосабливания к нему наших чувств…»

Свойства, присущие всем видам информации, разделяются на две крупные группы, внутри каждой из которых свойства тесно связаны между собой. Для одной группы ключевым свойством является *фиксируемость* информации Для другой группы определяющим свойством является ее *действенность*. Остальные свойства, входящие в эти группы, можно расценивать как раскрытие, проявление ключевых особенностей в формах, доступных для регистрации.

Фиксируемость информации – свойство, благодаря которому любая информация, не будучи ни материей, ни энергией, может существовать не в свободном виде, а *только в зафиксированном состоянии* – в виде записи на каком-либо физическом носителе. Способы записи информации на таком носителе всегда *условны*, т. е. не имеют никакого отношения к ее семантике (или содержательности). Например, одно и то же предложение может быть записано и на бумаге, и на электронном носителе.

Действенность информации. Вторая группа свойств информации объединяется ключевым свойством – действенностью (рисунок 2). Это свойство выявляется следующим образом: будучи включенной в свою информационную систему, информация может быть использована для построения того или иного *оператора*, который может совершать определенные целенаправленные действия. Оператор, таким образом, выступает в роли посредника, необходимого для *проявления* действенности информации.

Генерация информации – это выбор варианта, сделанный случайно (без подсказки извне) из многих возможных и равноправных (т.е. из принадлежащих одному множеству) вариантов. Если речь идет о возникновении новой информации, то выбор должен быть именно случайным.

Если выбор подсказан на основе уже имеющейся информации, то речь идет о восприятии, *рецепции* информации. «Запоминаемость» ассоциируется с рецепцией информации.

Методологические аспекты моделирования.

Понятие об аналогии. Термин аналогия употребляется в нескольких смыслах. Во-первых, под аналогией понимают отношение сходства между сопоставляемыми объектами. Во-вторых, иногда это понятие употребляется как синоним умозаключения по аналогии, т. е. определенного метода получения вероятностных заключений на основании указанного отношения сходства.

В последнем случае выводы, полученные путем подобного умозаключения, оказываются более достоверными, если общие свойства у сравниваемых объектов, предметов или явлений относятся к разряду наиболее существенных, а также являются достаточно многочисленными и разнообразными. Достоверность полученного вывода выше, если переносимое на другой объект свойство относится к тому же типу, что и те свойства, о которых

идет речь в посылках. В качестве примера ошибочного умозаключения по аналогии можно привести широко известный и вызывающий до сих пор горячие дискуссии вывод об обитаемости некоторых других планет на основе их сходства с Землей. Слабость подобного вывода заключается в том, что переносимое свойство – наличие жизни – относится к биологическому типу, тогда как в исходных посылках речь идет только о сходстве факторов геометрического и физико-химического характера.

С умозаключением по аналогии тесно связана так называемая ассоциативная аналогия, когда могут объединяться достаточно далекие по своей природе явления и предметы. Ярким примером проявления ассоциативной аналогии может служить история создания немецким химиком-органиком Ф. А. Кекуле структурной формулы бензола, после того как он увидел в зоопарке несколько обезьян, сцепившихся друг с другом в кольцо.

Аналогия в известном смысле смыкается с операцией обобщения, так как она строится в виде предположения о каком-то общем признаке.

Определение модели. Аналогия и моделирование. Моделирование является мощным методом научного исследования, позволяющим изучать те явления действительности, которые недоступны для непосредственного изучения из-за трудностей теоретического и экспериментального характера, а также в связи с невозможностью их выделения в чистом виде по тем же причинам. Поэтому построение модели связано с процессом идеализации – существенным звеном в цепи познания.

Модель (от лат. modus – копия, образ, очертание) можно определить как любую систему, мысленно представимую или реально существующую, которая находится в определенном отношении к другой системе (оригиналу или объекту). Как считает известный методолог В. А. Штофф, в процессе моделирования обязательно должны выполняться следующие условия:

1. Между моделью и оригиналом имеется отношение сходства, форма которого явно выражена и точно зафиксирована (условие отражения или уточненной аналогии).
2. Модель в процессе научного познания является заместителем изучаемого объекта (условие репрезентации).
3. Изучение модели позволяет получать информацию об оригинале (условие экстраполяции).

Поскольку модель служит цели объяснить или описать неизвестное явление при помощи сравнения его с другим, известным, корректное применение метода моделирования предполагает формулирование в ясном виде условий и границ, в пределах которых имеют место характерные аналогии отношения сходства и различия между моделью и оригиналом. Как отмечает современный немецкий философ Г. Клаус, сходство между системами может осуществляться на разных уровнях:

1. На уровне результатов, которые дают сравниваемые объекты.
2. На уровне поведения или функций, которые ведут к этим результатам.
3. На уровне структур, которые обеспечивают выполнение данных функций.
4. На уровне материалов или элементов, из которых состоят эти структуры.

В модели нет и не может быть совпадения на всех этих уровнях. Если бы удалось добиться полной тождественности модели и оригинала, то модель утратила бы свою специфику. Одновременно, поскольку модель должна помогать познанию оригинала, при создании модели необходимо стремиться к максимально возможному сходству.

Таким образом, отношение модели к моделируемому объекту есть отношение аналогии, а не тождества. При этом обычно в науке реализуются главным образом аналогии на уровне структур и на уровне функций. Аналогия на уровне структур более типична для моделей, применяемых в физических и химических исследованиях; в молекулярной биологии используется при моделировании и предсказании структуры и функции белков. Функциональная аналогия характерна для методов кибернетики, а потому также нашла, свое место в биоинформатике.

Типы моделирования. В основу выделения типов моделей можно положить различные отношения их к натурным объектам: характер воспроизводимых сторон оригинала; способ воспроизведения в модели действительности; характер реализации моделируемых объектов и т. п.

В своем изложении мы будем в основном следовать В. А. Штоффу, согласно которому все модели, применяемые в научном познании, можно разделить на два больших класса: материальные (вещественные) и мысленные (идеальные).

К числу первых относятся природные объекты, подчиняющиеся в своем функционировании естественным законам.

Вторые конструируются в форме мысленных образов, функционирующих по законам логики мышления.

Материальные модели можно разделить на три основные группы:

1. Натурные модели – реальные объекты, процессы и системы, над которыми выполняются эксперименты научные, технические и производственные.
2. Физические модели – макеты, муляжи, воспроизводящие физические свойства оригиналов (кинематические, динамические, гидравлические, тепловые, электрические, световые модели).
3. Математические модели (материальные) – аналоговые, структурные, геометрические, графические, цифровые и кибернетические модели.

Важную роль в познании играет мысленное, логическое моделирование, в процессе которого создается идеальная модель исследуемого явления. С точки зрения построения, мысленные модели можно разделить также на три группы:

1. Наглядные (образные) модели – это схемы, карты, чертежи, графики, графы, аналоги, структурные и геометрические модели.
2. Знаковые модели – символы, алфавит, языки программирования, упорядоченная запись, топологическая запись, сетевое представление.
3. Математические модели (идеальные) – это аналитические, функциональные, имитационные, комбинированные модели.

В настоящее время мысленное моделирование все чаще выступает в форме именно математического моделирования.

Проявлением наглядно-образного характера мысленных моделей можно считать следующее. Во-первых, элементы, из которых конструируются эти модели, представляют собой образы каких-то реальных явлений. Во-вторых, некоторые свойства и отношения моделируемых явлений выступают в форме, доступной чувственности, благодаря тому, что эти свойства и отношения присущи также и тем системам, которые выступают в качестве аналогов моделируемых явлений. В качестве примера можно привести работы Леонардо да Винчи – «отца бионики» по моделированию движения животных в рамках еще только зарождавшейся в XV в. механики. В XVIII в. выдающийся математик Л. Эйлер предложил механическую модель сердца, рассматривающую его как своеобразный насос, перекачивающий кровь по трубам – кровеносным сосудам. Более близки нам по времени работы по физиологии зрения и слуха знаменитого физика XIX в. Г. Гельмгольца – врача по образованию, также являющиеся примером использования в исследовании иконических моделей.

Знаковая, а тем более математическая модель уже не имеют по своей физической природе ничего общего с характером элементов моделируемого объекта.

Построить математическую модель – значит определить связи между теми или иными процессами и явлениями, создать (или подобрать) математический аппарат, позволяющий выразить количественно и качественно связь между элементами модели, т. е. теми или иными процессами и явлениями, между интересующими специалиста физическими величинами и факторами, от которых зависит конечный результат.

По принципам построения математические модели разделяют на аналитические, конструктивные (алгоритмические) и имитационные. В аналитических моделях развитие реальных объектов, процессов или систем записывается в виде явных функциональных выражений в зависимости от математической проблемы.

Однако с усложнением объекта моделирования построение аналитической модели превращается в трудноразрешимую проблему. Тогда исследователь вынужден переходить к конструктивным (алгоритмическим) моделям или использовать имитационное моделирование. О нем мы поговорим ниже, в разделе о мысленном эксперименте.

Применение метода аналогии: в тесной связи с другими приемами и методами научного исследования делает его ценным орудием поиска гипотезы, построения теории, эффективным средством научного открытия. Существование объектов и процессов, труднодоступных для непосредственного изучения, а также невозможность выделить некоторые явления в чистом виде диктуют применение методов моделирования. При этом основой для построения моделей служит наличие аналогии между сопоставляемыми объектами.

Материальное моделирование и модельный эксперимент. Одним из важнейших способов использования материальных моделей в научных исследованиях является их применение в так называемом модельном эксперименте. Его своеобразие заключается в том особом положении, в котором находится модель в таком эксперименте. Во-первых, модель является объектом изучения, так как замещает собой другой, подлинный: объект. Во-вторых, она служит экспериментальным средством в силу того, что используется для изучения свойств оригинала.

Так как в эксперименте с использованием модели непосредственному исследованию подвергается именно модель, а результаты переносятся на моделируемый объект, то теоретическое обоснование права на подобную экстраполяцию результатов является непременным условием всякого модельного эксперимента.

В механике, например, основой модельного эксперимента служит теория подобия, которая устанавливает соотношения между качественно однородными явлениями, а также между системами, относящимися к одной и той же форме движения. Вот почему нашли широкое применение аэродинамические трубы и гидродинамические бассейны, без которых невозможно себе представить автомобиле- и самолетостроение.

Иначе обстоит дело в биологии, где давно применяется метод культуры ткани, в том числе для воспроизведения специфических биохимических реакций. Однако изолированная клетка далеко не тождественна клетке в организме. В частности, для растительных клеток хорошо известно, что перевод их в культуру вызывает целый спектр мутаций. Это заставляет подходить с чрезвычайной осторожностью к интерпретации результатов, полученных в таком модельном эксперименте in vitro, и особенно к экстраполяции их на случай in vivo.

Мысленный эксперимент. В модельном эксперименте могут участвовать не только материальные модели. Характерной чертой мысленной модели является возможность ее использования в мысленном эксперименте. Структура последнего полностью соответствует структуре реального эксперимента с той только разницей, что тут и модель, и условия, к числу которых относятся также идеализированные «приборы» и «инструменты», являются мысленными построениями.

Мысленный эксперимент представляет собой такое воображаемое сочетание условий, которое может не наблюдаться или быть практически неосуществимым. Поэтому все типичные для такого умственного эксперимента способы рассуждений должны проводиться в рамках объективно возможного, и необходимо изгнание всяческих допущений или посылок, находящихся в противоречии с объективными законами действительности. Крупнейшие теоретики естествознания, широко применяя в своей научной деятельности метод умственного моделирования, всегда строго учитывали это положение. Примером тому может служить А.

Эйнштейн, любивший использовать мысленные модели для иллюстрации своих теоретических положений.

Характерной особенностью мысленного эксперимента является то обстоятельство, что в нем часто стирается различие между «внешними условиями» и «приборами», которые в конечном итоге сводятся к первым. Достигнутые этим упрощения делают структуру мысленного эксперимента такой, что она мало отличается от самой модели, т. е. в пределе понятия мысленного эксперимента и мысленной модели совпадают.

Особенно плодотворно мысленное экспериментирование применяется с середины XX в. в биологии в связи с развитием и широким внедрением ЭВМ и математических методов исследования. Распространение этого вида эксперимента при изучении живых объектов, особенно в форме компьютерного эксперимента, может быть объяснено их чрезвычайной сложностью, что приводит к необходимости теоретического исследования. Естественно, как и в любом другом виде мысленного эксперимента, в данном случае имеет место идеализация, как объектов изучения, так и условий, в которых они находятся.

Формой реализации компьютерного эксперимента является имитационная модель. Современное имитационное моделирование – это соединение традиционного математического моделирования с новыми компьютерными технологиями. В западной научной литературе этому направлению приблизительно соответствует термин simulation. Имитационная модель описывается набором алгоритмов или функций, которые имитируют реальные элементарные явления, составляющие изучаемые процесс или систему. При этом обязательно сохраняются их логическая структура и последовательность протекания во времени. Имитационное моделирование позволяет по исходным данным получить сведения о состояниях процесса или системы в определенные моменты времени, однако прогнозирование поведения объектов, процессов или систем здесь затруднено, и предсказательная сила этого метода невелика.

Литература: Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витебск. гос. ун-т имени П. М. Машерова, 2022. – 83 с.

Тема 1. 2. Принципы моделирования биологических систем, явлений и процессов.

Вопросы для рассмотрения:

1. Особенности моделирования в биологии.
2. Модели данных. Модели систем.
3. Применение некоторых информационных технологий в биологии. Компьютерная алгебра и биоинформатика.
4. Биологическое приложение теории нечётких множеств.

Краткое содержание лекции

Особенности моделирования в биологии. Сложность жизни проявляется в переплетении различных уровней организации живых объектов, что приводит к определенной специфичности при моделировании биологических явлений. Как и в других отраслях естествознания, в биологии в основе моделирования лежит представление о замещении какой-либо сложной системы более простой и доступной для изучения. Часто с этой целью в качестве упрощенных материальных моделей достаточно развитых явлений используются более простые формы живых организмов. Как отмечал Н. И. Вавилов, открытию сцепления генов наряду с коротким жизненным циклом дрозофилы содействовало наличие у нее по сравнению с более развитыми организмами небольшого количества хромосом. Успехам молекулярной биологии во многом способствовал удачный выбор объекта исследований (бактерий и вирусов). В практику биохимии вошло изучение отдельных биохимических реакций,

воспроизводимых в культуре клеток ткани ***in vitro.*** Эти и многие другие примеры дают представление о биологическом аналоге материальной, физической модели.

В современном биологическом модельном эксперименте широко применяются физические и кибернетические методы. Тесное переплетение подходов биофизики, биоинформатики, биокибернетики дает возможность использовать в исследовательских целях специальные конструкции, воспроизводящие отдельные функции живого организма на чуждой ему вещественной основе.

Однако далеко не всегда модель биологического объекта должна быть очень сложной. Иногда важные результаты можно получить с помощью достаточно простых приспособлений, как это было с механической моделью-макетом ДНК, созданной Дж. Уотсоном и Ф. Криком по результатам рентгеноструктурного анализа этой молекулы.

Говоря о математических моделях, необходимо подчеркнуть, что разнообразие биологических систем, процессов и явлений привело к тому, что для их моделирования привлекаются представления различных математических дисциплин. При этом в зависимости от характера и свойств изучаемых процессов и явлений для моделирования выбирается аппарат либо дискретной, либо непрерывной математики.

Методы дискретной математики (например, алгебра, теория вероятностей) оказались наиболее удобным и естественным средством для моделирования свойств уникальных объектов, количественные характеристики которых меняются скачкообразно, без промежуточных стадий. Например, для описания возрастной структуры популяций применяется матричное исчисление.

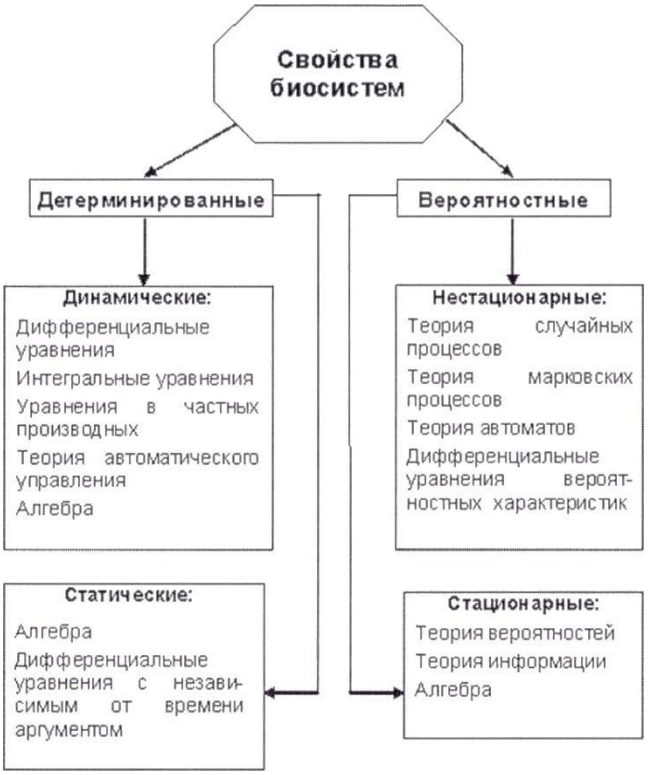


Рисунок 1 – Соотношение свойств биологических систем и используемых для их количественного описания математических методов

На теоретико-вероятностном подходе покоится все здание классической генетики. В последнее время для описания самых разнообразных классов биологических закономерностей все шире применяются более молодые отрасли математики, в частности теория конечных автоматов.

Когда поведение изучаемого объекта характеризуется непрерывными изменениями, адекватным средством моделирования являются методы непрерывной математики с их богатым аппаратом дифференциальных, интегральных и интегродифференциальных уравнений. Однако при определенных условиях эти методы можно применять и к процессам, дискретным по своей природе. В принципе процесс размножения микроорганизмов допустимо идеализированно рассматривать как непрерывный. Очень широко такие представления используются при моделировании процессов в экологии.

В целом детерминистский подход к моделируемым объектам, основанный на применении дифференциальных и интегральных уравнений, охватывает многие области современной теоретической биологии. На базе этого математического аппарата сформировалась новая научная дисциплина – биологическая кинетика. В ней на основе идей физической и химической кинетики строятся математические модели протекающих в клетке химических реакций и описываются многие аспекты клеточного метаболизма. Методы биологической кинетики наряду с теорией информации используются и в молекулярной генетике.

Однако возможна и несколько иная классификация математических моделей, предполагающая разбиение их на два класса.

1. Модели данных.
2. Модели систем.

Согласно этой классификации, статистика – основное средство для построения ***моделей данных,*** когда целью является поиск математической функции (полином, экспонента и др.), наиболее точно описывающей имеющийся набор экспериментальных данных. ***Модели систем*** основаны на конкретных гипотезах о структуре и физических принципах функционирования изучаемого явления. Такие модели предназначены для теоретического изучения механизмов явления или структуры системы, особенно если исследователь имеет дело с большой системой. Для построения теоретических моделей таких больших систем используются алгебраические, теоретико-множественные, логические, сетевые, графовые и другие подходы.

Текстовые модели представления знаний и данных наиболее широко применяются в биоинформатике для описания молекулярно-биологических и молекулярно-генетических данных (строение рибонуклеиновых кислот и белков, секвенирование ДНК, построение генных сетей и т. п.).

Применение некоторых информационных технологий в биологии. Компьютерная алгебра и биоинформатика. Говоря о биологическом приложении различных математических подходов, необходимо отметить в этом плане теорию информации. Попытки ее широкого использования в биологии, психологии и других дисциплинах, не имеющих прямого отношения к теории связи, относятся к началу 1950-х годов и связаны с именем Г. Кастлера. Однако к 1970-м годам появилось определенное разочарование в этом подходе за счет того, что статистическая теория информации фактически имеет дело только с количественной мерой информации, не выясняя глубинной природы самого этого явления, и начались попытки выхода за пределы исходных ограничений этой теории. Представляется перспективной разработка компьютерных технологий, в которых аппарат теории информации используется для оценки степени связи отдельных элементов биологической системы. В целом, наиболее успешно теоретико-информационный подход используется в молекулярной генетике и геномике.

Проблемы самоорганизации сложных динамических систем, к которым относятся и биологические, возникновения и ценности биологической/генетической информации рассматриваются в рамках достаточно молодой науки – синергетики (см., например, монографии Г. Хакена, Д. С. Чернавского).

Следует также вспомнить о **биологическом приложении теории нечетких множеств**, нашедшей широкое применение в дисциплинах медико-биологического профиля, где она

используется в прикладных разработках. Основное направление здесь – распознавание образов и таксономия, т. е. области исследований, в которых имеет место нечеткая кластеризация. Используется этот аппарат также в медицине как теоретическая основа для принятия решений в нечетких, размытых условиях. Весьма перспективным представляется этот подход и в нейробиологических исследованиях, особенно в связи с разработкой нейрокомпьютеров. Однако сочетание малого (недостаточного или уникального) количества экспериментальной информации и неконтролируемых (нечетких) условий, по нашему мнению, может служить критерием для применения этой теории не только в медицине, но и в экологии. Поэтому теория нечетких множеств может оказаться перспективной для описания самых разнообразных фундаментальных закономерностей живой природы.

В современной биологии нашли широкое применение не только теоретические, но и прикладные направления кибернетики и информатики. Дальнейшее развитие математического моделирования в биологии видится на пути применения современных средств компьютерной математики (или компьютерной алгебры, что означает то же самое) как инструмента подготовки высококвалифицированных специалистов, построения содержательных моделей, накопления, обработки и хранения информации, полученной в результате исследования этих моделей.

Литература: Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витебск. гос. ун-т имени П. М. Машерова, 2022. – 83 с.

Тема 2.1. Парное и множественное выравнивание. Алгоритм и программы выравнивания.

Вопросы для рассмотрения:

1. Выравнивание последовательностей.
2. Алгоритм Нидлмана-Вунша.

Краткое содержание лекции

Выравнивание последовательностей – это метод, основанный на размещении 2-х или более последовательностей мономеров ДНК, РНК или белков друг под другом таким образом, чтобы легко увидеть сходные участки в этих последовательностях.

Парное выравнивание используется для нахождения сходных участков 2-х последовательностей, различают (рисунок 2, 3):

– глобальное выравнивание предполагает, что последовательности гомологичны по всей длине, в глобальное выравнивание включаются обе входные последовательности целиком.

– локальное выравнивание применяется, если последовательности содержат как родственные (гомологичные), так и неродственные участки, результат локального выравнивания выбор участка в каждой из последовательностей и выравнивание между этими участками.

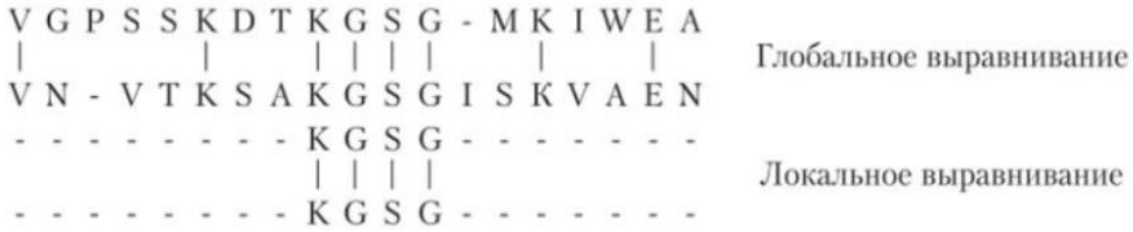


Рисунок 2 – Порядок глобального и локального выравнивания

Рисунок 3 – Парное выравнивание

Парное выравнивание последовательностей – это наиболее простой случай множественного выравнивания последовательностей, т.е. способ расположения нескольких последовательностей друг под другом путем внесения в них пропусков, чтобы одинаковые или близкие по своим свойствам мономеры, формировали столбцы данного выравнивания.

Множественное выравнивание последовательностей (англ. multiple sequence alignment, MSA) – выравнивание трех и более биологических последовательностей, обычно белков, ДНК или РНК. В большинстве случаев предполагается, что входной набор последовательностей имеет эволюционную связь. Используя множественное выравнивание, можно оценить эволюционное происхождение последовательностей, проведя филогенетический анализ (рисунок 4).

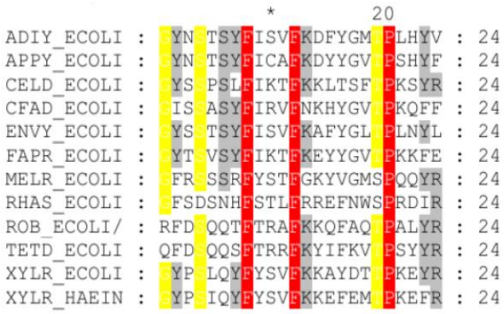


Рисунок 4 – Множественное выравнивание

Алгоритм Нидлмана-Вунша. Процент идентичности между двумя пептидными или нуклеотидными последовательностями является функцией количества аминокислот или нуклеотидных остатков, которые идентичны в двух последовательностях. Алгоритм Нидлмана-Вунша для глобального выравнивания последовательностей:

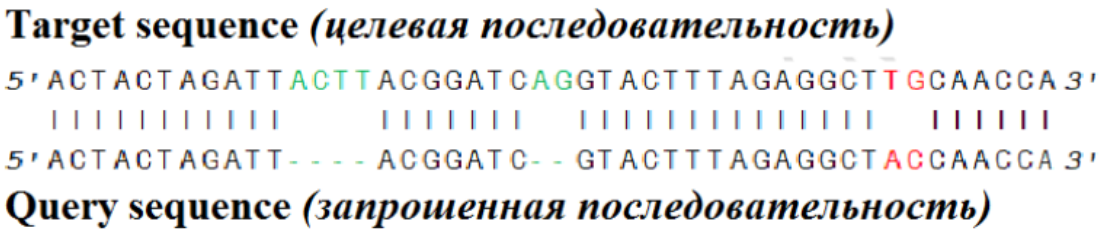


Рисунок 5 – Глобальное выравнивание последовательностей по алгоритму Нидлмана-Вунша

В указанной последовательности есть совпадения (match), они выделены черным цветом, красным цветом выделены участки, в которых нет совпадений (mismatch), зеленый – пробелы в последовательности (gap). Пробелы и несовпадения штрафуются (например, им присваивается значение -1). За каждый участок последовательности присваиваются веса (score):

Match +1 +5 Mismatch -1 -1 Gap -1 -2

Для заполнения каждой ячейки нужно оценить 3 веса (score), для первой клетки Y - E:

1) слева направо: gap (-2)
2) сверху вниз: gap (-2)
3) по диагонали mismatch (-1)

ИТОГО: выбираем максимальное значение -1
Для второй клетки E – E, есть совпадение поэтому:
• по диагонали mismatch (1)
• слева направо: gap (-2)
• сверху вниз: gap (-2)
Итого: выбираем максимальное значение -1.

В верхнем левом углу выставляется ноль, от которого производится заполнение таблицы выравнивания по 3-м направлениям: вправо, вниз и по диагонали. Движение по диагонали подразумевает, что мы продвинулись на один шаг. Движение вправо или вниз подразумевает, что мы делаем gap. Идентичные остатки определяются как остатки, которые являются одинаковыми в двух последовательностях в данной позиции выравнивания.

Процент идентичности последовательности рассчитывается из оптимального выравнивания путем взятия числа остатков, идентичного между двумя последовательностями, деления его на общее количество остатков в самой короткой последовательность и умножение на 100.

Оптимальное выравнивание – это выравнивание, при котором процент идентичности является максимально возможным. Разрывы могут быть введены в одну или обе последовательности в одном или нескольких положениях выравнивания, чтобы получить оптимальное выравнивание учитываются как неидентичные остатки для расчета процента идентичности последовательности.

Литература: Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витебск. гос. ун-т имени П. М. Машерова, 2022. – 83 с.

Тема 2.2. Эволюция молекул и организмов

Вопросы для рассмотрения:

1. Пути эволюции молекул и организмов.
2. Молекулярная эволюция, как наука.
3. Критерии сравнения нуклеотидных и белковых последовательностей

Краткое содержание лекции

Жизнь – это способ существования открытых коллоидных систем, обладающих свойствами самовоспроизведения, регуляции и обновления на основе преобразования потоков веществ, энергии и информации путем взаимодействия белков и нуклеиновых кислот. Решающую роль в превращении неживого вещества в живое играют белки. Они способны образовывать коллоидные гидрофильные комплексы, сливаться друг с другом и образовывать коацерваты – открытые системами, обладают упорядоченностью, способностью к самообновлению и поглощают вещества из окружающей среды.

Возникновение коацерватов привело к естественному отбору (движущий фактор биологической эволюции). Включение в состав коацерватов ионов металлов привело к образованию ферментов. В результате включения в состав коацерватов нуклеиновой кислоты и ферментов сформировались предбиологические системы, т.е. смеси ДНК и белка (ДНК способна мутировать, а белки – ускорять химические реакции). В результате произошел переход от химической к биологической эволюции. На границе между коацерватами и внешней средой появилась клеточная мембрана. С образованием мембраны появились протобионты – первые примитивные клетки.

В ходе эволюции наиболее вероятной последовательностью появления живых организмов является: анаэробные гетеротрофы → фотоавтотрофы аэробные гетеротрофы → автотрофы. Первые организмы – гетеротрофы (прокариоты), окаменелые остатки и следы

жизнедеятельности обнаружены в осадочных породах возрастом около 3,5 млрд. лет. Автотрофы возникли 3 млрд. лет назад (анаэробные бактерии, осуществляющие одностадийный фотосинтез).

Цианобактерии первые организмы, осуществившие 2-х стадийный фотосинтез с выделением кислорода. Постепенно атмосфера насытилась достаточным количеством кислорода (прекратилась химическая эволюция), появилась возможность кислородного типа обмена, что привело к появлению аэробов. Образование озонового экрана способствовало выходу организмов из водной среды на сушу.

2,5 млрд. лет назад появились протисты, а около 1,5 млрд. лет назад возникли многоклеточные организмы, которые усложнялись и сформировали типы животных и отделы растений.

Молекулярная эволюция – наука, изучающая изменения генетических макромолекул (ДНК, РНК, белков) в процессе эволюции, закономерности и механизмы этих изменений, а также реконструирующая эволюционную историю генов и организмов.

Объекты исследования молекулярной эволюции:

1. Последовательности НК как носителей генетической информации.
2. Последовательности белков.
3. Структура белков.
4. Геномы организмов.

Основными задачами молекулярной эволюции являются выявление закономерностей эволюции генетических макромолекул и реконструкция эволюционной истории генов и организмов. Молекулярная эволюция взаимосвязана с такими областями науки, как:

1) палеонтология (датировка эволюционных событий);
2) генетика (организация и передача наследственной информации);
3) молекулярная биология (строение генетических макромолекул);
4) эволюция (эволюционные закономерности);
5) биофизика (механизмы функционирования генетических макромолекул);
6) математика (построение моделей эволюции);
7) информатика (обработка и анализ данных);
8) биохимия (обмен нуклеиновых кислот и белков).

Разделами молекулярной эволюции как науки являются:

1. Эволюция макромолекул – изучает типы и скорости изменений, происходящих в генетическом материале (ДНК), а также созданных на его основе белков, и механизмов, ответственных за эти изменения.

2. Молекулярная филогения – изучает эволюционную историю макромолекул и организмов, получаемую на основе молекулярных данных. Эволюция макромолекул и молекулярная филогения тесно взаимосвязаны, и прогресс в одном из этих разделов способствует исследованиям в другом. Знание филогении нужно для определения последовательности изменений в изучаемых молекулах, а знание способов и темпов изменений изучаемой молекулы необходимо для восстановления эволюционной истории группы организмов.

3. Пребиотическая эволюция («происхождение жизни»), развитие этого раздела ограничивается тем, что в настоящее время неизвестны законы, направляющие процесс переноса информации в пребиотических системах.

Литература: Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витебск. гос. ун-т имени П. М. Машерова, 2022. – 83 с.

Тема 2.3. Гомология последовательностей. Ортологи и паралоги.

Вопросы для рассмотрения:

1. Понятие гомологии. Работы Оуэна Р. в сфере создания естественной системы видов.
2. Ортологи, паралоги: аллопаралоги, симпаралоги. Ксенология.

Краткое содержание лекции

При сравнительном описании любой формы или типа подобия используется термин «гомология» (от греч. ομολογειν – «согласующийся»). В биологии термин «гомология» используют для описания какого-либо подобия между органами, признаками, генами и геномами. Например, считают гомологичными крыло птицы и конечность крокодила, так как предполагается, что они представляют собой парные придатки, имеющие общий план строения, несмотря на их значительные структурные и функциональные различия.

Ричард Оуэном (Richard Owen, 1804–1892) предложил различать термины гомологию и аналогию для определения принципа, по которому можно ранжировать виды в естественной системе. Ученый исходил из того, что факта того, что конечности всех тетрапод имеют одинаковый план строения и общий план строения конечностей можно проследить, несмотря на различие их функций, таких, как ходьба, лазанье, плаванье, рытье или полет (рисунок 6).

Р. Оуэн описал и каталогизировал виды путем идентификации подобных частей у разных видов и их ранжированием по паттернам общих частей тела (т. е. гомологичным признакам), невзирая на их функцию. В результате возникла система, в которой гомологичные признаки являлись общими для совокупного набора видов («мутовки видов»). Р. Оуэн полагал, что гомологичные признаки идентичны из-за своего происхождения при сохранении общего плана строения тела, или архетипа. Архетип может быть понят как абстрактный или систематизирующий принцип живой природы. Согласно интерпретации Р. Оуэна, каждый вид есть частная реализация архетипа.

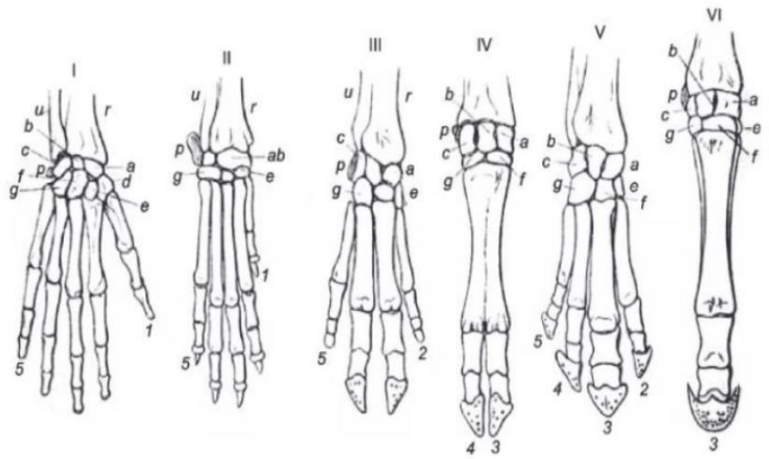


Рисунок 6 – Схема строения скелета с указанием гомологичных костей (из: Gegenbaur, 1870). I – человек, II – собака, III – свинья, IV – корова, V – тапир, VI – лошадь. *r – radius; u – ulna; a – scaphoid; B – lunare; c – triquetrum; d – trapezium; e – trapezoid; f – capitatum; g – hamatum; p – pisiforme.*

Из систематизирующего принципа архетипа Р. Оуэна следует, что между последовательностями ДНК, РНК, белков организмов близких рангов возможно установление степени родства путем их сравнения и оценки различий эволюционных или случайных отклонений. В БИ для определения значимости совпадений или различий введены свои определения для терминов сходство (подобие) и гомология: сходство – это наличие или измерение сходства или различия, независимо от источника сходства, гомология означает, что

последовательности или организмы, в которых они обнаружены, являются потомками общего предка.

О подобии последовательностей можно судить, проведя процедуру их выравнивания, а о гомологии организмов (или органов) – на основании наблюдаемого подобия. Здесь гомологии – это предположение, которое возникает из наблюдения подобия. Гомология между ДНК, РНК или белками обычно определяется по сходству их нуклеотидных или аминокислотных последовательностей. Вспомним, что выравнивания последовательностей используются, чтобы указать, какие участки каждой последовательности гомологичны. Значительное сходство является доказательством того, что две последовательности связаны эволюционными изменениями от общей предковой последовательности.

Причины гомологии: ортология, паралогия, аналогия

Два сегмента ДНК могут иметь общее происхождение из-за трех явлений: - ортологии – события видообразования, - паралогии – событие дублирования, - ксенологи – горизонтального (или латерального) переноса генов.

1. Термин «ортолог» придумал в 1970 году молекулярный эволюционист Уолтер Фитч. Гомологические последовательности являются ортологичными, если предполагается, что они произошли от одной и той же предковой последовательности, разделенной событием видообразования.

Ортологи – это гены у разных видов, которые произошли в результате вертикального спуска от одного гена последнего общего предка. Например, регуляторный белок гриппа растений есть и у растения Arabidopsis, и у хламидомонады Chlamydomonas. Вариант Chlamydomonas более сложен: он дважды пересекает мембрану, а не один раз, содержит дополнительные домены и подвергается альтернативному сплайсингу1. Однако он может полностью заменить гораздо более простой белок Arabidopsis, если перенести его из водорослей в геном растения с помощью генной инженерии. Значительное сходство последовательностей и общие функциональные домены указывают на то, что эти два гена являются ортологичными генами, унаследованными от общего предка (рисунок 7).

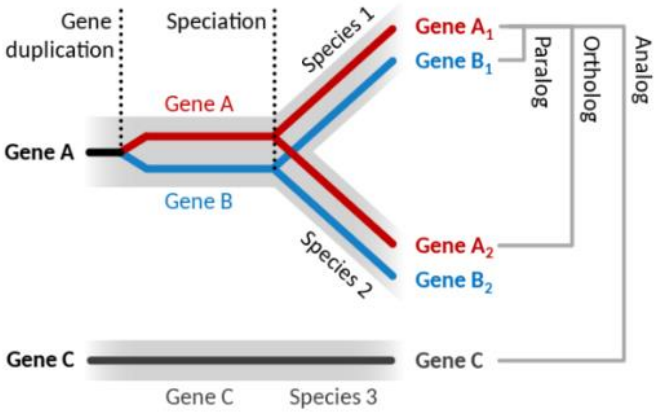


Рисунок 7 – Явления орто-, пара-, и аналогии

Примечание к рисунку: Вверху: наследственный ген дублируется, чтобы произвести два паралога (гены А и В). Событие видообразования приводит к появлению ортологов у двух дочерних видов. Внизу: у отдельного вида неродственный ген выполняет аналогичную функцию (Ген С), но имеет отдельное эволюционное происхождение и является аналогом.

Учитывая, что точное происхождение генов у разных организмов трудно установить из-за дупликации генов и событий перестройки генома, наиболее убедительные доказательства того, что два похожих гена являются ортологами, обычно обнаруживаются путем проведения филогенетического анализа происхождения генов. Ортологи часто, но не всегда, выполняют

одну и ту же функцию. Ортологические последовательности дают полезную информацию для таксономической классификации и филогенетических исследований организмов.

Паттерн генетической дивергенции может быть использован для отслеживания родства организмов. Два очень тесно связанных организма, вероятно, будут иметь очень похожие последовательности ДНК между двумя ортологами. Напротив, организм, который далее эволюционно отделен от другого организма, вероятно, будет демонстрировать большее расхождение в последовательности изучаемых ортологов.

2. Паралоги – это гены, которые связаны между собой посредством событий дупликации в последнем общем предке (Last common ancestor LCA) сравниваемых видов. Они возникают в результате мутации дублированных генов во время отдельных событий видообразования. Когда потомки от LCA имеют общие мутировавшие гомологи исходных дублированных генов, эти гены считаются паралогами.

Например, в LCA один ген (ген A) может быть продублирован, чтобы создать отдельный похожий ген (ген B), эти два гена будут продолжать передаваться последующим поколениям. Во время видообразования одна среда будет способствовать мутации в гене A (ген A1), создавая новый вид с генами A1 и B. Затем в отдельном событии видообразования одна среда будет благоприятствовать мутации в гене B (ген B1), приводящей к возникновению новый вид с генами A и B1 (рисунок 8).

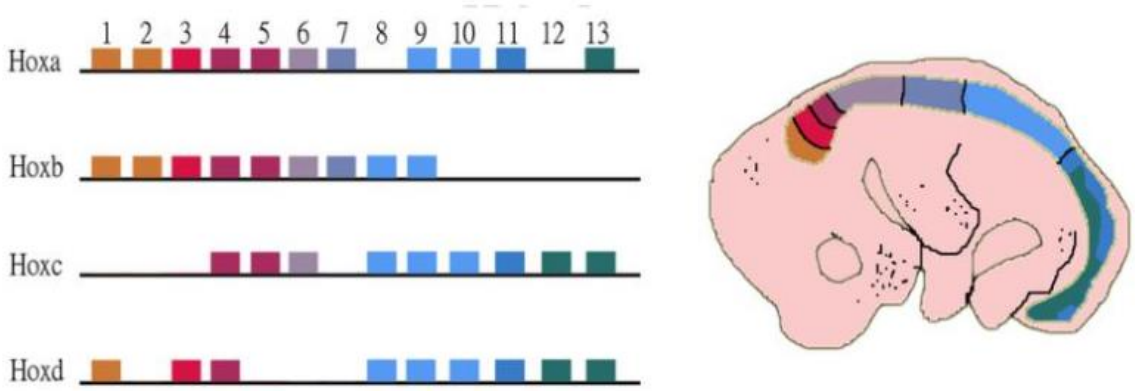


Рисунок 8 – Паралогия гена Hox

Гены потомков A1 и B1 паралогичны друг другу, потому что они являются гомологами, которые связаны посредством события дупликации у последнего общего предка двух видов. Дополнительные классификации паралогов включают:

– аллопаралоги (внешние паралоги),

– симпаралоги (внутренние паралоги).

Аллопаралоги – это паралоги, которые произошли от дупликаций генов, предшествовавших данному событию видообразования, они возникли в результате событий дублирования, которые произошли в LCA сравниваемых организмов.

Симпаралоги – это паралоги, возникшие в результате дублирования генов паралогов в последующих событиях видообразования. Если потомок с генами A1 и B претерпел другое событие видообразования, в котором дуплицировался ген A1, у нового вида были бы гены B, A1a и A1b – гены A1a и A1b являются симпаралогами.

Пример, гены Hox позвоночных организованы в наборы паралогов. Hox гены – это подмножество родственных гомеобоксных генов, задающих области плана тела в качестве зародыша вдоль головки хвоста оси животных. Каждый Hox-кластер (HoxA, HoxB и т.д.) находится на отдельной хромосоме.

Например, кластер HoxA человека находится на хромосоме 7, показанный на рисунке 4 кластер мыши HoxA имеет 11 паралоговых генов (2 отсутствуют).

Паралогичные гены могут формировать структуру целых геномов и, таким образом, в значительной степени объяснять эволюцию генома. Примеры включают гены Homeobox (Hox) у животных – эти гены претерпели не только дупликации генов в хромосомах, но и дупликации всего генома. В результате гены Hox у большинства позвоночных сгруппированы по множеству хромосом.

Другой пример – гены глобина, которые кодируют миоглобин и гемоглобин и считаются древними паралогами. Точно так же четыре известных класса гемоглобинов (гемоглобин A, гемоглобин A2, гемоглобин B и гемоглобин F) являются паралогами друг друга. Хотя каждый из этих белков выполняет одну и ту же основную функцию переноса кислорода, они уже немного разошлись по функциям: гемоглобин плода (гемоглобин F) имеет более высокое сродство к кислороду, чем гемоглобин взрослого человека.

3. Ксенология – возникновение гомологичных ДНК-последовательностей в геномах разных видов при «горизонтальном» (ненаследственном) переносе генов между организмами. Горизонтальный перенос происходит при физическом контакте клеток, обменивающихся генетическим материалом, т.е. в паразитарных, симбиотических, ассоциативных системах. Ксенологичные гены (ксенологи) обнаруживаются у филогенетически отдаленных, но территориально близких групп клеток или организмов.

В качестве носителей ксенологичной ДНК выступают ретровирусы, захватывающие фрагменты оттранслированной в РНК ДНК клеткихозяина одного вида и встраивающих эти последовательности в геном клеток-хозяев другого вида: плазмиды при конъюгации, бактериофаги при трансдукции, содержащаяся в среде свободная ДНК при трансформации.

Литература: Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витебск. гос. ун-т имени П. М. Машерова, 2022. – 83 с.

Тема 2.4. Биотехнология, генная инженерия и биобезопасность Горизонтальный перенос генов. Филогенетическое дерево и методы его построения.

Вопросы для рассмотрения:

1. Генетический перенос. Вертикальный и горизонтальный переносы генов.

2. Филогенетическое древо и методы его построения (UPGMA, NEIGHBOR-JOINING, MINIMAL EVOLUTION).

3. Биотехнология и генная инженерия. Генная инженерия микроорганизмов.

4. Перенос генов в растения. Перенос генов животных.

5. Биотехнология и биобезопасность.

Краткое содержание лекции

Генетический перенос – это передача генетической информации от одного организма другому. Это происходит при помощи двух механизмов: вертикального переноса генов и горизонтального переноса генов. *Вертикальный перенос генов* происходит, когда генетическая информация передается от одного поколения следующему, что происходит гораздо чаще, чем горизонтальный перенос генов. И половое, и бесполое размножение являются формами вертикального переноса генов, когда один или несколько организмов передают часть генома или весь свой геном потомству. Кроме того, вертикальный перенос генов происходит как у прокариотических, так и у эукариотических видов.

Горизонтальный перенос генов происходит, когда генетическая информация передается представителю того же поколения, и чаще всего встречается у прокариотических видов. В то время как среди эукариот межвидовой горизонтальный перенос генов чрезвычайно редок, у

прокариот он часто встречается. Горизонтальный перенос генов между различными видами – важный источник генетического разнообразия у прокариот.

Большинство прокариотических видов размножаются бесполым путем. Хотя это позволяет быстрее производить потомство, полученные потомки обладают ограниченным генетическим разнообразием. Горизонтальный перенос генов, таким образом, играет жизненно важную роль в обеспечении генетического разнообразия прокариот. Посредством горизонтального переноса генов прокариоты могут делиться небольшой частью своего генома с другими организмами, конспецифичными (принадлежащими тому же виду) или гетероспецифичными (принадлежащими другому виду), в одном поколении. Многие ученые полагают, что горизонтальный перенос генов и мутации являются наиболее значительными источниками генетической изменчивости прокариот. Таким образом, горизонтальный перенос генов обеспечивает некий исходный материал, на который действует естественный отбор. Ярким примером этого является появление устойчивых к антибиотикам штаммов бактерий. Гены, придающие устойчивость к антибиотику, могут передаваться между различными видами и штаммами бактерий, что дает бактериям-реципиентам избирательное преимущество, например, устойчивые к пенициллину штаммы *Neisseria gonorrhoeae*, вызывающие гонорею. Более того, по некоторым оценкам, по меньшей мере 18% генома *E. coli* было приобретено посредством горизонтального переноса генов в течение миллионов лет эволюции.

Филогенетика – наука, изучающая процессы образования биоразнообразия.

Кладистика – направление филогенетической систематики. Характерные особенности кладистической практики состоят в использовании так называемого кладистического анализа (строгой схемы аргументации при реконструкции родственных отношений между таксонами), строгом понимании монофилии и требовании взаимно-однозначного соответствия между реконструированной филогенией и иерархической классификацией. Кладистический анализ – основа большинства принятых в настоящее время биологических классификаций, построенных с учетом родственных отношений между живыми организмами. Кладистика относится к числу трех ведущих таксономических школ, доминирующих в современной биологической систематике. Ей противостоят фенетика, основанная на количественной оценке так называемого общего сходства, и эволюционная таксономия, которая, подобно кладистике, при построении системы опирается на эволюционную близость (то есть общность происхождения), однако не требует строгого соответствия системы и филогении (в частности, это выражается в признании права на существование в системе парафилетических групп).

Филогенетическое дерево (**эволюционное дерево**, **дерево жизни**) – дерево, отражающее эволюционные взаимосвязи между различными видами или другими сущностями, имеющими общего предка.

Вершины филогенетического дерева делятся на три класса: листья, узлы и (максимум один) корень. Листья – это конечные вершины, то есть те, в которые входят ровно по одному ребру; каждый лист отображает некоторый вид живых организмов (или иной объект, подверженный эволюции, например, домен белка). Каждый узел представляет эволюционное событие: разделение предкового вида на два или более, которые в дальнейшем эволюционировали независимо. Корень представляет общего предка всех рассматриваемых объектов. Ребра филогенетического дерева принято называть «ветвями» (см. рисунок 5).

Идея «дерева» появилась в ранних взглядах на жизнь, как на процесс развития от простых форм к сложным. Современные эволюционные биологи продолжают использовать деревья для иллюстрации эволюции, так как они наглядно показывают взаимосвязи между живыми организмами (рисунок 9).

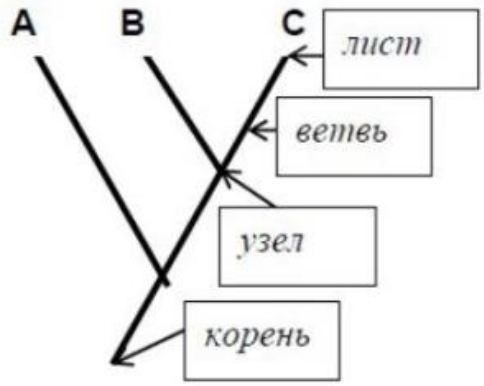


Рисунок 9 – Филогенетическое дерево

Примечание: листья – рассматриваемые объекты (A, B, C); узлы – точки схождения ветвей (узел, указанный стрелкой, объединяет объекты B и C, а нижележащий узел уже объединяет группу B+C и объект A); ветви – линии, соединяющие листья с узлами и узлы друг с другом; корень – узел, объединяющий все рассматриваемые объекты в одну группу.

Биотехнология и биобезопасность. Если экономическая выгода от использования ГМО в целом очевидна, то их безопасность по-прежнему вызывает жаркие споры, давно вышедшие за пределы лабораторий и научных форумов. Особенно это касается генетически модифицированных растений, бесконтрольное широкомасштабное использование которых может быть, в принципе, чревато неблагоприятными последствиями для окружающей среды и здоровья человека. К числу таких потенциальных опасностей мировое сообщество относит:

- разрушительное воздействие на биологические сообщества и утрату ценных биологических ресурсов в результате засорения местных видов генами, перенесенными от ГМО;
- создание новых более вредоносных паразитов, прежде всего сорняков, и усиление вредоносности уже существующих;
- выработку веществ, которые могут быть токсичными для организмов, не являющихся их мишенями, живущих или питающихся на ГМО;
- неблагоприятное воздействие на экосистемы токсичных веществ, производных неполного разрушения опасных химикатов (поскольку, как мы видели, большинство создаваемых в настоящее время ГМО – это формы, устойчивые к гербицидам).

Во избежание негативных последствий от бесконтрольного применения генетически модифицированных организмов 29 января 2000 г. в Монреале (Канада) более 130 стран приняли Протокол по биологической безопасности. Он называется ***«Картахенский протокол по биологической безопасности»*** по имени г. Картахена (Колумбия), который принимал участников Чрезвычайной конференции, подписавших в 1999 г. Конвенцию по биологическому разнообразию (КБР). Цель этого первого Протокола к КБР состояла в том, чтобы внести вклад в безопасную передачу живых модифицированных организмов, пересекающих международные границы, безопасное обращение с ними и их безопасное применение. К числу таких организмов относятся растения, животные и микробы, полученные с помощью методов генной инженерии. Кроме того, Протокол по биологической безопасности имеет своей целью предотвращение неблагоприятного воздействия на охрану природы и устойчивое использование биологического разнообразия без необходимого нарушения мировой торговли продовольственными товарами.

Наибольшее беспокойство вызывает вероятность переноса генетического материала трансгенного растения в геномы других дикорастущих или сельскохозяйственных организмов в результате их скрещиваний с ГМО. Для предотвращения таких ситуаций используют растения-самоопылители, изолируют посевы трансгенных растений, тщательно анализируют

возможные последствия такого переноса и вероятность фиксации трансгена в природных популяциях и т. д.

Другой «больной вопрос» – это опасность внедрения трансгенов в геномы почвенных микроорганизмов, организмов-симбионтов желудочно-кишечного тракта животных (в том числе человека) и, наконец, в геном самого человека. Человек, как и другие гетеротрофные организмы, постоянно сталкивается с огромным количеством чужеродной ДНК. Часть ее способна попадать в клетки человека. В частности, в 2003 г. немецкие исследователи из Кельна и Эрлангена опубликовали результаты экспериментов, в которых в рацион мышей добавляли препараты, содержащие маркерные ДНК, – их фрагменты были найдены в ядрах некоторых клеток эпителия желудка и кишечника, а также клеток крови (лейкоцитов), печени, почек и селезенки мышей. В других экспериментах, проведенных в университете Уппсала (Швеция) в клетках крови людей-добровольцев, которые питались ***приготовленным*** мясом кролика, были найдены небольшие фрагменты как геномной, так и митохондриальной ДНК кролика.

Пока еще никому не удалось обнаружить экспрессию проникших с пищей фрагментов генов или какие-то негативные последствия их присутствия. Однако ряд специалистов-биоинформатиков и молекулярных биологов считает, что в исследованиях по ГМО не дается оценка риска от неизученных пока механизмов действия РНК и ДНК. По их мнению, человек является самым уязвимым видом среди живых существ именно потому, что доля внегенной ДНК у него составляет до 98% генома, а механизмы действия малых и микро-РНК еще недостаточно изучены.

В последние годы появились публикации о связи РНК с канцерогенезом – как с развитием рака, так и с его предотвращением. С эффектами внеклеточных РНК и ДНК могут быть, по последним данным, связаны некоторые эндокринные нарушения при беременности (см. материалы Рабочего совещания «Оценка эффектов ГМО на здоровье человека», Вена, 23-24 ноября 2003 г.). В целом, по словам Эрика Нейманна, вице-президента по биоинформатике фирмы Beyond Genomics Inc.: «На самом деле мы имеем плохое представление о том, что ген на самом деле делает и где и когда ему это следовало бы делать. Вы можете понимать весь геном в целом и все же понимать менее 1% того, что происходит в клетке».

Опасения противников генетической трансформации растений по части непредсказуемого влияния трансгена на метаболизм и биохимию самого растения (новый ген может вызвать нарушение работы генов, что теоретически способно привести к синтезу метаболита с токсическим или онкогенным эффектом) заставили по-новому взглянуть даже на обычную селекцию. В частности, во многих схемах традиционной селекции используется химический или радиационный мутагенез, ведущий к появлению организмов с непредсказуемыми свойствами, проверка опасности которых не проводится. Как отметил новосибирский генетик А. В. Кочетов, «даже если мутагенез не используется, любой новый сорт представляет собой оригинальную уникальную комбинацию аллелей (природных вариантов генов), полученную при скрещивании различных (непохожих друг на друга) представителей данного вида.

Вероятность того, что в результате этих манипуляций в метаболизме растения нового сорта произойдут сдвиги с негативным эффектом, точно такая же, как при трансгенозе или даже выше. Однако в данном случае никто не обсуждает эту потенциальную опасность всерьез, поскольку селекция сельскохозяйственных растений практикуется человеком тысячи лет и рассматривается в качестве одного из ключевых достижений нашей цивилизации».

Национальный координационный центр биобезопасности. Продолжая обзор применения информационных технологий белорусскими биологами, следует отметить деятельность Национального координационного центра биобезопасности, который проводит большую работу по разъяснению реального положения дел в сфере создания и использования генетически модифицированных организмов. Центр организован на базе Института генетики и цитологии НАН Беларуси в 1998 г. для предоставления информации заинтересованным

министерствам и республиканским органам государственного управления, международным и национальным организациям по биобезопасности и средствам массовой информации по законодательству, научным исследованиям, полевым испытаниям, ввозе/вывозе и коммерческом использовании ГМО и их продуктов в Беларуси.

Среди прочих информационных материалов на сайте центра представлены разные национальные и международные документы, относящиеся к проблемам биобезопасности, содержатся ответы на наиболее часто задаваемые вопросы по биобезопасности и генно-инженерной деятельности, а также приведены ссылки на национальные и международные Web-сайты организаций и учреждений по биобезопасности, другие ссылки по вопросам биобезопасности и биотехнологии. Кроме того, с 2006 г. при центре действует хозрасчетная лаборатория по тестированию наличия генетически модифицированных компонентов в сельскохозяйственной продукции и продуктах питания, аккредитованная Госстандартом для проведения таких работ.

Литература: Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витебск. гос. ун-т имени П. М. Машерова, 2022. – 83 с.

Тема 3.1. Гены и регуляция их экспрессии.

Вопросы для рассмотрения:

1. Механизмы регуляции экспрессии генов прокариотов и эукариотов.
2. Модель оперона. Позитивная и негативная регуляции экспрессии генов.
3. Этапы процесса экспрессии генов.

Краткое содержание лекции

Ген – это участок ДНК, кодирующий всего один белок или РНК, кроме непосредственно кодирующей части, он также включает в себя регуляторные и иные структурные элементы, имеющие разное строение у прокариот и эукариот.

Экспрессия генов – процесс, в ходе которого наследственная информация от гена (последовательности нуклеотидов ДНК) преобразуется в функциональный продукт – РНК или белок. Некоторые этапы эксперессии генов могут регулироваться: это транскрипция, трансляция, спайсинг РНК и стадия посттрансляционных модификации белков. Процесс активации экспрессии генов короткими двуцепочечными РНК называется **активацией РНК.**

Регуляция экспрессии генов позволяет клеткам контролировать основой дифференцировки клеток, морфогенеза и адаптации. Экспрессия генов является субстратом для эволюционных изменений, так как контроль над временем, местом и количественными характеристиками экспрессии одного гена может иметь влияние на функции других генов в целом организме.

Механизм регуляции экспрессии генов.

Экспрессия генов – это реализация заложенной в них информации, то есть синтез РНК и белков. Другими словами, под экспрессией генов понимают их активность.

В клетках живых организмов экспрессия генов регулируется: одни гены могут быть реализованы, другие – нет. Причем регуляция может осуществляться на разных этапах: может выполняться или нет транскрипция, из пре-мРНК в результате альтернативного сплайсинга могут образовываться разные мРНК, может блокироваться трансляция и др.

У эукариот, обладающих отграниченным от цитоплазмы ядерным содержимым и более сложным геномом, регуляция экспрессии генов намного разнообразнее и сложнее, чем у прокариот.

У прокариот пока молекула РНК синтезируется на участке ДНК, она тут же может транслироваться (начиная с уже синтезированного конца). Поэтому у них регуляция экспрессии

(активности) генов осуществляется почти исключительно на уровне ДНК, так как в РНК часто невозможно внести какие-нибудь изменения до ее трансляции.

В 1961 г. Жакобом и Моно была предложена **модель оперона** как системы регуляции генов у бактерий. **Оперон состоит из промотора**, **оператора, структурных генов оперона (их может быть разное количество) и терминатора.** В области промотора прикрепляется фермент РНКполимераза. В области оператора присоединяется белок-репрессор, который кодируется отдельно отстоящим от оперона геном-регулятором (может быть сцеплен со своим опероном, а может находиться на расстоянии).

Если белок-репрессор соединяется с оператором, то транскрипция всех структурных генов оперона становится невозможной, так как РНКполимераза не может перемещаться по цепи ДНК.

В свою очередь активность белка-репрессора может блокироваться определенным для него низкомолекулярным соединением – индуктором (тем или иным питательным веществом бактерий). В результате взаимодействия с индуктором белок-репрессор видоизменяется и уже не может присоединиться к оператору своего оперона. В этом случае гены оперона экспрессируются (т.е. на них идет синтез).

Бывает обратная ситуация, когда индуктор активирует белок-репрессор.

Таким образом, в зависимости от того, какие индукторы находятся в цитоплазме, у прокариот экспрессируются те или иные генные группы.

Вышеописанный механизм экспрессии генов относится к **негативной регуляции**, так как гены транскрибируются, если они не выключены репрессором. И наоборот: не транскрибируются, если выключены.

Кроме негативной регуляции у бактерий существует также **позитивная.** В этом случае вместо белка-репрессора действие оказывает белок-активатор. На эти белки также действуют индукторы, активируя или инактивируя их.

Также у прокариот были выявлены опероны, которые актируются двумя регуляторными белками, соединенными друг с другом.

У многоклеточных организмов в клетках разных тканей экспрессируются разные гены, т. е. для **эукариот** характерна *дифференциальная экспрессия*.

У эукариот, также как и у прокариот, существуют регуляторные белки с похожим механизмом действия. При этом для эукариот не характерна регуляция по типу оперона. Цистроны (транскрибируемые участки) эукариот обычно содержат по одному гену (это не касается геномов хлоропластов и митохондрий).

Кроме регуляторных белков, взаимодействующих с ДНК, у эукариот существуют и другие способы регуляции экспрессии генов.

Конденсация и деконденсация хроматина. Это наиболее универсальный метод регуляции транскрипции. Когда нужно экспрессировать определенные гены, хроматин в этом месте деконденсируется.

Альтернативные промоторы. У гена может быть несколько промоторов, каждый из которых начинает транскрипцию с разных его экзонов в зависимости от типа клетки. В конечном итоге будут синтезированы разные белки.

Метилирование и деметилирование ДНК. Метилирование ДНК происходит в регуляторных областях гена. Метилируется цитозин в последовательности ЦГ, после чего ген инактивируется. При деметилировании активность гена восстанавливается. Процесс регулируется ферментом метилтрансферазой.

Гормональная регуляция. При гормональной регуляции гены активируются в ответ на внешний химический сигнал (поступление в клетку определенного гормона). Этот гормон запускает те гены, которые имеют специфические последовательности нуклеотидов в регуляторных областях.

Геномный импринтинг. Это малоизученный способ регуляции экспрессии генов у эукариот. Он возможен только у диплоидных организмов и выражается в том, что активность генов зависит, от какого из родителей они были получены. Выключение генов осуществляется путем метилирования ДНК.

Альтернативный сплайсинг. Это регуляция на уровне процессинга. При альтернативном сплайсинге порядок сшивки экзонов может быть различным. Отсюда следует, что на основе одной и той же нуклеотидной последовательности ДНК могут быть синтезированы разные белки. Хотя их отличие друг от друга будет в основном заключаться лишь в разных сочетаниях одних и тех же аминокислот.

Тканеспецифическое редактирование РНК также протекает на уровне процессинга. Выражается в замене отдельных нуклеотидов в РНК в определенных тканях организма.

Кроме того, у эукариот иРНК часто не подвергается процессингу вообще (а распадается) или подвергается с задержкой. Это также можно рассматривать как способ регуляции экспрессии генов.

Регуляция стабильности иРНК. У эукариот существует регуляция и на уровне трансляции, когда готовые иРНК не «допускаются» к рибосомам или разрушаются. Другие же иРНК могут дополнительно стабилизироваться для многократного использования.

Посттрансляционная модификация белка. Чтобы молекула полипептида превратилась в активную молекулу белка, в ней должны произойти различные модификации определенных аминокислот, должны быть сформированы вторичная, третичная и возможно четверичная структуры. На этом этапе также можно повлиять на реализацию генетической информации, например, не дав молекуле сформироваться.

Риборегуляторы. Были обнаружены РНК, выполняющие регуляторные функции путем ослабления работы отдельных генов.

Для высокоорганизованных животных отмечается существование надклеточного уровня регуляции экспрессии генов.

Генная экспрессия – это совокупность молекулярных механизмов реализации наследственной информации, благодаря которому, ген проявляет свой потенциал в конкретном фенотипическом признаке организма. Все этапы экспрессии генов протекают с использованием энергии и обслуживаются десятками разнообразных ферментов. Процесс экспрессии гена состоит из нескольких этапов:

транскрипция → процессинг → трансляциия → модификация → экспрессия

а) на основе гена ДНК синтезируется про-мРНК. Первый этап экспрессии называется «транскрипцией»;

б) крупная молекула про-мРНК подвергается «процессингу», в результате этого значительно уменьшается в размерах. Образуется «зрелая» мРНК, считывание информации с которой упрощается. Биологический смысл процессинга – облегчение доступа к генетической информации;

в) мРНК при участии тРНК «выбирает» необходимые аминокислоты, которые связываются на рибосоме в строго определенную последовательность полипептида. Процесс переноса информации с мРНК на полипептид называется «трансляцией»;

г) синтезированный полипептид подвергается «модификации» и превращается в активный белок;

д) функционируя, белок делает свой вклад в морфологический или функциональный признак (фенотип) клетки или организма. Это процесс называется «экспрессией».

В процессе транскрипции участвует не только смысловая часть гена, но и другие регуляторные и структурные части. Образуемая про-мРНК содержит все элементы, характерные для гена ДНК. Процессинг существенно модифицирует про-мРНК, которая превращается в мРНК и содержит намного меньше структурно-функциональных элементов. На основе мРНК трансляция создает молекулы совершенно другой природы – полипептиды, ничего не имеющие общего с нуклеиновыми кислотами и обладающими совершенно другими свойствами и организацией.

Модификация полипептидов приводит к еще одному природному явлению – появлению сложной пространственной организации молекулы белка. Происходит переход линейной информации ДНК и РНК в пространственную организацию протеина, которая, в свою очередь, является основой специфического пространственного взаимодействия молекул в живом организме, что и лежит в основе жизни и всех жизненных явлений. В данном случае процесс модификации обеспечивает пространственную организацию – объединение четырех субъединиц гемоглобина в единый комплекс. В результате всех этапов экспрессии проявляется признак – способность к транспорту газов (O_2 и CO_2).

Оперон – это последовательность специальных, функциональных сегментов ДНК, а также структурных генов, которые кодируют и регулируют синтез определенной группы белков одной метаболической цепи, например, ферментов гликолиза. *Оперон (регулируемая единица транскрипции)* состоит из следующих структурных частей (специальных последовательностей нуклеотидов) (рисунок 10):

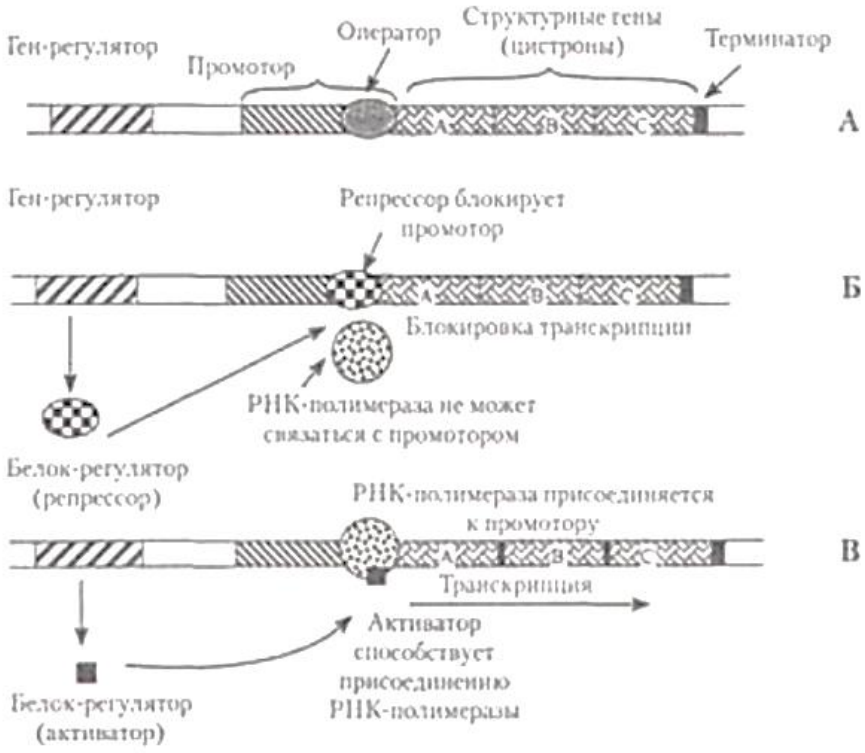


Рисунок 10 – Структурные части оперона

1. Ген-регулятор, контролирующий образование белка-регулятора.

2. Промотор – участок ДНК, к которому присоединяется РНКполимераза и начинается транскрипция.

3. Оператор – участок промотора, связывающий белок-регулятор.

4. Структурные гены (цистроны) – участки ДНК, кодирующие мРНК конкретных белков.

5. Терминаторный участок ДНК несет сигнал об остановке транскрипции.

Литература: Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витебск. гос. ун-т имени П. М. Машерова, 2022. – 83 с.

Тема 3.2. Анализ последовательностей ДНК

Вопросы для рассмотрения:

1. Значение анализа последовательностей ДНК.
2. Секвенаторы. Автоматическое секвенирование ДНК.
3. Секвенирование ДНК по Сэнгеру Ф. Секвенирование ДНК по Максаму и Гилберту: метод химической деградации.
4. Проект «Геном человека».

Краткое содержание лекции

Несмотря на научный прогресс, ДНК-анализ остается технологически сложной процедурой, ведь речь идет об исследовании микроскопических структур, состоящих из миллионов «букв», последовательность которых следует установить. Это называется секвенированием ДНК.

Для того чтобы «прочитать» молекулу, ее для начала нужно выделить, затем многократно скопировать, а после – «нарезать» на небольшие кусочки, удобные для анализа. Азотистые основания, составляющие «алфавит» генетического кода, при этом окрашиваются особым флуоресцентным красителем, благодаря которому в ходе дальнейшего анализа их можно будет распознать при просвечивании лазером.

Некоторые современные методы секвенирования ДНК напоминают компьютерную томографию, в ходе которой одна-единственная нить ДНК проходит через нанопору, а компьютер фиксирует изменения ионного тока в единицу времени, что позволяет распознать каждую «букву» генетического кода.

Технологии продолжают совершенствоваться, что позволяет ускорить и удешевить процесс ДНК-анализа. Это достигается за счет одномоментного «чтения» сразу нескольких участков ДНК, разработки нового программного обеспечения и усложнения конструкции приборов, предназначенных для автоматического секвенирования, - секвенаторов.

Так, с момента появления первых секвенаторов в 90-е годы минувшего столетия ученым удалось снизить стоимость ДНК-анализа в 13 раз.

Наибольшую известность получили следующие методики секвенирования ДНК:

– метод Сэнгера (метод терминации цепи);
– пиросеквенирование (секвенирование путем синтеза);
– секвенирование на основе лигирования;
– секвенирование ДНК одиночных молекул;
– нанопоровое секвенирование (тот самый «ДНК-томограф»).

Что показывает анализ ДНК? Возможности ДНК-анализа выходят за пределы подтверждения уже установленного другими методами диагноза - с его помощью можно узнать о заболеваниях, которые еще никак себя не проявили, но при стечении определенных обстоятельств могут серьезно пошатнуть здоровье пациента.

С помощью анализа ДНК определяется:

– наследственная предрасположенность к конкретным патологиям, которые уже встречались в семье (например, раку или психическим заболеваниям);

– общий «генетический анамнез» человека, который желает знать наверняка, какие заболевания могут возникнуть у него в будущем;

– причина неясных симптомов в отсутствии возможности поставить диагноз иным путем (особенно актуально для детей с редкой генетической патологией);

– индивидуальная непереносимость определенных лекарственных препаратов;

– степень генетического родства с предполагаемыми членами семьи;

– вероятность осложнений во время беременности;

– склонность к алкоголизму или наркомании (на основании выявления генов, ответственных за синтез ферментов, способных расщеплять алкоголь и другие соединения);

– риски при наличии серьезных физических нагрузок (важно для профессиональных спортсменов);

– возможные причины бесплодия и т.д.

Благодаря ДНК-анализу будущие родители могут оценить вероятность рождения ребенка с наследственной патологией, а в случае, если заболевание уже выявлено у малыша, - с первых месяцев его жизни выработать оптимальный план лечения, который позволит избежать осложнений.

Сроки и стоимость ДНК-анализа пропорциональны объему исследования: чем больше информации вы хотите получить, тем больший объем ДНК предстоит «прочитать» специалистам, сопоставив полученные данные с имеющимися сведениями о функциях обнаруженных генов.

Срок проведения анализа составляет в среднем 3–4 недели - в зависимости от того, где проводится тест (образцы могут быть отправлены в другую лабораторию, где есть соответствующее оборудование), какие реактивы потребуются для его проведения и каков объем предполагаемой работы.

Наиболее широко используемый метод для ДНК-анализа – секвенирование по Сангеру.

В основе метода секвенирования ДНК, разработанного Сэнгером и соавт., называемого также методом секвенирования путем терминации цепи, лежал принцип ферментативного построения комплементарной цепи ДНК по существующей одноцепочечной матрице при происходящем в разных местах цепи ДНК ингибировании ее дальнейшего роста.

Первый метод секвенирования ДНК, предложенный Ф. Сэнгером и А. Коулсоном в 1975 г, основан на ферментативных реакциях и носит название плюс-минус -метод. Данный подход предполагает выделение одноцепочечного фрагмента ДНК, соответствующего исследуемому участку генома. Этот фрагмент используют затем в реакции полимеразного копирования в качестве матрицы, а в качестве праймера - синтетические олигонуклеотиды или природные субфрагменты, получаемые после гидролиза определенными рестриктазами.

В настоящее время существует множество вариантов метода Сэнгера. Главное, этот метод удалось полностью автоматизировать. Так, например, при секвенировании ДНК по Сэнгеру на 5-конец праймера вводят флуоресцентные метки, причем для каждого из четырех анализируемых нуклеотидов используются флуоресцирующие агенты с различными спектральными характеристиками. После электрофоретического разделения гель сканируется при четырех различных длинах волн и полученная информация сразу обрабатывается на ЭВМ. При этом все биохимические операции также проводятся роботом.

Хотя сам принцип специфической терминации, положенный в основу первоначального метода секвенирования ДНК с помощью дидезокситерминаторов, остался неизменным, само секвенирование ДНК по Сэнгеру все же стало в значительной степени другим.

Наиболее широко сейчас применяется метод ферментативного секвенирования, или метод секвенирования путем терминации (остановки синтеза) цепи, предложенный Ф. Сэнгером в 1977 г.

Секвенирование ДНК по Сэнгеру. Это наиболее употребительный метод секвенирования ДНК, поскольку по сравнению с химическим методом он позволяет

анализировать более крупные фрагменты ДНК с меньшей вероятностью ошибки. Секвенируемую ДНК сначала клонируют в одноцепочечный бактериофаговый вектор. Чаще всего используют фаг М13 и его производные. Одноцепочечная ДНК взаимодействует как субстрат с ДНК-полимеразой, которая точно копирует первую цепь. Однако если нормальный дезокси-нуклеозидфосфат заменить его аналогомдидезоксинуклеозидфосфатом, то дальнейшее функционирование полимеразы прекращается и рост синтезируемой цепи останавливается.

Для инициирования этого процесса используют химически синтезированный универсальный нуклеотидный праймер. Реакцию проводят в присутствии четырех аналогов dNTP, один из которых содержит метку Р. Исходно имеется четыре реакционные смеси, в каждой из которых понижена концентрация одного из аналогов меченого нуклеотида. Поэтому в результате реакции получаются смеси меченных радионуклидами фрагментов ДНК, один конец у которых одинаков, но при этом они имеют разную длину и разные основания на другом конце. После инкубации ДНК в каждой смеси превращается в одноцепочечную форму. Затем смеси подвергают электрофорезу в полиакриламидном геле. Длину и последовательность фрагментов ДНК можно считывать непосредственно с гелевого слоя. Этот метод позволяет секвенировать в одной серии реакции от 250 до 500 пар оснований. Полученные данные часто анализируют с помощью компьютера.

Ключевым моментом ферментативного метода секвенирования ДНК, разработанного Сэнгером и соавт., является образование специфически терминированных меченых фрагментов вновь синтезированной ДНК. Такая терминация построения комплементарной цепи ДНК происходит при включении ДНК-полимеразой в растущую цепь ДНК-модифицированных аналогов природных.

В ферментативном методе секвенирования ДНК по Сэнгеру каждая дорожка соответствует только одному конкретному типу оснований и поэтому для правильного прочтения радиоавтографа секвенирующего геля необходимо следить, чтобы какая-либо полоса содержалась только в одной дорожке, поскольку в противном случае точное установление данного нуклеотида будет затруднено или скорее просто невозможно. В методе Сэнгера нет необходимости сопоставления наличия полос пуриновых или пиримидиновых оснований в одной или двух дорожках, то можно считать, что чтение радиоавтографа секвенирующего геля при этом будет проще.

Однако, существует ряд ограничений, из-за которых метод секвенирования ДНК гибридизацией так и не стал основным. Главными препятствиями этому служат серьезные проблемы при секвенировании повторяющихся элементов генома и чтение относительного малого числа нуклеотидов в ходе одного эксперимента.

Секвенирование ДНК по Максаму и Гилберту: метод химической деградации. Метод основан на специфической химической деградации фрагмента ДНК, радиоактивно меченного с одного конца. Препарат меченой ДНК разделяют на четыре аликвоты и каждую обрабатывают реагентом, модифицирующим одно или два из четырех оснований. А. Максам и У. Гилберт предложили модифицировать пуриновые основания диметилсульфатом. При этом происходит метилирование адениновых остатков по азоту в положении 3, а гуаниновых – по азоту в положении 7. Обработка образца ДНК соляной кислотой при 0 ^{0}C приводит к вьпщеплению метиладенина. Последующая инкубация при температуре 90 °С в щелочной среде вызывает разрыв сахарно- фосфатной цепи ДНК в местах выщепления оснований. Обработка пиперидином приводит к гидролизу образца по остаткам метилгуанина. Пиримидиновые основания модифицируют гидразином. Если реакцию вести в бессолевой среде, то модифицируются как цитозин, так и тимидин; если обработку вести в присутствии 2М NaCl, то модифицируется лишь цитозин. Расщепление цепи ДНК на фрагменты и в этом случае осуществляется пиперидином. Условия реакций авторы подбирали таким образом, чтобы в итоге получить полный набор субфрагментов разной длины. Последующий

электрофорез в полиакриламидном геле позволяет восстановить полную структуру исследуемого фрагмента.

Проект «Геном человека»

По материалам статьи: https://www.krugosvet.ru/enc/nauka_i_tehnika/biologiya/GENOM_CHELOVEKA.html

ГЕНОМ ЧЕЛОВЕКА – международная программа, конечной целью которой является определение нуклеотидной последовательности (секвенирование) всей геномной ДНК человека, а также идентификация генов и их локализация в геноме (картирование). Исходная идея проекта зародилась в 1984 среди группы физиков, работавших в Министерстве энергетики США и желавших заняться другой задачей после завершения работ в рамках ядерных проектов.

В 1988 Объединенный комитет, куда входили Министерство энергетики США и Национальные институты здоровья, представили обширный проект, в задачи которого – помимо секвенирования генома человека – входило всестороннее изучение генетики бактерий, дрожжей, нематоды, плодовой мушки и мыши (эти организмы широко использовались в качестве модельных систем в изучении генетики человека). Кроме того, предусматривался детальный анализ этических и социальных проблем, возникающих в связи с работой над проектом. Комитету удалось убедить Конгресс выделить на проект 3 млрд. долларов (один нуклеотид ДНК – за один доллар), в чем немалую роль сыграл ставший во главе проекта Нобелевский лауреат Дж.Уотсон. Вскоре к проекту присоединились другие страны (Англия, Франция, Япония и др.). В России в 1988 с идеей секвенирования генома человека выступил академик А.А.Баев, и в 1989 в нашей стране был организован научный совет по программе «Геном человека».

В 1990 была создана Международная организация по изучению генома человека (HUGO), вице-президентом которой в течение нескольких лет был академик А.Д.Мирзабеков. С самого начала работ по геномному проекту ученые договорились об открытости и доступности всей получаемой информации для его участников независимо от их вклада и государственной принадлежности. Все 23 хромосомы человека были поделены между странами-участницами. Российские ученые должны были исследовать структуру 3-й и 19-й хромосом. Вскоре финансирование этих работ в нашей стране было урезано, и реального участия в секвенировании Россия не принимала. Программа геномных исследований в нашей стране была полностью перестроена и сконцентрирована на новой области – биоинформатике, которая пытается с помощью математических методов понять и осмыслить все, что уже расшифровано.

Закончить работу предполагалось через 15 лет, т.е. примерно к 2005. Однако скорость секвенирования с каждым годом возрастала, и если в первые годы она составляла несколько миллионов нуклеотидных пар за год по всему миру, то на исходе 1999 частная американская фирма «Celera», возглавляемая Дж.Вентером (J.Venter), расшифровывала не менее 10 млн. нуклеотидных пар в сутки. Этого удалось достичь благодаря тому, что секвенирование осуществляли 250 роботизированных установок; они работали круглосуточно, функционировали в автоматическом режиме и сразу же передавали всю информацию непосредственно в банки данных, где она систематизировалась, аннотировалась и становилась доступной ученым всего мира. Кроме того, фирма «Celera» широко использовала данные, полученные в рамках Проекта другими его участниками, а также разного рода предварительные данные. 6 апреля 2000 состоялось заседание Комитета по науке Конгресса США, на котором Вентер заявил, что его компания завершила расшифровку нуклеотидной последовательности всех существенных фрагментов генома человека и что предварительная работа по составлению нуклеотидной последовательности всех генов (предполагалось, что их 80 тыс. и что они содержат примерно 3 млрд. нуклеотидов) будет завершена через 3–6 недель, т.е. гораздо раньше, чем планировалось.

Доклад был сделан в присутствии представителя HUGO, крупнейшего специалиста по секвенированию д-ра Р.Уотерсона. Расшифрованный фирмой «Celera» геном принадлежал анонимному мужчине, т.е. содержал как X-, так и Y-хромосомы, а HUGO использовали в своих исследованиях материал, полученный от разных людей. Между Вентером и HUGO велись переговоры о совместной публикации результатов, однако они закончились безрезультатно из-за разногласий по поводу того, что считать завершением расшифровки генома. По мнению компании «Celera», об этом можно говорить лишь в том случае, если гены полностью секвенированы и известно, как расшифрованные сегменты располагаются в молекуле ДНК. Этому требованию удовлетворяли результаты «Celera», в то время как результаты HUGO не позволяли однозначно определить взаимное положение расшифрованных участков. В результате в феврале 2001 в специальных выпусках двух авторитетнейших научных журналов, «Science» и «Nature», были раздельно опубликованы результаты исследований «Celera» и HUGO и приведены полные нуклеотидные последовательности генома человека, охватывающие около 90% его длины.

ОБЩЕБИОЛОГИЧЕСКОЕ ЗНАЧЕНИЕ ИССЛЕДОВАНИЙ, ПРОВЕДЕННЫХ В РАМКАХ ПРОЕКТА.

Исследования генома человека «потянули» за собой секвенирование геномов огромного числа других организмов, гораздо более простых; без геномного проекта эти данные были бы получены гораздо позже и в гораздо меньшем объеме. Их расшифровка ведется все возрастающими темпами. Первым крупным успехом стало полное картирование в 1995 генома бактерии Haemophilus influenzae, позже были полностью расшифрованы геномы более 20 бактерий, среди которых – возбудители туберкулеза, сыпного тифа, сифилиса и др. В 1996 картировали геном первой эукариотической клетки (клетки, содержащей оформленное ядро) – дрожжевой, а в 1998 впервые секвенировали геном многоклеточного организма – круглого червя Caenorhabolits elegans (нематоды). Завершена расшифровка генома первого насекомого – плодовой мушки дрозофилы и первого растения – арабидопсиса. У человека уже установлено строение двух самых маленьких хромосом – 21-й и 22-й. Все это создало основы для создания нового направления в биологии – сравнительной геномики.

Знание геномов бактерий, дрожжей и немтоды дает биологам-эволюционистам уникальную возможность сравнения не отдельных генов или их ансамблей, а целиком геномов. Эти гигантские объемы информации только начинают осмысливаться, и нет сомнения, что нас ждет появление новых концепций в биологической эволюции. Так, многие «личные» гены нематоды, в отличие от генов дрожжей, скорее всего связаны с межклеточными взаимодействиями, характерными именно для многоклеточных организмов. У человека генов только в 4–5 раз больше, чем у нематоды, следовательно, часть его генов должна иметь «родственников» среди известных теперь генов дрожжей и червя, что облегчает поиск новых генов человека. Функции неизвестных генов нематоды изучать гораздо проще, чем у аналогичных генов человека: в них легко вносить изменения (мутации) или выводить их из строя, одновременно прослеживая изменения свойств организма. Выявив биологическую роль генных продуктов у червя, можно экстраполировать эти данные на человека. Другой подход – подавление активности генов с помощью особых ингибиторов и отслеживание изменений в поведении организма.

Весьма интересным представляется вопрос о соотношении кодирующих и некодирующих областей в геноме. Как показывает компьютерный анализ, у C.elegans примерно равные доли – 27 и 26% соответственно – занимают в геноме экзоны (участки гена, в которых записана информация о структуре белка или РНК) и интроны (участки гена, не несущие подобной информации и вырезаемые при образовании зрелой РНК). Остальные 47% генома приходится на повторы, межгенные участки и т.д., т.е. на ДНК с неизвестными функциями. Сравнив эти данные с дрожжевым геномом и геномом человека, мы увидим, что доля кодирующих участков в расчете на геном в ходе эволюции резко уменьшается: у дрожжей она

очень высока, у человека очень мала. Налицо парадокс: эволюция эукариот от низших форм к высшим сопряжена с «разбавлением» генома – на единицу длины ДНК приходится все меньше информации о структуре белков и РНК и все больше информации «ни о чем», на самом деле просто непонятой и непрочитанной нами. Много лет назад Ф.Крик, один из авторов «двойной спирали» – модели ДНК, – назвал эту ДНК «эгоистической», или «мусорной». Возможно, какая-то часть ДНК человека действительно относится к такому типу, однако теперь ясно, что основная доля «эгоистической» ДНК сохраняется в ходе эволюции и даже увеличивается, т.е. почему-то дает ее обладателю эволюционные преимущества. Никаких объяснений такого феномена в настоящее время не существует, и без детального анализа нуклеотидных последовательностей геномных ДНК их дать невозможно.

Еще один важный результат, имеющий общебиологическое (и практическое) значение – вариабельность генома. Вообще говоря, геном человека высококонсервативен. Мутации в нем могут либо повредить его, и тогда они приводят к тому или иному дефекту или гибели организма, либо оказаться нейтральными. Последние не подвергаются отбору, поскольку не имеют фенотипического проявления. Однако они могут распространяться в популяции, и если их доля превышает 1%, то говорят о полиморфизме (многообразии) генома. В геноме человека очень много участков, различающихся всего одним-двумя нуклеотидами, но передающихся из поколения в поколение. С одной стороны, этот феномен мешает исследователю, поскольку ему приходится разбираться, имеет ли место истинный полиморфизм или это просто ошибка секвенирования, а с другой – создает уникальную возможность для молекулярной идентификации отдельного организма. С теоретической точки зрения вариабельность генома создает основу генетики популяций, которая ранее основывалась на чисто генетических и статистических данных.

ПРАКТИЧЕСКИЕ ПРИЛОЖЕНИЯ.

Самые большие надежды и ученые, и общество возлагают на возможность применения результатов секвенирования генома человека для лечения генетических заболеваний. К настоящему времени в мире идентифицировано множество генов, ответственных за многие болезни человека, в том числе и такие серьезные, как болезнь Альцгеймера, муковисцидоз, мышечная дистрофия Дюшенна, хорея Гентингтона, наследственный рак молочной железы и яичников. Структуры этих генов полностью расшифрованы, а сами они клонированы. Еще в 1999 была установлена структура 22-й хромосомы и определены функции половины ее генов. С дефектами в них связано 27 различных заболеваний, в том числе шизофрения, миелолейкоз и трисомия 22 – вторая по распространенности причина спонтанных абортов. Самым эффективным способом лечения таких больных была бы замена дефектного гена здоровым. Для этого, во-первых, необходимо знать точную локализацию гена в геноме, а во-вторых – чтобы ген попал во все клетки организма (или хотя бы в большинство), а это при современных технологиях невозможно. Кроме того, даже попавший в клетку нужный ген мгновенно распознается ею как чужой, и она пытается избавиться от него. Таким образом, «вылечить» удается только часть клеток и только на время. Еще одно серьезное препятствие на пути применения генной терапии – мультигенная природа многих заболеваний, т.е. их обусловленность более чем одним геном. Итак, массового применения генной терапии в ближайшем будущем вряд ли стоит ожидать, хотя успешные примеры такого рода уже есть: удалось добиться существенного облегчения состояния ребенка, страдающего тяжелым врожденным иммунодефицитом, путем введения ему нормальных копий поврежденного гена. Исследования в этой области ведутся по всему миру, и, может быть, успехи будут достигнуты раньше, чем предполагается, как это и произошло с секвенированием генома человека.

Еще одно важное применение результатов секвенирования – идентификация новых генов и выявление среди них тех, которые обусловливают предрасположенность к тем или иным заболеваниям. Так, есть данные о генетической предрасположенности к алкоголизму и наркомании, открыто уже семь генов, дефекты в которых приводят к токсикомании. Это

позволит проводить раннюю (и даже пренатальную) диагностику заболеваний, предрасположенность к которым уже установлена.

Широкое применение несомненно найдет и еще один феномен: обнаружилось, что разные аллели одного гена могут обусловливать разные реакции людей на лекарственные препараты. Фармацевтические компании планируют использовать эти данные для производства лекарств, предназначенных разным группам пациентов. Это поможет избежать побочных эффектов терапии, снизить миллионные затраты. Возникает целая новая отрасль – фармакогенетика, которая изучает, как те или иные особенности строения ДНК могут повлиять на эффективность лечения. Появятся совершенно новые подходы к созданию лекарственных средств, основанные на открытии новых генов и изучении их белковых продуктов. Это позволит перейти от неэффективного метода «проб и ошибок» к целенаправленному синтезу лекарственных веществ.

Важный практический аспект вариабельности генома – возможность идентификации личности. Чувствительность методов «геномной дактилоскопии» такова, что достаточно одной капли крови или слюны, одного волоса, чтобы с абсолютной достоверностью (99,9%) установить родственные связи между людьми. После секвенирования генома человека этот метод, использующий теперь не только специфические маркеры в ДНК, но и однонуклеотидный полиморфизм, станет еще более надежным. Вариабельность генома породила направление геномики – этногеномику. Этнические группы, населяющие Землю, имеют некоторые групповые генетические признаки, характерные для данного этноса. Получаемая информация в ряде случаев может подтвердить или опровергнуть те или иные гипотезы, циркулирующие в рамках таких дисциплин, как этнография, история, археология, лингвистика. Еще одно интересное направление – палеогеномика, занимающаяся исследованием древней ДНК, извлеченной из останков, найденных в могильниках и курганах.

ПРОБЛЕМЫ.

Финансирование «геномной гонки» и участие в ней тысяч специалистов основывались прежде всего на постулате, что расшифровка нуклеотидной последовательности ДНК сможет решить фундаментальные проблемы генетики. Оказалось, однако, что лишь 30% генома человека кодируют белки и участвуют в регуляции действия генов в ходе развития. Каковы функции остальных участков ДНК и есть ли они вообще – остается совершенно неясным. Около 10% генома человека составляют так называемые Alu-элементы длиной 300 п.н. Они появились неизвестно откуда в ходе эволюции у приматов, и только у них. Попав к человеку, они размножались до полумиллиона копий и распределились по хромосомам самым причудливым образом, то образуя сгустки, то прерывая гены.

Другая проблема касается самих кодирующих участков ДНК. При чисто молекулярно-компьютерном анализе возведение этих участков в ранг генов требует соблюдения сугубо формальных критериев: есть в них знаки пунктуации, необходимые для прочитывания информации, или нет, т.е. синтезируется ли на них конкретный генный продукт и что он собой представляет. В то же время роль, время и место действия большинства потенциальных генов пока неясны. По мнению Вентера, для определения функций всех генов может потребоваться не меньше ста лет.

Далее необходимо договориться, что вкладывать в само понятие «геном». Часто под геномом понимается лишь генетический материал как таковой, однако с позиции генетики и цитологии его составляет не только структура элементов ДНК, но и характер связей между ними, который определяет, как гены будут работать и как пойдет индивидуальное развитие при определенных условиях среды. И наконец, нельзя не упомянуть о феномене так называемой «неканонической наследственности», привлекшем к себе внимание в связи с эпидемией «коровьего бешенства». Эта болезнь стала распространяться в Великобритании в 1980-х годах после того, как в корм коровам стали добавлять переработанные головы овец, среди которых встречались овцы, больные скрэпи (нейродегенеративное заболевание). Сходная болезнь стала

передаваться людям, употреблявшим в пищу мясо больных коров. Обнаружилось, что инфекционным агентом являются не ДНК или РНК, а белки прионы (от англ. prions, protein infections particles, белковые инфекционные частицы). Проникая в клетку-хозяина, они изменяют конформацию нормальных белков-аналогов. Феномен прионов обнаружен также у дрожжей.

Таким образом, попытка представить расшифровку генома как чисто научно-техническую задачу несостоятельна. А между тем такой взгляд широко пропагандируется даже весьма авторитетными учеными. Так, в книге Код кодов (The Code of Codes, 1993) У.Гилберт, открывший один из методов секвенирования ДНК, рассуждает о том, что определение нуклеотидной последовательности всей ДНК человека приведет к изменениям в наших представлениях о самих себе. «Три миллиарда пар оснований могут быть записаны на одном компакт-диске. И любой может вытащить из кармана свой диск и сказать: «Вот он – Я!» Между тем необходимо знать не только порядок следования звеньев в цепи ДНК и не только взаимное расположение генов и их функции. Важно выяснить характер связей между ними, который определяет, как гены будут работать в конкретных условиях – внутренних и внешних. Ведь многие болезни человека обусловливаются не дефектами в самих генах, а нарушениями их согласованных действий, системы их регуляции.

Расшифровка генома человека и других организмов не только привела к прогрессу во многих областях биологии, но и породило множество проблем. Одна из них – идея «генетического паспорта», в котором будет указано, несет ли данный человек опасную для здоровья мутацию. Предполагается, что эти сведения будут конфиденциальными, но никто не может гарантировать, что не произойдет утечки информации. Прецедент уже был в случае «генетической паспортизации» афроамериканцев с той целью, чтобы определить, являются ли они носителями гена гемоглобина, содержащего мутацию, которая связана с серповидноклеточной анемией. Эта мутация распространена в Африке в малярийных районах, и если она присутствует в одном аллеле, то обеспечивает носителю устойчивость к малярии, обладатели же двух копий (гомозиготы) умирают в раннем детстве. В 1972 в рамках борьбы с малярией на «паспортизацию» было истрачено более 100 млн. долл., а после выполнения программы выяснилось, что а) у здоровых людей, носителей мутации, возникает комплекс вины, эти люди чувствуют себя не совсем нормальными, и такими их начинают воспринимать окружающие; б) появились новые формы сегрегации – отказ в приеме на работу. В настоящее время некоторые страховые компании выделяют средства на проведение ДНК-тестов в отношении ряда заболеваний, и если будущие родители, носители нежелательного гена, не соглашаются на прерывание беременности и у них рождается больной ребенок, им могут отказать в социальной поддержке.

Другая опасность – эксперименты по трансгенозу, созданию организмов с пересаженными от других видов генами, и распространению таких «химер» в окружающей среде. Здесь особую опасность представляет необратимость процесса. Если атомную станцию можно закрыть, использование ДДТ и аэрозолей прекратить, то изъять из биологической системы новый организм невозможно. Мобильные гены, открытые МакКлинток у растений, и сходные с ними плазмиды микроорганизмов передаются в природе от вида к виду. Ген, вредный или полезный (с точки зрения человека) для одного вида, может со временем перейти к другому виду и непредсказуемым образом изменить характер своего действия. В Америке мощная биотехнологическая компания «Монсанто» создала сорт картофеля, в клетки которого включен бактериальный ген, кодирующий токсин, который убивает личинок колорадского жука. Утверждается, что этот белок безвреден для человека и животных, однако страны Европы не дали разрешения на выращивание у себя этого сорта. Картофель испытывается в России. Опыты с трансгенными растениями предусматривают строжайшую изоляцию делянок с подопытными растениями, однако на охраняемых полях с трансгенными растениями Института фитопатологии в Голицыне под Москвой ремонтные рабочие выкопали картошку и

тут же ее съели. На юге Франции ген устойчивости к насекомым «перескочил» от культурных растений к сорнякам. Другой пример опасного трансгеноза – выпуск в озера Шотландии лосося, который набирает вес в 10 раз быстрее, чем обычный лосось. Существует опасность, что этот лосось попадет в океан и нарушит сложившееся популяционное равновесие у других видов рыб.

Литература:
Киселев, Л. Л. Геном человека и биология XXI века. – Вестник Российской АН, 2000, т. 20, № 5.
Янковский, Н. К. ГенЭтика: что заботит Европу, а что Россию. – Химия и жизнь, 2000, № 8.
Янковский, Н. К., Боринская С. А. Наша история, записанная в ДНК. – Природа, 2001, № 6.

Тема 3.3. Нуклеиновый состав (изохоры, GC-острова) ДНК

Вопросы для рассмотрения:

1. Композиционная гетерогенность ДНК.
2. Изохоры. Связь семейств изохор с кодирующими последовательностями ДНК и их экспрессией.

Краткое содержание лекции

Неоднородность распределения по геному ГЦ- и АТ-обогащенных районов, другими словами, композиционная гетерогенность ДНК – одна из наиболее важных характеристик молекулярной организации геномов.

Более сорока лет назад Бернарди и сотрудники при исследовании генома мыши *Mus musculus* обнаружили, что комплекс ДНК и серебра может быть разделен с помощью равновесного центрифугирования в градиенте плотности Cs_2SO_4 (сульфат цезия) по частоте сайтов на молекуле ДНК, связавших серебро. Это открытие позволило с высокой точностью разделять ДНК на фракции. Дальнейшее изучение этих фракций ДНК привело к открытию четкой композиционной гетерогенности ДНК. Композиционно гомогенные сегменты ДНК, принадлежащие к небольшому числу семейств, различающихся по плавучей плотности, были названы **изохорами**. Фракционирование ДНК по плавучей плотности при центрифугировании фрагментов ДНК в градиенте Cs2SO4 (или сахарозы) было выявлено у большого числа видов животных. Это отражает гетерогенность ДНК по нуклеотидному составу: АТ-богатые последовательности обладают большей плавучей плотностью, чем ГЦ-богатые. Относительные количества ДНК в семействах изохор формируют так называемый композиционный изохорный паттерн генома (также он называется геномным фенотипом), т.е. характерный «рисунок» из изохор, являющийся специфичным для каждого отряда или семейства.

У человека были выявлены два «легких» семейства изохор: L1 (1,698 г/см3) и L2 (1,700 г/см3), и три «тяжелых»: H1 (1,704 г/см3), H2 (1,708 г/см3) и H3 (1,712 г/см3). У человека семейства L1 и L2 составляют свыше 62% всего генома; H1 – 22%, H2 – 9%, семейство H3 составляет около 3% генома.

Наибольшее значение с точки зрения структурно-функциональной организации генома имеет вопрос о связи различных семейств изохор с кодирующими последовательностями ДНК и особенностями их экспрессии.

В первых исследованиях этого вопроса было показано, что большинство из 40 взятых в анализ генов человека расположены в ГЦ-богатых семействах. Впоследствии локализация in silico (путем компьютерного анализа) более 14000 генов человека привела авторов к тому же самому выводу, позднее подтвержденному на еще больших выборках кодирующих последовательностей. На рисунке 10 представлена диаграмма, демонстрирующая распределение плотности генов в каждом из семейств изохор у человека (рисунок 11).

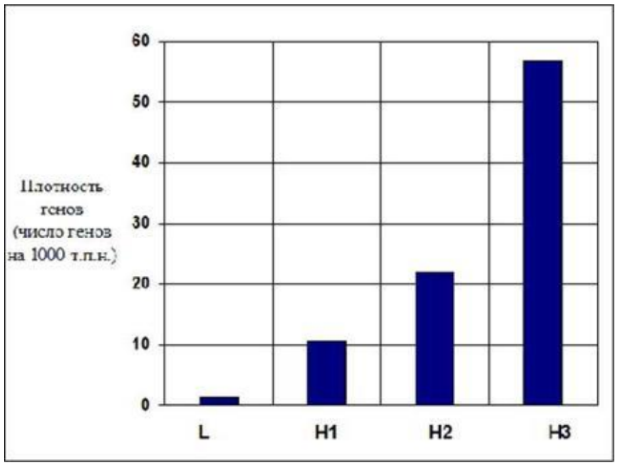


Рисунок 11 – Плотность генов в различных семействах изохор у человека

Выявление связи временных и межтканевых различий уровня экспрессии генов с ГЦ-уровнем, иными словами, распределения тканеспецифических генов и генов домашнего хозяйства относительно изохор, является не менее важным аспектом характеристики функционального значения композиционной гетерогенности геномной ДНК. Обобщая данные по структуре хроматина и распределению генов и изохор, Дж. Бернарди в 1993 г. писал: «Скорее всего, наибольший уровень транскрипции встречается в семействе изохор H3, поскольку там концентрация генов, прежде всего генов домашнего хозяйства, является наибольшей». Гипотеза о высокой транскрипционной активности генов, локализованных в семействе H3, подтверждалась и исследованиями нуклеотидного контекста стартовых АУГ-кодонов. Привлечение дополнительного критерия – процентного содержания гуанина или цитозина в третьем положении кодона (ГЦ3-уровень) – помимо молярного отношения ГЦ и АТ (ГЦ-уровень), и статистический анализ последовательностей ДНК из геномных баз данных человека и шпорцевой лягушки (*Xenopus laevis*) позволили установить, что **гены домашнего хозяйства**:

1) преимущественно локализованы в ГЦ-богатых семействах изохор;
2) не составляют большинства генов в ГЦ-богатых семействах изохор;
3) являются не менее ГЦ3-обогащенными, чем тканеспецифические гены.

Повышенное содержание кодирующих последовательностей, более высокий уровень рекомбинации, большое число ГЦ-обогащенных коротких интерспесных повторов (SINE) в тяжелых изохорах, в то время как в легких локализовано значительно меньше генов, ниже уровень рекомбинации и находятся почти исключительно ГЦ-обедненные длинные интерсперсные повторы (LINE), дает основание говорить об **изохорах как структурно-функциональных единицах организации генома.** Границы между тяжелыми и легкими изохорами на примере детально исследованного в отношении композиционного состава на молекулярном уровне кластера генов гистосовместимости (MHC) человека являются более чем композиционными структурами – временные границы репликации в фазе S клеточного цикла практически точно соответствуют физическим границам локализации изохор.

Литература: Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витебск. гос. ун-т имени П. М. Машерова, 2022. – 83 с.

Тема 3.4. Подбор праймеров для ПЦР

Цель:

Вопросы для рассмотрения:

1. Значение праймеров для процесса амплификации и ПЦР.
2. Принципы подбора праймеров. Требования, предъявляемые к праймерам.

Краткое содержание лекции

Праймеры являются чрезвычайно важной составляющей любого процесса амплификации, включая ПЦР. Для того чтобы химически синтезированные олигонуклеотидные праймеры (или иным способом приготовленные) могли служить затравочными молекулами для ферментативного синтеза комплементарных цепей ДНК, необходим их отжиг на подходящей одноцепочечной нуклеиновой кислоте с образованием одно/двухцепочечного стартового комплекса с наличием способного к удлинению по одноцепочечной ДНК/РНК-матрице спаренного 3’-конца такого праймера, несущего на нем гидроксильную группу. Отжиг праймеров на матрице может происходить в широком диапазоне температур, который зависит от длины конкретного праймера, а также от его химического состава и в этой связи можно напомнить, что A-T-пары образуют между собой две водородные связи, а G-C-пары – три водородные связи, что вносит существенный вклад в температурные оптимумы при молекулярной гибридизации нуклеиновых кислот с разными GC- и AT-составами.

Как имеется множество вариаций ПЦР, решающих часто весьма разные задачи, так существуют и довольно сильно отличающиеся принципы подбора олигонуклеотидных праймеров. Например, в целях диагностики необходимо выявлять наличие подходящего фрагмента ДНК или РНК (по сути – любого, но специфичного именно для выявляемого организма) и поэтому в таких случаях праймеры подбираются, исходя из задач обнаружения выбранного участка с максимальной достоверностью и экономией (времени, денежных и прочих ресурсов), тем более, если такие анализы должны носить массовый характер. При проведении научных исследований часто возникают задачи по клонированию конкретного гена, у которого есть начало и конец и в этом случае праймеры для его амплификации подбираются не с целью достичь максимальной эффективности процесса, а так чтобы было удобно вести дальнейшие эксперименты, но при этом праймеры могут оказаться далеко не самыми оптимальными по целому ряду параметров. Наконец, с помощью ПЦР изучают полиморфизм ДНК разных организмов – видов, сортов/пород, отдельных особей или индивидов. В таких случаях праймеры подбираются с таким расчетом, чтобы вариабельный участок оказывался между ними, что также накладывает определенные ограничения. Во многих подходах для генотипирования организмов вообще используются праймеры с произвольными последовательностями, не заботясь о присутствии для них мест отжига в геномах, поскольку они выбираются весьма короткими (часто от 10 до 12 нуклеотидов) и отжигаться на множестве мест теоретически должны. Подробные обзоры таких способов амплификации неопределенных фрагментов ДНК были сделаны нами не так давно и поэтому здесь на праймерах для таких вариантов ПЦР останавливаться не будем.

Относительно взаимного расположения мест отжига праймеров, определяющих размер(ы) ампликонов, которые на самом деле диктуются многими привходящими обстоятельствами, то этот вопрос имеет больше отношения к эффективности процесса амплификации. В данной статье будут изложены главные принципы подбора праймеров, различные основные требования, предъявляемые к ним с учетом решаемых в каждом конкретном случае своих задач. Значительное внимание будет уделено различным модификациям праймеров, включая вырожденные праймеры, тогда как специальных праймеров для различных вариаций ПЦР коснемся довольно кратко. Есть также масса и других задач, решаемых с помощью ПЦР, под которые подбираются специализированные праймеры. Однако ни в одной даже очень большой статье невозможно рассмотреть все многообразие праймерных систем, применяемых в ПЦР. Что касается праймеров, используемых в различных

способах изотермической амплификации нуклеиновых кислот, имеющих свои особенности, то они также останутся за пределами рассмотрения.

Требования, предъявляемые к праймерам:

1. Размер праймера должен быть 16–25 нуклеотидов. Меньше 16-ти – слабая связь с целью.
2. Разница в температуре плавления праймеров – не более 6 градусов.
3. Ц+Г должно быть 50-60%.
4. Для улучшения качества отжига рекомендуется подбирать праймеры так, чтобы последние несколько нуклеотидов 3’-конца праймера содержали GC-основания.
5. Проверить сбалансированность PCR по нуклеотидам и праймерам.
6. Отсутствие внутренней вторичной структуры (праймеры не должны быть само- и взаимнокомплиментарными).
7. Отсутствие комплиментарности между 3’-концами (чтобы не образовывалось праймер-димеров).
8. Оптимальная концентрация праймеров подбирается эмперически, но не должна быть более 50 пикомолей по пробирку – иначе начнется неспецифический отжиг праймеров.
9. Упрощенный расчет оптимальной температуры отжига праймера:

Tm = [(F+N)*20C] + [(G+C)*40C] (если суммарная длина олигонуклеотида не превышает 20 оснований);

Tm = 22+1.46([2*(G+C)] + (F+A)) (если суммарная длина олигонуклеотида составляет 20-30 оснований).

10. Область отжига праймеров должна находиться вне зон мутаций, делеций или инсерций в пределах видов или иной, взятой в качестве критерия при выборке праймеров, специфичности. При попадании на такую зону, отжига праймеров происходить не будет, и как следствие – ложноотрицательный результат.

Литература: Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витебск. гос. ун-т имени П. М. Машерова, 2022. – 83 с.

Тема 3.5. Анализ частоты использования кодонов

Вопросы для рассмотрения:

1. Частота транзиций и трансверсий. Определение соотношения транзиций и трансверсий.
2. ГН-насыщенность общая и отдельных положений кодона. Зависимость ГЦ-насыщенности отдельных положений кодона от общей ГЦ-насыщенности.
3. Анализ аминокислотного состава белков.
4. Системы классификации сайтов нуклеиновых кислот в связи с их положением в кодоне - система вырожденности и синонимичности / несинонимичности.
5. Методики изучения стратегии кодирования белков.

Краткое содержание лекции

Определение нуклеотидного состава нуклеиновых кислот и аминокислотного состава белков преимущественно проводится с помощью специальных компьютерных программ (например, MEGA). При анализе нуклеотидного состава ДНК и РНК наиболее часто анализируются следующие показатели:

1. **Частота транзиций**

Транзиция - это мутация, обусловленная заменой одного пуринового основания на другое (А ↔ Г) или одного пиримидинового на другое (У, Т↔Ц).

Транзиции - простые замены (не происходит изменения ориентации пурин-пиримидин в мутантном сайте двухцепочечной молекулы ДНК).

Частота транзиций (*P*) вычисляется по формуле: $P=np/L$, где np – число наблюдаемых транзиций; L – общее число нуклеотидных сайтов, по которым сравниваются последовательности.

2. **Частота трансверсий**

Трансверсия - это мутация, обусловленная заменой пуринового основания на пиримидиновое и наоборот (А, Г ↔У, Т, Ц).

Трансверсии - сложные или перекрестные замены (происходит изменение ориентации пурин–пиримидин в мутантном сайте двухцепочечной молекулы ДНК). Частота трансверсий (*Q*) определяется по формуле: $Q=nq/L$

где *nQ* - число наблюдаемых трансверсий; *L* - общее число нуклеотидных сайтов, по которым сравниваются последовательности.

3. **Соотношение наблюдаемых трансверсий и транзиций (*q*)**, определяемое по формуле: $q=nq/np/$

4. **Общая ГЦ-насыщенность** - это общее содержание гуанина и цитозина. Повышение ГЦ-насыщенности свидетельствует об увеличении термодинамической стабильности ДНК за счет образования большего количества водородных связей (трех между гуанином и цитозином вместо двух между аденином и тимином).

5. **ГЦ-насыщенность отдельных положений кодона** - ГЦ1, ГЦ2 и ГЦ3-насыщенность. Наибольшее значение имеет определение ГЦ3- насыщенности как маркера мутационного давления.

6. **Зависимость ГЦ-насыщенности отдельных положений кодона от общей ГЦ-насыщенности**. Эта зависимость, изученная С. Кумаром и М. Неем, наиболее часто имеет вид, показанный на рисунке 12.

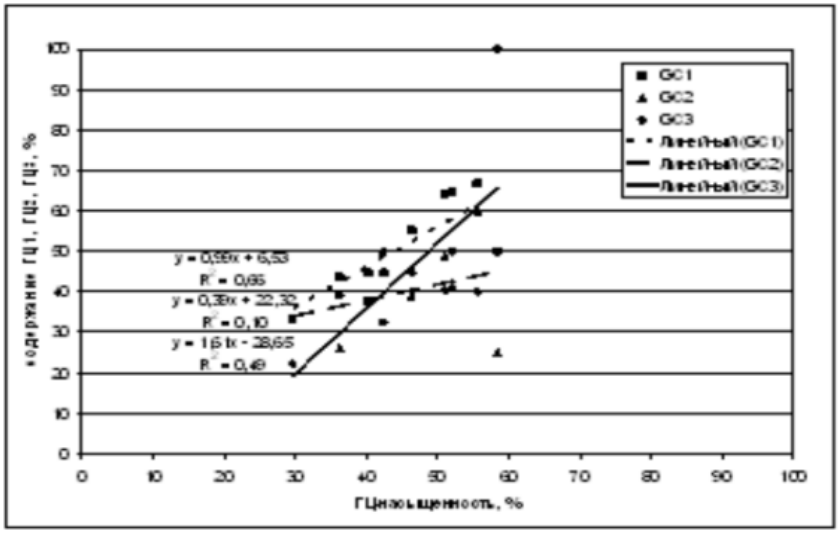


Рисунок 12 – Классическая зависимость содержания гуанина и цитозина в отдельных положениях кодона от общей ГЦ-насыщенности на примере экзонов гена, кодирующего алкогольдегидрогеназу класса 3 человека

Данный график показывает:

1. С увеличением общей ГЦ-насыщенности наблюдается различной степени линейный рост содержания гуанина и цитозина во всех положениях нуклеотида в кодоне.

2. Наибольший рост характерен для значений ГЦ3 (наклон тренда -1,61), что свидетельствует о выраженном влиянии на замены в данном положении кодона мутационного давления и слабом влиянии отрицательного отбора. Это объясняется тем, что 72% замен по третьему положению не приводят к изменению кодируемой аминокислоты.

3. Меньший наклон тренда по значениям ГЦ1, равный 0,99, связан с меньшей долей синонимичных замен (5%).

4. Несмотря на то, что все замены по второму положению несинонимичны, наклон тренда по значениям ГЦ2 является положительным (0,39). Объяснить этот факт можно тем, что часть замен аминокислот является нейтральной, т.е. не приводит к изменению функции белка. В этом случае отрицательный отбор не оказывает никакого влияния на данную мутацию, и она закрепляется, отражая основное направление нуклеотидных замен.

Таким образом, можно утверждать, что в экзонах изучаемого гена АДГ «разрешены» преимущественно замены в третьем положении кодона, в меньшей степени - в первом, а в наименьшей степени – во втором, что связано с вероятностью синонимичных и несинонимичных замен. Выраженное предпочтение синонимичных замен свидетельствует о структурно-функциональных ограничениях, налагаемых на молекулу, а получение такого графика может выступать в качестве одного из критериев ее эволюционной сформированности.

При анализе аминокислотного состава белков чаще анализируются:

1. В связи с ГЦ-насыщенностью = *содержание* аминокислот *GARP* (кодируемых ГЦ-богатыми кодонами) и *FYMINK* (кодируемых ГЦ-бедными кодонами, таблица 1).
2. В связи с зарядом аминокислот – *содержание положительнозаряженных, отрицательнозаряженных и нейтральныхаминокислот*.
3. В связи с химической структурой - *содержание ароматических,серосодержащих, гидроксильных аминокислот и др.*

Таблица 1 – Аминокислоты группы GARP и FYMINK и соответствующие им кодоны мРНК

Аминокислоты группы GARP			Аминокислоты группы FYMINK		
G	глицин	ГГУ			
		ГГЦ			
		ГГА			
		ГГГ	F	фенилаланин	УУУ
A	аланин	ГЦУ			УУЦ
		ГЦЦ	Y	тирозин	УАУ
		ГЦА			УАЦ
		ГЦГ	M	метионин	АУГ
R	аргинин	ЦГУ	I	изолейцин	АУУ
		ЦГЦ			АУЦ
		ЦГА			АУУ
		ЦГГ	N	аспарагин	ААУ
		АГА			ААЦ
		АГГ	K	лизин	ААА
P	пролин	ЦЦУ			ААГ
		ЦЦЦ			
		ЦЦА			
		ЦЦГ			

Сайт - это участок небольшого размера молекулы нуклеиновой кислоты или белка, обычно равный одному нуклеотиду или одному аминокислотному остатку соответственно.

Константные сайты - это нуклеотидные или аминокислотные сайты выравненных последовательностей, в которых не наблюдаются замены.

Вариабельные сайты - это нуклеотидные или аминокислотные сайты выравненных последовательностей, в которых наблюдаются замены.

Синглетонные сайты - это вариабельные сайты не менее пяти выравненных последовательностей, в которых преобладает частота одного нуклеотида или аминокислотного остатка. Пример: сайты № 192, 207 на (рисунок 13).

Парсимоничные сайты – это вариабельные сайты не менее пяти выравненных последовательностей, в которых как минимум два нуклеотида или аминокислотных остатка встречаются хотя бы два раза. Пример: сайты № 191, 192 на рисунке 13.

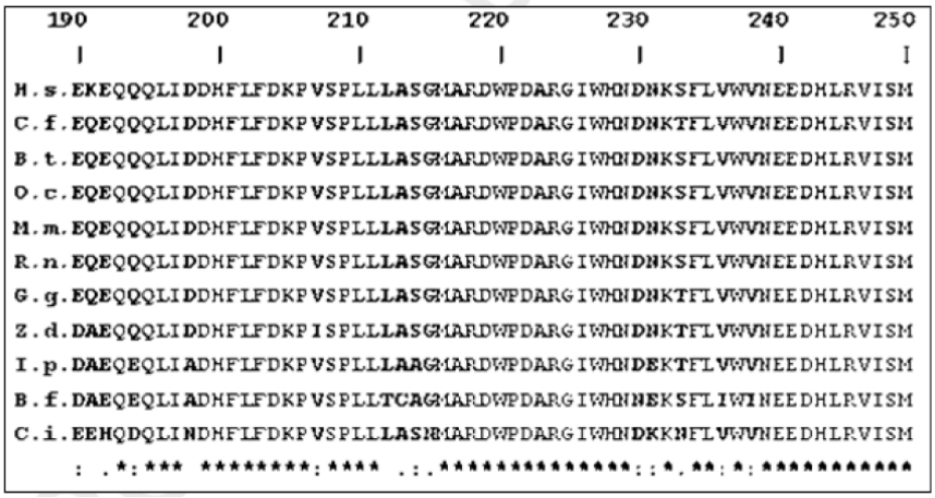


Рисунок 13 – Выравненные участки каталитических доменов М-изоферментов креатинкиназ позвоночных и соответствующие участки креатинкиназ ланцетника и оболочника

Существуют две системы классификации сайтов нуклеиновых кислот в связи с их положением в кодоне - система вырожденности и синонимичности/несинонимичности (таблица 2).

Таблица 2 – Соотношение систем вырожденности и синонимичности / несинонимичности сайтов

Система классификации сайтов	
По вырожденности	**По синонимичночности/ несинонимичности**
невырожденный	абсолютно несинонимичный
двухкратновырожденный	несинонимичный на 2/3 синонимичный на 1/3
трехкратновырожденный	несинонимичный на 1/3 синонимичный на 2/3
четырехкратновырожденный	абсолютно синонимичный

Нолькратно вырожденный сайт (невырожденный сайт) – это нуклеотидный сайт, в котором любая возможная замена является несинонимичной. Пример: в первом положении кодона ЦУУ (Лей) возможны три несинонимичные замены:

1. ЦУУ→УУУ (Фен).
2. ЦУУ→АУУ (Иле).
3. ЦУУ→ГУУ (Вал).

Вывод: первый сайт кодона ЦУУ является невырожденным.

Двухкратно вырожденный сайт - это нуклеотидный сайт, в котором одна из трех возможных замен является синонимичной. Пример: в третьем положении кодона УУУ (Фен) возможны две несинонимичные и одна синонимичная замена:

1. УУУ →УУЦ (Фен).
2. УУУ →УУА (Лей).
3. УУУ →УУГ (Лей).

Вывод: третий сайт кодона УУУ является двухкратно вырожденным.

Трехкратно вырожденный сайт - это нуклеотидный сайт, в котором две замены из трех возможных синонимичны. Пример: в третьем положении кодона АУУ (Иле) возможны одна несинонимичная и две синонимичные замены:

1. АУУ→АУЦ (Иле).
2. АУУ→АУА (Иле).
3. АУУ→АУГ (Мет).

Вывод: третий сайт кодона АУУ является трехкратно вырожденным.

Четырехкратно вырожденный сайт - это нуклеотидный сайт, в котором возможны лишь синонимичные замены. Пример: в третьем положении кодона ЦУУ (Лей) возможны три синонимичные замены:

1. ЦУУ→ЦУЦ (Лей).
2. ЦУУ→ЦУА (Лей).
3. ЦУУ→ЦУГ (Лей).

Вывод: третий сайт кодона ЦУУ является четырехкратно вырожденным.

Несинонимичный сайт - это сайт, в котором возможна несинонимичная замена.

Синонимичный сайт - это сайт, в котором возможна синонимичная замена.

Если все замены в данном сайте несинонимичны, сайт называется абсолютно несинонимичным. Если все замены в данном сайте синонимичны, сайт называется абсолютно синонимичным. Если одна из замен в данном сайте несинонимична, сайт называется несинонимичным на ⅓ (соответственно, синонимичным на ⅔). Если две замены в данном сайте несинонимичны, сайт называется несинонимичным на ⅔ (соответственно, синонимичным на ⅓).

Кодон (триплет) - это наименьшая функциональная единица гена, состоящая из трех рядом расположенных нуклеотидов, кодирующая одну аминокислоту.

Серия кодонов - это группа синонимичных кодонов.

Синонимичные (эквивалентные, изоакцепторные) кодоны – это кодоны, кодирующих одну и ту же аминокислоту.

Двухкратно вырожденная серия кодонов - это серия кодонов, состоящая из двух синонимичных триплетов.

Четырехкратно вырожденная серия кодонов - это серия кодонов, состоящая из четырех триплетов.

Шестикратно вырожденная серия кодонов - это серия кодонов, состоящая из шести триплетов.

Претерминальные кодоны - это кодоны, которые могут стать терминальными в результате замены одного нуклеотида и, следовательно, прервать синтез пептидной цепочки.

ГЦ3-кодоны - это кодоны, содержащие в третьем положении гуанин или цитозин, исключая невырожденные кодоны (АУГ- метионин, УГГ- триптофан) и стоп-кодоны.

Стратегия кодирования белка определяется картиной использования кодонов в соответствующих ему мРНК и ДНК. Многообразие возможных стратегий кодирования связано с вырожденностью генетического кода (в среднем на каждую из 20 аминокислот приходится три синонимичных кодона). Одним из наиболее часто анализируемых при изучении стратегии кодирования показателей является картина использования синонимичных кодонов.

Методики изучения стратегии кодирования белков:

1. Сравнения количества (доли) кодонов в изучаемых последовательностях нуклеиновых кислот.

2. Сравнения показателей относительного использования синонимичных кодонов (RSCU - relative synonymous codon's usage), которые производятся по формуле: **RSCUК= (RSCUСК*nК)/nСК,**

где RSCUК - показатель относительного использования кодона А,

RSCUСК - количество кодонов в серии, *n*К - частота использования кодона А,

*n*СК - частота использования всех кодонов серии.

Рассчитаем в качестве примера RSCU для кодона УГУ в мРНК, кодирующей НАДН-дегидрогеназу 6 человека (таблица 3). RSCUСК = 2 (серия представлена двумя кодонами - УГУ и УГЦ), *n*К = 1 (кодон УГУ используется в данной мРНК 1 раз), *n*СК = 1 (кодон УГУ используется 1 раз, кодон УГЦ не используется), тогда: RSCUУГУ= (2*1)/2=2.

3. Сравнения суммарных показателей RSCU серии кодонов, соответствующих одной и той же аминокислоте (рисунок 14).

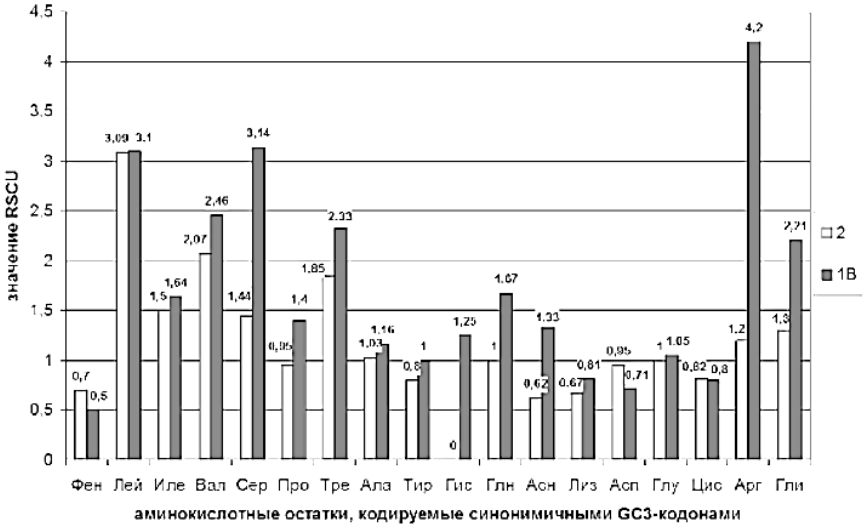


Рисунок 14 – Показатель относительного использования синонимичных ГЦ3-кодонов для алкогольдегидрогеназ классов 1В и 2 человека

4. Суммарные показатели RSCU соответствующих аргинину ГЦ3-кодонов мРНК, кодирующих алкогольдегидрогеназы классов 1В и С человека, значительно отличаются (4,2 и 1,2 соответственно), а соответствующих цистеину – сходны (0,8 и 0,82).

Таблица 3 – Показатели RSCU и количество кодонов УГУ и УГЦ в мРНК, кодирующих НАДН-дегидрогеназу 6 человека, трихинеллы и цианорабдитис

Организм/показатель	Кодон УГУ	Кодон УГЦ
Человек	n=1 RSCU=2,0	n=0 RSCU=0
Трихинелла	n=0 RSCU=0	n=2 RSCU=2,0
Цианорабдитис	n=3 RSCU=2,0	n=0 RSCU=0

5. Вычисление дистанции Дж. МакАйнерни (*Djk*):

$$D_{jk} = \sum_{i-1}^{n} \frac{\text{abs}(RSCU_{ji} - RSCU_{ki})}{n}, \quad (1)$$

Литература: Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витебск. гос. ун-т имени П. М. Машерова, 2022. – 83 с.

Тема 4.1. Фолдинг и транспорт белков у про- и эукариот

Вопросы для рассмотрения:

1. Фолдинг, белки шапероны
2. Секреция белков у прокариот:
3. Sec-аппарат и сигнальный пептид.
4. Распределение белков по компартментам клетки эукариот.
5. Деградация белков

Краткое содержание лекции

Фолдинг, белки шапероны

Кроме последовательности аминокислот полипептида (первичной структуры), крайне важна трехмерная структура белка, которая формируется в процессе фолдинга (от англ. folding - «сворачивание»).

Фолдинг – это процесс формирования «правильной» третичной структуры из полипептидной цепочки.

Универсальным механизмом, обеспечивающим быстрый и безошибочный фолдинг, является участие в сворачивании белковых цепей **шаперонов** – специализированных белков, обнаруженных во всех органеллах всех организмов от бактерий до приматов. Шапероны бывают двух типов.

1. **Молекулярные шапероны**, которые связываются с белковой нитью, предотвращая ее агрегацию или деградацию.

2. **Шаперонины**, которые обеспечивают фолдинг белков.

Молекулярные шапероны состоят из белков **Hsp70** и их гомологов:

– **Hsp70** в цитозоле и матриксе митохондрий,

– **BiP** в эндоплазматическом ретикулуме,

– **DnaK** в бактериях.

Белки Hsp70 называют белками теплового шока (heat shock proteins), поскольку они активно синтезируются клеткой при нагревании. Связанный с АТФ Hsp70 (Hsc70 это постоянно синтезируемый гомолог Hsp70) присоединяется к гидрофильному участку белковой цепи. Гидролиз АТФ→АДФ разрешает фолдинг белковой цепи. Замена АДФ на АТФ приводит к диссоциации Hsp70 со свернутой белковой цепи (рисунок 15 вверху).

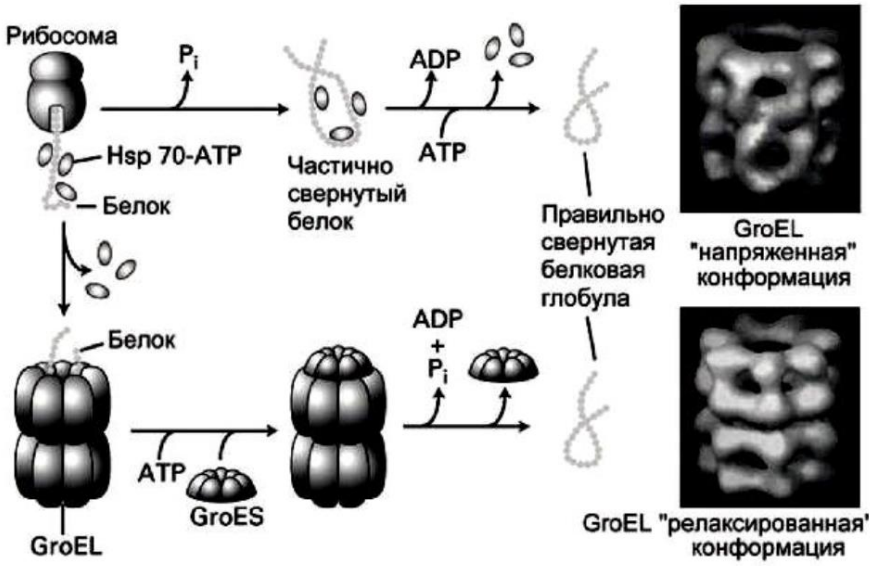


Рисунок 15 – Фолдинг белков с участием шаперонов и шаперонинов: а – схема фолдинга; б – конформации шаперонина

Шаперонины – большие цилиндрические макромолекулярные комплексы, состоящие из двух олигомерных колец (по восемь мономеров в шаперонине эукариот TriC и по семь мономеров в шаперонинах бактерий, митохондрий и хлоропластов GroEL) – обеспечивают изоляцию белковой нити на время фолдинга (рисунок 15 внизу). Связывание GroEL с АТФ высвобождает свернутый белок. Функционирование GroEL осуществляется с участием ко-шаперонина GroES, который закрывает (кэпирует) полость шаперона на время фолдинга.

После правильного сворачивания, крышка открывается и готовая функциональная молекула белка покидает шаперонин.

Шапероны и шаперонины могут исправлять поврежденную структуру белка. Если исправить не удается, белок должен быть разрушен в протеасоме – это тоже сложный белок, образующий полость.

Результатом нарушения фолдинга белков являются конформационные болезни. К ним относятся:

- Прионовые заболевания;
- Болезнь Альцгеймера;
- Синдром Марфана;
- Куриная слепота;
- Злокачественные опухоли (нарушение фолдинга p53).

Прионы – это инфекционные белки. Вызывают тяжелые заболевания центральной нервной системы у человека и ряда высших животных («медленные инфекции»):

✓ у людей – куру («смеющаяся смерть»), болезнь Крюцельда-Якоба, семейная фатальная бессонница;

✓ у коров, норок, оленей, козы – бешенство (губчатая энцефалопатия);

✓ у овец – почесуха (scrappy).

Секреция белков у прокариот: Sec-аппарат и сигнальный пептид

Большинство бактерий способны транспортировать синтезируемые белки в окружающую среду. Для этого бактериальные клетки используют различные системы секреции в зависимости от строения и конечной локализации белка. Все эти системы должны специфически распознавать свои субстраты и облегчать секрецию без нарушения целостности клеточной оболочки. Однако для достижения такой цели эти системы используют существенно разные механизмы и отличаются друг от друга по своей сложности.

Грамотрицательные бактерии имеют две мембраны – цитоплазматическую мембрану (IM) и внешнюю мембрану (OM), разделенные периплазматическим пространством. Такая организация клеточной стенки делает процесс секреции топологически сложным. В грамотрицательных бактериях биологические молекулы, выделяемые во внешнюю среду, должны пересечь два гидрофобных барьера. Для обеспечения транспорта через клеточную стенку в указанных микроорганизмах функционируют как минимум шесть специализированных систем секреции (т.н. системы секреции типов 1–6).

Системы секреции можно классифицировать по тому, за один или два этапа происходит транспортировка белковой молекулы. Один класс систем секреции, включающий типы 1, 3 и 4, обеспечивает одноэтапную секрецію из цитоплазмы сразу во внеклеточную среду. Второй класс систем секреции (типов 2, 5 и 6) предполагает экспорт в два этапа (из цитоплазмы в периплазматическое пространство, и из периплазмы в среду), причем в последовательных этапах транслокации работают разные молекулярные механизмы. Первый этап – транспорт в периплазму – происходит при участии одного из трех разных путей: Sec, Tat или SRP. Основные особенности этих путей – Sec (SecB-зависимый путь), Tat (твин-аргинин зависимый путь, twin-arginine translocation) и SRP (signal recognition particle) – схематически показаны на рисунке 16.

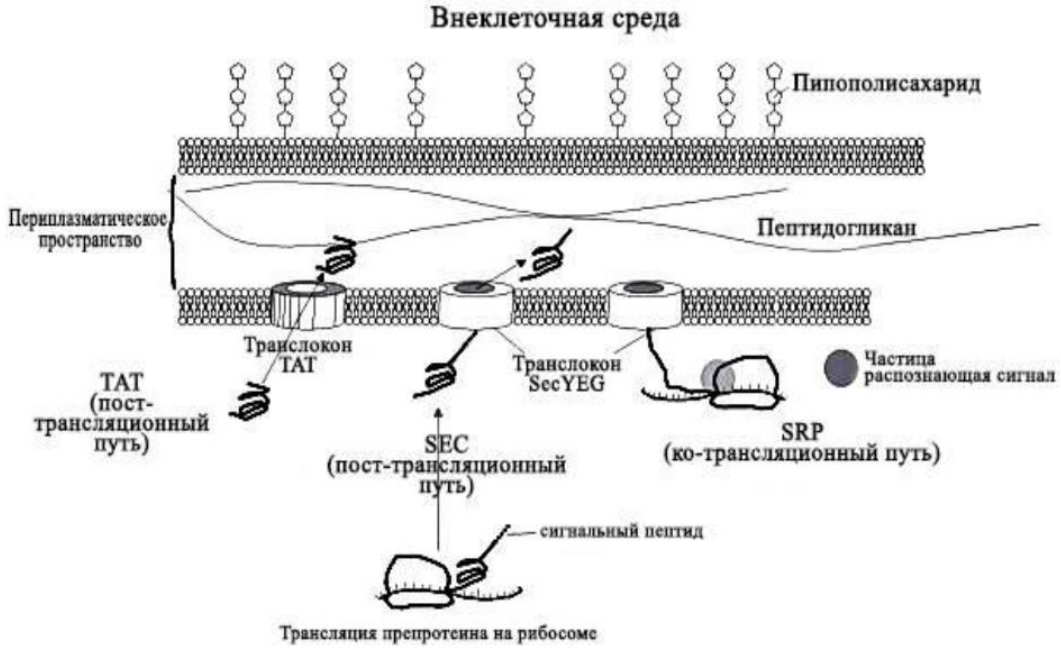


Рисунок 16 – Пути транслокации через плазматическую мембрану

Все три перечисленных пути используют разные механизмы и приспособлены для транслокации разных белков-мишеней. Транслокация по пути Sec или Tat является пост-трансляционным процессом, в то время как путь SRP является ко-трансляционным, т.е. осуществляет экспорт полипептидной цепи, которая продолжает синтезироваться на рибосоме. Белкимишени пути Sec для транслокации должны быть развернуты полностью до состояния вытянутой полипептидной цепи (т.е. теряют третичную структуру в ходе транслокации), эти белки подвергаются рефолдингу в периплазме. Белки-мишени пути Tat транслоцируются в свернутом состоянии, не теряя третичной структуры, которую они приобретают в цитоплазме. С использованием пути Tat возможно экспортировать функционально активные белки, имеющие сложную третичную структуру.

Sec-аппарат. Наибольшая доля секретируемых и мембранных белков у прокариот используется для транслокации через путь Sec, в связи с чем исторически путь Sec получил название основного секреторного пути (general secretory pathway, GSP). Природные белки-мишени пути Sec, как правило, локализуются в периплазматическом пространстве или интегрируются во внешней мембране. Секреторный **аппарат Sec** гомологичен транслокону в эндоплазматическом ретикулуме высших эукариот и транслокону Sec 61 у дрожжей. Центральную роль в транслоконе Sec играет комплекс из трех белков SecYEG, который образует канал в цитоплазмотической мембране. В состав транслоказного комплекса также входят другие белки: мембранная АТРаза SecA, цитоплазматический шаперон SecB и акцессорные белки SecD и SecF, YidC и YajC. Белки-мишени пути Sec синтезируются в цитоплазме в виде

предшественников, содержащих на N-конце короткий (15-30 аминокислотных остатков) пептид, называемый **сигнальным пептидом**. SecA распознает сигнальный пептид и в кооперации с шапероном SecB, транспортирует препротеин к транслокону SecYEG (рисунок 17).

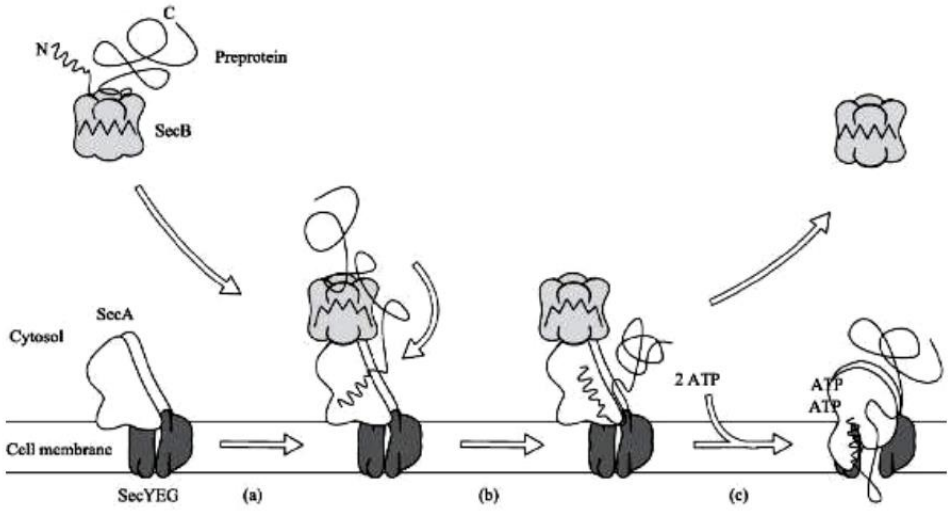


Рисунок 17 – Sec-аппарат

В ходе транслокации белка-мишени, как только С-конец сигнального пептида достигает поверхности цитоплазмотической мембраны, сигнальный пептид отщепляется при участии сигнальных пептидаз (пептидазы I (LepB) или пептидазы II (LSPA).

Сигнальные пептиды имеют сильно различающиеся аминокислотные последовательности, но демонстрируют и общие структурные характеристики. Типичный сигнальный пептид имеет на N-конце короткий участок, т.н. N-домен, длиной от 2 до 10 а.о., богатый положительно заряженными аминокислотами; в центре сигнального пептида находится участок (длиной 10–20 а.о.), богатый аминокислотными остатками с гидрофобными боковыми цепями (H-домен); на С-конце сигнального пептида находится С-домен, менее гидрофобный, чем H-домен, и содержащий сайт узнавания сигнальной пептидазы. Во время транспорта белков из цитоплазмы сигнальный пептид отщепляется сигнальной пептидазой с высвобождением зрелого белка.

Вторичная структура экспортируемого белка играет важную роль в том, насколько эффективно будет отщепляться сигнальный пептид. Получение эффективной транслокации требует подбора сигнального пептида к выбранному белку-мишени.

Распределение белков по компартментам клетки эукариот

Мембранные структуры клетки активно участвуют во внутриклеточном транспорте белков. Существует три основных мехнизма, с помощью которых клетка решает эту задачу (рисунок 18).

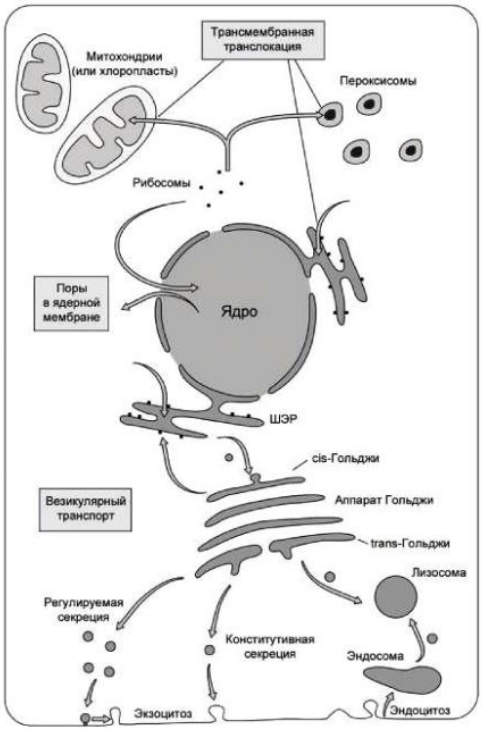


Рисунок 18 – Механизмы внутриклеточного транспорта белков в эукариотической клетке

1. Белок после синтеза и фолдинга в неизменном виде доставляется в нужную органеллу через специализированные поры в мембранах. Такой вид доставки называется управляемый транспорт (gated transport).

2. Трансмембранная транслокация белков в ходе которой полипептид сначала денатурируется, затем полипептидная цепочка протягивается через одну или несколько мембран, а затем снова происходит фолдинг функционального белка.

3. Везикулярное движение (vesicular trafficking) белков, в ходе которого от мембраны отпочковывается везикула, в состав которой входят транспортируемые вещества. С помощью управляемого транспорта, например, в ядро клетки через ядерные поры доставляются все вещества. Трансмембранная транслокация обеспечивает доставку синтезированных в цитозоле белков в пероксисомы, митохондрии и хлоропласты. Везикулярное движение обеспечивает доставку веществ в лизосомы и секрецию веществ из клетки.

Распределение белков обеспечивается наличием в их структуре специальных сортировочных сигналов (sorting signals). Сразу после окончания синтеза белка рибосомой в цитозоле, когда он является просто полипептидом, сортировочными сигналами являются последовательности аминокислот на концах белковой цепи, которые называют нацеливающими.

Для белков, которые синтезируются на мембранах шероховатого эндоплазматического ретикулума, дополнительные сортировочные сигналы (такие, как сахара или фосфатные группы) могут быть добавлены с помощью специализированных ферментов в цистернах аппарата Гольджи в ходе

посттрансляционной модификации белков. Такие сигналы обычно представляют собой специфические лиганды, которые распознаются рецепторными белками, а эти рецепторные белки с присоединенными к ним транспортируемыми белками, в свою очередь, присоединяются к мембранным транслокационным комплексам соответствующего компартмента. Нацеливающие последовательности белка, которые представляют собой цепочку из 3–80 аминокислот, тоже распознаются специализированными рецепторами, которые доставляют данный белок к соответствующим транслокационным комплексам. После доставки в нужный компартмент нацеливающие последовательности обычно отрезаются от белковой цепи специализированными ферментами.

Одними из наиболее изученных нацеливающих последовательностей являются сигнальные пептиды (или сигнальные последовательности) – цепочки из 5–15 преимущественно гидрофобных аминокислот. Наличие такой сигнальной последовательности в синтезируемом белке вынуждает рибосому присоединиться к эндоплазматическому ретикулуму, и синтезируемая белковая цепь направляется не в цитозоль, а в люмен ретикулума. Другой пример нацеливающей последовательности – сигнал импортирования для белков, которые должны быть перемещены из цитозоля в матрикс митохондрии.

Этот сигнал представляет собой цепочку из 20–80 аминокислот, формирующих полярную α-спираль, у которой положительно заряженные аминокислоты выстроены с одной стороны спирали, а гидрофобные аминокислоты – с другой. Для транспорта белков в ядро клетки определена последовательность из пяти положительно заряженных аминокислот. Для переноса белков в пероксисому служит пероксимальная нацеливающая последовательность Ser-Lys-Lys-COOH – С-концевой трипептид. Существуют также сортировочные сигналы, которые не способствуют перемещению белка, а наоборот, служат сигналом о том, что белок уже доставлен к месту назначения и никуда далее его не следует перемещать. Например, белки с так называемой KDEL-последователь-остью Lys-Asp-Glu-Leu-COOH на С-конце остаются в эндоплазматическом ретикулуме и не должны удаляться из него везикулярным транспортом.

Иллюстрацией вышесказанного может быть кальций-связывающий белок гладкого эндоплазматического ретикулума калретикулин (calciumbinding protein of the endoplasmatic reticulum – calreticulin). Первые 17 аминокислот на N-конце калретикулина являются сигнальной последовательностью, которая инициирует транслокацию белка в люмен эндоплазматического ретикулума, а последние 4 аминокислоты – последовательность KDEL – не позволяет белку уйти из ретикулума. Между этими сортировочными сигналами располагается первичная структура функционального белка.

Деградация белков

Активность ферментов зависит от их концентрации, а значит, определяется балансом процессов синтеза и деградации белков в клетке. Диапазон времени жизни белков – от нескольких минут (например, белки обеспечивающие митоз,

митотические циклины) до времени жизни всего организма (например, белки в хрусталике глазного яблока). Эукариотические клетки имеют несколько внутриклеточных протеолитических механизмов деградации (утилизации) в основном следующих трех типов белков:

1) неверно свернутые (misfolded) или денатурированные (развернутые) белки, 2) «нормальные» белки, чья концентрация должна быть снижена, 3) внеклеточные белки, захваченные клеткой в результате, например, эндоцитоза или фагоцитоза.

Главный внутриклеточный механизм деградации – это ферментативный протеолиз в лизосомах, чья «кислая» среда заполнена гидролитическими ферментами (рисунок 19).

Деградация в лизосомах предназначена главным образом для протеолиза чужих белков, попадающих в клетку в результате: 1) эндоцитоза, 2) фагоцитоза, 3) протеолиза нефункциональных органелл клетки, которые отторгаются и «перевариваются» клеткой (аутофагия).

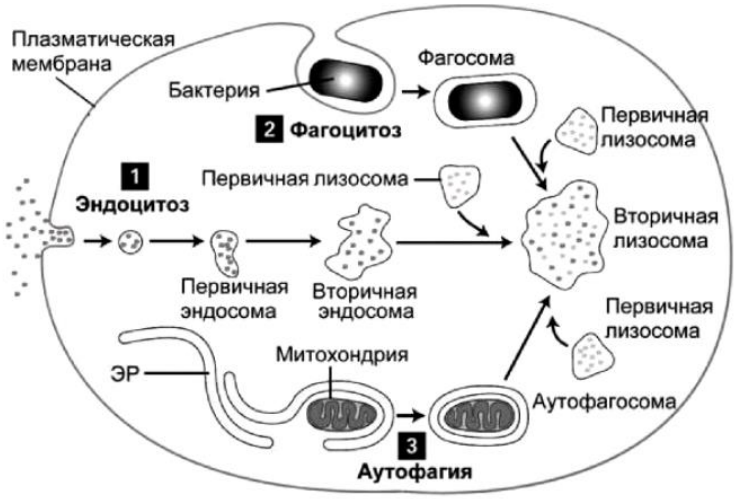


Рисунок 19 – Протеолиз в лизосомах

Убиквитирование. Другой, отличный от лизосомального, цитозольный механизм деградации белков реализуется путем модификации лизина в белке добавлением **убиквитина (или юбикитина, ubiquitin)**, полипептида из 76 аминокислотных остатков. Убикитированные белки деградируют затем в протеосомах. Убиквитирование осуществляется в три этапа тремя ферментами.

1. АТФ-зависимая активация убиквитин-активирующего фермента E1 (ubiquitin-activating enzyme) посредством присоединения к нему молекулы убиквитина.

2. Перенос молекулы убиквитина на цистеин убиквитинсопряженного фермента E2 (ubiquitin-conjugating enzyme).

3. Формирование пептидной связи между молекулой убиквитина, находящейся на ферменте E2, и лизином на белке, который должен

деградировать, реакция катализируется ферментом убиквитин лигаза E3 (ubiquitin ligase).

Эти три этапа повторяются много раз, новые молекулы убиквитина присоединяются к уже присоединенным, в результате чего полиубиквитиновая цепь распознается протеосомой. Протеосома имеет полое цилиндрическое ядро, закрытое с двух сторон кэпирующими белками. Внутри протеосомы убиквитированные белки разрезаются на короткие (7–8 остатков) пептиды множественными протеазами, расположенными на стенках протеосомы, а молекулы убиквитина отделяются от белков.

Убиквитиновая деградация белков используется в двух случаях.

1. Удаление цитозольных белков, чья концентрация должна быть снижена. Например, циклин должен присутствовать в клетке только на определенном этапе клеточного цикла. Фосфорилирование циклина изменяет его конформацию, и внутренняя, исходно погруженная внутрь белковой глобулы, последовательность rg-X-X-Leu-Gly-X-Ile-Gly-Asp/Asn (где X – любая аминокислота) становится доступной ферментам юбиквитинизации.

2. Протеолиз белковых молекул неправильно свернутых (misfolding) в эндоплазматическом ретикулуме.

Мисфолдинг, так же как и в предыдущем случае, делает доступными те гидрофильные части белковой нити, которые спрятаны внутрь глобулы при правильном фолдинге. Эти белки переносятся в цитозоль, где распознаются ферментами юбиквитинизации.

В обоих случаях белки содержат последовательности аминокислот, распознаваемые ферментами убиквитинизации.

Литература: Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витебск. гос. ун-т имени П. М. Машерова, 2022. – 83 с.

Тема 4.2. Мотивы и домены

Цель:

Вопросы для рассмотрения:

1. Мотивы белков.
2. Домены, их состав.
3. Разница между мотивом и доменом.

Краткое содержание лекции

В ключевое отличие между мотивом и доменом заключается в том, что мотив не является независимо стабильным, в то время как домен является независимо стабильным.

Белки – важные биологические макромолекулы, присутствующие в нашем организме. С другой стороны, генетический код гена определяет аминокислотную последовательность белка. Более того, белки имеют первичную, вторичную и третичную структуры. Первичная структура - это аминокислотная последовательность полипептидной цепи. Когда полипептидные цепи складываются друг с другом, образуется вторичная структура белка. Альфа-

спирали, бета-листы и супервторичные структуры относятся к вторичным структурам. Определенные группы супервторичных элементов известны как белковые мотивы. Третичная структура белка относится к его трехмерной структуре, которая определяет функцию белка. Домен - это складчатая часть белковой молекулы, которая является глобулярной и выполняет дискретную функцию. Это фундаментальная функциональная и трехмерная структура белка.

Мотив – это определенная группа супервторичных элементов белков, таких как альфа-спирали и бета-структуры. Это своего рода узоры, присутствующие в разных белках. Мотивы описывают схемы складывания вторичных структурных элементов и их взаимодействия. Эти схемы складывания стабилизируются с помощью аналогичных связей, которые присутствуют в третичных структурах. Однако они не такие сложные, как третичные структуры.

Это простые комбинации вторичных структур белков. Мотив сам по себе нестабилен. Более того, мотивы объясняют структуру белка, но не предсказывают функцию белка. Примерами белковых мотивов являются бета-альфа-бета-мотив, греческий ключевой мотив, бета-бочка, бетамеандровый мотив и т.д.

Домен – это фундаментальная, функциональная и трехмерная единица белка. Он выполняет определенную функцию. Один белок может иметь несколько различных доменов. Каждый домен – это независимая единица. Это шаровидная структура. Он отвечает за конкретную функцию или взаимодействие. Домены могут использоваться для предположения о функции не охарактеризованного белка. При анализе белка это важно учитывать, поскольку домены являются функциональными единицами белка (рисунок 20).

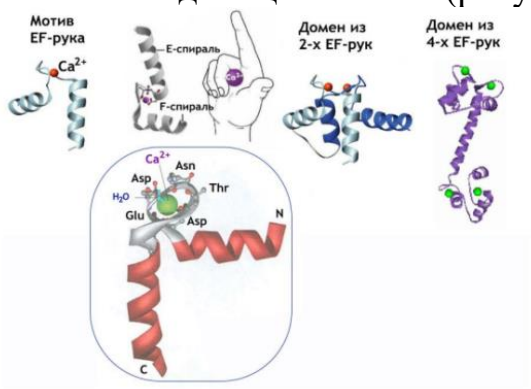


Рисунок 20 – Домен

Домены - это очень стабильные и компактные структуры. Их легко отличить от других регионов. Пируваткиназа имеет три различных домена, как показано на рисунке 20. Длина домена может варьироваться, и средняя длина составляет 100 аминокислот. Каждый домен содержит гидрофобное ядро, построенное из вторичных структурных единиц. Несколько доменов вместе составляют третичную структуру белка.

Сходство между мотивом и предметной областью:

• Мотив и домен представляют собой единицы, присутствующие в самих белковых молекулах.

• Они полезны при классификации семейств белков.

В чем разница между мотивом и доменом?

Мотив - это определенная группа супервторичных элементов белков, таких как альфа-спирали и бета-структуры, а домен – это функциональная единица белка. Кроме того, мотив является вторичной структурой, а домен отвечает за третичную структуру белка. Более того, домен – это самостоятельная единица, в отличие от мотива. Кроме того, домен отображает функцию белка, а мотив – нет. Это основные различия между мотивом и доменом. В приведенной ниже инфографике в табличной форме представлена разница между мотивом и доменом.

Мотив – это расположение вторичных структур белковой молекулы. Сам по себе он обычно нестабилен, в отличие от домена. Домен – это независимо стабильная структура белка. Следовательно, это может быть часть или целая молекула белка. Это трехмерная фундаментальная функциональная единица белка. Более того, у него есть функция, и это самостоятельная единица. Мотив может быть частью домена. Но домен не может быть частью мотива. В этом разница между мотивом и доменом.

Литература: Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витебск. гос. ун-т имени П. М. Машерова, 2022. – 83 с.

Тема 4.3. Сворачивание белков, предсказание структуры белка, предсказание функции и клеточной локализации белков

Вопросы для рассмотрения:

1. Сворачивание белков.
2. Цель и методы предсказания структуры белка.
3. Алгоритмы предсказания вторичной структуры белков. Предсказание функции и клеточной локализации белков.
4. Методы прогнозирования структуры белков, основанные на гомологии, мотивах последовательностей, структуре белка, геномном контексте.

Краткое содержание лекции

Сворачивание белков (фолдинг) – это процесс сворачивания полипептидной цепи в правильную пространственную структуру. Для обеспечения фолдинга используется группа вспомогательных белков под названием шапероны.

Предсказание структуры белка – направление молекулярного моделирования, предсказание по аминокислотной последовательности трехмерной структуры белка (вторичной, третичной или четвертичной). Данная задача является одной из самых важных целей биоинформатики и теоретической химии. Данные, полученные при помощи предсказания, применяются в медицине (например, в фармацевтике) и биотехнологии при создании новых ферментов.

Цели предсказания:

– предсказание вторичной структуры проще чем, предсказание третичной;

– аккуратное предсказание может упростить предсказание третичной структуры;

– на основе вторичной структуры можно предположить функцию белка;
– классификация белков;
– предсказание изменения структуры при функционировании белка.

Основные методы:

• статистические: Chou-Fasman method, GOR I-IV
• методы ближайших соседей: NNSSP, SSPAL
• нейронные сети: PHD, Psi-Pred, J-Pred
• HMM.
73

Основные предположения:

• последовательность содержит достаточно информации для предсказания;
• боковые группы определяют структуру;
• окно в 13-17 остатков достаточно для предсказания;
• основы для выбора размера окна:
• α-спирали 5-40 остатков
• β-тяжи 5-10 остатков.

Алгоритмы предсказания вторичной структуры – это набор методов предсказания локальной вторичной структуры белков, основанных только на знании об их аминокислотной последовательности. Для белков предсказание состоит в соотнесении отдельных участков аминокислотной последовательности с наиболее вероятными классами вторичных структур, таких, как α-спирали, β-тяжи или петли. Точность предсказания определяется, как соотношение количества аминокислот, для которых предсказанный структурный класс совпал со структурным классом, определенным для этой аминокислоты алгоритмом DSSP (или похожим алгоритмом, к примеру, алгоритмом STRIDE), к общему числу аминокислот в последовательности. Эти алгоритмы производят разметку аминокислотной последовательности белка в соответствии с принадлежностью аминокислот к одному из классов вторичной структуры, различающихся специфическими паттернами водородных связей и наборами двугранных углов. Для DSSP это 8 классов, которые можно объединить в три группы: 3 класса спиралей (α-спираль, π-спираль и 3-спираль), два класса β-структур (изолированные β-мостики и β-листы) и три вида петли (повороты, изгибы и неклассифицированные элементы, отвечающие характеристикам петли). Чаще всего для оценки качества структуры используют упрощенную классификацию, в которой классы внутри этих трех групп считаются тождественными. Алгоритмы предсказания вторичной структуры белка можно условно разделить на группы, основываясь на принципах, лежащих в их основе. Эти группы включают в себя статистические методы, методы ближайших соседей, методы, использующие нейронные сети, методы опорных векторов и методы, основанные на скрытых марковских моделях.

Ниже рассмотрим некоторые из этих алгоритмов.

Статистический метод Чоу-Фасмана основан на расчете оценки вероятности принадлежности определенной аминокислоты к определенному классу вторичной структуры в базах данных. Предсказание делается относительно трех классов вторичных структур: петли, β-листа и поворота. Цель алгоритма - найти участок из определенного для каждого класса вторичной структуры количества идущих подряд аминокислот, для каждой из которых оценка вероятности принадлежности к этому классу вторичной структуры больше заданного значения. На выход такие алгоритмы выдают предсказанные таким образом участки для каждого из трех основных классов вторичных структур, картированные на последовательность.

Первый этап метода ближайших соседей (алгоритм NNSSP) заключается в поиске гомологичной последовательности, для которой известна трехмерная структура. Учитывая локальные структурные особенности определенного аминокислотного остатка в трехмерной структуре гомологичной последовательности, такие, как доступность для растворителя,

полярность и вторичная структура, каждому аминокислотному остатку присваивается «класс окружения». Оценка вероятности принадлежности аминокислоты в центре исследуемого сегмента длиной n аминокислот к определенному классу вторичной структуры рассчитывается как логарифм частоты нахождения этой аминокислоты в окружении, к которому относится большинство ее соседей, в базах данных.

Один из алгоритмов, использующих нейронные сети, PSIPRED, включает в себя четыре основных этапа: генерация позиционной весовой матрицы с помощью PSI-BLAST, первичное предсказание вторичной структуры и дальнейшая фильтрация предсказаний. Второй и третий этапы задействуют две нейросети. Для определения принадлежности аминокислоты к определенному классу вторичной структуры на вход первой нейронной сети подается фрагмент позиционной весовой матрицы размером 33x21, соответствующий фрагменту исходной последовательности в 33 аминокислоты с аминокислотой интереса по центру. Эта сеть имеет два скрытых слоя и три выходных узла, соответствующих трем предсказываемым классам вторичной структуры. Вторая нейронная сеть используется для фильтрации предсказаний первой сети и также обладает тремя выходного узлами для каждого класса вторичной структуры в центральной позиции исследуемого окна. На выход алгоритм выдает разметку аминокислотной последовательности по элементам вторичной структуры.

Помимо вышеописанного, классические алгоритмы с использованием скрытых марковских моделей, такие как алгоритм прямого-обратного хода, алгоритм Витерби и алгоритм Баума-Велша, могут быть оптимизированы для соотнесения аминокислотной последовательности с классами вторичных структур.

Наилучшие современные методы определения вторичной структуры белка достигают около 80% точности. Точность ныне существующих методов предсказания вторичных структур оценивается такими еженедельно обновляющимися ресурсами, как LiveBench и EVA.

Предсказание функции белка - определение биологической роли белка и значения в контексте клетки. Предсказание функций проводится для плохо изученных белков или для гипотетических белков, предсказанных на основе данных геномных последовательностей. Источником информации для предсказания могут служить гомология нуклеотидных последовательностей, профили экспрессии генов, доменная структура белков, интеллектуальный анализ текстов публикаций, филогенетические и фенотипические профили, белок-белковые взаимодействия.

Функция белка - очень широкий термин: роли белков варьируются от катализа биохимических реакций до передачи сигнала и клеточного транспорта, и один белок может играть определенную роль в нескольких клеточных процессах. В целом, функцию можно рассматривать как «все, что происходит с белком или с его помощью». Проект «Генная Онтология» предложил полезную классификацию функций, в основе которого лежит список (словарь) четко сформулированных терминов, разделенных на три основные категории – *молекулярные функции, биологические процессы* и *клеточные компоненты*. Из этой базы данных можно по названию белка или его идентификационному номеру найти присвоенные ему термины «Генной Онтологии» или аннотации, сделанные на основе расчетных или экспериментальных данных.

Несмотря на то что на сегодняшний день для экспериментального доказательства функций белка используются такие современные методы, как анализ микрочипов, РНК-интерференция и двугибридный анализ, технологии секвенирования продвинулись настолько, что темпы экспериментально доказательной характеристики открытых белков сильно отстают от темпов открытия новых последовательностей. Поэтому аннотирование новых белковых последовательностей будет в основном осуществляться путем предсказания на основе вычислительных методов, так как таким образом можно осуществлять характеристику последовательностей гораздо быстрее и одновременно по нескольким генам/белкам. Первые методики предсказания функций были основаны на сходстве гомологичных белков с известными функциями (так называемое *предсказание функций, основанное на гомологии*).

Дальнейшее развитие методов привело к появлению *предсказаний на основе геномного контекста* и *на основе структуры белковой молекулы*, что позволило расширить спектр получаемых данных и комбинировать методики, основанные на разных типах данных, для получения наиболее полной картины роли белка. Ценность и производительность вычислительного предсказания функции генов подчеркивает тот факт, что по состоянию на 2010 год 98% аннотаций Генной Онтологии были сделаны на основе автоматического извлечения из других баз аннотаций и только 0,6% - на основе экспериментальных данных.

Методы, основанные на гомологии

Белки, имеющие сходные последовательности, как правило, являются гомологичными и, стало быть, имеют сходную функцию. Поэтому в недавно секвенированных геномах белки обычно аннотируют по аналогии с последовательностями схожих белков из других геномов. Однако не всегда близкородственные белки выполняют одну и ту же функцию, например, дрожжевые белки Gal1 и Gal3 являются паралогами с 73% и 92% сходства, приобретшие в ходе эволюции очень разные функции: так, Gal1 является галактокиназой, а Gal3 - индуктором транскрипции. К сожалению, нет четкого порога степени сходства по последовательности для безопасного предсказания функций; многие белки с одинаковой функцией имеют едва обнаруживаемые сходства, тогда как встречаются очень схожие по последовательности, но совершенно разные по функциям (рисунок 21).

Рисунок 21 – Часть множественного выравнивания белковых последовательностей гемоглобина из четырых разных организмов. Белки, сходные по последовательности, могут иметь также и сходную функцию

Методы, основанные на мотивах последовательностей

Развитие таких баз данных белковых доменов, как Pfam позволяет находить в искомой последовательности уже известные домены для предположения возможных функций. В ресурсе dcGO содержатся аннотации как к отдельным доменам, так и супра-доменам (т.е. комбинациям из двух или более последовательно расположенных доменов), что позволяет сделать предсказание более приближенным к реальности. Также, внутри самих белковых доменах содержатся более короткие характерные последовательности, связанные с определенными функциями (так называемые мотивы), наличие которых в искомом белке можно определить поиском в базах данных мотивов, таких как PROSITE. Мотивы также могут быть использованы для предсказания внутриклеточной локализации белка: наличие особых коротких сигнальных пептидов предопределяет, в какие органеллы белок будет транспортирован после синтеза, и было разработано множество ресурсов для определения таких сигнальных последовательностей, например, SignalP, который обновлялся несколько раз по мере развития методов. Таким образом, некоторые особенности функции белков можно предсказать без сравнения с полноразмерными гомологичными последовательностями.

Методы, основанные на структуре белка

Поскольку 3D-структура белка, как правило, является более консервативной, чем белковая последовательность, сходство структур может указывать на сходство и функций белков. Было разработано много программ для поиска похожих укладок внутри базы данных белковых структур (Protein Data Bank), например, FATCAT, CE, DeepAlign. В случае, когда для искомой белковой последовательности нет решенной структуры, сначала составляют

вероятную трехмерную модель последовательности, на основе которой в дальнейшем делается предсказание функции белка; так работает, например, сервер по предсказанию функции белка RaptorX. Во многих случаях вместо структуры всего белка, поиск ведется по структурам отдельных мотивов, содержащим, например, сайт связывания лиганда или активный сайт фермента. Для аннотации последних в новых белковых последовательностях была разработана база данных Catalytic Site Atlas (рисунок 22).

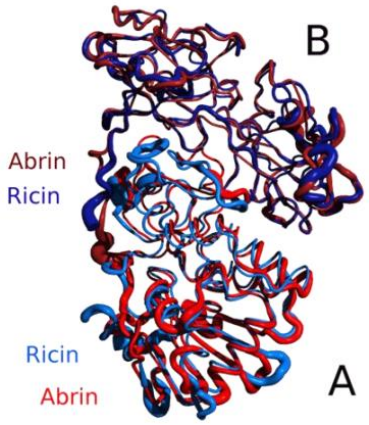


Рисунок 22 – Пространственное выравнивание двух беловых токсинов рицина и абрина

Методы, основанные на геномном контексте

Многие из недавно появившихся методов прогнозирования основаны не на сравнении последовательностей или структуры, как описанные ранее, а на корреляции между новыми генами/белками и уже аннотированными: для каждого гена составляется филогенетический профиль (по наличию или отсутствию в различных геномах), которые затем сравнивают для установления функциональных связей (предполагается, что гены с одинаковыми профилями функциональны связаны друг с другом). В то время, как методы на основе гомологии часто используются для установления молекулярных функций, предсказание на основе геномного контекста может быть использовано для предположения биологического процесса, в котором участвует белок. Например, белки, участвующие в одном и том же пути передачи сигнала, имеют общий для всех видов геномный контекст.

Литература: Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витебск. гос. ун-т имени П. М. Машерова, 2022. – 83 с.

Тема 6.1. Иерархические и диффузные системы мозга

Вопросы для рассмотрения:

1. Кратковременные и долговременные изменения в синапсах.
2. Иерархические и диффузные системы мозга.
3. Стереотаксис.

Краткое содержание лекции

Кратковременные и долговременные изменения в синапсах

Говоря о длительности хранения информации, т. е. памяти, будем опираться на механизмы этого процесса: электрофизиологические процессы, биохимические реакции и структурные изменения в нейронах и синапсах ЦНС. При этом значение временного параметра (длительность хранения информации) также сохраняется, т. е. можно говорить о кратковременной, промежуточной и долговременной памяти.

Кроме того, в раннем онтогенезе следует выделить память по механизму импринтинга, которая связана с экспрессией в нейронах мозга специфических ранних генов cfos и cjun, функцией которых является перестройка работы генетического аппарата нервных клеток под влиянием запечатлеваемого воздействия. По механизму импринтинга у взрослых животных

запечатляется действие жизненно значимых подкрепляющих факторов. По мере индивидуального развития механизм импринтинга все больше уступает место другим механизмам памяти.

Кратковременная (электрофизиологическая) память. Этот вид памяти формируется на основе непосредственного сенсорного отпечатка внешнего мира. Запоминание любой информации начинается с электрофизиологических процессов в нейронных сетях головного мозга (возникновение ВПСП – возбуждающего постсинаптического потенциала, ПД – потенциала действия, выделение различных медиаторов в синаптическом аппарате мозга). Это свидетельствует о том, что в основе механизма кратковременной памяти лежит импульсная активность нейронов и, в частности, циркуляция возбуждения по замкнутым нейронным цепям. Считают, что в замкнутых нейрональных цепочках циркуляция длится минутами, сохраняя информацию в виде последовательности импульсов, передающихся от нейрона к нейрону. Известно, что, пока циркуляция продолжается, сохраняется нейроналъный след о воздействии того или иного раздражителя на организм в прошлом.

Промежуточная (нейрохимическая) память. Это процесс перевода кратковременной памяти в долговременную (консолидация памяти), который продолжается несколько часов. Следы кратковременной памяти становятся устойчивыми примерно через 4 ч. В консолидации памяти (промежуточная память) важную роль играют нейропептиды. Известно, что пептиды могут находиться в пресинаптических терминалях в качестве сопутствующего медиатора. Например, вместе с норадреналином часто выделяются нейропептид У, опиоидные пептиды, соматостатин. Дофамин обычно выделяется окончаниями аксонов вместе с холецистокинином, энкефалином; ацетилхолин с вазоактивным интестинальным пептидом (ВИП), энкефалином, люлиберином; серотонин с веществом Р, тиреолиберином, холецистокинином.

Нейропептиды оказывают пре- и постсинаптическое модулирующее действие. Доказано, что выделение пептидов в пресинаптических окончаниях зависит от частоты работы нейрона, при этом избыточное выделение пептида-спутника всегда наблюдается при усилении активности нейронов. Пептид-спутник может значительно повысить сродство рецептора постсинаптической мембраны к основному медиатору. Например, ВИП усиливает сродство к ацетилхолину более чем в 10 000 раз.

Повышение содержания нейропептидов обнаружено в цереброспинальной жидкости у животных при обучении. Это дает основания ставить вопрос о возможности переносов выработанных навыков от одного животного к другому (Г. Унгар) с помощью образованных в ходе обучения и «отвечающих» за соответствующий навык специфических пептидов. Гипотеза Г. Унгара основана на фактах накопления в мозге особого пептида, кодирующего страх темноты и способного транспортировать этот страх от обученных доноров к необученным реципиентам. Пептид с такими свойствами скотофобин (в пер. с греч. – боязнь темноты) был синтезирован и использовался в различных лабораториях для проверки способности переноса химическим путем определенной информации. Было установлено, что скотофобин действительно вызывает стрессовую реакцию у животных, когда они находятся в темноте.

Установлено, что животные с врожденным дефектом выработки вазопрессина не способны к образованию оборонительных навыков. Выработка навыка восстанавливается у них только при дополнительном введении вазопрессина. Характерно, что при этом у животных страдает не сам процесс обучения, а именно консолидация сформировавшихся энграмм (следов, оставляемых в мозге тем или иным событием, в частности при обучении). Другой олигопептид окситоцин нарушает сохранение выработанных навыков независимо от типа обучения животных. Нейропептиды в механизмах памяти тесно взаимодействуют в синапсах с медиаторами, и после обучения кругооборот катехоламинов в мозге увеличивается. Заметно улучшают память эндогенные опиатные пептиды эндорфины и энкефалины.

Долговременная (нейроструктурная) память. Основой долговременной памяти на морфофизиологическом уровне являются структурные изменения в нейронах. Ее отличают

длительность (часы, дни и на протяжении всей жизни при повторении информации) и практически безграничный объем. Долговременная память по своему механизму качественно отличается от кратковременной и промежуточной памяти, так как не нарушается при многих экстремальных воздействиях на мозг – механической травме, электрошоке, наркозе и др. Долговременная память формируется на основе кратковременной и промежуточной памяти, при этом важную роль играют синаптические процессы. Под влиянием обучения в ЦНС образуются новые синапсы, увеличиваются их размеры и количество выделяемых медиаторов, разрастаются дендриты, увеличивается число шипиков на них, количество коллатералей аксонов – все эти явления в течение месяца исчезают.

В Белорусском государственном университете на кафедре биофизики физического факультета под руководством академика НАН Беларуси С. Н. Черенкевича проводилось исследование информационных и нейросенсорных аспектов функционирования нейронных ансамблей in vitro их многоканальной стимуляции. Выдвинуто предположение, что для синапсов так называемых коллатералей Шаффера в области СА1 гиппокампа реализуется правило обучения, основанное на эффекте синаптического депотенциирования, что приводит к возникновению конкурентных взаимодействий между различными группами синапсов. Эти данные могут быть использованы при разработке новых теорий, описывающих роль гиппокампа в формировании памяти, а также открывают возможность моделирования в условиях in vitro базовых процессов обучения нейросети.

Одни и те же медиаторы могут оказывать разные эффекты в процессах усвоения и хранения информации. Серотонин, в частности, ускоряет обучение и удлиняет сохранение навыков при положительном эмоциональном подкреплении, например пищевом, и блокирует у животных выполнение оборонительных навыков. Норадреналин ускоряет обучение в условиях применения отрицательного подкрепления (электрокожного).

После открытия способа кодирования генетической информации в ДНК (генетической памяти) и успешного изучения иммунологической памяти были предприняты попытки отыскать молекулярные основы нейронной памяти, т. е. возможного нервного субстрата энграммы. Открытие структурных изменений нейронов в ходе формирования памяти дало основание предполагать ключевую роль синтеза белка в консолидации памяти.

К отличительным особенностям генома нервных клеток относится их высокий по сравнению с другими тканями уровень функционально активных (транскрибируемых) последовательностей ДНК. Для нервных клеток характерно также прогрессирующее увеличение в них (в течение индивидуальной жизни) числа открытых для синтеза уникальных кодонов ДНК, чего не происходит в тканях других органов. В частности, у эмбриона человека в возрасте 22 недель число активных генов в нервной клетке составляет 8,2%, у взрослого человека эта величина достигает 24,6%, а в некоторых зонах мозга – 38%, тогда как в мышцах с возрастом она не меняется. Транскрибируемость ДНК и ее синтез в нейронах увеличиваются при обучении животных и содержании их в условиях информационно обогащенной среды. Блокада синтеза ДНК или РНК препятствует переходу кратковременной памяти в долговременную в результате нарушения синтеза белка.

Иерархические системы мозга. Мозг – это упрощенное название центральной нервной системы (ЦНС). Различают спинной мозг и головной мозг. Существует также термин костный мозг, который не имеет отношения к нервной системе. Головной мозг – центральный орган нервной системы. Говорить о наличии головного мозга в строгом смысле можно только применительно к позвоночным, начиная с рыб. Однако несколько вольно этот термин используют для обозначения аналогичных структур высокоорганизованных беспозвоночных – так, например, у насекомых головным мозгом называют скопление ганглиев окологлоточного нервного кольца. При описании более примитивных организмов говорят о ***головных ганглиях,*** а не о мозге. Структурная организация нервной системы в различных иерархических масштабах приведена на рисунке 23 из статьи члена-корреспондента РАН Г. Р. Иваницкого и соавт.

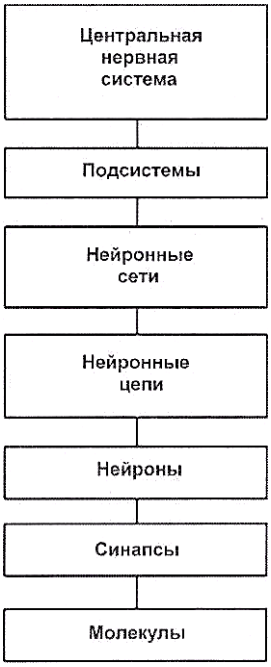


Рисунок 23 – Структурная организация нервной системы в различных иерархических масштабах

Головной мозг высших позвоночных организмов состоит из ряда структур: коры больших полушарий, базальных ганглиев, таламуса, мозжечка, ствола мозга. Эти структуры соединены между собой нервными волокнами (проводящие пути). Часть мозга, состоящая преимущественно из клеток, называется серым веществом, из нервных волокон – белым веществом. Белый цвет – это цвет миелина, вещества, покрывающего волокна. Демиелинизация волокон приводит, как мы видели выше, к тяжелым нарушениям (в спинном мозге – амиотрофический боковой склероз, в головном – рассеянный склероз).

Клетки мозга включают ***нейроны*** (клетки, генерирующие и передающие нервные импульсы) и ***глиальные клетки***, выполняющие важные дополнительные функции. (Можно считать, что нейроны являются паренхимой мозга, а глиальные клетки – стромой.) Нейроны делятся на ***возбуждающие*** (т. е. активирующие разряды других нейронов) и ***тормозные*** (препятствующие возбуждению других нейронов). Форма и размеры нейронов головного мозга очень разнообразны, в каждом его отделе разные типы клеток. Различают ***принципиальные нейроны,*** аксоны которых передают импульсы другим отделам, и ***интернейроны***, осуществляющие коммуникацию внутри каждого отдела. Примерами принципиальных нейронов являются пирамидные клетки коры больших полушарий и клетки Пуркинье мозжечка. Примерами интернейронов являются корзиночные клетки коры. Кроме нейромедиаторов нейроны мозга реагируют на гормоны. Одна из основных эндокринных желез организма, гипофиз, находится в головном мозге. Гормоны гипофиза управляют функциями других желез.

Функции мозга включают обработку сенсорной информации, поступающую от органов чувств, планирование, принятие решений, управление движениями, положительные и отрицательные эмоции, внимание, память. Мозг человека выполняет высшую функцию – мышление. Одной из важнейших функций мозга человека является восприятие и генерация речи. Поток сигналов к головному мозгу и от него осуществляется через спинной мозг, управляющий телом, и через черепномозговые нервы. Сенсорные (или афферентные) сигналы поступают от органов чувств в подкорковые (т. е. предшествующие коре полушарий) ядра, затем в таламус, а оттуда в высший отдел – кору больших полушарий. Кора состоит из двух полушарий, соединенных между собой пучком нервных волокон. Левое полушарие

ответственно за правую половину тела, правое – за левую. У человека правое и левое полушарие имеют разные функции.

Зрительные сигналы поступают в зрительный отдел коры (в теменной доле), тактильные в соматосенсорную кору (в теменной доле), обонятельные – в обонятельную кору и т. д.

В ассоциативных же областях коры происходит интеграция сенсорных сигналов разных типов (модальностей). Моторные области коры (первичная моторная кора и другие области лобных долей) ответственны за регуляцию движений. Префронтальная кора (развитая у приматов) отвечает за мыслительные функции. Области коры взаимодействуют между собой и с подкорковыми структурами – таламусом, базальными ганглиями, ядрами ствола мозга и спинным мозгом. Каждая из этих структур, хоть и более низкая по иерархии, выполняет важную функцию, а также может действовать автономно. Так, в управлении движениями задействованы базальные ганглии, красное ядро ствола мозга, мозжечок и другие структуры, в эмоциях – амигдала, в управлении вниманием – ретикулярная формация, в краткосрочной памяти – гиппокамп.

С одной стороны, существует локализация функций в отделах головного мозга, с другой – все они соединены в единую сеть.

Исследования, проведенные в последние годы в Институте мозга человека РАН (Санкт-Петербург), позволили определить, какие области мозга отвечают за осмысление различных особенностей воспринимаемой человеком речи: за грамматику, синтаксис, орфографию и др.

Стереотаксис. Одно из самых современных направлений в нейробиологии – ***стереотаксис*** – это медицинская технология, обеспечивающая возможность малотравматичного, щадящего, прицельного доступа к глубоким структурам головного мозга и дозированное воздействие на них. В Институте мозга человека РАН, по словам его директора члена-корреспондента РАН С. В. Медведева, родилось самое современное направление этой нейрохирургии будущего – компьютерный стереотаксис с программно-математическим обеспечением, которое осуществляется на ЭВМ. Совместно с коллегами из ЦНИИ «Электроприбор» была создана и впервые в России серийно выпускалвсь компьютеризированная стереотаксическая система, которая по ряду основных показателей превосходит аналогичные зарубежные образцы. В Институте мозга человека РАН стереотаксис применяется при лечении больных, страдающих двигательными нарушениями (болезнью Паркинсона, хореей Гентингтона и др.), эпилепсией, неукротимыми болями (в частности, фантомно-болевым синдромом), некоторыми психическими нарушениями. Кроме того, стереотаксис используется для уточнения диагноза и лечения некоторых опухолей головного мозга, для лечения гематом, абсцессов, кист мозга.

Распределенная (диффузная) система мозга. ***Согласно эволюционной теории мозг развивается в филогенезе путем последовательного прибавления передних отделов.*** По этой теории, каждое новое добавление или увеличение сопровождалось выработкой все более сложного поведения и в то же время создавало необходимость в регуляции деятельности более каудальных и примитивных отделов и, по-видимому, более примитивного поведения, которым они управляют. Как полагают, ***распад этой иерархии выявляется при заболеваниях или повреждениях головного мозга*** у человека и при повреждениях или перегрузке нервной оси у экспериментальных животных.

Достижения последних десятилетий требуют новых формулировок, которые включали бы иерархический принцип организации мозга и в то же время выходили за его пределы. Среди них выделяется представление о том, что головной мозг является комплексом широко взаимосвязанных систем и что динамическое взаимодействие нервной активности в пределах систем и между ними составляет самую сущность функции мозга. Межсистемные и внутрисистемные макро- и микросвязи образовались в филогенезе, как полагают, на основе эволюционных принципов. Они детерминированы генетически, но на уровне ультраструктуры

и молекулярных процессов могут быть до известной степени модифицированы в ходе постнатального развития.

Согласно этой идее, крупные структуры в головном мозге, известные как ядра (или области) новой коры, лимбическая доля, дорсальный таламус и т. п., сами состоят из повторяющихся локальных нервных цепей. Эти цепи образуют модули, которые в разных местах варьируют по числу клеток, внутренним связям и способу обработки, но которые в пределах данной структуры в основном сходны. Каждый модуль – это локальная нервная цепь, которая обрабатывает информацию, передает ее со своего входа на выход и при этом подвергает ее трансформации, определяемой общими свойствами структуры и ее внешними связями. Модули объединяются в структуры – например, в ядра или в области коры – общей или доминирующей связью, потребностью в наложении функции на определенное топографическое представительство или каким-нибудь иным фактором. Группа модулей, составляющая структуру, сама может быть разбита на подгруппы разными связями с обособленными таким же образом подгруппами в других крупных структурах. Тесно и многократно взаимосвязанные подгруппы модулей в разных и часто далеко отстоящих друг от друга структурах образуют, таким образом, точно связанные, но распределенные системы. ***Такая распределенная (диффузная) система предназначена для обслуживания распределенной функции.*** Один модуль структуры может быть членом нескольких систем. Только в пограничном случае все модули совокупности могут иметь одинаковые связи.

Литература: Дромашко, С. Е. Очерки биоинформатики / С. Е. Дромашко ; НАН Беларуси, Ин-т генетики и цитологии ; науч. ред. Л. В. Хотылева. – Минск : Беларус. навука, 2009. – 400 с.

Тема 6.2. Нейроинформатика. Моделирование процессов мышления и человеко-машинного общения. Искусственный интеллект

Цель:

Вопросы для рассмотрения:

1. Математический нейрон Маккалока-Питтса.
2. Линейная передаточная функция. Сигмоидальная передаточная функция. Логистическая функция.
3. Гиперболический тангенс. Моделирование формальных логических функций.
4. Различия между биологическим и искусственным нейроном. Нейронные сети.
5. Приложение искусственных нейронных сетей к распознаванию образов. Будущее нейрокомпьютинга.

Краткое содержание лекции

Нейроинформатика. Моделирование процессов мышления и человеко-машинного общения. Искусственный интеллект

Нейроинформатика – это наука, изучающая нейроподобные способы обработки информации при помощи компьютеров. На Западе ее еще называют ***computational neuroscience*** (вычислительная нейронаука). Свой вклад в становление нейроинформатики внесли биология и физиология высшей нервной деятельности, психология восприятия, дискретная математика, статистическая физика, синергетика и, конечно, кибернетика и компьютерное моделирование.

Искусственный нейрон называют иначе ***математическим нейроном Маккалока–Питтса.*** Согласно определению, это – узел искусственной нейронной сети, представляющий собой упрощенную модель естественного нейрона. С математической точки зрения – это сумматор всех входящих сигналов, применяющий к полученной сумме некоторую простую, в

общем случае, нелинейную функцию, называемую функцией срабатывания или передаточной функцией. Полученный результат посылается на единственный выход. Такие искусственные нейроны объединяют в сети – соединяют выходы одних нейронов с входами других, при этом полученные связи ранжируются (т. е. коэффициент связи может быть разным). На рисунке 23 проведено сопоставление биологического и искусственного нейронов.

Математическая модель искусственного нейрона была предложена У. Маккалоком и В. Питтсом вместе с моделью сети, состоящей из этих нейронов. Авторы показали, что сеть на таких элементах может выполнять числовые и логические операции. Практически сеть была реализована Фрэнком Розенблаттом в 1958 г. как компьютерная программа, а впоследствии как электронное устройство – перцептрон. Первоначально нейрон мог оперировать только с сигналами логического нуля и логической единицы, поскольку был построен на основе биологического прототипа, который может пребывать только в двух состояниях – возбужденном или невозбужденном. Развитие нейронных сетей показало, что для расширения области их применения необходимо, чтобы нейрон мог работать не только с бинарными, но и с непрерывными (аналоговыми) сигналами. Такое обобщение модели нейрона было сделано Б. Уидроу и М. Е. Хоффом, которые предложили в качестве функции срабатывания нейрона использовать логистическую кривую.

Связи, по которым выходные сигналы одних нейронов поступают на входы других, часто называют синапсами по аналогии со связями между биологическими нейронами. Каждая связь характеризуется своим весом. Связи с положительным весом называются возбуждающими, а с отрицательным – тормозящими. Нейрон имеет один выход, часто называемый аксоном по аналогии с биологическим прототипом. С единственного выхода нейрона сигнал может поступать на произвольное число входов других нейронов.

Математически нейрон представляет собой, так называемый, ***взвешенный сумматор***, единственный выход которого определяется через его входы и матрицу весов следующим образом (рис. 24):

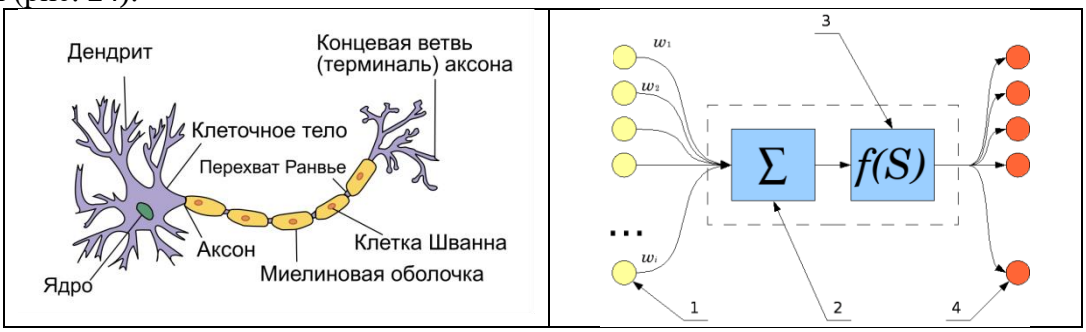


Рисунок 24 – Биологический и искусственный нейроны. Типичная структура биологического нейрона (а). Биологический нейрон состоит из тела диаметром от 3 до 100 мкм, содержащего ядро и органеллы, и отростков (дендритов и аксонов). Аксон – обычно длинный отросток, приспособленный для проведения возбуждения от тела нейрона. Дендриты – как правило, короткие и сильно разветвленные отростки, служащие главным местом образования влияющих на нейрон возбуждающих и тормозных синапсов (разные нейроны имеют различное соотношение длины аксона и дендритов). Нейрон может иметь несколько дендритов и обычно только один аксон. Один нейрон может иметь связи с 20 тысячами других нейронов. Кора головного мозга человека содержит 10–20 миллиардов нейронов. Схема искусственного нейрона (б): 1 – нейроны, выходные сигналы которых поступают на вход данному; 2 – сумматор входных сигналов; 3 – вычислитель передаточной функции; 4 – нейроны, на входы которых подается выходной сигнал данного; 5 – w_1 – веса входных сигналов.

Здесь x_1 и w_1 – соответственно сигналы на входах нейрона и веса входов. Возможные значения сигналов на входах нейрона всегда лежат в интервале [0,1], они могут быть либо

дискретными (ноль или единица), либо аналоговыми. Дополнительный вход x_0 и соответствующий ему вес используется для инициализации нейрона. Под инициализацией подразумевается смещение активационной функции нейрона по горизонтальной оси, т. е. формирование порога чувствительности нейрона. Кроме того, иногда к выходу нейрона специально добавляют некую случайную величину.

Передаточная функция определяет зависимость сигнала на выходе нейрона от взвешенной суммы сигналов на его входах. В большинстве случаев она является монотонно возрастающей и имеет область значений [-1,1] или [0,1], однако существуют исключения. Также для некоторых алгоритмов обучения сети необходимо, чтобы она была непрерывно дифференцируемой на всей числовой оси. Искусственный нейрон полностью характеризуется своей передаточной функцией. Использование различных передаточных функций позволяет вносить нелинейность в работу нейрона и в целом нейронной сети.

В основном нейроны классифицируют на основе их положения в топологии сети. При этом различают следующие.

- ***Входные нейроны*** – принимают исходный вектор, кодирующий входной сигнал. Как правило, эти нейроны не выполняют вычислительных операций, а просто передают полученный входной сигнал на выход, возможно, усилив или ослабив его.
- ***Выходные нейроны*** – представляют из себя выходы сети. В выходных нейронах могут производиться какие-либо вычислительные операции.
- ***Промежуточные нейроны*** – выполняют основные вычислительные операции.

К основным типам передаточных функций относятся следующие.

Линейная передаточная функция. Сигнал на выходе нейрона линейно связан со взвешенной суммой сигналов на его входе:

$$f(x) = Ax. \quad (2)$$

В искусственных нейронных сетях со слоистой структурой нейроны с передаточными функциями такого типа, как правило, составляют входной слой. Кроме простой линейной функции могут быть использованы ее модификации. Например, полулинейная функция (если ее аргумент меньше нуля, то она равна нолю, а в остальных случаях ведет себя как линейная) или шаговая (линейная функция с насыщением), которую можно выразить формулой:

$$f(x) = \begin{cases} 0 & \text{if } x \le 0, \\ 1 & \text{if } x \ge 1, \\ x & \text{else.} \end{cases} \quad (3)$$

Недостатками шаговой и полулинейной активационных функций относительно линейной можно назвать то, что они не являются дифференцируемыми на всей числовой оси, а значит, не могут быть использованы при обучении по некоторым алгоритмам.

Пороговая передаточная функция. Представляет собой перепад. До тех пор пока взвешенный сигнал на входе нейрона не достигает некоторого уровня, Г-сигнал на выходе равен нолю. Как только сигнал на входе нейрона превышает указанный уровень – выходной сигнал скачкообразно изменяется на единицу. Самый первый представитель слоистых искусственных нейронных сетей – перцептрон состоял исключительно из нейронов такого типа. Математическая запись этой функции выглядит так:

$$f(x) = \begin{cases} 1 & \text{if } x \ge T, \\ 0 & \text{else.} \end{cases} \quad (4)$$

Здесь $T = -w_{0}x_{0}$ – сдвиг функции активации относительно горизонтальной оси, соответственно под x следует понимать взвешенную сумму сигналов на входах нейрона без учета этого слагаемого. Ввиду того что данная функция не является дифференцируемой на всей оси абсцисс, ее нельзя использовать в сетях, обучающихся по алгоритму обратного распространения ошибки и другим алгоритмам, требующим дифференцируемоемости передаточной функции.

Сигмоидальная передаточная функция. Один из самых часто используемых на данный момент типов передаточных функций. Введение функций сигмоидального типа было обусловлено ограниченностью нейронных сетей с пороговой функцией активации нейронов – при такой функции активации любой из выходов сети равен либо нолю, либо единице, что ограничивает использование сетей не в задачах классификации. Использование сигмоидальных функций позволило перейти от бинарных выходов нейрона к аналоговым. Функции передачи такого типа, как правило, присущи нейронам, находящимся во внутренних слоях нейронной сети.

Логистическая функция. Математически эту функцию можно выразить так:

$$f(x) = \frac{1}{(1 + \exp(-Ax))}. \quad (5)$$

Здесь A – это параметр функции, определяющий ее крутизну. Когда A стремится к бесконечности, функция вырождается в пороговую. При A = 0 сигмоида вырождается в горизонтальную прямую на уровне 0,5. Очевидно, что область значений данной функции находится в интервале [0; 1], никогда не достигая нуля или единицы. Важным достоинством этой кривой является простота ее производной:

$$\frac{\partial f(x)}{\partial x} = Af(x)(1 - f(x)). \quad (6)$$

То, что производная этой функции может быть выражена через ее значение, облегчает использование этой функции при обучении сети по алгоритму обратного распространения. Особенностью нейронов с такой передаточной характеристикой является то, что они усиливают сильные сигналы существенно меньше, чем слабые, поскольку области сильных сигналов соответствуют пологим участкам характеристики. Это позволяет предотвратить насыщение от больших сигналов.

Гиперболический тангенс. Математически его можно выразить так:

$$f(x) = \frac{\exp(Ax) - \exp(-Ax)}{\exp(Ax) + \exp(-Ax)}. \quad (7)$$

Существенным отличием гиперболического тангенса от логистической кривой является то, что его область значений лежит в интервале [–1; 1]. Использование гиперболического тангенса также обусловлено простотой его производной.

Перечисленные выше функции составляют лишь часть от множества передаточных функций, используемых на данный момент. В число других передаточных функций входят также, как:

Экспонента $f(x) = \exp(-Ax)$;

тригонометрический синус;

модульная $f(x) = |x|$;

квадратичная и др.

Моделирование формальных логических функций. Нейрон с пороговой передаточной функцией может моделировать различные логические функции. Изображения иллюстрируют, каким образом можно, задав веса входных сигналов и порог чувствительности, заставить нейрон выполнять конъюнкцию (логическое «И») и дизъюнкцию (логическое «ИЛИ») над входными сигналами, а также логическое отрицание входного сигнала. Этих трех операций достаточно, чтобы смоделировать абсолютно любую логическую функцию любого числа аргументов (рис. 25).

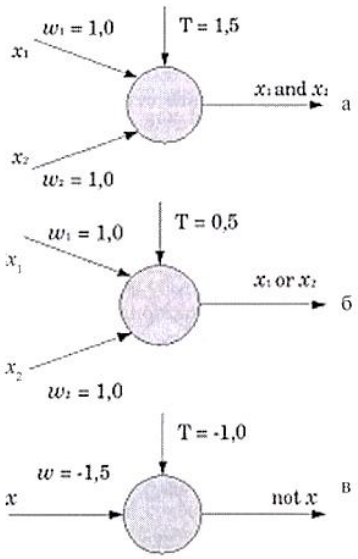


Рисунок 25 – Схемы нейронов, настроенных на моделирование логических «И» (а), «ИЛИ» (б), «НЕТ» (в)

Различия между биологическим и искусственным нейроном. Нейронные сети, построенные на искусственных нейронах, обнаруживают некоторые признаки, которые позволяют сделать предположение о сходстве их структуры со структурой мозга живых организмов. Тем не менее, даже на низшем уровне искусственных нейронов имеются существенные различия. Например, искусственный нейрон является безынерционной системой, т. е. сигнал на выходе появляется одновременно с возникновением сигналов на входе, что совсем нехарактерно для биологического нейрона.

Нейронные сети. Один нейрон способен выполнять простейшие процедуры распознавания, однако сила нейронных вычислений заключается в соединении нейронов в сети, которые обладают, как правило, большими вычислительными возможностями. Существуют сети различных конфигураций, какие только можно себе представить. Однако многослойная организация нейронов копирует слоистые структуры определенных отделов мозга. Именно для таких сетей в последние годы были разработаны алгоритмы для их обучения. Многослойные сети могут образовываться каскадами слоев. Выход одного слоя является входом для последующего слоя. Пример двухслойной сети приведен на рисунке 26.

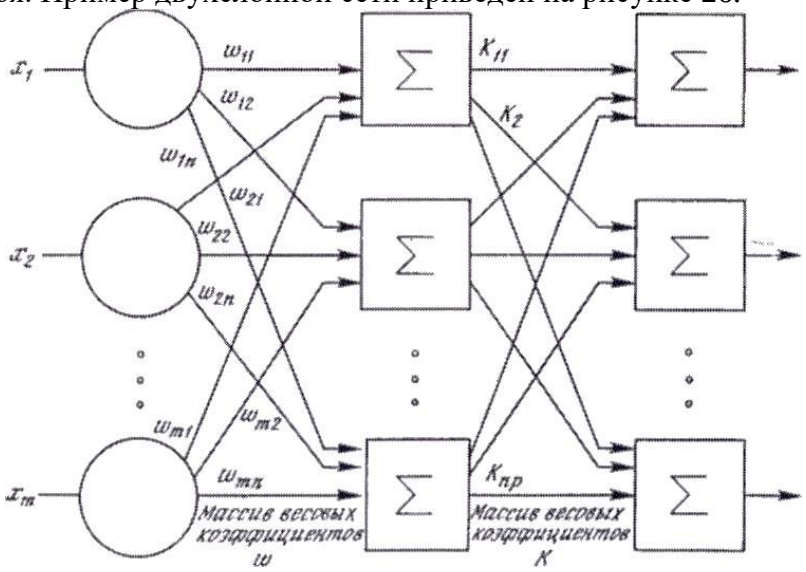


Рисунок 26 – Двухслойная нейронная сеть

Можно показать, что любая многослойная линейная сеть может быть заменена эквивалентной однослойной сетью. Таким образом, для расширения возможностей многослойных сетей по сравнению с однослойной сетью необходима нелинейная активационная функция. Только это может дать увеличение вычислительной мощности по сравнению с однослойной сетью.

Нами рассмотрены сети, не имеющие обратных связей, т. е. соединений, идущих от выходов одного слоя к входам этого же слоя или предшествующих слоев. Этот класс сетей называется ***сетями без обратных связей*** или ***сетями прямого распространения.*** У сетей без обратных связей нет памяти, их выход полностью определяется текущими входами и значениями весов. Сети более общего вида, имеющие соединения от выходов к входам, называются ***сетями с обратными связями.*** В некоторых конфигурациях сетей с обратными связями предыдущие значения выходов возвращаются на входы, т. е. выход определяется как текущим входом, так и предыдущими выходами. Поэтому сети с обратными связями могут обладать свойствами, сходными с кратковременной памятью человека.

Наиболее интересное свойство искусственных нейронных сетей – это их способность к обучению. Этот процесс очень похож на процесс интеллектуального развития человеческой личности. Возможности обучения искусственных нейронных сетей ограничены, однако уже получены убедительные достижения, например «говорящая сеть Сейновского». Возникают и другие практические применения.

Различают стратегии обучения: с «учителем» и «без учителя». ***«Обучение с учителем»*** предполагает, что для каждого входного вектора существует целевой вектор, представляющий собой требуемый выход. Вместе оба таких вектора называются ***обучающей парой.*** Обычно нейросеть обучается на некотором числе таких обучающих пар. Предъявляется выходной вектор, вычисляется выход нейросети и сравнивается с соответствующим целевым вектором, разность (ошибка) с помощью обратной связи подается в сеть и веса изменяются в соответствии с алгоритмом, стремящимся минимизировать ошибку. Векторы обучающего множества предъявляются последовательно, вычисляются ошибки и веса подстраиваются для каждого вектора до тех пор, пока ошибка по всему обучающему массиву не достигнет приемлемо низкого уровня.

Обучение с учителем имеет многочисленные прикладные достижения, для очень многих практических задач удается добиться на удивление высокого качества работы сети (порядка 95% и выше), однако с биологической точки зрения является мало правдоподобным. Вряд ли в реальном мозге можно найти обучающий механизм, который бы сравнивал желаемые и действительные значения выходов, выполняя коррекцию с помощью обратной связи**.** Большинство современных алгоритмов обучения основано на концепции Д. Хэбба, представляющей модель обучения без учителя, в которой синаптическая сила (вес) возрастает, если активированы оба нейрона: источник и приемник. Отсюда следует, что часто используемые пути в сети усиливаются; это дает объяснение реально существующего в природе феномена привычки и обучения через повторение.

Приложение искусственных нейронных сетей к распознаванию образов (нейрокомпьютинг). Распознавание образов – это одна из трудноформализованных задач, решение которой можно выполнить, используя нейросеть. Нейронные сети – мощный метод для решения задач распознавания образов в ситуациях, когда в экспериментальных данных отсутствуют значительные фрагменты информации, а имеющаяся информация предельно зашумлена. Компьютеры с нейронными сетями представляют собой машины, работающие аналогично тому, как по нашим современным представлениям функционирует мозг. О нейрокомпьютинге впервые заговорили в 1940-е годы, затем после разработок 1950–1960-х годов наступил период затишья, расцвет этого научного направления пришелся на 1983–1986 гг. При этом важную роль сыграли работы группы ***PDP (Parallel Distributed Processing***. В них

рассматривались нейронные сети, названные многослойными перцептронами, которые оказались весьма эффективными для решения задач распознавания, управления и предсказания. Высокая степень параллельности, допускаемая при реализации нейросистем, обеспечивает обработку недоступных оператору объемов информации за время, меньшее или сравнимое с допустимым временем измерений.

Пусть существует конечное множество графических образов, которые нужно распознать, и соответствующие им двоичные коды желаемых выходов (идентификаторов). В совокупности мы получили обучающее множество, в котором каждому графическому образу соответствует двоичный идентификатор. Спроецируем каждый образ на панель (рис. 27) и сопоставим ему двоичный код – это код графического образа.

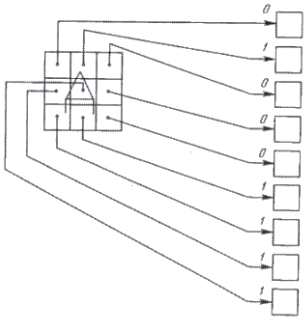


Рисунок 27 – Распознавание изображения

Возьмем какую-нибудь многослойную нейросеть обратного распространения и проведем обучение этой нейросети: подадим на вход нейросети считанный с панели двоичный код графического образа, а на выход – желаемый идентификатор; будем повторять этот процесс до тех пор, пока нейросеть не обучится распознавать все обучающее множество.

После этого нейросеть можно считать способной распознавать образы из обучающего множества без искажений, а также образы, похожие на те, которые предъявлялись нейросети при обучении. Решение этой задачи действительно интересно, так как ранее такие действия мог выполнять только человек.

Будущее нейрокомпьютинга. Можно сказать, что мы становимся свидетелями небывалого взрывного роста в сфере нейрокомпыотерных технологий. Сегодня уже открыто немало новых возможностей в таких областях, как аппроксимация функций, идентификация систем и анализ данных. Эти работы являются важным вкладом в науку и имеют большое экономическое значение. Однако, несмотря на успехи в прикладной области, создание искусственного интеллекта, принципиально похожего на человеческий, пока потерпело неудачу. Несмотря на то, что обширные работы по нейронному моделированию ведутся уже более 50 лет, нет ни одной области мозга, где процесс обработки информации был бы понятен до конца. Также ни для одного нейрона в мозге пока не определен код, посредством которого передается информация в виде последовательности импульсов.

По словам члена-корреспондента РАМН К. В. Анохина (Институт нормальной физиологии им. П. К. Анохина РАМН), неудачи классического искусственного интеллекта, или, как его называют на Западе, GOFAI – Good Old Fashioned Artificial Intelligence (старый добрый искусственный интеллект), связаны с серьезными изъянами методологического характера. Разработчики искусственного интеллекта игнорировали тот факт, что человеческий интеллект является продуктом преемственной эволюции усложнявшегося адаптивного поведения. Мозг человека состоит из 10^{11} нейронов, образующих между собой примерно 10^{14} связей и имеющих гиперастрономическое число возможностей взаимодействия. Поэтому начинать надо было не с моделирования такой сверхсложной системы, а с более простой, взяв за основу мозг обезьяны, кошки или даже дрозофилы. При этом фундаментальные механизмы психологических функций мозга должны быть сформулированы в терминах, не зависящих от

конкретного строения мозга того или иного организма, несущего эти функции. С компьютерной точки зрения теория высших функций мозга должна быть построена на базе языка, независимого от какой-либо конкретной платформы (см. интервью с К. В. Анохиным в журнале «Компьютера онлайн» от 23 октября 2002 г.

Споры между сторонниками и противниками применимости моделирования деятельности мозга идут много лет. В табл. 4 из работы сведены их аргументы и контраргументы.

Таблица 4 – Аргументация сторон в споре о роли моделирования в нейронауках

Тезис	Антитезис
Поскольку материал, на основе которого сделан компьютер, не имеет отношения к вычисляемым им функциям (процедура вычисления зафиксирована лишь в программе), то модель-программа в принципе может воспроизводить информационную работу мозга с любой заданной программистом точностью. Все существующие ограничения и дефекты программ – явление временное, базирующееся на недостаточности наших знаний и неизбежных пока упрощениях	Программа компьютера может манипулировать лишь с символами. Мозг же работает не с символами, а со смысловыми ситуациями. Элементы сознательного разума обладают семантическим (смысловым) содержанием, обеспечивающим выживание организма. Например, мозг распознает контуры хищника в зашумленной среде и мгновенно выбирает правильную реакцию на его появление: проводит различия между съедобными и несъедобными объектами мира; между половым партнером и другими животными: короче, выбирает смысловое поведение в сложной внешней среде. Машина же оперирует с символами, перекладывая их по определенным правилам из ячейки в ячейку. Это есть синтаксис без семантики
Утверждение о том, что система «искусственного интеллекта» должна непременно обладать всеми свойствами нашего мозга, абсурдна. Требовать полного соответствия по всем параметрам – все равно, что добиваться от искусственного летательного аппарата (на том основании, что он должен летать) необходимости нести яйца. Нам пока малоизвестно о том, в чем именно состоит процесс мышления и семантика, поэтому всякая уверенность по поводу того, какие свойства здесь существенны, преждевременна	Мозг – это прежде всего «биохимическая машина», манипулирующая с молекулами. Именно особые свойства биохимических молекул закладывают основу эффекта сознания, описанную в определенных специфических молекулярных гормонально-рецепторных терминах. Деятельность мозга строится на основе таких понятий, как боль, жажда, радость, возбуждение. Эта деятельность формируется в иерархической системе «снизу вверх» от молекулярного уровня до уровня целостного мозга. Например, чувство жажды, по крайней мере в некоторых случаях, обусловлено срабатыванием нейронов определенных типов в гипоталамусе, которое, в свою очередь, вызвано действием специфического пептида ангиотензина II

«Биохимический мозг» совершенно не обязательно должен быть единственной физической системой, способной на мыслительную деятельность. Компьютерные программы, моделирующие мозговые процессы, должны отражать лишь информационный аспект этих процессов. Моделирование не следует смешивать с полным воспроизведением. Самолет летает не потому, что в нем скопирован принцип полета птиц. Хотя можно (если нужно) представить себе компьютерную модель, отражающую воздействие пептидов на гипоталамус, которая будет точна вплоть до каждого синапса	С таким же успехом мы можем представить себе компьютерное моделирование процесса окисления углеводородов в автомобильном двигателе или пищеварительного процесса в желудке. Модель процессов, протекающих в мозгу, будет ничуть не реальнее моделей, описывающих процессы сгорания топлива или пищеварительные процессы. Нельзя привести автомобиль в движение, моделируя на компьютере окисление бензина, нельзя переварить обед, выполняя программу, моделирующую пищеварение. Моделирование мышления также не произведет нейрофизиологического эффекта мышления
Искусственный мозг» может и не пользоваться биохимическими молекулами и достичь того же эффекта. Например, можно создать микропроцессоры (и они уже созданы), которые будут по входу и выходу моделировать работу сетчатки или, например, улитки уха. Они в реальном времени реагируют на реальные сигналы: свет, звук. Такие схемы основаны на известных анатомических и физиологических свойствах сетчатки глаза кошки и ушной улитки сипухи, и их выход по выбранным параметрам сигнала чрезвычайно близок выходам органов, которые они моделируют. В микросхемах не используются никакие нейромедиаторы, следовательно, нейромедиаторы могут и не являться необходимыми элементами для достижения желаемых результатов	Добиться конкретных, а не абстрактных свойств мозга только за счет выполнения формальной программы операции с символами невозможно. Чтобы такое стало возможным, структура логических элементов должна начинаться с биологически важных молекул. Они могли эволюционно возникнуть случайно, но, раз возникнув, теперь они уже определяют мышление живых систем, так как оно основывается на структурных изменениях именно этих молекул. Любая система другой природы, возможно, и сможет мыслить, но совсем иначе, чем биологическая. Суть не в том, что современные программы делают в воспроизведении процессов мышления первые шаги, они просто находятся на другой дороге

Можно не сомневаться, что будущий прорыв в понимании биологических основ обработки информации вызовет невиданную активность в построении искусственного мозга. Этот проект очень быстро даст огромный экономический эффект, позволив человечеству строить «умные» ЭВМ, известные прежде лишь по фантастическим романам, где, в частности, описаны миниатюрные устройства, которые вживляются в головной мозг и дают возможность человеку с помощью мыслей управлять приборами и машинами. Уже сейчас первый вариант подобного устройства создан массачусетской компанией «Cyberki- netics» и даже допущен к клиническим испытаниям американским Управлением по лекарствам и пищевым продуктам (FDA). Устройство – небольшая пластинка с электродами, которая вживляется в лобную долю головного мозга человека и трансформирует нервные импульсы в электрические сигналы, передающиеся на персональный компьютер. С помощью этого чипа человек может работать за компьютером или даже управлять сложными электромеханическими устройствами, например протезами. Предполагается, что микрочип будет использоваться для улучшения качества жизни людей, больных церебральным параличом или другими заболеваниями, сопровождающимися нарушением нервно-мышечной передачи.

Литература: Дромашко, С. Е. Очерки биоинформатики / С. Е. Дромашко ; НАН Беларуси, Ин-т генетики и цитологии ; науч. ред. Л. В. Хотылева. – Минск : Беларус. навука, 2009. – 400 с.

2. ПРАКТИЧЕСКИЙ РАЗДЕЛ

2.1. Тематика и задания для практических занятий

Занятие 1.3 Статистические методы обработки медико-биологических данных.

Вопросы для изучения:

1. Биологическая статистика или биометрия.
2. Сложные системы и оптимизация эксперимента.
3. Полиномиальная модель и её адекватность.
4. Понятие линейного программирования.
5. Особенности биометрической культуры.
6. Статистика ДНК, как характеристика генома.

Учебно-методические материалы.

Биологическая статистика, или биометрия. *Биометрия* – раздел биологии, занимающийся планированием и обработкой результатов количественных экспериментов и наблюдений методами математической статистики. При проведении биологических исследований экспериментатор всегда имеет дело со статистическими вариациями частоты встречаемости или степени проявления различных признаков и свойств. Поэтому необходимо знать, каковы возможные пределы случайных колебаний изучаемой величины и являются ли наблюдаемые различия между вариантами опыта случайными или достоверными.

Математико-статистические методы, применяемые в биологии, разрабатываются иногда вне зависимости от биологических исследований, но чаще в связи с задачами, возникающими в биологии, сельском хозяйстве и медицине. Таковы работы Ф. Гальтона, внесшего большой вклад в создание корреляционного и регрессионного анализа, и К. Пирсона – основателя крупнейшей биометрической школы, проанализировавшего основные типы распределений, встречающиеся в биологии. Он также предложил один из самых распространенных статистических методов – критерий «хи-квадрат» и развил теорию корреляции. Методология современной биометрии создана главным образом Р. Фишером, который впервые показал, что планирование экспериментов и наблюдений и обработка их результатов – неразрывно связанные задачи статистического анализа. Он заложил основы теории планирования эксперимента, предложил ряд эффективных статистических методов (прежде всего, дисперсионный анализ).

При обработке результатов экспериментов и наблюдений возникают три основные статистические задачи:

- оценка параметров распределения – среднего, дисперсии и т. д.;
- сравнение параметров разных выборок;
- выявление статистических связей – корреляция.

Белорусские исследователи внесли достойный вклад в биометрию. Подтверждением тому является учебник П. Ф. Рокицкого, на котором воспитаны многие поколения биологов.

Сложные системы и оптимизация эксперимента. Со времен И. Ньютона и до начала XX в. ученые привыкли работать лишь с так называемыми хорошо организованными системами. Они были названы так потому, что в них можно легко выделить и описать с помощью небольшого числа переменных все связи между отдельными частями системы. Однако сейчас большинство таких систем, характерных для механики, физики и, отчасти, химии, уже изучено. Все чаще исследователь имеет дело с большими, или, как часто говорят, плохо организованными, системами. К системам такого типа относятся системы сложные, чье

поведение характеризуется очень большим числом переменных – такие системы, в которых не всегда можно выделить явления и процессы одной физической природы. Все биологические системы – от клетки до биогеоценоза относятся как раз к этой категории. Для количественного описания сложных систем и управления ими обычно используются два подхода – статистический и кибернетический.

Наиболее часто статистический подход применяется в научных исследованиях в уже упоминавшихся выше моделях данных, а также для оптимизации эксперимента, получения максимально достоверной информации из имеющихся данных. Возникла даже целая отрасль математической статистики – теория планирования эксперимента, часто использующая полиномиальные модели.

Фундаментальный шаг в развитии статистического направления в биологии сделал Р. Фишер, заложивший в 1920-е годы основы дисперсионного анализа и факторного планирования эксперимента. Другой подход к планированию эксперимента демонстрируют науки химико-технологического профиля, где с конца 1940-х годов начала развиваться теория оптимального эксперимента, т. е. эксперимента, поставленного в целях нахождения оптимальных условий процесса. В этом случае функция наилучшего приближения ищется в виде пол

$$Y = a_0 + \Sigma a_i x_i + \Sigma a_{ij} x_i x_j + \Sigma a_{ijk} x_i x_j x_k, \qquad (8)$$

где знак Σ означает суммирование по одному, двум или трем индексам. Отметим, что в случае многофакторного эксперимента обычно ограничиваются полиномом третьей степени, т. е. учитывают взаимодействие максимум трех факторов.

Традиционным методом изучения хорошо организованных систем является однофакторный эксперимент. В этом случае считается, что можно с любой степенью точности стабилизировать все независимые переменные, характеризующие систему. Поэтому их можно менять поочередно и изучать влияние на систему каждой из них независимо от всех остальных. Принципиально иначе обстоит дело в случае плохо организованных систем. Для них единственно верной оказывается методология многофакторного эксперимента. При этом исследователь одновременно изменяет сразу большое число переменных, но делает это не наугад, а по строго определенным правилам. Эти правила позволяют ему выбрать наилучшую стратегию эксперимента и получить оптимальные результаты.

Развитие теории планирования эксперимента диктуется стремлением повысить эффективность экспериментальных работ, культуру эксперимента. Применение методов планирования эксперимента позволяет формализовать большинство действий исследователя, подчинив их оптимальной стратегии эксперимента. Придерживаясь такой стратегии, экспериментатор получает данные об изучаемом процессе с минимальным числом опытов и максимальной степенью достоверности. Справедливость, однако, требует отметить, что по поводу логических оснований теории планирования эксперимента и правомерности ее повсеместного применения для оценки достоверности экспериментальных результатов до сих пор ведутся дискуссии. В частности, такая точка зрения излагается в брошюре Ю. И. Алимова «Альтернатива методу математической статистики».

Адекватность полиномиальной модели Полиномиальная модель не обязательно должна быть адекватной, т. е. точно соответствовать описываемой системе. Ведь назначение такой модели – оптимизация условий эксперимента. Сугубо практическая направленность теории планирования эксперимента позволяет обходиться достаточно грубыми моделями, отражающими только самые общие черты моделируемой системы. Стремление же к получению адекватной модели можно расценивать как излишнюю роскошь, не являющуюся непременным условием решения оптимизационной задачи. Иллюстрацией может служить специально сконструированный пример, взятый из нашей работы (рис. 28).

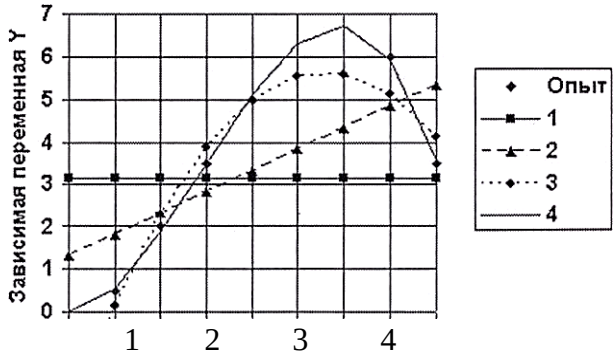


Рисунок 28 – Моделирование неизвестной функции *y{x)* полиномами разных степеней: 1) y = 3,147; 2) y = 1,315 + 1,009x; 3) y = –2,539 + 5,858x - 1,049x^2; 4) y = 0,007 + 0,130x + 2,234x^2 - 0,511x^3. Среднее отклонение полинома в заданных точках равно: 1) 1,812; 2) 1,280; 3) 0,503; 4) 0,089

Из рис. 26 видно, что уже при n = 3 дальнейшее усложнение модели за счет повышения степени аппроксимационного полинома делается неэффективным. Расчет показывает, что количество вычислений увеличивается, а точность почти не растет: так, при n = 3 среднеквадратичная ошибка аппроксимации равна 0,089, при n = 4 – 0,010, при n = 5 – 0,008. К тому же в узловых точках расхождение между экспериментальными и расчетными значениями наблюдается только во втором- третьем знаках после запятой, что много меньше точности самих экспериментальных данных. Для практических целей вполне достаточно было бы ограничиться n = 3, когда уже видна главная особенность аппроксимирующей кривой.

Линейное программирование. Теория планирования эксперимента – относительно молодая отрасль математической статистики, насчитывающая менее сотни лет. Еще более молодо линейное программирование – отрасль прикладной математики, позволяющая решать многие оптимизационные задачи медицины и биологии с так называемыми ограничениями.

В биомедицине задача линейного программирования часто формулируется как задача о диете или оптимальном рационе. Пусть имеется n видов продуктов, в которых содержится в разных количествах m видов питательных веществ. Обозначим через y_i количество купленного продукта *i*-го вида (j = 1,..., n), b_i – цену единицы *i*-го продукта, c_i – необходимый минимум *j*-го питательного вещества (j = 1,..., m), через a_{ji} – количество питательного вещества в единице *i*-го продукта. Тогда получаем систему

$$\Sigma y_i\, a_{ij} \geq c_j \ (j = 1,\ldots, m),$$
$$y_i \geq 0 \ (i = 1,\ldots, n),$$
$$\Sigma b_i\, y_i \rightarrow \min. \qquad (9)$$

Примерно так же описывается в общем виде задача об оптимизации лечения, которую мы подробно рассмотрим на примере из учебника Ю. И. Гильдермана.

Пусть имеются две возможности лечения рака – лучевая терапия и химиотерапия, и эффективность каждого из этих методов лечения можно оценить количественно в некоторых условных единицах (у. е.). Например, химиотерапевтический препарат обладает эффективностью в 1000 у. е. на единицу веса, а рентгеновское облучение – 1000 у. е. в минуту. При этом пациенту для выздоровления требуется получить не менее 3000 у. е. эффективности. Однако хорошо известно, что каждая медаль имеет свою оборотную сторону – и лекарственные препараты, и ионизирующая радиация достаточно токсичны. Примем, что токсичность лекарства равна 400 у. е. на единицу веса, а токсичность облучения – 1000 у. е. в минуту. Таким образом, ни один из двух этих методов нельзя применять неограниченно – суммарная

максимальная токсическая доза не должна превышать, допустим, 2000 у. е. Примем также, что введение больному одной весовой единицы лекарственного препарата причиняет ему в 3 раза больше неудобств, чем облучение в течение одной минуты.

Займемся формализацией модели. Если мы пациенту ввели x_t единиц лекарственного препарата (в принятых весовых единицах) и облучали его в течение x_2 минут, то причинили ему общее неудобство, равное

$$Z=3x_1+x_2. \quad (10)$$

При этом должны выполняться условия

$$1000x_1 + 1000x_2 \geq 3000, \quad (11)$$

$$400x_1 + 1000x_2 \leq 2000 \quad (12)$$

По вполне понятным причинам $x_1 \geq 0$ и $x_2 \geq 0$. Задача состоит в отыскании такого сочетания обоих методов, которое минимизировало бы неудобства, т. е. функцию z, и при этом удовлетворяло сформулированным выше ограничениям. Очевидно, что решение (искомая точка (X_1, X_2) должно принадлежать заштрихованной области на рис. 29, а если точнее, то находиться в одной из вершин треугольника. Вычислим значения функции z в этих вершинах, т. е. в точках (3, 0), (5, 0) и (5/3, 4/3). Получим $z = 9$, $z = 15$ и $z = 6{,}3$. Последнее значение и есть искомый оптимальный результат.

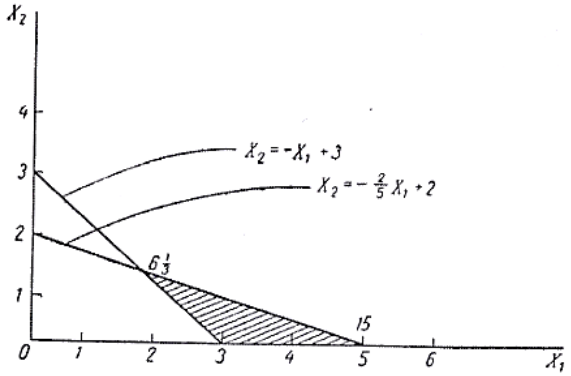


Рисунок 29 – Решение задачи о совместном применении лекарства и облучения

Разумеется, приведенный пример намеренно упрощен. В реальной ситуации может быть не один, а несколько лекарственных препаратов, а также разные способы облучения. В соответствии с этим возрастет и число возможных ограничений. Кроме того, критерием оптимальности лечения может служить его эффективность – тогда ищется не минимум функции z, а ее максимум. Однако всегда функция z (целевая, как ее называют) описывается линейным полиномом вида

$$z = a_1x_1 + a_2x_2 + \ldots + a_nx_n \quad (13)$$

Линейны и все ограничивающие ее равенства или неравенства. В этом случае мы имеем дело с ***задачей линейного программирования.*** Существуют разные методы поиска решения, в том числе, так называемый симплекс-метод, позволяющий перебирать вершины в определенном порядке, так чтобы все время двигаться в направлении оптимума (от хорошего к лучшему). Практически без помощи компьютера такую задачу решить сложно.

Особенности биометрической культуры. На рубеже 1980–1990 годов в западной биомедицине сформировалось новое научное направление – клиническая эпидемиология. Основной методологический постулат клинической эпидемиологии – ***«evidence-based medicine»,*** что буквально переводится как *«медицина, основанная на фактах»,* либо ***«научно-доказательная медицина».*** В рамках этой концепции существенная роль отводится статистическим методам на этапах планирования исследований и анализа полученных данных. При этом необходимо придерживаться ряда правил, которые действуют в большинстве ведущих медицинских и биологических журналов относительно применения математических

методов и описания результатов с использованием статистических параметров. Этой проблеме посвящено много работ, том числе на цитировавшемся выше сайте Биометрика (Томск). Важно четко представлять характер распределения, которому подчиняются экспериментальные данные, и в случае его несовпадения с нормальным применять статистические методы, на нем не основанные. В частности, нельзя применять такие статистические показатели, как среднее и среднеквадратичное отклонение. Некорректно также для описания дисперсии применять стандартную ошибку среднего и т. д. Имеются свои правила для описания точности количественных данных и сравнения результатов парных наблюдений, а также для графической демонстрации статистических зависимостей.

Статистика ДНК как характеристика генома

В основе статистического анализа ДНК лежат методы математической оценки сходства и различия анализируемых последовательностей. При этом, как при проведении любого типа статистического анализа, для получения обоснованных выводов требуется некая совокупность оцениваемых параметров, то есть объемная выборка. В данном случае выборка представляет собой набор последовательностей одного и того же гена (или белка), принадлежащих разным организмам. Обычно в выборку включают последовательности не только одного гена, но и одного ранга, т.е. последовательности разных особей одного вида или последовательности разных видов.

Формирование такой выборки может быть выполнено разными способами. Новые последовательности исследователи получают в результате специфичной амплификации (в процессе полимеразной цепной реакции или ПЦР) целевого участка генома интересующего их объекта и последующей расшифровкой (секвенированием) продукта ПЦР. Дополнительно, для проведения широких сравнений, последовательности могут быть получены из одной из публичных баз данных. Чаще всего с этой целью используется база данных GenBank NCBI, которая содержит наибольшее число депонированных последовательностей.

Для работы с последовательностями ДНК используют специализированное программное обеспечение, позволяющее просматривать последовательности, сравнивать их, проводить разнообразные статистические расчеты. Количество таких программ значительно, и большинство из них предоставляют исследователю сходный набор возможностей, различаясь только качествами интерфейса и количеством дополнительных опций. Чрезвычайно часто используемой является программа MEGA (Molecular Evolutionary Ge-netics Analysis), которая, благодаря ее авторам, доступна в интернете для свободной и бесплатной загрузки.

Статистический анализ последовательностей или, **«статистика ДНК»**, это сравнительно новый подход к исследованию геномов, возникший на стыке генетики, математики и информатики. Проведение статистического анализа последовательностей ДНК невозможно без использования компьютерной техники и специального программного обеспечения, созданного специально для работы с большими объемами генетических текстов.

Анализ нуклеотидной последовательности ДНК с использованием методов математической статистики предоставляет исследователю колоссальный объем информации, касающейся частот встречаемости конкретных нуклеотидов и кодонов, скорости и особенностей нуклеотидных замещений в гене, сходства и различия «рисунка» нуклеотидных заменив последовательностях ДНК разных организмов. Однако, как и при использовании любого аналитического метода, статистический анализ ДНК требует тщательного выбора применяемых подходов, а также внимательной трактовки полученных результатов.

Основной задачей статистического анализа последовательностей ДНК традиционно считается оценка уровня сходства двух сравниваемых последовательностей с тем, чтобы сделать заключение о степени их эволюционной близости, т.е. родства. Сравнению могут подвергаться как идентичные гены разных организмов, так и гены, сходные по кодируемому продукту, внутри одного или нескольких геномов. В большинстве случаев наблюдаемое сходство последовательностей какого-либо участка генома или геномов говорит об общности

их происхождения (гомологии). Благодаря сравнениям такого рода удается выявить эволюционное родство организмов даже при наличии значительных различий в морфологии и образе жизни, а сделать предположения о происхождении генов и их семейств, в том числе и при изменении функции продукта этих генов.

Занятие 1.4 Пакеты компьютерных программ

Вопросы для изучения:

1. Программное обеспечение для статистической обработки биомедицинских исследований.
2. Базы и банки данных. Первичные, вторичные и курируемые базы данных.
3. Автоматическое аннотирование последовательности. Идентификаторы записей в базах данных.
4. Аналитические программы в биоинформатике.
5. Энциклопедия KEGG и ее использование.

Учебно-методические материалы.

Программное обеспечение для статистической обработки биомедицинских исследований.

MS Excel. Самое часто упоминаемое (и используемое) в отечественных статьях приложение из пакета офисных программ MS Office. Однако MS Excel – это электронная таблица с достаточно мощными математическими возможностями, где некоторые статистические функции являются просто дополнительными встроенными формулами. Расчеты, сделанные при ее помощи, не признаются авторитетными биомедицинскими журналами.

STADIA. Программа отечественной разработки. Включает в себя все необходимые статистические функции. Она прекрасно справляется со своей задачей – статистическим анализом. Однако программа внешне фактически не изменяется с 1996 г. Графики и диаграммы, построенные при помощи STADIA, выглядят в современных презентациях архаично.

SPSS (Statistical Package for Social Science). Самый часто используемый пакет статистической обработки данных. Отличается гибкостью, мощностью. Применим для всех видов статистических расчетов, используемых в биомедицине.

STATISTICA. Производителем программы является фирма StatSoft Inc. (США), которая выпускает статистические приложения начиная с 1985 г. STATISTICA включает большое количество методов статистического анализа (более 250 встроенных функций), объединенных следующими специализированными статистическими модулями: Основные статистики и таблицы, Непараметрическая статистика, Дисперсионный анализ, Множественная регрессия, Нелинейное оценивание, Анализ временных рядов и прогнозирование, Кластерный анализ, Факторный анализ, Дискриминантный функциональный анализ, Анализ длительностей жизни, Каноническая корреляция, Многомерное шкалирование, Моделирование структурными уравнениями и др. Несложный в освоении этот статистический пакет может быть рекомендован для биомедицинских исследований любой сложности. По пользованию пакетом имеется русскоязычная книга О. Ю. Ребровой, включенная ВАК России в перечень обязательной литературы при подготовке к сдаче кандидатского экзамена по биоинформатике.

SYSTAT. Статистическая система для персональных компьютеров. Последняя 11-я версия обладает неплохим интуитивно понятным интерфейсом. Компания Systat Software также разрабатывает популярные у отечественных исследователей SigmaStat и SigmaPlot, которые

являются соответственно программой статистической обработки и программой построения диаграмм. При совместной работе становятся единым пакетом для статистической обработки и визуализации данных.

STATGRAPHICS PLUS. Довольно мощная статистическая программа. Содержит более 250 статистических функций, генерирует понятные, настраиваемые отчеты. Последняя доступная версия – 5.1. Ранние версии этой программы были весьма популярны у отечественных исследователей.

Пакеты прикладных генетико-статистических программ для персональных компьютеров.

В настоящее время на рынке информационных продуктов отсутствуют современные объектно-ориентированные программные средства для обработки генетико-селекционных данных и оптимизации и ускорения процесса количественной оценки нового генофонда по показателям продуктивности. Не учитывают они и влияния факторов среды и генетико-средовых взаимодействий (общая и специфическая комбинационная способность, коэффициенты наследуемости, зависимость урожайности от эколого-генетических факторов, устойчивость к основным биотическим и абиотическим стрессам, минимизация приемов интенсификации выращивания). Имеющиеся статистические пакеты, например SYSTAT, STATGRAPHICS PLUS, STATISTICA, хорошо приспособлены к обработке медицинских и фармакологических данных, но не включают блока генетико-статистического анализа, учитывающего специфику требований селекционера.

Пакет РИШОН. К 1995 г. под операционную систему MS DOS был создан пакет прикладных генетико-статистических программ для персональных компьютеров РИШОН, в который вошло около 40 программ по различным видам биометрического анализа:

- **элементарный статистический анализ** – первичная обработка, вычисление критериев Стьюдента и Фишера, сравнение распределений, разбиение по классам;
- **корреляционный анализ** – выбор уравнения регрессии (17 различных аппроксимирующих формул, включая полином степени N), определение множественной нелинейной регрессии, вычисление корреляционного отношения, нахождение линейных корреляций, вычисление корреляций по Спирмену и т. п.;
- **дисперсионный анализ** – однофакторный, двухфакторный и трехфакторный (в том числе учет неполноблочных планов, расчет коэффициентов наследуемости);
- **многомерный анализ** – построение дендрограммы, компонентный анализ, разные виды кластерного анализа;
- **генетический анализ** – вычисление общей и специфической комбинационной способности (по четырем методам Гриффинга), оценка комбинационной способности при скрещивании с тестерами, нахождение генетических параметров по методу Хеймана, определение экологической стабильности и пластичности по Эберхарту и Расселу, вычисление путевых коэффициентов Райта и целый ряд других методов.

Пакет АБ-СТАТ. Пакет прикладных программ АБ-Стат предназначен для статистического анализа результатов селекционных, генетических и медико-биологических экспериментов и является продолжением и развитием пакета программ «Сигма», созданного Б. Ю. Аношенко для СМ ЭВМ в 1986 г. Удобство и простота пакета привели разработчика в 1991 г. к необходимости перевода его на IBM-совместимые персональные компьютеры, несмотря на наличие большого количества разнообразных «фирменных» программных продуктов по статистической обработке результатов экспериментов (Framework, Statgraf, MSTAT, Systat и др.).

АБ-Стат не является идеальным или универсальным пакетом, но имеет ряд преимуществ. В частности, простая структура обрабатываемого файла данных (обычный текстовый файл, иногда называемый ASCII файл), который может быть подготовлен в любом текстовом редакторе (Лексикон, Norton editior, Multi edit. Word и т. д.) или «экспортирован»

почти из любых других программных продуктов (Dbase, Paradox, Stat- graphics и т. п.), позволяет быстро и оперативно проводить как предварительную, так и основную обработку данных. Файл данных или результаты его обработки также могут быть быстро и легко переведены в другие программные продукты для дальнейшего анализа, графического представления или создания баз данных. Файл данных содержит краткую поясняющую информацию, что позволяет хорошо ориентироваться даже в старых файлах данных. Каждый файл может содержать по несколько задач, отдельно обрабатываемых программами, причем задача, как и сам файл, содержит поясняющую информацию.

К числу преимуществ пакета следует отнести также его диалоговые средства. В частности, все программы работают в диалоговом режиме (вопрос – ответ, краткое меню – выбор варианта). В пакете предусмотрена защита от неверного ответа. Кроме того, каждая задача файла может обрабатываться отдельно в диалоговом режиме.

Результаты по желанию пользователя выводятся на печать, экран или в файл. Форма представления результатов – в основном таблицы и графики. Ширина выводной строки – обычно 80 символов, т. е. ширина стандартного листа формата А4. Однако по указанию пользователя она может быть расширена до 250 символов. Если таблица больше указанного количества символов, то она печатается частями.

Для всех критериев сравнения (t-, *F*-, *U*-, χ^2-) и других статистических показателей (коэффициенты корреляции, асимметрии, эксцесса и т. д.), требующих определения их достоверности, находится их статистическая вероятность. Чтобы не загружать расчетные таблицы лишними цифрами, их достоверность указывается звездочками.

В данную версию пакета АБ-Стат включены следующие блоки программ, всего около 30:

- **предварительная обработка файлов данных** (проверка данных, конкатенация (объединение) данных, печать данных);
- **вычисление элементарных показателей** (средние и корреляции, ранговая корреляция по Спирмену);
- **дисперсионный анализ** (многофакторный (до 6) анализ, восстановление пропущенных данных, одно- и двухфакторный дисперсионный анализ неравномерных комплексов);
- **сравнение** (по t-критерию Стьюдента, по (U-критерию Манна–Уитнея, сравнение частот по критерию χ^2);
- **графическое представление данных** (гистограммы и другие одномерные графики, двумерные графики рассеивания);
- **регрессионный анализ** (регрессии от одного аргумента, множественная полиномиальная регрессия, путевые коэффициенты Райта);
- **многомерная классификация** (кластерный анализ признаков, кластерный анализ объектов, линейный дискриминантный анализ);
- **сервисные программы** (очистка экрана, транслитерация текстов);
- **специальные программы** (учет пестроты почвенного плодородия, анализ общей (ОКС) и специфической (СКС) комбинационной способности, определение экологической стабильности);
- **программы оптимизации селекционного процесса** (планирование селекционных скрещиваний по методу «белорусского квадрата», составление схем посева и печати журнала полевых наблюдений, создание «bsd» и «fid» файлов и проверка соответствия журнала схеме посева, оценка образцов по комплексу признаков в селекционных питомниках, создание файла средних значений по линиям, комбинациям скрещивания, родительским формам).

Программа BIODIS. При обработке экспериментальных данных, как правило, предполагается, что они подчиняются нормальному, или Гауссову, распределению. В крайнем

случае используются еще биномиальное распределение, а для редких событий или малых выборок – соответственно распределения Пуассона и Стьюдента. Однако необходимо учитывать, что в целом ряде случаев биологические данные могут отвечать другим распределениям, например распределению Максвелла, Шарлье и т. д. Создана программа *BIODIS (BIO- metrical Distribution)* для персональных ЭВМ, дающая в руки биологов удобный в пользовании инструмент для быстрой и надежной оценки вида распределения экспериментальных данных вне зависимости от их характера. Программа написана на Паскале, размер ее exe-файла 136 КБ.

Программа позволяет сделать выбор между семью следующими распределениями: нормальное, биномиальное, Пуассона, t-распределение (Стьюдента), Максвелла, геометрическое, равномерное. При этом учитывается характер экспериментальных данных, т. е. величина выборки (больше или меньше 20 измерений в обрабатываемом массиве) и наличие так называемых выбросов (или грубых ошибок измерений), так что экспериментатор может задать соответствующий режим обработки.

Для оценки достоверности гипотезы о виде распределения на выбор предлагаются три критерия: χ^2, Колмогорова и ω^2. При этом в программу встроены рекомендации по применению того или иного критерия согласия.

В частности, отмечается, что критерий χ^2 является стандартным в биометрии для проверки гипотезы согласия. Недостатком метода является то, что предусмотренная в нем группировка данных по классам (интервалам) приводит к некоторой потере информации. К числу его преимуществ помимо универсальности относится то, что при его применении нет необходимости учитывать точные значения наблюдений.

Однако применять этот критерий рекомендуется для выборок, чей объем превышает 50 значений.

Критерий согласия Колмогорова применяется в случаях, когда имеет место непрерывное распределение (нормальное, Максвелла, Стьюдента, равномерное). Он наиболее удобен для малых выборок, но его использование обычно затруднено большим объемом вычислений. Применение ЭВМ снимает это ограничение.

В отличие от критерия Колмогорова тест ω^2 применим как для непрерывных, так и дискретных распределений. Поскольку он работает с каждым измерением, его рекомендуется применять для выборок малого объема. Степень близости экспериментального и теоретического распределений измеряется более слабой метрикой, поэтому он слабее реагирует на экстремальные данные, т. е. он предпочтительнее для обработки данных с «выбросом».

Перспективы модернизации. В России в 1993–2000 гг. под руководством д. б. н. С. П. Мартынова был разработан пакет AGROS, содержащий более 60 специализированных программ статистического и биометрико-генетического анализа в растениеводстве и селекции, написанных под операционную систему MS DOS. В 2009 г. с российскими коллегами достигнута принципиальная договоренность о разработке на более современной платформе совместного программного продукта, объединяющего достоинства ранее созданных пакетов. В случае реализации всех планов новый пакет программ явится хорошим примером использования информационных технологий для оптимизации и ускорения селекционного процесса. Он также будет востребован в учебных, научно-исследовательских и селекционных учреждениях Беларуси и стран СНГ при подготовке специалистов, магистров и аспирантов биологического и сельскохозяйственного профилей.

Система компьютерной алгебры Mathematica. В настоящее время на Западе разработано несколько универсальных компьютерных систем, дающих специалистам-нематематикам возможность решать ряд сложных задач в области своих исследований «в формульном представлении», не вдаваясь в математические тонкости. К ним можно отнести системы Maple, MathCAD, MatLab, Mathematica. Наиболее мощной и эффективной из них является система Mathematica, разработанная в США коллективом авторов под руководством

профессора С. Вольфрама и выпускаемая фирмой Wolfram Research Inc. Эта система обладает целым набором дополнительных инструментов и специализированных пакетов, таких, как Dynamic Visualizer, Wavelet Explorer, Experimental Data Analyst, Time Series, Fuzzy Logic. Несмотря на то, что основное назначение системы – символьные вычисления, она может быть использована и как «очень большой калькулятор», и для проведения численных вычислений с любой заданной точностью. На русском языке имеется пособие по компьютерной системе Mathematica 3.0.

Компьютерные программы, используемые для анализа секвенированных последовательностей генов

Появление быстрых методов секвенирования ДНК ферментативным построением новой цепи ДНК в условиях терминации по Сэнгеру привело к резкому увеличению числа секвенированных фрагментов ДНК. Определяемые в ходе секвенирования нуклеотидные последовательности, в виде так называемых ДНКовых текстов, повлекли за собой разработку специализированных компьютерных программ по их анализу, поскольку обработка таких больших массивов данных без помощи компьютеров стала просто невозможна. Однако первые появившиеся компьютерные программы обращения с нуклеотидными последовательностями и анализа ДНК характеризовались минимальным набором сервисных функций. Стремительный рост числа разнообразных программ, рассчитанных на проведение тех или иных операций с нуклеотидными последовательностями ДНК, даже потребовал выделения отдельных номеров журнала «Nucleic Acids Research», целиком посвященных данной проблеме.

Многие компьютерные программы тех лет представляли собой небольшие программки для решения конкретных задач, вроде поиска сайтов рестрикционных эндонуклеаз, определения размеров получающихся фрагментов после расщепления ими молекул ДНК или определения молекулярной массы таких фрагментов, их нуклеотидного состава. С целью некоторого упорядочения информации обо всех этих программах и лучшей ориентации исследователей был подготовлен специальный указатель, вобравший в себя максимально возможное число известных к тому времени компьютерных программ и дающий краткое описание их возможностей. Однако становилась очевидной насущная потребность создания так называемых пакетов прикладных программ, которые бы позволяли проводить целый необходимых операций по всевозможному анализу секвенированных фрагментов ДНК, начиная от занесения последовательности нуклеотидов в компьютер до выявления особенностей кодируемых ими белков.

Другим аспектом компьютерного анализа секвенированных молекул ДНК стал вопрос хранения полученных данных и необходимость обеспечения широкого доступа ученых к уже известным нуклеотидным последовательностям. Это привело к образованию специализированных белковых данных, сначала одного, потом нескольких и уже затем многочисленных. В настоящее время, кроме трех основных, так называемых первичных банков данных (GebBank, EMBL, DDBJ), главной целью ставящих сбор нуклеотидных последовательностей, существует еще множество баз данных, преследующих какую-либо цель.

Чтобы выявить какие-то особенности или характерные черты исследуемого гена или фрагмента ДНК, необходимо провести анализ его нуклеотидной последовательности, ставшей известной в результате секвенирования. Причем, зачастую требуется всесторонний анализ, который более удобно осуществить с помощью интегрированного пакета специализированных программ. В настоящее время имеется широкий выбор различных пакетов таких программ, отличающихся требованиями, предъявляемыми ими к самой компьютерной технике, к операционным системам. В силу особенностей компьютерного парка нашей страны, представленного подавляющим количеством IBM-совместимых компьютеров, программы, написанные для других типов машин, здесь, за редкими исключениями, упоминаться не будут.

Еще один уже упоминавшийся выше модуль MapDraw из пакета программ Lasergene рассчитан на поиск в секвенированной последовательности сайтов различных рестрикционных

эндонуклеаз, поиск открытых рамок считывания и их трансляцию в белковые продукты. Построение как линейных, так и кольцевых карт сопровождается возможностью добавления большого числа поясняющих символов различных элементов. Важной чертой этого модуля является создание рисунков, пригодных для публикации. Широкие возможности анализа белков заключены в модуле Protean. Так, данная программа позволяет выявлять в исследуемых белках участки с *a*-спиральной структурой и *P*-складками и прочие элементы их структурной организации, рассчитывать гидрофильные и гидрофобные области белков, предсказывать физико-химические и электрофоретические свойства анализируемых с помощью компьютера белков.

Важным элементом анализа секвенированных последовательностей ДНК является их сравнение друг с другом или так называемое множественное выравнивание. Ранее большинство компьютерных программ позволяло одновременно сравнивать друг с другом только две последовательности, что снижало ценность получаемых результатов. В настоящее время уже многие программы анализа ДНК и белков рассчитаны на одновременный анализ большого числа родственных последовательностей. Такая же возможность реализована в модуле Meg Align пакета программ Lasegene. Причем данный пакет позволяет проводить множественное выравнивание как последовательностей ДНК, так и белков. Повышению достоверности проведенного анализа способствует возможность ручного редактирования и окончательного доведения таких данных, поскольку случается так, что компьютер в некоторых случаях дает определенные систематические ошибки. Затем на основе уже выравненных последовательностей можно реконструировать филогенетические деревья, с вычисленным процентом сходства и генетическими расстояниями. Полученные результаты возможно сохранить в виде специальных файлов, что весьма удобно.

Аналитические программы в биоинформатике

Приведем примеры основных программ сравнения аминокислотных и нуклеотидных последовательностей (рисунок 30).

Рисунок 30 – Форматы файлов, используемых в биоинформатике

1. ACT – (Artemis Comparison Tool) – геномный анализ;
2. Bio Edit – редактор множественного выравнивания аминокислотных и нуклеотидных последовательностей;
3. Bio Numerics – коммерческий универсальный пакет программ по биоинформатике;
4. BLAST – поиск родственных последовательностей в базе данных аминокислотных и нуклеотидных последовательностей;
5. ClustaIW – множественное выравнивание аминокислотных и нуклеотидных последовательностей;
6. FASTA – набор алгоритмов определения схожести аминокислотных и нуклеотидных последовательностей;
7. Mesquite – программа для сравнительной биологии на языке Java;
8. Muscle – множественное сравнение аминокислотных и нуклеотидных последовательностей. Более быстрая и точная программа в сравнении ClustaIW;

9. Pop Gene – анализ генетического разнообразия популяций;

10. Populations – популяционно-генетический анализ. Примером интегрированного инструмента биолога является также Unipro UGENE. Это свободно распространяемое программное обеспечение для работы молекулярного биолога. Пользовательский интерфейс этого продукта обеспечивает:

– с последовательностями;

– визуализацию хроматограмм;

– использование редактора множественного выравнивания последовательностей;

– просмотр трехмерных моделей PDB и MMDB с поддержкой стереорежима;

– просмотр филогенетических деревьев;

– применение конструктора вычислительных схем, автоматизирующего процесс анализа;

– поддержку сохранения изображений в векторные форматы для удобства публикаций.

Базы и банки данных

База данных – это файл специального формата, содержащий информацию, структурированную определенным образом.

Одной из базовых задач БИ является хранение и организация доступа к накопленным массивам биологической информации. Реализуют данную задачу с применением технологий баз данных (БД). Базы данных представляют собой информационные модели, содержащие данные об объектах и их свойствах. Они хранят информацию о группах объектов с одинаковыми наборами свойств. Простыми бытовыми примерами БД можно считать любые справочники, энциклопедии, записные книжки и каталоги.

Информация в базах данных хранится в упорядоченном виде, что позволяет обеспечить удобный доступ к нужным фрагментам хранимой информации.

Технически БД можно представить как набор таблиц, каждая их которых предназначена для хранения информации об объектах одного типа. Каждая строка таблицы содержит данные одного объекта и называется записью. При этом столбцы, формирующие строку, называются полями, и каждое поле описывают какую-либо характеристику объекта. Поскольку для каждой записи в БД должна существовать возможность уникальной идентификации, часто выделяют отдельный тип записи для хранения подобной информации, а данное поле называют ключевым. Также следует отметить, что возможно существование полей, содержащих в качестве значения ссылку на объект того же или другого типа, позволяя, таким образом, хранить информацию об иерархических и сетевых связях объектов содержащихся в БД.

По характеру хранимых данных биологические БД можно разделить:

• на *первичные БД*, хранящие результаты молекулярно-биологических исследований. Как правило, это последовательности и структуры биологических полимеров (Genbank, EMBL, DDBJ, SWISS-PROT, TREMBL, PIR, PDB);

• в*торичные*, данные в которых являются результатом обработки первичной биологической информации. Типичными примерами являются БД, хранящие информацию о паттернах, обнаруживаемых в последовательностях, разного рода классификации последовательностей и структур (PROSITE, Pfam, BLOCKS, PRINTS, DSSP, SCOP);

• *составные (композитные) БД*, агрегируют информацию из первых двух видов, предоставляя расширенные по сравнению с отдельными БД возможности по поиску и навигации в данных (NRDB, OWL, GO).

По механизму наполнения базы данных можно разделить:

• *на архивные базы данных*, фактически являются хранилищем файлов определенного формата, предоставляемых учеными. Как правило, это первичные базы данных наподобие GenBank, EMBL, PDB:

GenBank – база данных генетических последовательностей, основанная в 1982 г. – это аннотированная коллекция всех общедоступных последовательностей ДНК, РНК, белков, снабженных литературными ссылками. Эта база является частью объединения International Nucleotide Sequence Database Collaboration, которое объединяет 3 крупных банка нуклеотидных последовательностей:

1. DDBJ (DNA Data Bank of Japan),
2. EMBL (European Molecular Biology Laboratory)
3. GenBank (National Center for Biotechnology Information).

Эти организации ежедневно обмениваются новой информацией. Большинство журналов требуют посылки новых секвенированных последовательностей в любую из этих 3 баз данных до опубликования статьей. В статьях, посвященных очередной порции последовательностей, должен упоминаться лишь номер последовательности в базе данных GenBank.

Адрес DDBJ: http://www.ddbj.nig.ac.jp/

Адрес GenBank: http://www.ncbi.nlm.nih.gov/Genbank/

EMBL (European Molecular Biology Laboratory) – эта база данных содержит информацию о каждом фрагменте последовательностей, включая литературные ссылки, перекрестные ссылки на документы других баз данных и др.

Адрес EMBL: http://www.ebi.ac.uk/embl/

PDB (Brookhaven Protein DataBank) – содержит данные о коллекции экспериментально определенных трехмерных структур биологических макромолекул (белков и нуклеиновых кислот). С 2002 года в основном депозитарии PDB хранятся структуры, экспериментально определенные с помощью рентгеноструктурного, ядерно–магнитного резонансного и др. методов. Теоретические структуры выделены в отдельную подбазу PDB.

Адрес: http://www.rcsb.org/pdb/

• *автоматические базы данных*, представляющие результат работы какого-либо метода. Часто по предыдущей классификации их можно отнести ко вторичным (DSSP);

В таких базах данных записи генерируются (моделируются) компьютерными программами. Например, **TrEMBL (Translated EMBL)** – автоматическая база предсказаний последовательностей белков. Это формальная трансляция всех кодирующих нуклеотидных последовательностей из банка EMBL. В 2002 году в результате объединения SwissProt, TrEMBL и PIR был создан банк данных **UniProt (Universal Protein Resourse)**. Это основное хранилище белковых последовательностей и их функций. UniProt состоит из трех частей:

– UniProt Knowlegebase – является центральной базой данных и обеспечивает доступ к обширной курируемой информации по белкам, включая их функцию, классификацию и перекрестные информационные ссылки;

– UniProt Archive – UniParc. – отражает хронологию данных определения о всех белковых последовательностях;

– UniProt Reference – UniRef. – содержит базы данных, которые объединяют последовательности в кластеры для ускорения поиска.

Адрес UniProt: http://www.ebi.uniprot.org/index.shtml

• *курируемые базы данных*, наполнение которых контролируется группой/лабораторией/исследовательским центром, их поддерживающим. Типичный пример - SWISS-PROT.

Информацию для курируемых баз данных отбирают эксперты из архивных баз. К курируемым базам относятся, например, SwissProt. Эта база данных белковых последовательностей существует с 1986 года и поддерживается двумя институтами: Swiss Institute of Bioinformatics (SIB) и European Bioinformatics Institute (EBI).

Адрес: http://www.ebi.ac.uk/swissprot/

Поскольку БИ ориентирована на автоматическую обработку данных, основу большинства первичных и вторичных биологических баз данных составляют файлы

определенного формата. В каждом подобном файле хранится информация об одном основном объекте данной БД, например данные о пространственной структуре одного комплекса в случае БД PDB. Обычно пользователь редко работает с самим файлом, поскольку вебинтерфейс сайта БД предоставляет более удобное для человека представление информации об объектах в виде различного рода сводных таблиц, последовательностей символов, рисунков и ссылок на другие сайты, содержащие дополнительную связанную информацию. Однако всегда следует помнить, что, как правило, данные файлы доступны для скачивания (при необходимости).

• *Производные базы данных*. Получаются в результате компьютерной обработки данных из архивных и курируемых баз данных. Это, например, SCOP, PFAM, GO и др. SCOP (Structural Classification Of Proteins) – база данных по структурной классификации белков.

Адрес: http://scop.protres.ru/

PFAM (Protein families database of alignments and HMMs) – это большая коллекция семейств белков и доменов, построенных на основании экспертной оценки множественных выравниваний. В банке существуют две основные части: PFAMA, содержащая подробно аннотированные белковые семейства, и PFAMB, содержащая различные множественные выравнивания.

Адрес: http://www.sanger.ac.uk/Pfam/

GO (Gene Ontology consortium database). Целью создателей базы было установление контроля за единообразием в описаниях функций, биологических процессов и клеточных компонентов, относящихся к продуктам генов. Унификация описаний в различных базах данных облегчает поиск в них нужного гена. GO – независимая база данных: другие базы данных сотрудничают с ней, помещая ссылки на унифицированные термины GO, либо поддерживают поиск с использованием терминов базы GO, а также стимулируют ее дополнение и уточнение.

Адрес: http://www.geneontology.org/

• *Интегрированные базы данных***.** Объединяют информацию из разных баз. Например, введя имя гена, можно найти всю, связанную с ним информацию. К таким базам относится **ENTREZ (Molecular Biology**

DataBase and Retrieval System). Эта интегрированная база данных содержит нуклеотидные и аминокислотные последовательности, которые собираются из крупнейших специализированных хранилищ – баз данных. Основой является **GenBank**, кроме того, информация пополняется из **dbEST, dbSTS, SwissProt, PIR, PDB, PRF, GSDB**. Данные из перечисленных ресурсов поступают в интегрированную базу данных после:

1) присвоения уникального идентификатора последовательности,

2) перевода документов в единый стандарт хранения,

3) проверки данных,

4) проверки всех ссылок по базе данных MedLine,

5) проверки названий организмов по таксономической классификации GenBank Taxonomy. Адрес ENTREZ: http://www.ncbi.nlm.nih.gov/Database/index.html Описания многих баз данных по БИ можно найти на русскоязычном сайте, который находится по адресу: http://www.jcbi.ru/index.html

Автоматическое аннотирование последовательности – это процесс, в котором идентифицируются гены, их регуляторные области и функции генов, также определяют гены, которые не кодируют белки (в частности, гены рРНК, тРНК и малой ядерной РНК); обнаруживают и характеризуют мобильные генетические элементы и семейства повторов, которые могут также присутствовать в геноме.

После определения нуклеотидной последовательности встает следующая задача по ее аннотации, которая заключается в идентификации всех генов и кодируемых белков, мобильных элементов и семейств повторов, которые могут присутствовать в геноме.

Гены, кодирующие белки, обнаруживаются при анализе нуклеотидной последовательности самим исследователем или при помощи компьютерных программ. Гены, кодирующие белки, содержат так называемую **открытую рамку считывания,** которая начинается с **инициирующего кодона** АТГ и заканчивается одним из трех **терминирующих кодонов –** ТАА, ТАГ или ТГА. Сканирование последовательности ДНК для обнаружения открытой рамки считывания, ограниченной АТГ с одной стороны и стоп-кодоном с другой, является одной из *стратегий поиска генов*. Однако этот метод высоко эффективен только для аннотации бактериальных геномов. В случае же геномов эукариот продуктивность метода резко снижается, поскольку большинство эукариотических генов состоят из **экзонов** (кодирующих участков гена) и **интронов** (некодирующих участков гена), и программа часто интерпретирует экзоны, как отдельные гены, т.к. стопкодоны часто встречаются в интронах.

Следует отметить, что последние версии программ настроены на поиск специфических черт открытых рамок: интрон-экзонных сочленений, 3'полиА-сигналов и **преимущественных кодонов**. Например, аланин может кодироваться четырьмя кодонами, но в геноме человека кодон ГЦЦ встречается в 41% аланиновых кодонов, а ГЦГ только в 11%. Наиболее часто встречающиеся кодоны присутствуют в экзонах, но не встречаются в интронах и пространствах между генами.

После обнаружения предполагаемых открытых рамок считывания для определения гена проводят поиск гомологичных последовательностей среди расшифрованных генов других организмов в базах данных (например, в Genbank).

Идентификаторы записей в базах данных

Для каждой записи в БД должна существовать возможность уникальной идентификации. В качестве уникального имени обычно используется идентификатор или инвентарный помер. Подобная двойственность является следствием того, что на ранних этапах становления биоинформатики, когда число последовательностей было невелико, имена последовательностям старались давать в удобочитаемой форме, закладывая в аббревиатуру указание на биологическую функцию последовательности. Так, в идентификаторах БД EMBL и Genbank первые две (три) буквы указывают на биологический вид организма, а оставшиеся – на функцию. Однако с увеличением объема БД и возрастанием скорости добавления последовательностей подобная схема именования перестала удовлетворять потребности научного сообщества из-за отсутствия возможности автоматической генерации имен. На смену (в дополнение) к идентификаторам пришли инвентарные номера – уникальные символьные (буквы и цифры) последовательности. Следует иметь в виду, что три наиболее известные базы данных EMBL, GenBank и SwissProt используют общую схему нумерации последовательностей, т.е. единый инвентарный номер однозначно идентифицирует последовательность в этих трех базах данных.

Основные базы данных последовательностей биологических полимеров. База данных Genbank содержит аннотации последовательностей ДНК различных организмов. Каждая запись включает ряд полей, список которых несколько отличается для прокариотических и эукариотических последовательностей. Так, для записей, характеризующих прокариотические гены, свойственны следующие описания:

• LOCUS - название локуса (произвольное имя), длина нуклеотидной последовательности, тип молекулы (ДНК) и ее топология (линейная, кольцевая);

• DEFINITION - содержит короткое описание гена;

• ACCESSION - номера (идентификаторы) данного объекта в других БД;

• VERSION - перечислены синонимы и предыдущие идентификаторы;

• KEYWORDS - список терминов, характеризующих запись;

• SOURCE - общее название организма, являющегося источником данной последовательности;

• ORGANISM - полная таксономическая идентификация организма источника;

• REFERENCE - ссылки на статьи, связанные с выделением и определением функций последовательности;

• COMMENT - комментарии, не подходящие по формату другим полям.

Секция, описывающая открытую рамку считывания гена:

• координаты стартового и стоп-кодона;

• тип таблицы кодонов, используемой для трансляции;

• translation – декодированная аминокислотная последовательность Вторая важная БД, о которой стоит упомянуть, – база белковых последовательностей SWISS-Prot (wvvw.expast.org/sprot). Данная БД, в отличие БД Genbank, ориентирована на белковые последовательности, в том числе последовательности, «транскрибированные in silico». Также база SWISS-Prot содержит подробные аннотации известных последовательностей и тесно интегрирована с другими БД. Например, если для последовательности из SWISS-Prot доступна структурная информация, то данные об этой последовательности будут содержать и ее PDB-идентификатор. Формат записей в SWISS-Prot состоит из следующих полей:

• ID/AC (accession number) - название записи и инвентарный номер. Иногда в данном поле могут присутствовать несколько различных номеров;

• DT - даты создания/обновления информации о записи;

• DE - поле описания (description) перечисляет все известные имена данного белка;

• GN (gene) - название гена (генов), кодирующих данный продукт;

• OS/OC/OX - содержат название организма, таксономическую классификацию и уникальный таксономический идентификатор организма, являющегося источником данной последовательности. Секция ссылок (RN/RP/RX/RA/RT/RL) включает все литературные ссылки, использованные для аннотации данной записи;

• CC - блок комментариев. Состоит из текста, разделенного на различные «темы» и описывающие: функцию белка, его внутриклеточную локализацию, посттрансляционные модификации, возможные связи с различными заболеваниями и т.д.;

• поле DR содержит кросс-ссылки на идентификаторы данного белка в других Бд (например, PDB);

• KW - поле, включающее ключевые слова (keywords), характеризующие данную запись.

• FT (features) - самое важное поле содержит список доменов и важных сайтов последовательности с указанием номеров (интервалов) аминокислотных остатков: описания посттрансляционных модификаций, вариантов последовательностей (известные замены остатков), доменную структуру, повторы, элементы вторичной структуры и т.д.

• SQ поле содержит саму аминокислотную последовательность белка

Энциклопедия KEGG и ее использование

KEGG (Kyoto Encyclopedia of Genes and Genomes) - веб-ресурс, предоставляющий доступ к ряду биологических баз данных и инструментам для анализа биологических и медицинских данных, созданный в 1995 году в рамках проекта «Геном человека». С момента создания интегрированная база данных KEGG значительно расширилась и на данный момент насчитывает шестнадцать баз данных, для удобства поиска разделенных на четыре категории: системная информация, геномная информация, химическая информация и информация, связанная непосредственно со здоровьем человека. Также KEGG предоставляет ряд инструментов для удобной работы с базами данных и анализа хранящейся в них информации (рис. 31):

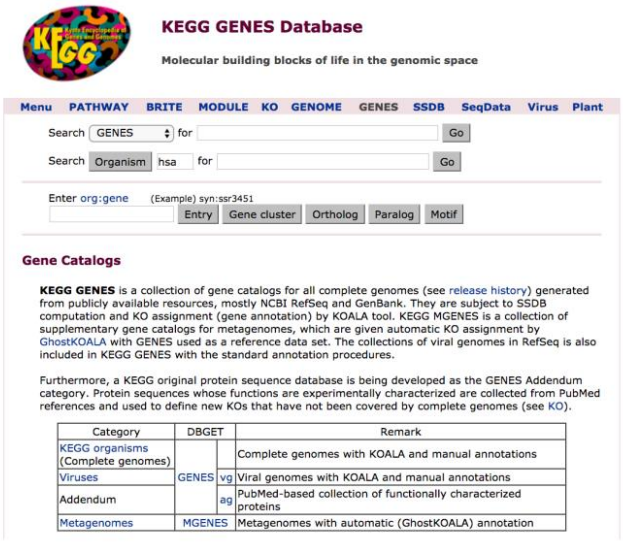


Рисунок 31 – Изображение страницы сайта энциклопедии KEGG

Веб-ресурс KEGG был создан в 1995 году в Японии при поддержке Kanehisa Laboratories. Базы данных KEGG непрерывно обновляются и дополняются. Главной целью проекта KEGG является интеграция полученной геномной информации, данных о биологических и химических процессах, происходящих в живых организмах, знаний о человеческих болезнях и открытых лекарствах в единое целое для понимания высокоуровневой организации различных биологических систем, таких как клетка, организм или целая экосистема.

Доступ к данным, хранящимся на KEGG, осуществляется с помощью веб-сайта KEGG. Главная страница сайта содержит список ссылок на основные базы данных KEGG, вторичные базы данных, созданные для удобного поиска, и различные инструменты для анализа биологических и медицинских данных. Представленные ссылки указывают на страницы с подробным описанием каждой базы данных/инструмента и с интерфейсом поиска/работы. По ссылке KEGG2 располагается страница с полным перечнем всех баз данных и программных средств ресурса KEGG, в том числе те, которые доступны на сайте GenomeNet. Поиск данных на сайте KEGG можно осуществлять разными способами: непосредственно в основных базах данных таблица 2, по субъектам таблица 3 и по организмам. Поиск по субъектам и по организмам осуществляется с помощью интерфейсов, специально созданных для упрощения работы с базами данных.

KEGG BRITE – это тотальное структурированное формализованное описание объектов и явлений биологии, отраженных в базах KEGG. До 2005 года BRITE существовал как отдельная база данных, впоследствии включенная в проект KEGG. База данных KEGG BRITE отражает онтологию - иерархическую классификацию биологических сущностей, к числу которых относятся гены, белки, организмы, патологии, лекарственные препараты, химические соединения и т. п.

KEGG MODULE –это коллекция оформленных вручную функциональных единиц, называемых модулями KEGG, которые используются для аннотации и биологической интерпретации отсеквенированных геномов. В этой базе лежат метаболические схемы с высоким разрешением, изображающие функциональные подпути, характерные для определенных таксонов, и молекулярные комплексы, встречающиеся в этих процессах. Представлены 4 типа модулей.

KEGG GENOME – это коллекция организмов KEG с полногеномной последовательностью, каждый из которых идентифицирован трех или четырехбуквенным кодом, и некоторых вирусов, имеющих отношение к болезням.

Занятие 2.1 Парное и множественное выравнивание. Алгоритм и программы выравнивания

Вопросы для изучения:

1. Выравнивание последовательностей.
2. Алгоритм Нидлмана-Вунша.

Учебно-методические материалы.

Литература для изучения

1. Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витеб. гос. ун-т, 2022. – 83 с.
2. Биоинформатика : учеб.-метод. комплекс / сост.: Д. Н. Дроздов, С. А. Зятьков, Г. Г. Гончаренко. – Гомель : Гомел. гос. ун-т, 2020. – 86 с.
3. Дромашко, С. Е. Мозг, интеллект, нейроинформатика : учеб.-метод. пособие / С. Е. Дромашко, Р. В. Телятников ; НАН Беларуси, Ин-т подгот. науч. кадров, Каф. естественнонауч. дисциплин. – Минск : ИПНК НАН Беларуси, 2010. – 54 с.

Занятие 2.2 Эволюция молекул и организмов

Вопросы для изучения:

1. Пути эволюции молекул и организмов.
2. Молекулярная эволюция, как наука.
3. Критерии сравнения нуклеотидных и белковых последовательностей.

Учебно-методические материалы.

Литература для изучения

1. Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витеб. гос. ун-т, 2022. – 83 с.
2. Биоинформатика : учеб.-метод. комплекс / сост.: Д. Н. Дроздов, С. А. Зятьков, Г. Г. Гончаренко. – Гомель : Гомел. гос. ун-т, 2020. – 86 с.
3. Дромашко, С. Е. Мозг, интеллект, нейроинформатика : учеб.-метод. пособие / С. Е. Дромашко, Р. В. Телятников ; НАН Беларуси, Ин-т подгот. науч. кадров, Каф. естественнонауч. дисциплин. – Минск : ИПНК НАН Беларуси, 2010. – 54 с.

Занятие 3.3 Нуклеиновый состав (изохоры, GC-острова) ДНК

Вопросы для изучения:

1. Композиционная гетерогенность ДНК.
2. Изохоры. Связь семейств изохор с кодирующими последовательностями ДНК и их экспрессией.

Учебно-методические материалы.

Литература для изучения

4. Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витеб. гос. ун-т, 2022. – 83 с.
5. Биоинформатика : учеб.-метод. комплекс / сост.: Д. Н. Дроздов, С. А. Зятьков, Г. Г. Гончаренко. – Гомель : Гомел. гос. ун-т, 2020. – 86 с.

Дромашко, С. Е. Мозг, интеллект, нейроинформатика : учеб.-метод. пособие / С. Е. Дромашко, Р. В. Телятников ; НАН Беларуси, Ин-т подгот. науч. кадров, Каф. естественнонауч. дисциплин. – Минск : ИПНК НАН Беларуси, 2010. – 54 с.

Занятие 3.4 Подбор праймеров для ПЦР

Вопросы для изучения:

1. Значение праймеров для процесса амплификации и ПЦР.
2. Принципы подбора праймеров.
3. Требования, предъявляемые к праймерам

Учебно-методические материалы.

Литература для изучения

1. Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витеб. гос. ун-т, 2022. – 83 с.
2. Биоинформатика : учеб.-метод. комплекс / сост.: Д. Н. Дроздов, С. А. Зятьков, Г. Г. Гончаренко. – Гомель : Гомел. гос. ун-т, 2020. – 86 с.
3. Дромашко, С. Е. Мозг, интеллект, нейроинформатика : учеб.-метод. пособие / С. Е. Дромашко, Р. В. Телятников ; НАН Беларуси, Ин-т подгот. науч. кадров, Каф. естественнонауч. дисциплин. – Минск : ИПНК НАН Беларуси, 2010. – 54 с.

Занятие 3.5 Анализ частоты использования кодонов

Вопросы для изучения:

1. Частота транзиций и трансверсий. Определение соотношения транзиций и трансверсий.
2. ГН-насыщенность общая и отдельных положений кодона. Зависимость ГЦ-насыщенности отдельных положений кодона от общей ГЦ-насыщенности.
3. Анализ аминокислотного состава белков. Системы классификации сайтов нуклеиновых кислот в связи с их положением в кодоне - система вырожденности и синонимичности / несинонимичности.
4. Методики изучения стратегии кодирования белков.

Учебно-методические материалы.

Литература для изучения

1. Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витеб. гос. ун-т, 2022. – 83 с.
2. Биоинформатика : учеб.-метод. комплекс / сост.: Д. Н. Дроздов, С. А. Зятьков,

Г. Г. Гончаренко. – Гомель : Гомел. гос. ун-т, 2020. – 86 с.
3. Дромашко, С. Е. Мозг, интеллект, нейроинформатика : учеб.-метод. пособие / С. Е. Дромашко, Р. В. Телятников ; НАН Беларуси, Ин-т подгот. науч. кадров, Каф. естественнонауч. дисциплин. – Минск : ИПНК НАН Беларуси, 2010. – 54 с.

Занятие 4.3 Сворачивание белков, предсказание структуры белка, предсказание функции и клеточной локализации белков

Вопросы для изучения:

1. Сворачивание белков. Цель и методы предсказания структуры белка.
2. Алгоритмы предсказания вторичной структуры белков.
3. Предсказание функции и клеточной локализации белков.
4. Методы прогнозирования структуры белков, основанные на гомологии, мотивах последовательностей, структуре белка, геномном контексте.

Учебно-методические материалы.

Литература для изучения

1. Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витеб. гос. ун-т, 2022. – 83 с.
2. Биоинформатика : учеб.-метод. комплекс / сост.: Д. Н. Дроздов, С. А. Зятьков, Г. Г. Гончаренко. – Гомель : Гомел. гос. ун-т, 2020. – 86 с.
3. Дромашко, С. Е. Мозг, интеллект, нейроинформатика : учеб.-метод. пособие / С. Е. Дромашко, Р. В. Телятников ; НАН Беларуси, Ин-т подгот. науч. кадров, Каф. естественнонауч. дисциплин. – Минск : ИПНК НАН Беларуси, 2010. – 54 с.
4. Дромашко, С. Е. Очерки биоинформатики / С. Е. Дромашко ; НАН Беларуси, Ин-т генетики и цитологии ; науч. ред. Л. В. Хотылева. – Минск : Беларус. навука, 2009. – 400 с.

Занятие 5.1 Закономерности дифференцировки соматических клеток. Органогенез. Старение и раковая трансформация. Моделирование клеточной дифференцировки

Вопросы для изучения:

1. Типы клеточных популяций. Схема развития эукариотического организма.
2. Понятие о диффероне. Апоптоз.
3. Онтогенез и старение. Новая компьютерная технология прижизненного наблюдения клеток.
4. Бета-гал тест как маркер старения.
5. Злокачественные новообразования как следствие нарушения молекулярно-генетических и клеточных регуляторных механизмов.

6. Моделирование работы генов. Два типа моделей клеточной дифференцировки.

7. К-генетические сети. Модель Сендова-Цанева.

Учебно-методические материалы.

Закономерности дифференцировки соматических клеток

Типы клеточных популяций. Клеточная популяция – группа клеток одного или нескольких типов, которая может быть охарактеризована в понятиях пространства и времени. Такое расширительное понимание термина может быть отнесено к любому клеточному сообществу – от суммы всех клеток организма на протяжении всей жизни человека, начиная от зиготы до, скажем, клеток желтого тела яичника в одной из фаз менструального цикла. Другими словами, для выделения конкретных клеточных популяций необходим дифференцирующий критерий. Термин клеточная популяция применяют по крайней мере в двух значениях – в расширительном понимании и по отношению к леблоновским популяциям.

Выделяют следующие сообщества клеток:

Клеточный тип. Характеризуется реальной экспрессией (фенотип) или потенциальной возможностью (сумма фенотипов) экспрессировать конкретный спектр генов, что и выделяет данный тип клетки (эритроидный, макрофаги и т. д.) из других типов. Эта концепция включает понятие клеточные фенотипы, а также пластичность и границы нормы клеточного типа. Клетки с идентичным набором разрешенных к экспрессии генов (вне зависимости от того, транскрибируются ли эти гены) относятся к одному клеточному типу. В организме человека насчитывают более 200 клеточных типов.

Леблоновские клеточные популяции. На основании способности к клеточному обновлению (в том числе путем пролиферации) К. Леблон в 1964 г. выделил 4 категории клеточных популяций: эмбриональная, статическая, растущая и обновляющаяся. Следует помнить, что каждая из них неоднородна.

Образующие популяцию клетки могут находиться на различных стадиях дифференцировки фазах клеточного цикла, в различном функциональном состоянии.

Статическая популяция. Ее составляет гомогенная группа клеток, не проявляющих митотической активности (например, нейроны).

Растущая популяция. В растущей популяции клетки делятся, митотическая активность постепенно затухает.

Обновляющаяся популяция. Обновляющаяся клеточная популяция характеризуется множественными митозами и быстрой гибелью клеток. При этом количество вновь образованных клеток слегка превышает клеточные потери (эпидермис, эпителий кишки, клетки тканей внутренней среды). При неоплазиях клеточная продукция намного превосходит гибель клеток, что обеспечивает быстрый рост опухоли.

Эмбриональная популяция. Развитие организма человека начинается с оплодотворения яйцеклетки и образования зиготы, которая дает начало целому организму. Оплодотворенная яйцеклетка тотипотентна – она обладает неограниченным потенциалом в том смысле, что ее одной достаточно для формирования и развития нормального плода при соответствующих условиях. В первые часы после оплодотворения она делится с образованием идентичных тотипотентных клеток, и любая из них, будучи имплантирована в матку женщины, способна дать начало развитию плода.

Схема развития эукариотического организма. Миллиарды клеток растущего организма происходят из одной клетки (зиготы), которая образуется в результате слияния мужской (сперматозоид) и женской (яйцеклетка) половых клеток (гамет). Эта единственная клетка содержит не только информацию об организме, но и схему его последовательного развития (рис. 32).

Рисунок 32 – Схема развития организма

Примерно через четверо суток после оплодотворения, когда проходит несколько циклов клеточного деления, тотипотентные клетки начинают специализироваться с образованием сферической структуры, называемой **бластоцистой**. У бластоцисты есть наружный слой и внутренняя полость, где образуется внутренняя клеточная масса. Из наружного слоя развивается плацента и другие поддерживающие структуры, необходимые для формирования плода, а из внутренней клеточной массы – практически все органы и ткани самого плода. Клетки внутренней клеточной массы плюрипотентны – их наличие является необходимым, но не достаточным условием формирования плода. Если их имплантировать в матку женщины, то беременность не наступит.

Плюрипотентные клетки подвергаются дальнейшей специализации с образованием стволовых клеток, которые дают начало еще более специализированным клеткам, обладающими специфическими функциями. Так, из кроветворных (гемопоэтических) стволовых клеток развиваются эритроциты, лейкоциты и тромбоциты, а из стволовых клеток кожи – различные типы клеток этой ткани.

Клеточный клон – группа клеток, происходящая от одной родоначальной клетки-предшественника. Представление о клоне возникло в иммунологии. При попадании в организм антигена одна иммунокомпетентная клетка усиленно размножается, и образуется большое количество одинаковых клеток (клон), способных синтезировать антитела к этому антигену Согласно клональной теории развития, структуры зародыша формируются из ограниченного количества клонов. Наконец, опухоли также развиваются как клоны, происходящие от одной трансформированной клетки.

Понятие о диффероне. **Дифферон** (гистогенетический ряд) – совокупность клеточных форм, составляющих ту или иную линию дифференцировки. В диффероне последовательно различают: стволовые клетки, клетки-предшественники, зрелые клетки, достигшие состояния окончательной (терминальной) дифференцировки.

Стволовые клетки – самоподдерживающаяся популяция клеток, способных дифференцироваться в нескольких направлениях и формировать различные клеточные типы. Стволовые клетки обладают высокими пролиферативными потенциями, но, как правило, делятся редко.

Клетки-предшественники. По мере дифференцировки их пролиферативные потенции

постепенно уменьшаются. Выделяют наиболее раннюю стадию клеток-предшественников – коммитированные, или полустволовые, клетки.

Зрелые клетки. Ими заканчивается гистогенетический ряд. Способность к пролиферации полностью исчезает.

Знание основ кинетики клеточных популяций необходимо для понимания теории регенерации, т. е, восстановления структуры биологического объекта после ее разрушения. Соответственно уровням организации живого различают клеточную (или внутриклеточную), тканевую, органную регенерацию.

Различают ***регенерацию физиологическую***, которая совершается постоянно в здоровом организме, и ***репаративную*** – вследствие повреждения. У разных тканей возможности регенерации неодинаковы.

В ряде тканей гибель клеток генетически запрограммирована и совершается постоянно (в многослойном ороговевающем эпителии кожи, в однослойном каемчатом эпителии тонкой кишки, в крови). За счет непрерывного размножения, в первую очередь полустволовых клеток-предшественников, количество клеток в популяции пополняется и постоянно находится в состоянии равновесия.

Наряду с запрограммированной физиологической гибелью клеток во всех тканях происходит и незапрограммированная – от случайных причин: травмирования, интоксикаций, воздействий радиационного фона. Хотя в ряде тканей запрограммированной гибели нет, но в течение всей жизни в них сохраняются стволовые и полустволовые клетки. В ответ на случайную гибель возникает их размножение, и популяция восстанавливается.

У взрослого человека в тканях, где стволовых клеток не остается, регенерация на тканевом уровне невозможна, она происходит лишь на клеточном уровне.

«Лимит Хайфлика». Основываясь на опытах с культурами раковых (иммортализированных) клеток, биологи долгое время были уверены, что в оптимальных условиях бесконечно долго могут делиться и нормальные клетки животных и человека как в культуре, так и в организме. Однако в начале 1960-х годов Л. Хайфлик и П. С. Мурхед установили, что в клеточных культурах нормальные диплоидные (соматические клетки) человека способны делиться лишь ограниченное число раз. Предельное число делений («лимит Хайфлика») сильно зависит от возраста индивидуума, которому эти клетки изначально принадлежали. Так, клетки новорожденных делились в культуре 80–90 раз, а 70-летнего человека – только 20–30 раз. Достигнув «лимита Хайфлика», клетки переходят в состояние одряхления – *сенесенса* (от англ. senescence), которое характеризуется резким изменением метаболизма, прежде всего нарушением репликации ДНК. Вслед за этим состоянием обычно следует гибель клеток.

Запрограммированная гибель клеток (апоптоз). Апоптоз (по греч. – опадание листьев) – явление программируемой клеточной смерти, форма гибели клетки, проявляющаяся в уменьшении ее размера, конденсации и фрагментации хроматина, уплотнении наружной и цитоплазматической мембран без выхода содержимого клетки в окружающую среду. Кроме проявления указанного выше набора характерных цитологических признаков (маркеров апоптоза) сопровождается рядом молекулярных процессов. Апоптоз запускается специфическими сигналами и избавляет организм от нежизнеспособных либо патологических клеток. В частности, в геноме имеются специальные гены, контролирующие программу быстрого и эффективного разрушения клетки, индукция апоптоза связана с особым *геном-«киллером» p53*, а подавление – с геном-*«спасителем» bcl-2*. Продукт *p53* гена следит за целостностью генома при митозе. При нарушении целостности генома клетка переключается на апоптоз, тогда как белок *bcl-2* ингибирует апоптоз. Таким образом, недостаток *p53* или избыток *bcl-2* приводит к накоплению клеток – такие нарушения наблюдаются в различных опухолях. Ежедневно около 5% клеток организма подвергаются апоптозу, а их место занимают новые клетки. В процессе апоптоза клетка исчезает бесследно в течение 15– 120 мин. Апоптоз

имеет ряд существенных отличий от **некроза** – патологического процесса местной гибели ткани в живом организме в результате какого-либо экзо- или эндогенного ее повреждения (рис. 33).

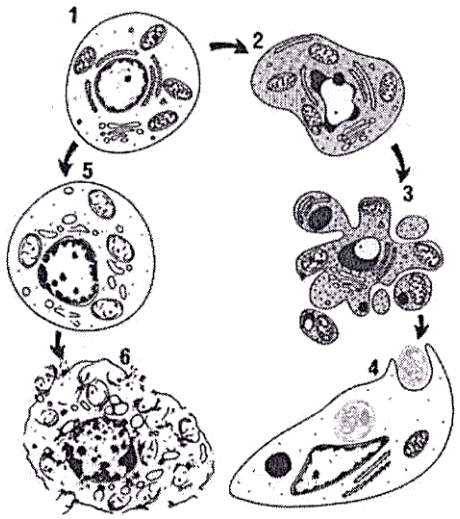


Рисунок 33 – Последовательность ультраструктурных изменений при апоптозе (справа) и некрозе (слева): 1 – нормальная клетка; 2 – начало апоптоза; 3 – фрагментация апоптотической клетки; 4 – фагоцитоз апоптотических телец окружающими клетками; 5 – гибель внутриклеточных структур при некрозе; 6 – разрушение клеточной мембраны (согласно сайту лаборатории клеточной иммунологии Нижегородского НИИ эпидемиологии и микробиологии им. академика И. Н. Блохина Минздрава Российской Федерации)

Исследования по математическому моделированию клеточных процессов с учетом апоптоза и некроза проводит в Санкт-Петербургском университете профессор И. Б. Токин. На основе экспериментальных данных о действии радиации на клеточном и тканевом уровнях организации им была создана модель реорганизации клеток после облучения. Модель описывается системой дифференциальных уравнений 6-го порядка и дает возможность анализа зависимости клеточной устойчивости и тканевой жизнеспособности при различных параметрах системы. Определено также время, необходимое для завершения клеточной репарации при повреждениях различной интенсивности. Следующий этап моделирования реакций кишечного эпителия на популяционном и тканевом уровнях в норме и при действии радиации связан с исследованиями апоптоза как физиологической (запрограммированной) гибели клеток (рис. 34).

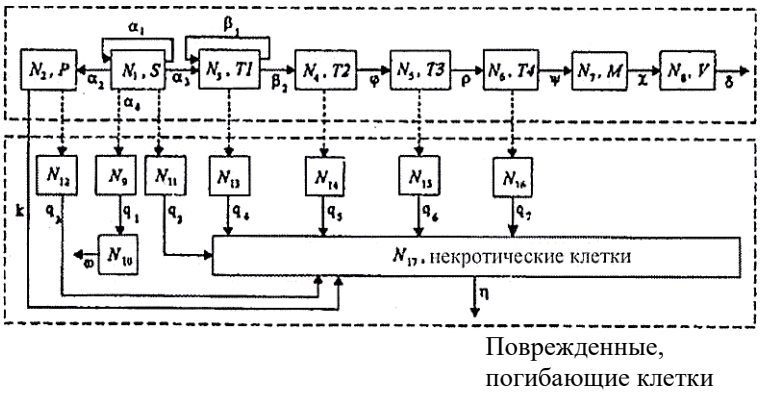


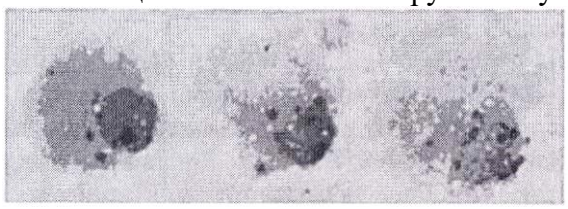
Рисунок 34 – Модель кинетики нормально функционирующей и восстанавливающейся после облучения клеточной популяции кишечного эпителия: N – среднее число клеток разных популяций; P, S, T – типы клеток; M – зрелые клетки крипты; K – клетки ворсинки. Стрелками

указаны интенсивности переходов клеток из одного состояния в другое

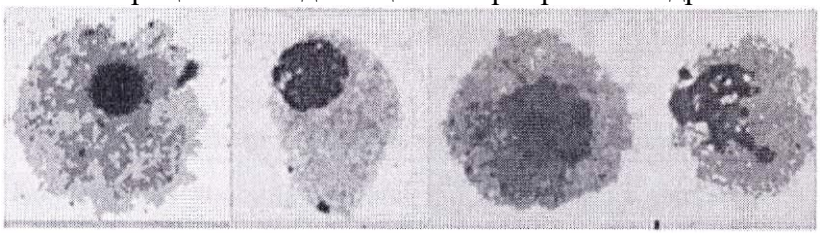
Недавно на созданной И. Б. Токиным кафедре управления медико-биологическими системами построена новая модель кинетики восстанавливающейся после облучения клеточной популяции кишечного эпителия, учитывающая присутствие возможных вариантов гибели клеток – некроза и апоптоза. В модели рассмотрены связи между уровнями выживаемости и длительностью задержки деления облученных клеток, а также учтено действие регулирующих численность популяции функций, основанное на двойном механизме обратной связи (в качестве примера таких моделей (см. рис. 34).

Лаборатория моделирования генетических процессов ГНУ «Институт генетики и цитологии НАН Беларуси» ведет работу над электронным атласом морфологических форм клеток, отображающим процессы дифференцировки, апоптоза и цитогенетических нарушений под влиянием загрязнения окружающей среды. В этом направлении по инициативе к. б. н. В. Ю. Афонина на основе результатов собственных исследований уже создан широкий набор видеоматериалов, отражающих различные аспекты указанных процессов у животных отдельных таксономических групп (рис. 35), в том числе такие:

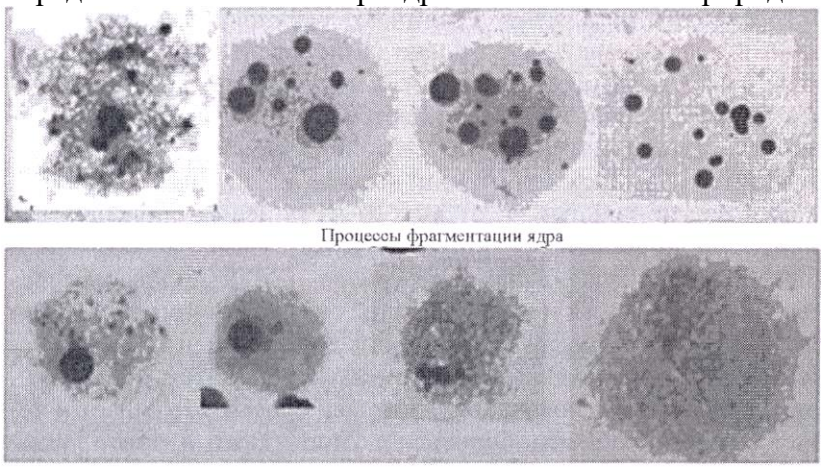
- клетки белой крови амфибий;
- цитогенетические нарушения в эритроцитах бурых лягушек;
- эмбриональная гибель и цитогенетические нарушения у моллюсков (прудовики);



Процессы конденсации и кариорексиса ядра



Процесс фрагментации и конденсации, когда одиночные фрагменты представляют собой микроядра апоптотической природы



Апоптотичеокис тела с минимальным весом ядерного материала

Рисунок 35 – Процессы клеточной гибели на цитогенетических препаратах метафаз костного мозга амфибий (данные В. Ю. Афонина)

- цитогенетические повреждения эритроидной ткани лабораторных мышей;
- дифференцировка и гибель фибробластов культуры клеток мыши;
- апоптоз в различных тканях европейской рыжей полевки;
- гибель клеток карпа *ex vivo*.

Поскольку ткани филогенетически разных животных обновляются с разной скоростью, пособие позволит выбирать комплекс критериев для оценки повреждений кроветворной ткани взаимозаменяемых модельных видов, а также отбора тест-объектов, наиболее подходящих для конкретной экологической ситуации, что важно для объективного планирования и проведения мониторинговых исследований.

Органогенез, старение и раковая трансформация

Онтогенез и старение. Согласно он-лайновой Гистологической энциклопедии, органогенез (от греч. organ + genesis –зарождение, образование) – это процесс образования органов при развитии организма. Различают органогенез онтогенетический и филогенетический; изучение последнего составляет задачу сравнительной анатомии. Онтогенез, или индивидуальное развитие организма, начинается с момента оплодотворения и делится на два периода: эмбриональный – от образования зиготы до рождения или выхода из яйцевых оболочек и постэмбриональный – от выхода из яйцевых оболочек или рождения до смерти организма. Отсюда видно, что составной частью онтогенеза является старение – как процесс постепенного нарушения и потери важных функций организма или его частей, в частности способности к размножению и регенерации. Вследствие старения организм становится менее приспособленным к условиям окружающей среды, уменьшает и теряет свою способность бороться с хищниками и противостоять болезням и травмам. Наука, которая изучает старение, называется геронтологией. Максимальная продолжительность жизни некоторых многоклеточных животных приведена в табл. 5 (согласно AnAge Database).

Таблица 5 – Продолжительность жизни (ПЖ) многоклеточных животных (согласно AnAge Database)

Организм	Максимальная ПЖ, лет	Организм	Максимальная ПЖ, лет
Млекопитающие		Пресмыкающиеся и земноводные	
Человек (Homo sapiens)	122	Галапагосская черепаха (Geochelone nigra)	177
Слон (Elephas maximus)	69-86	Нильский крокодил (Crocodylus niloticus)	44-68
Шимпанзе (Pan troglodytes)	37-75	Японская гигантская саламандра (Andrias japonicus)	55
Собака (Canis familiaris)	29-34	Большая зеленая жаба (Lithobates catesbeianus)	30
Корова/домашний бык (Bos taurus)	30	Европейская черная саламандра (Salamandra at га)	17
Кошка (Felis catus)	29	Храмовая черепаха (Hieremys annandalii)	9
Дикий кабан (Sus scrofa)	27	Рыбы	
Домашняя коза (Cap ra hi reus)	18-20	Озерный осетр (A cipenser fuh 'escens)	152
Белка (Sciurus vulgaris)	15-16	Сом (Silurus glanis)	60
Домовая мышь (Mus musculus)	4	Миссисипская гамбузия (Gambusia affinis)	2-3
Птицы		Моллюски	
Гриф-индейка (Cathartes aura)	118	Моллюск (Arctica islandica)	до 500
Лебедь-шипун (Cygnus olor)	70	Насекомые	
Суринамский амазон (Amazona ochrocephala)	56	Периодическая цикада (несколько видов рода Magic ica da)	17
Сизый голубь (Columba livid)	35	Муравей-королева (Formicidae)	15
Воробей (Passer domesticus)	23	Дрозофила меланогастер (Drosophila melanogaster)	30 дней

До исследований Леонарда Хайфлика считалось, что смертным, стареющим является целый организм, а отдельные его клетки – практически бессмертны. После его открытия начались исследования в области генетики соматических клеток, которые внесли заметный вклад в современный уровень познания процессов старения и раковой трансформации. Так, например, недавно группа исследователей из Греции и Италии изучила способность к репликативному удвоению фибробластов кожи молодых (18 и 28 лет), старых (75 и 80 лет) и столетних (101, 102 и 103 года) здоровых женщин. Эмбриональные фибробласты подвергались клеточному старению через 39–42 удвоения (пассажа). Фибробласты от 28-летнего донора и четырех столетних женщин прошли 25–35 удвоений, тогда как полученные от 80-летней женщины – только 15 пассажей. Старческий фенотип был подтвержден также различными молекулярными и клеточными маркерами, включая окраску на бета-галактозидазу, экспрессию генов и длину теломер.

Новая компьютерная технология прижизненного наблюдения клеток. Обычно исследования в этой сфере выполняются на большой совокупности клеток в один из моментов роста клеточной популяции. Для того чтобы проследить всю последовательность изменений, происходящих в отдельных клетках клеточной популяции, в лаборатории моделирования генетических процессов ГНУ «Институт генетики и цитологии НАН Беларуси» к. б. н. О. В. Квитко, к. б. н. Я. И. Шейко и к. б. н. И. И. Коневой разработана компьютерная технология, позволяющая вести непрерывное изучение отдельных клеток и их потомства на протяжении многих поколений.

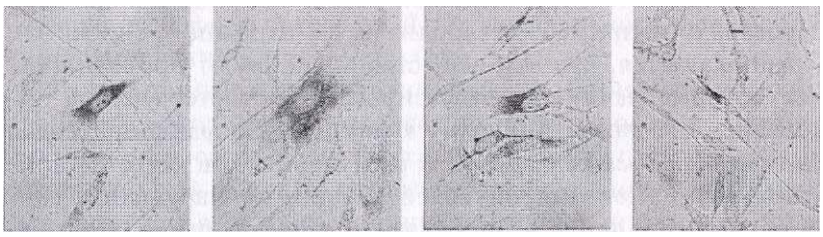
Компьютерная система для видеонаблюдения живых клеток состоит из инвертированного микроскопа с присоединенной видеокамерой, которая подключена к компьютеру. В инвертированном микроскопе объектив расположен снизу от объекта, что позволяет визуализировать клетки, которые растут на дне культурального сосуда. Для продолжительной видеосъемки живых клеточных культур в камере инвертированного микроскопа поддерживается постоянная температура, оптимальная для жизнедеятельности клеток (например, для клеток человека – это 37 °С). В термостатируемую камеру микроскопа помещается культуральный сосуд, на дне которого растут клетки в питательной среде. Для нормального роста клеток в атмосфере культурального сосуда должна быть 5%-ная концентрация углекислого газа ($C0_2$). Для видеозаписи и анализа полученных видеофильмов наряду со стандартным программным обеспечением используются оригинальные компьютерные программы, разработанные Я. И. Шейко. Накопленный массив видеофильмов в реальном масштабе времени является ценным материалом для научных исследований и биологического образования.

В частности, формирование клеточного клона можно при анализе видеозаписей схематически отразить в виде клеточной генеалогии, напоминающей родословные в генетике человека. На генеалогии наносятся такие данные, как площадь и форма клетки, митозы и клеточная гибель, аномальные митозы, размер и количество ядер, скорость и характер перемещений клеток и другие характеристики, что позволяет изучать изменения в клеточном клоне в течение ряда поколений.

Анализ накопленных в ходе многолетних исследований данных привел О. В. Квитко к оригинальной развитийной теории антистарения, согласно которой причиной старения является накопление хаотических эпигенетических изменений, нарушающих экспрессию генов, т. е. нормальное течение информационных процессов в клетке и организме (см. его статью «Развитийная доминанта как механизм антистарения» в российском альманахе «Геронтология и гериатрия», а также список цитируемой в ней литературы, в котором приведены основные работы других авторов в этой области). В период эмбриогенеза и постнатального роста особый сигнальный механизм управляет исправлением эпигенетических ошибок, омолаживая клетки и блокируя возрастное ослабление функций до тех пор, пока организм не становится способным производить потомство. Вот почему виды с протяженными

периодами развития и роста обладают и более продолжительной жизнью. Отсюда и совершенно другой видится задача геронтологии: не борьба с «болезнями старости» для продления зрелого и пожилого периодов жизни, а использование средств интенсификации интеллектуального развития после завершения процесса физического развития организма, что обеспечит генерацию новых сигналов антистарения, превращая жизнь человека в «вечную юность».

Бета-гал тест как маркер старения. Известно, что при старении в тканях организма и в клеточных культурах повышается количество клеток с экспрессией гена бета-галактозидазы, что, как мы видели выше, дало основание причислить это явление к числу маркеров старения. Однако в литературе имеются данные о проявлении этого маркера не только в стареющих, но и в иммортализированных клеточных популяциях *in vitro*, а также в опухолях в организме. Я. И. Шейко проведено сравнение экспрессии гена бета- галактозидазы в нормальных и иммортализированных клеточных культурах фибробластов мыши и человека (рис. 36). Как и следовало ожидать, при увеличении времени культивирования нормальных фибробластов человека происходило резкое увеличение количества бета-галпозитивных клеток. В иммортализированной линии фибробластов мыши активность бета-галактозидазы проявляла значительная часть клеток (около 25%), с интенсивностью, не зависящей от сроков культивирования.



а б в г

Рисунок 36 – Гетерогенность клеток иммортализированной линии из эмбриона мыши по активности бета-галактозидазы (буквами обозначены различные образцы, отличающиеся по уровню экспрессии)

Наличие и постоянство доли бета-галпозитивных клеток в нестареющей культуре свидетельствуют о присутствии в ней двух субпопуляций – стволовой линии и фракции стареющих клеток, что открывает определенные практические перспективы. Подобно старению клеток в иммортализированной культуре, увеличение доли постаревших опухолевых клеток может коррелировать с темпами роста опухоли. Следовательно, использование бета-галактозидазного теста как маркера клеточного старения может быть перспективным для оценки пролиферативной активности опухолевых клеточных популяций в онкологической диагностике.

Злокачественные новообразования как следствие нарушения молекулярно-генетических и клеточных регуляторных механизмов. Продолжая разговор о раковой трансформации, необходимо вспомнить определение новообразования, или опухоли. Согласно одному из он-лайновых словарей, **опухоль** (neoplasm, tumor) – это локальная клеточная популяция, характеризующаяся неконтролируемой пролиферацией клеток, которые становятся атипичными с точки зрения дифференцировки и характера роста и могут передавать измененные свойства своим потомкам.

Различают доброкачественные опухоли (например, липома), при которых группа аномальных клеток как бы инкапсулируется, не образуя метастазов, и злокачественные (раковые) опухоли, когда происходят инфильтрация, метастазирование и рецидивирование, приводящие к разрушению соседних тканей. Для злокачественных опухолей характерен как клеточный, так и тканевой **атипизм**. При клеточном атипизме наблюдается утолщение и необычность строения базальной мембраны, изменение соотношения объемов цитоплазмы и

ядра, изменение ядерной оболочки, увеличение объема, а иногда и числа ядрышек, увеличение числа фигур митоза, нарушения митоза. Тканевой атипизм проявляется в нарушении пространственных и количественных соотношений между компонентами ткани, например стромой и паренхимой, сосудами и стромой и т. д.

Этиология опухолей до настоящего времени неизвестна, как не существует единой теории канцерогенеза. Согласно вирусно-генетической теории, решающая роль в развитии опухолей принадлежит онкогенным вирусам, к которым относят герпесоподобный вирус Эпштейна–Барр (лимфома Беркитта), вирус герпеса (рак шейки матки), вирусы гепатитов В и С (рак печени). По этой теории интеграция генома вируса в генетический аппарат клетки может дать толчок опухолевой трансформации клетки. Совершив это переключение генетической программы клетки на путь злокачественного роста, при дальнейшем размножении опухолевых клеток вирус перестает играть существенную роль.

Физико-химическая теория в качестве основной причины развития опухолей рассматривает воздействие различных физических и химических факторов (рентгеновское и гамма-излучение, канцерогенные вещества) на клетки организма, которое приводит к их трансформации в опухолевые. Помимо химических канцерогенов окружающей среды в этой теории определенная роль в возникновении опухолей отводится эндогенным канцерогенам (например, метаболитам триптофана и тирозина). Предполагается, что эти вещества активируют протоонкогены, которые начинают синтезировать соответствующие онкобелки, что также приводит к онкотрансформации. Теория дисгормонального канцерогенеза рассматривает в качестве причины возникновения опухолей различные нарушения гормонального баланса в организме. Существует также дизонтогенетическая теория, по которой причиной развития опухолей считается нарушение эмбриогенеза тканей, которое в дальнейшем под действием провоцирующих факторов может привести к онкотрансформации клеток ткани. Наконец, полиэтиологическая теория объединяет все вышеперечисленные теории.

Долгое время считалось, что все раковые клетки обладают одинаковым пролиферативным потенциалом и любая опухолевая клетка, оставшаяся в организме после лечения, способна запустить патологический процесс снова. Поэтому все усилия клиницистов были направлены на уничтожение максимально возможного их числа. Однако при многих видах рака такими свойствами оказывается наделенной лишь небольшая часть клеток опухоли. Клетки, находящиеся у истоков раковой трансформации, и стволовые клетки имеют много общих черт; в частности, они обладают неограниченной продолжительностью жизни и способностью давать начало клеткам других типов. Такие аналогии позволяют называть их раковыми стволовыми клетками. Злокачественные родоначальные клетки, по-видимому, появляются в результате сбоя в регуляторной системе поврежденных стволовых клеток или их прямых потомков. Поэтому раковые стволовые клетки являются основной мишенью для противоопухолевых препаратов.

Как видим, ключевая роль в онкогенной трансформации отводится нарушениям в работе генов, т. е. сбоям в их регуляции. В связи с открытием явления РНК-интерференции и роли малых РНК в регуляции экспрессии генов натуральные и синтетические двуспиральные полинуклеотиды (ДНК и РНК) представляются перспективной основой для создания антираковых препаратов. Еще в начале 1990-х годов к. б. и. О. В. Квитко с сотр. удалось обнаружить резко выраженный ингибирующий эффект препарата ДНК, добавляемого в среду для культивирования клеток *HeLa* (рис. 37).

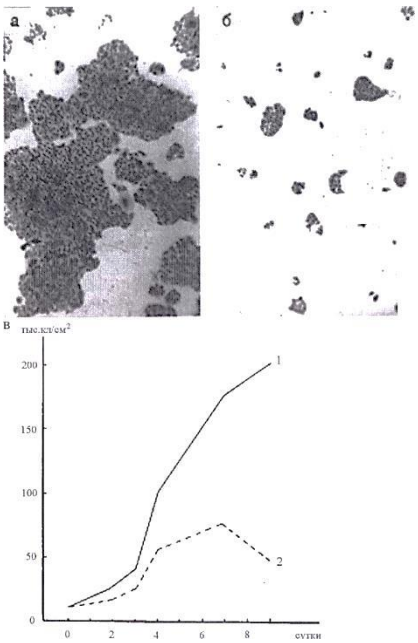


Рисунок 37 – Антираковый эффект ДНК эритроцитов цыплят: а – культура *HeLa* (посевная плотность 5000 клеток на квадратный сантиметр); б – культура *HeLa* с добавлением ДНК (100 мкг/мл); культивирование в течение 7 суток; в – динамика плотности культуры *HeLa* (посев 10000 кл/см^2) в контроле (1) и с добавлением ДНК эритроцитов цыплят (2)

Разработанная в лаборатории моделирования генетических процессов методология позволяет по-новому подойти к изучению механизмов раковой трансформации. Группой под руководством О. В. Квитко предполагается создание клеточных тест-систем для подбора полинуклеотидных препаратов, различающихся по нуклеотидной последовательности и размеру фрагментов, которые бы способствовали переводу генетической изменчивости раковых стволовых клеток в сторону пролиферативного ограничения и гибели клеточных клонов.

Моделирование клеточной дифференцировки

Моделирование работы генов. Механизм регуляции синтеза белков был сначала изучен на модельном объекте – бактериях. Французские микробиологи Ф. Жакоб и Ж. Моно показали, что по своим функциям гены разделяются на два иерархических класса: структурные и управляющие. Структурные гены содержат информацию о строении белков.

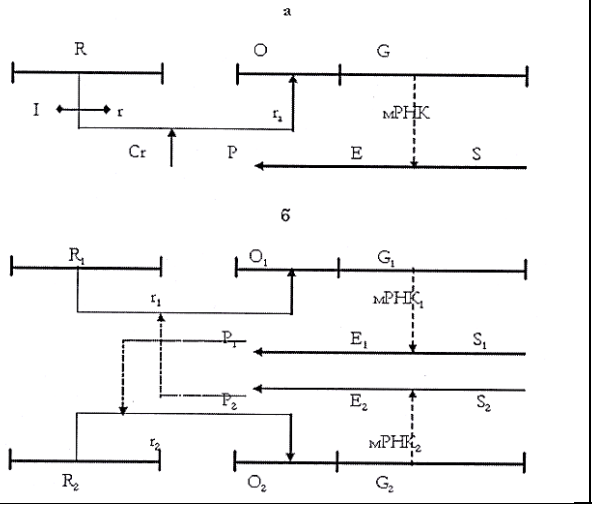


Рисунок 38 – Управление синтезом белка в бактериальной клетке: а – оперон; б – генетический триггер; R – регулятор; O – оператор; S – структурный ген; E – синтезируемый фермент, катализирующий образование продукта P из субстрата S; r – репрессор; ra – активная форма репрессора; Cr – корепрессор; I – индуктор

Управляющие гены, в свою очередь, делятся на два типа: операторы, расположенные рядом со структурными генами и служащие выключателями, и регуляторы, которые включают

и выключают их. Регуляция осуществляется посредством так называемых белков-репрессоров, синтезируемых под контролем регуляторных генов и инактивирующих операторы. Группа структурных генов, контролируемая одним оператором, называется опероном (рис. 38, а).

Управление биосинтезом осуществляется по кибернетическому принципу обратной связи. В случае отрицательной обратной связи избыточный продукт, соединяясь с репрессором, переводит его в активное состояние, в котором репрессор блокирует синтез. Конечный продукт, подавляя собственный синтез, выступает в роли так называемого корепрессора. Наблюдается в биосинтетических системах и управление по типу положительной обратной связи, так как наряду с репрессорами существуют особые вещества – индукторы, способные переводить репрессор в неактивное состояние, разблокируя оператор. Например, в случае лактозного оперона таким индуктором служит сама лактоза, управляющая собственным сбраживанием.

Ф. Жакоб и Ж. Моно показали, что отдельные опероны, как правило, не работают изолированно. Часто они могут быть связаны крест-накрест, образуя так называемый генетический триггер (рис. 38, б). В этом случае продукт, вырабатываемый под контролем одного оперона, управляет состоянием другого и наоборот.

Работа и отдельного оперона, и генетического триггера может быть описана математическими моделями в форме дифференциальных уравнений. Простейшая модель генетического триггера, разработанная московским биофизиком Д. С. Чернавским с сотр., имеет вид

$$dP_1/dt = k_1/\left(B + E_2^2\right) - k_2 P_1,\ dP_2/dt = k_1/\left(B + E_1^2\right) - k_2 P_2,$$

$$dE_1/dt = k_3 P_1 - k_4 E_1,\ dE_2/dt = k_3 P_2 - k_4 E_2, \qquad (14)$$

где E_1 и E_2 – концентрации белков-ферментов, синтез которых регулируется молекулами мРНК; P_1 и P_2 – продукты, синтезируемые с помощью E, и E_2 из субстратов S_1, и S_2; B, k_1, $к_2$, $к_3$ и $к_А$ – параметры, характеризующие скорости реакций. В модели предполагается, что ресурсы субстратов не ограничены (S_1 = const и S_2 = const), а концентрации продуктов-ко репрессоров пропорциональны концентрациям мРНК.

В простейшем случае $P_1 = k_4E_1/k_3$ и $P_2 = k_4E_2/k_3$ и можно от системы из четырех уравнений перейти к редуцированной системе из двух уравнений только для E_1 и E_2.

$dE^l/dt = L/(1 + E_2^2) – E_l$, $dE_2/dt = L/(l + E_1^2) – E_2$. (15)

Эта редуцированная система имеет два устойчивых решения. Переход от одного к другому осуществляется скачком при наличии внешних возмущений, т. е. действительно триггерным способом.

Два типа моделей клеточной дифференцировки. Для моделирования процессов морфогенеза и дифференцировки, в результате которых образуются сложные многоклеточные организмы, развиваются модели двух типов. Один тип связан с изучением диссипативных структур, которые могут возникнуть в континууме связанных диффузией метаболитов живых клеток, в каждой из которых существует, по крайней мере, один генетический триггер. Математическая модель таких дифференцированных тканей, состоящих из слоев, в каждом из которых состояние клеток однотипно, была построена Д. С. Чернавским с соавт. Эта модель в простейшем случае одномерного пространства представляет собой систему дифференциальных уравнений в частных производных (уравнений нелинейной диффузии).

Исследование модели показывает, что ферменты E_1 и E_2 распределяются по координате так, что клетки с преобладающим синтезом E_1 (либо E_2) располагаются слоями с резко очерченными границами. Субстраты S_1 и S_2 распределяются более плавно. Таким образом, решение описывает контрастную диссипативную структуру.

Другим перспективным путем описания реализации генетической информации в структурах дифференцированных клеток является применение теории конечных автоматов, в качестве элементов в которой выступают генетические триггеры. Широкий круг вопросов,

связанный с построением математических моделей молекулярно-генетических систем, разрабатывается в Новосибирске В. А. Ратнером, Р. Н. Чураевым и др. Дискретные модели регуляторных процессов у эукариогических организмов наиболее успешно развиты в разработках американского математика С. Кауфмана, болгарских исследователей Р. Цанева и Б. Сендова.

К-генетические сети. Модель К-генетических сетей Кауфмана целиком основана на идеях теории конечных автоматов. Как известно, конечный автомат – это набор клеток, образующих некоторую периодическую решетку с заданными правилами перехода, которые определяют состояние клетки в следующий момент времени через состояние клеток, находящихся от нее на расстоянии не больше некоторого заранее заданного, в текущий момент времени. Согласно модели, геном клетки определяется набором генов, каждый из которых может находиться только в одном из двух состояний: «1» – ген работает, «О» – не работает. Гены связаны между собой, причем i-й ген имеет s(i) входов и только один выход. Состояние входов i-го гена в некоторый дискретный момент времени t_k характеризуется набором функций $g_1^1(t_k)$, $g_1^2(t_k)$…, $g_i^{s(i)}(t_k)$а состояние выхода – одной функцией $g_i(t_k)$. Так как ген может находиться только в одном из двух режимов, состояние выхода i-го гена g_i в момент t_{k+1}, является двоичной функцией G_i от функции входов g_i в момент t_k:

$$g_i(t_{k+1}) = G_i(g_i^1(t_k),\ g_i^2(t_k),\ldots,\ g_i^{s(i)}(t_k)). \tag{16}$$

К-генетическая сеть будет задана, если кроме N генов с их функциями g.заданы матрицы связи Ее элементами

$$e_{ij} = \begin{cases} 1, & \text{если выход } i\text{-го гена является входом } j\text{-го гена,} \\ 0 - & \text{в противном случае.} \end{cases} \tag{17}$$

Пусть, например, имеется сеть из трех генов v, y и z, каждый из которых связан с двумя другими (рис. 39, а). Тогда матрица связей Е имеет вид

$$E = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \tag{18}$$

Действительно, выход гена x является входом генов *y* и *z*, поэтому в первую строку матрицы идут единицы во второй и третьей позициях. Аналогично заполняются и остальные строки матрицы. Режим работы каждого гена под влиянием двух других задают матрицы, также показанные на рис. 39, а. Исходя из них, для изучаемой К-генетической сети можно составить таблицу, характеризующую ее состояния в последовательные моменты t и t + 1 (табл. 6).

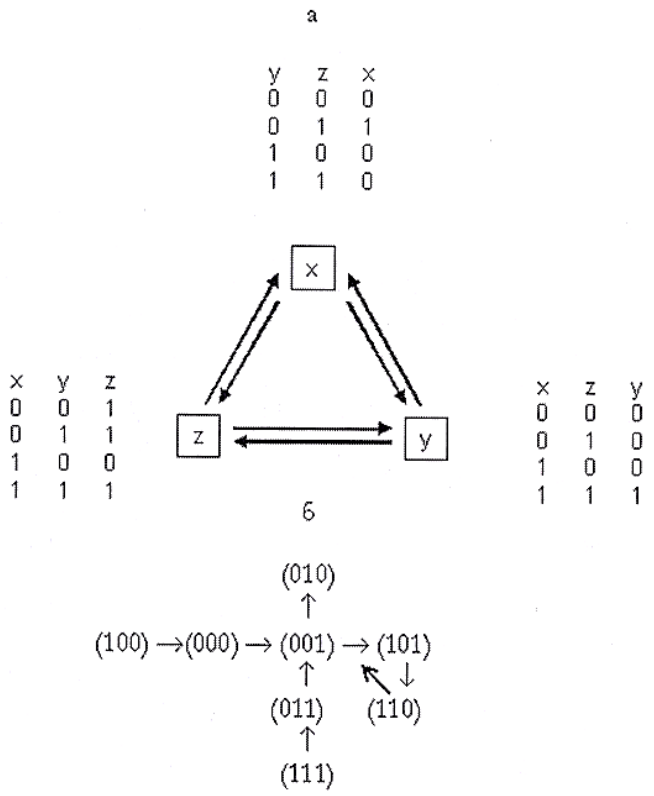


Рисунок 39 – К-генетическая сеть из трех генов: а – схема связей между составляющими сеть генами; б – граф возможных переходов между состояниями сети

Таблица 6 – Поведение К-генетической сети из трех генов

t			$t+1$		
x	y	z	x	y	z
0	0	0	0	0	1
0	0	1	1	0	1
0	1	0	0	0	1
0	1	1	0	0	1
1	0	0	0	0	0
1	0	1	1	1	0
1	1	0	0	0	1
1	1	1	0	1	1

Из табл. 6 видно, что эта сеть образует замкнутый цикл длиной в три шага:
001–>101–>110–>001 (рис. 39, б).

При этом состояние 001 является своеобразным устойчивым «стоком»: из всех других состояний К-генетическая сеть переходит именно в него, а затем переключение генов идет уже только в рамках цикла.

В общем случае К-генетическая сеть из N генов имеет 2^N разных состояний. Поскольку набор их конечен, то с течением времени любая К-генетическая сеть обязательно начинает повторять уже принимавшиеся ранее состояния, т. е. входит в цикл. В принципе таких циклов может быть несколько. В этом случае переход от одного цикла к другому возможен только

вследствие внешних воздействий. Работа генома клетки в цикле определяет, по существу, особое динамически устойчивое состояние клетки. Разные циклы будут, по-видимому, отвечать разным типам дифференцировки. Используя этот подход, С. Кауфман провел не только теоретическое исследование, но и обширный машинный эксперимент. Он хотел выяснить, какой длины циклы образуются при произвольном выборе матрицы связей и функций генов.

Оказалось, что в полностью связанной сети каждый ген связан со всеми остальными генами. Для N = 200 возникают циклы с длиной порядка 1030. Чтобы представить себе эту величину, предположим, что переход гена из одного состояния в другое требует 1 мксек. Тогда длительность одного цикла составит 1017 лет, что примерно в 10 миллионов раз больше возраста нашей Вселенной. В односвязной сети каждый ген имеет вход только от одного гена. В этом случае при том же N = 200 длина циклов превышает несколько миллионов шагов, что также лишено физического смысла. Иначе обстоит дело с двухсвязной сетью, когда каждый ген имеет два входа от разных генов. В этом случае даже при N = 1000 образуются замкнутые циклы из 10–12 состояний. Поэтому именно двухсвязную К-генетическую сеть можно использовать для моделирования клеточной дифференцировки.

Модель Сендова–Цанева. Хотя в настоящее время не существует единой и общепринятой модели управления генами многоклеточных организмов, в основе большинства из них лежат следующие положения.

Клеточная дифференцировка обусловлена выражением (активностью) разных частей генома в разных клетках.

Группы функционально связанных генов часто активируются вместе.

Выделяются два типа таких групп генов: один определяет процесс деления клетки («митотический» оперон), другой – синтез тех или иных веществ.

Кроме рассмотренного Жакобом и Моно механизма репрессии-дерепрессии имеется еще механизм блокировки-де- блокировки генов. Большинство генов в дифференцированной клетке заблокировано. Блокировка соответствующих участков ДНК осуществляется с помощью специальных белков – гистонов. Деблокировка осуществляется специальными молекулами РНК – продуктом особых деблокирующих генов. Таким образом, группа генов активна тогда и только тогда, когда она деблокирована и не репрессирована.

Наиболее последовательно эти положения переводятся на язык математики в модели Б. Сендова и Р. Цанева. Она основана, с одной стороны, на теории дискретных генетических сетей, а с другой – учитывает непрерывную внутриклеточную кинетику, подобно тому как это делается в модели Жако-ба–Моно.

Согласно модели Сендова–Цанева, различные типы клеток определяются разными наборами деблокированных генов, а различная функциональная активность клеток данного типа определяется разными наборами дерепрессированных оперонов среди деблокированных оперонов. Учет непрерывной кинетики внутри клеток позволяет объяснить в рамках этой модели влияние соседних клеток друг на друга через диффузию и действие на генетический аппарат индукторов и других внешних агентов.

Дискретность процесса учитывается с помощью RB-генетической сети. Каждый ген в ней управляется двумя типами входов: R-входами (входами репрессии) и B-входами (входами деблокировки). И те, и другие могут быть только в состоянии «О» или «1». Поэтому всякий ген-оперон характеризуется четырьмя состояниями: «00»,«01», «10» или «11».

Всего оперонов в сети N + 1 (в том числе один из них – митотический оперон). Связи между ними задаются специальными матрицами порядка (N + l)(N + 1), определяющими взаимодействие генов по R- и B-входам. После деления каждая из дочерних клеток развивается самостоятельно. Однако соседние клетки могут оказывать влияние друг на друга благодаря диффузии репрессоров. В модели возможна имитация мутаций, которая осуществляется дискретными случайными воздействиями в момент митоза.

Модель достаточно сложна и позволяет «проиграть» гипотетические варианты развития

модельных эмбрионов. Математический аналог двумерного эмбриона был назван Б. Сендовым и Р. Цаневым «килиндрос». С его помощью проведен обширный машинный эксперимент в разных конкретных случаях, в частности построена модель канцерогенеза.

Результаты исследований конца XX в. по моделированию развития клеток обобщены в монографии, молекулярные аспекты клеточной биологии рассматриваются в книге.

Литература: Дромашко, С. Е. Очерки биоинформатики / С. Е. Дромашко ; НАН Беларуси, Ин-т генетики и цитологии ; науч. ред. Л. В. Хотылева. – Минск : Беларус. навука, 2009. – 400 с.

Занятие 5.2 Стволовые клетки

Вопросы для изучения:

1. История открытия стволовых клеток.
2. Получение плюрипотентных клеток.
3. Молекулярные основы тотипотентности эмбриональных стволовых клеток. Стволовые клетки взрослого организма.
4. Различия в потенции стволовых клеток разного происхождения.
5. Механизмы коммитирования стволовых клеток.
6. Значение микроокружения для самоподдержания популяции стволовых клеток.
7. Молекулярные маркеры стволовых клеток.
8. Источники стволовых клеток у взрослого организма. Стволовые клетки в медицине.
9. Сайты и базы данных по стволовым клеткам и клеточным технологиям.

Учебно-методические материалы.

Стволовые клетки эмбриона и взрослого организма

Из истории стволовых клеток. Открытие стволовой клетки стоит в одном ряду с такими великими достижениями человечества, как открытие двуспиральной цепочки ДНК. Однако про стволовые клетки известно уже почти сто лет. В 1908 г. собственную концепцию их развития предложил русский гематолог А. А. Максимов. Первые в мировой науке работы по стволовым клеткам костного мозга в 1960–1970-е годы выполнили советские ученые И. Л. Чертков и А. Я. Фриденштейн, но обратили внимание на них только в 1980-х годах, когда стволовые клетки заново «открыли» в США. Американские исследователи сумели выделить эмбриональные стволовые клетки и доказать, что они могут при соответствующих воздействиях превращаться в любые другие. В 1999 г. журнал «Science» признал открытие эмбриональных стволовых клеток третьим по значимости событием в биологии после расшифровки двойной спирали ДНК и программы «Геном человека». Когда стало ясно, что стволовые клетки могут быть внесены в организм искусственно, начался настоящий штурм проблемы, первые итоги которого отражает Нобелевская премия 2007 г.,

Mario R. Capecchi	sir Martin J. Evans	Oliver Smithles

Рис. 40. Лауреаты Нобелевской премии 2007 г. по физиологии и медицине

присужденная англичанину Мартину Эвансу (Martin J. Evans) и американцам Марио Капеччи (Mario R. Capecchi) и Оливеру Смитису (Oliver Smithles) «за их открытия принципов введения специфических генных модификаций у мышей с помощью эмбриональных стволовых клеток» (рис. 40).

Получение плюрипотентных клеток. В настоящее время линии плюрипотентных клеток человека получают из двух источников с помощью методов, отработанных на животных моделях (рис. 41, а).

1. Плюрипотентные клетки выделяют непосредственно из внутренней клеточной массы эмбриона человека на стадии бластоцисты. Сам эмбриональный материал получали в больших количествах в клинических, а не исследовательских целях для осуществления экстракорпорального оплодотворения, всякий раз испрашивая разрешение на его использование у обоих доноров. Клетки внутренней клеточной массы культивировали и получали линию плюрипотентных клеток

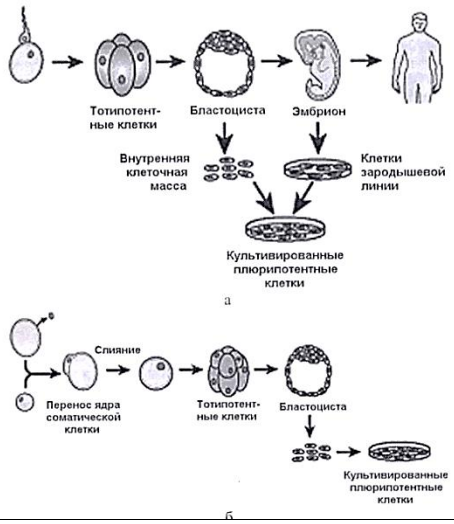


Рисунок 41 – Различные способы получения плюрипотентных эмбриональных клеток: из клеточной массы эмбриона или ткани плода (а); путем переноса в яйцеклетку ядра соматической клетки (б).

2. Другая группа исследователей выделяла плюрипотентные клетки из ткани плода. Разрешение на это давалось обоими супругами уже после того, как они сами приняли решение прервать беременность. Клетки отбирались из той области плода, которая должна была развиться в яичники или семенники.

Несмотря на то, что плюрипотентные клетки в двух указанных случаях происходили из разных источников, полученные клеточные линии были идентичными.

Еще одним способом получения плюрипотентных клеток может стать метод, основанный на переносе в энуклеированную (лишенную ядра) яйцеклетку ядра соматической клетки (рис. 41, б). Соответствующие опыты уже проведены на животных. Сама яйцеклетка с новым ядром и ее непосредственные «потомки» способны при соответствующих условиях развиться в полноценный организм, т. е. являются тотипотентными. Из них формируется бластоциста, которая и служит источником плюрипотентных клеток.

Молекулярные основы тотипотентности эмбриональных стволовых клеток (ЭСК). Геном зиготы и ЭСК находится в так называемой нулевой точке, откуда стартуют две главные программы эмбриогенеза:

гаструляция + органогенез;

рестрикционное созревание дифференцированных клеточных линий.

В реализации этих двух программ участвуют примерно 5000 генов. Каким образом реализуется тотипотентность генома, т. е. способность выбирать одну из многих траекторий развития? Серийный анализ генной экспрессии (SAGE) в культуре ЭСК, мезенхимальной, нейрональной и гематогенной стволовой клетки (МСК, НСК, ГСК соответственно) выявил

важнейшую закономерность. Во всех перечисленных стволовых клетках присутствовали сходные наборы предсинтезированных мРНК следующих классов генов:

1) ранних генов экто-, мезо- и эндодермы;

2) набор мРНК гомеотических (Нох) генов, контролирующих сегментацию и трехмерную карту зародыша;

3) пять наборов мРНК генов тотипотентности. Каждый набор факторов тотипотентности увязан с комплементарным набором факторов и рецепторов пролиферации на прогениторных клетках в клоне. Особенно существенны спаренные блоки генов, контролирующие пролиферацию соседних слоев прогениторных клеток в клоне (типа Delta/Notch);

4) мРНК генов апоптоза, контролирующие численность прогениторных клеток в клоне. Баланс клеток достигается за счет миграции и частичной гибели клеток в клоне. Пролиферация постоянно уравновешивается миграцией и апоптозом клеток;

5) набор мРНК контрольных генов (master genes) рестрикционного созревания специализированных клеточных линий;

6) наборы мРНК генов миграции и навигации миграторных клеток.

Таким образом, в основе тотипотентности ЭСК, МСК и других типов стволовых клеток выявлен универсальный механизм: направленная активация и последующий импорт мРНК в ядро. Детали импорта/экспорта информации между ядром и цитоплазмой в ЭСК плохо изучены. Эта концепция тотипотентности подтверждается новыми наблюдениями, выполненными на одиночных клонах мезенхимальных стволовых клеток специальной техникой MICRO-SAGE. Микрометод верифицировал присутствие мРНК 1200 ключевых генов эмбриогенеза в одном клоне МСК. Полученные данные поражают воображение.

Стволовые клетки взрослого организма. В клонах МСК из взрослой гематогенной ткани идентифицированы практически все наборы мРНК зародышевых листков, органогенеза, а также мРНК master-genes, контролирующих рестрикционное созревание линий мезенхимального, мезодермального происхождения, экто- и эндодермы. Клоны МСК по сути представляли собой библиотеки мРНК ранних этапов органогенеза и рестрикционного созревания специализированных линий.

Если программа органогенеза обеспечивает направленный импорт мРНК в ядро для включения новых котранскриптаз, гранскриптаз и транскрипционных комплексов, то рестрикционное созревание клеточных линий связано с экспортом мРНК из ядра в цитоплазму с последующей модификацией белкового фенотипа прогениторных клеток.

Различия в потенции стволовых клеток разного происхождения. В зародыше и взрослом организме потенции генома стволовых клеток существенно варьируют по «ассортименту» фенотипа специализированных клеток. Более ранние, тотипотентные ЭСК дифференцируются в любую из 250 линий специализированных клеток органов. Необходимо подчеркнуть, что ЭСК *in vitro* не продуцируют клеток трофобласта, плаценты, т. е. потенции генома ЭСК меньше зиготы. Соответственно биологический статус ЭСК меньше статуса раннего зародыша. Плюрипотентные ЭСК дают более ограниченный спектр фенотипов. Например, мезенхимальные стволовые клетки (МСК), локализованные в опорно-сосудистом каркасе органов, дифференцируются в культуре только в клетки хряща, кости, кардиомиоциты и миоциты. Монопотентные стволовые клетки (мышц, жировой ткани, периферических нервов) созревают до одного преобладающего фенотипа клеток.

Стволовые региональные клетки взрослого организма наделены мультипотентностью – пластичной плюрипотентностью, которая сильно варьирует в контексте органа-реципиента. Так, пересадки гематогенных стволовых клеток в мозг, сердечную или скелетную мышцу приводили к образованию тканеспецифичных ростков донорской ткани в органах реципиента. Однократное переливание 0,5 л донорской женской крови мужчине-реципиенту химеризует женскими клетками все паренхиматозные органы. Большое внимание уделяется сейчас

мультипотентным МСК взрослых органов, поскольку эти клетки хорошо мигрируют и многократно химеризуют ткани. В свою очередь, пересадки нейральных стволовых клеток в печень, мышцу или иммунную систему сопровождались тканеспецифичной перестройкой фенотипа донорских клеток.

Гематопоэзные стволовые клетки (ГСК). Клетки, способные производить все клетки крови, находятся в красном: костном мозге.

Мышечные стволовые клетки (МСК).

Обонятельные стволовые клетки (ОСК). Дают начало обонятельным клеткам. Располагаются в слизистой оболочке носа.

Стволовые клетки печени (СКП). В нормальном состоянии клетки печени делятся очень медленно, но под влияниям повреждений или инфекции пролиферация клеток усиливается многократно. Имеется 3 популяции клеток, способных восстановить печень.

Нейральные стволовые клетки (НСК). Нейрогенез в мозге происходит в двух местах: субвентрикулярной зоне (СВЗ), из которой были изолированы первые стволовые клетки, и зубчатой извилине гиппокампа. Зрелые нейроны образуются в обонятельной луковице – области, в которую мигрируют клетки-предшественники из СВЗ различными путями, называемыми ростральной миграционной системой. СВЗ содержит эпендимальные клетки и астроциты со схожей ролью с клетками стромы в костном мозге. Эпендимальные клетки и астроциты образуют каналы – так называемые глиальные трубки, по которым происходит миграция нейробластов к обонятельной луковице, где дифференцируются в перигломерулярные или гранулярные нейроны, которые выстраиваются цепочкой. Астроциты в трубках обеспечивают питание клеток.

Механизмы коммитирования стволовых клеток. Коммитирование – это ограничение возможных путей развития вследствие детерминации. Коммутирование совершается ступенчато. Сначала соответствующие преобразования генома касаются крупных его участков. Затем все более детализируются, поэтому вначале детерминируются наиболее общие свойства клеток, а затем и более частные.

Как известно, на этапе гаструляции возникают эмбриональные зачатки. Клетки, которые входят в их состав, еще не окончательно детерминированы, так что из одного зачатка возникают клеточные совокупности, обладающие разными свойствами. Следовательно, один эмбриональный зачаток может служить источником развития нескольких тканей.

Последовательная ступенчатая детерминация и коммитирование потенций однородных клеточных группировок – дивергентный процесс. Развитие тканей в эмбриогенезе происходит в результате дифференцировки клеток. Под дифференцировкой понимают изменения в структуре клеток в результате их функциональной специализации, обусловленные активностью их генетического аппарата. Различают четыре основных периода дифференцировки клеток зародыша – оотипическую, бластомерную, зачатковую и тканевую дифференцировку. Проходя через эти периоды, клетки зародыша образуют ткани (гистогенез).

Каждая ткань имеет или имела в эмбриогенезе стволовые клетки – наименее дифференцированные и наименее коммитированные. Они образуют самоподдерживающуюся популяцию, их потомки способны дифференцироваться в нескольких направлениях под влиянием микроокружения (факторов дифференцировки), образуя клетки-предшественники и, далее, функционирующие дифференцированные клетки. Таким образом, стволовые клетки полипотентны. Они делятся редко, пополнение зрелых клеток ткани, если это необходимо, осуществляется в первую очередь за счет клеток следующих генераций (клеток-предшественников). По сравнению со всеми другими клетками данной ткани стволовые клетки наиболее устойчивы к повреждающим воздействиям.

Если одна из стволовых клеток вступает на путь дифференциации, то в результате последовательного ряда коммитирующих митозов возникают сначала полустволовые, а затем и дифференцированные клетки со специфической функцией. Выход стволовой клетки из

популяции служит сигналом для деления другой стволовой клетки по типу некоммитирующего митоза. Общая численность стволовых клеток в итоге восстанавливается. В условиях нормальной жизнедеятельности она сохраняется приблизительно постоянной.

Совокупность клеток, развивающихся из одного вида стволовых клеток, составляет стволовой дифферон. Часто в образовании ткани участвуют различные дифферoны. Так, в состав эпидермиса кроме кератиноцитов входят клетки, развивающиеся в нейральном гребне и имеющие другую детерминацию (меланоциты), а также клетки, развивающиеся путем дифференциации стволовой клетки крови, т. е. принадлежащие уже к третьему дифферону (внутриэпидерминальные макрофаги, или клетки Лангерганса).

Значение микроокружения для самоподдержания популяции стволовых клеток. Дифференцированные клетки наряду с выполнением своих специфических функций способны синтезировать особые вещества – кейлоны, тормозящие интенсивность размножения клеток-предшественников и стволовых клеток. Если в силу каких-либо причин количество дифференцированных функционирующих клеток уменьшается (например, после травмы), тормозящее действие кейлонов ослабевает и численность популяции восстанавливается. Кроме кейлонов (местных регуляторов) клеточное размножение контролируется гормонами; одновременно продукты жизнедеятельности клеток регулируют активность желёз внутренней секреции. Если какие-либо клетки под воздействием внешних повреждающих факторов претерпевают мутации, они элиминируются из тканевой системы вследствие иммунологических реакций.

Выбор пути дифференциации клеток определяется межклеточными взаимодействиями. Влияние микроокружения изменяет активность генома дифференцирующейся клетки, активируя одни и блокируя другие гены. У клеток, уже дифференцированных и утративших способность к дальнейшему размножению, строение и функция тоже могут изменяться (например, у гранулоцитов начиная со стадии метамиелоцита). Такой процесс не приводит к возникновению различий среди потомков клетки и для него больше подходит название «специализация».

Даже следовые концентрации эндотоксина в среде культивирования вызывают массовую гибель незрелых стволовых клеток. Впрочем, получить клетки трофобласта из ЭСК человека или приматов удается. Недавние совместные исследования израильских и американских биологов выявили комбинации из 8 ростовых факторов, которые направляют дифференцировку эмбриоидных телец в сторону экто-, мезо- или эндодермы. Так, активин в сочетании с TGF-beta предпочтительно стимулировал развитие мезодермы, тогда как комбинация ретиноевой кислоты, bFGF, EGF, BMP-4 стимулировали одновременное развитие экто- и мезодермы. Возникновение эндодермы в культуре не шло без добавления HGF и NGF.

Доказано, что сигналы микроокружения играют решающую роль в судьбе трансплантированных ЭСК/МСК in situ. Существенно, что ЭСК/МСК при дифференцировке в культуре дают лишь «природные» линии дифференцированных клеток, которые имеются в организме взрослого животного и человека: новых типов клеток или неизвестных линий дифференцированных клеток из ЭСК *in vitro* не возникает. Наблюдения подтверждены многочисленными работами, что снижает риск осложнений и повышает безопасность клеточной терапии стволовыми клетками.

Молекулярные маркеры стволовых клеток

Маркеры эмбриональных стволовых клеток (ЭСК). Классическими маркерами ЭСК являются изозимы щелочной фосфатазы, транскрипционный фактор Oct-4, высокая теломеразная активность и ряд маркеров клеточной поверхности, например GCTM-2, TRA-1-60, SSEA-3 и SSEA-4, распознаваемые моноклональными антителами к специфическим эмбриональным или опухоль-определяющим антигенам. Физиологическое значение большинства маркеров остается неясным, за исключением Oct-4. Исследования, проведенные на ЭСК и

эмбрионах мыши, выявили критическую роль Oct-4 в поддержании тотипотентности ранних эмбриональных клеток и клеток зародышевого пути. Дифференцировка клеток внутренней массы сопровождается понижением уровня Oct-4, а изменение уровня синтеза Oct-4 в ЭСК в свою очередь приводит к потере тотипотентности и переходу к дифференцировке.

Кроме Oct-4 имеется еще ряд транскрипционных факторов, синтезируемых в основном недифференцированными ЭСК. например Nanog, который занимает важное место в иерархии факторов, определяющих недифференцированную природу ЭСК и происхождение.

В последние годы бессмертные тотипотентные линии ЭСК были выделены из эмбрионов крыс, коровы, приматов, свиней. Показательно, что все ЭСК имели не только близкое строение, но и практически универсальный набор антигенов-маркеров (SSEA-3, SSEA-4, TRA-1-81, TRA-1-60).

Маркеры гематопоэзиых стволовых клеток (ГСК). Клетки, из которых образуются клетки крови, имеют на своей поверхности маркеры CD34, CD59 и Thyl, по которым могут быть идентифицированы.

Маркеры нейральных стволовых клеток (ИСК). Известны молекулярные маркеры, позволяющие идентифицировать как стволовые нервные клетки, так и последовательные фазы их развития, – это нестин для стволовой клетки, виментин для клетки-предшественника, бета-тубулин для нейробласта, GFAP (кислый глиальный фибриллярный белок) для клетки, «движущейся» в направлении глиального развития и т. д.

Источники стволовых клеток у взрослого организма

Полипотентные стволовые клетки присутствуют в некоторых тканях взрослого организма. Стволовые клетки в организме взрослого человека вырабатывает костный мозг. Это основной их источник, но далеко не единственный. Также стволовые клетки обнаружены и в жировой ткани, коже, мышцах, печени, легких, сетчатке глаза, практически во всех органах и тканях организма. Они обеспечивают восстановление поврежденных участков органов и тканей, служат источником клеток различных тканей, естественным образом выбывающих из строя. Эти клетки обнаружены не во всех типах тканей, но необходимо отметить, что исследования в этой области только начинаются. Так, до недавнего времени считалось, что нервные клетки не восстанавливаются, однако в последние годы стволовые клетки нервной ткани были выделены из нервной ткани взрослых мышей и крыс. У человека такие клетки обнаружены в соответствующей ткани плода; кроме того, клетки, сходные со стволовыми клетками нервной ткани, обнаружены в мозге больного эпилепсией, часть которого была удалена в ходе операции.

С возрастом количество стволовых клеток становится все меньше, и соответственно восстановительные возможности организма снижаются. Так, когда мы рождаемся, у нас в костном мозге на 10 тыс. кроветворных клеток приходится одна стволовая клетка. У растущих подростков стволовых клеток уже в 10 раз меньше. К 50 годам на 0,5 млн обычных клеток приходится 1 стволовая, в 70 лет – 1 стволовая клетка на 1 миллион. В результате страдает способность ткани к физиологической регенерации и к восстановлению после болезни или травмы.

Обладают ли стволовые клетки взрослого организма таким же потенциалом, как и плюрипотентные клетки? До недавнего времени практически не было данных о том, что полипотентные стволовые клетки млекопитающих, например гемопоэтические стволовые клетки, могут изменить направление своего развития и дать начало клеткам кожи, печени или любым другим специализированным клеткам, отличным от форменных элементов крови. Однако проведенные в последние годы опыты на животных показали, что здесь еще рано ставить точку. Обнаружилось, что некоторые стволовые клетки животных, ранее считавшиеся строго специализированными, при некоторых условиях могут менять свою специализацию. Так, стволовые клетки нервной ткани мыши, введенные в костный мозг, оказались способными дифференцироваться в разные клетки крови, а стволовые клетки, обнаруженные в костном

мозге крыс, могут дифференцироваться в клетки печени. Таким образом, показано, что при определенных условиях стволовые клетки проявляют большую гибкость, чем это считалось ранее.

Почему нельзя использовать «взрослые» стволовые клетки в медицинских целях прямо сейчас? Стимулом к изучению стволовых клеток человека служит то, что технология их использования открывает огромные возможности в клеточной терапии, прежде всего при трансплантации. Если бы удалось получить стволовую клетку от взрослого индивида, стимулировать ее деление и изменить специализацию, ее можно было бы ввести в организм донора, не опасаясь отторжения. Такой подход избавил бы от необходимости использовать стволовые клетки человеческого эмбриона или плода – эта практика вызывает неприятие общественности по этическим соображениям.

Однако, несмотря на всю перспективность, этот метод сталкивается с серьезными проблемами. Во-первых, стволовые клетки обнаружены далеко не во всех типах тканей взрослого человека. Так, не найдены стволовые клетки сердечной мышцы и островков поджелудочной железы. Во-вторых, уже обнаруженные клетки присутствуют в тканях в очень малых количествах, их трудно выделить и очистить, а с возрастом их становится еще меньше.

Чтобы стволовые клетки взрослого человека можно было использовать для его же лечения, нужно, прежде всего, получить их от данного пациента, затем культивировать до достижения достаточно большой плотности, чтобы их хватило для терапии. Однако бывают случаи, когда болезнь просто не дает времени на проведение всех этих процедур; кроме того, если заболевание имеет генетическую природу, пораженными, скорее всего, будут и стволовые клетки. Есть указания на то, что стволовые клетки взрослого организма делятся не так быстро, как стволовые клетки плода, а их ДНК, по-видимому, содержит больше нарушений. Кроме того, из одной линии «взрослых» стволовых клеток можно получить не более 3–4 типов тканей.

Стволовые клетки в медицине

Применение стволовых клеток для восстановления органов. Одно из самых впечатляющих применений плюрипотентных клеток человека – это так называемая клеточная терапия. Многие заболевания человека обусловливаются нарушением функционирования клеток или целых органов, и сегодня для устранения дефекта в таких случаях используется метод трансплантации. К сожалению, нередко повреждения носят множественный характер, и заменить все затронутые ими органы не представляется возможным. Плюрипотентные клетки, стимулированные к дифференцировке с образованием строго специализированных клеток, могут служить возобновляемым источником клеток, замещающих дефектные клетки. Это открывает широкие возможности для лечения самых разных заболеваний человека, включая такие серьезные, как болезнь Паркинеона, болезнь Альцгеймера, сердечно-сосудистые заболевания, ревматоидный артрит, диабет и др. (рис. 42). Не зря одна из статей в журнале «Cell» («Клетка») называется «Стволовые клетки: Возрождение надежд» (одно из значений американского словосочетания «new lease on life»).

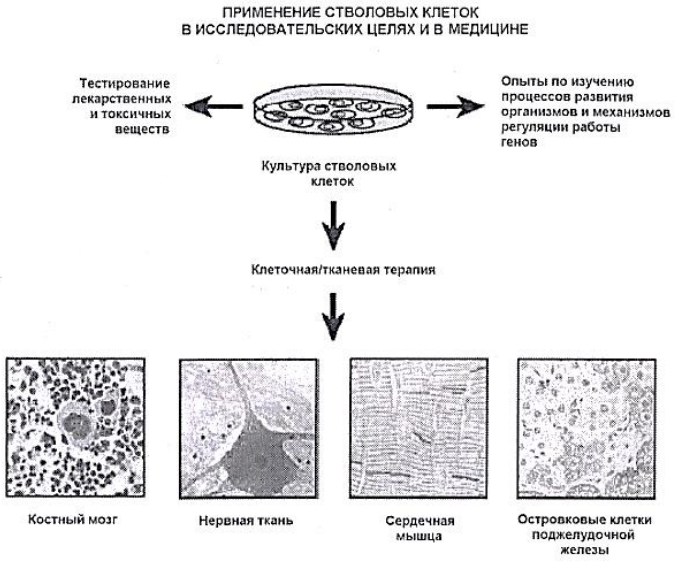


Рисунок 42 – Основные направления использования стволовых клеток в научных исследованиях и медицинской практике

Изолированные плюрипотентные клетки человека – ценный материал для исследователей и клиницистов. Эксперименты с их использованием могут помочь разобраться в сложнейших процессах развития человеческого организма, в частности в том, что именно влияет на переход клетки от стадии роста и деления к стадии дифференцировки. Известно, что ключевым моментом здесь является «включение» и «выключение» специфических генов, но в понимании деталей этого процесса еще много неясного. Изучив функционирование клетки в норме, можно будет понять, какие сбои в ее работе приводят к фатальным для организма последствиям.

Выделение плюрипотентных клеток человека открывает также новые возможности при поиске новых лекарственных веществ и их тестировании. Разнообразные клеточные линии (например, линии раковых клеток) используются в этих целях уже сейчас, а культура плюрипотентных клеток позволяет проводить тестирование сразу на нескольких типах клеток. Это не заменяет тестирование на уровне целого организма, но значительно облегчает и ускоряет поиск новых лекарственных веществ.

С помощью ЭСК разрабатываются эффективные технологии лечения многих наследственных заболеваний, которые не удалось решить методами молекулярной генетики, включая средства генной терапии. В целях экстренной помощи (например, при острой недостаточности органа) обоснованы и допустимы трансплантации аллогенных стволовых клеток, которые в определенных условиях микроокружения способны дать ростки донорской ткани в органе-реципиенте в количестве, достаточном для компенсации генетического дефекта. Всего 3–5% генетически нормальных клеток достаточно, чтобы восстановить утраченную биохимическую функцию поврежденной печени, скелетной мышцы, эндокринной железы. Проблемы использования эмбриональных стволовых клеток обсуждаются в многочисленных статьях, в частности, можно рекомендовать обзор в журнале «The New England Journal of Medicine», который имеется в свободном он-лайновом доступе.

Что тормозит применение методов стволовых клеток в клинике? Во-первых, необходимо выяснить, какие события предшествуют переходу клетки в организме человека к стадии дифференцировки, – тогда можно будет управлять ходом событий, получая из плюрипотентных клеток именно те клетки, которые нужны для трансплантации. Во-вторых, следует решить проблему иммунологического отторжения. Поскольку плюрипотентные клетки, взятые из бластоцисты или ткани плода, вряд ли будут идентичны клеткам реципиента,

необходимо научиться модифицировать их для минимизации этого различия или создать банк тканей.

В некоторых случаях проблему несовместимости удается решить, используя метод переноса ядра соматической клетки. Предположим, что пациент страдает прогрессирующей сердечной недостаточностью. Если взять у него любую соматическую клетку и ввести ее ядро в лишенную собственного ядра яйцеклетку-реципиент, мы получим химерную яйцеклетку, у которой практически весь генетический материал идентичен таковому у пациента. Из нее можно получить бластоцисту, а затем, отобрав клетки внутренней клеточной массы, – плюрипотентные клетки. Последние можно стимулировать к образованию клеток сердечной мышцы, идентичных в генетическом: отношении нормальным клеткам пациента, и имплантировать их больному без необходимости подвергать его иммуносупрессорной терапии, чреватой серьезными последствиями. Однако на практической реализации этого пути тоже встречается большое количество подводных камней. В частности, необходимо научиться перепрограммировать соматическое донорское ядро, возвращая такие его характеристики, как метилирование ДНК, гистонная модификация и, прежде всего, структура хроматина, к состоянию, свойственному ядру недифференцированной эмбриональной клетки.

Лечение заболеваний сердца. Предварительные эксперименты на мышах показали, что трансплантация нормальных клеток сердечной мышцы животным, страдающим сердечной недостаточностью, приводит к частичной замене пораженных клеток и восстановлению сердечной деятельности. При этом не наблюдалось никакого отторжения донорных клеток клетками-хозяевами. Аналогичной процедуре можно подвергнуть и человека, страдающего сердечной недостаточностью, используя в качестве трансплантата клетки сердечной мышцы, полученные из плюрипотентных клеток человека.

Лечение заболеваний костной системы. Введение полученных из стволовых клеток остеобластов (предшественники клеток кости) в зону перелома значительно ускоряет процесс сращивания кости, а введение хондробластов (предшественники клеток хряща) в суставную сумку стимулирует регенерацию хряща на суставных поверхностях.

Лечение тканевых дефектов. Эпидермис – классический пример ткани, самообновляющейся благодаря функционированию стволовых клеток. Здесь разработаны методы получения чистой популяции базальных клеток в культуре и реализовано разделение различных субпопуляций по их фенотипу и поведению во время регенерации. Для изучения стволовых и прогениторных клеток использованы предполагаемые маркеры стволовых клеток рб3 и К19, а также выброс ядерного красителя Hoechst: 33342.

Полученные результаты позволяют заключить, что в культуре базальных кератиноцитов сохраняется популяция клеток, которые по ряду признаков могут быть отнесены к стволовым. В базальных кератиноцитах человека обнаружена экспрессия нестина. Кроме этого разработаны методы для изучения в культуре длительно сохраняющих метку клеток. Обнаружена значительная гетерогенность клеток базального слоя эпидермиса по кинетическим и морфологическим параметрам. Разработаны тканевые трансплантаты, содержащие базальные кератиноциты человека, которые успешно использованы для лечения различных тканевых дефектов.

Лечение диабета. У больных, страдающих диабетом I типа, нарушена секреция инсулина специализированными, так называемыми островковыми клетками поджелудочной железы, и такие больные нуждаются в инъекциях инсулина. Есть данные о том, что трансплантация таким больным целой поджелудочной железы или островковых клеток позволяет существенно снизить дозу препарата. Островковые клетки, полученные из плюрипотентных клеток человека, можно использовать как в исследовательских целях для изучения диабета, так и для клеточной терапии.

Мобилизация донорских и эндогенных стволовых клеток. Мобилизация стволовых клеток – стимуляция миграции (выхода) стволовых клеток в кровоток с применением ростовых

факторов, в частности специальных цитокинов. Для гемопоэтических стволовых клеток факторы мобилизации – гранулоцитарный колониестимулирующий фактор (Г-КСФ) и гранулоцитарномакрафагальный колониестимулирующий фактор (ГМ-КСФ). Например, в клинической кардиомиопластике при лечении ишемической болезни сердца перспективным представляется применение в качестве мобилизующих факторов Г-КСФ и интерлейкина ИЛ-2.

ЭСК: законодательство и биоэтика. После создания методов генной инженерии, рекомбинантных технологий переноса ДНК между клетками разных видов, завершения программы «Геном человека», получения первых линий ЭСК человека возникла новая биоэтика фундаментальных процессов, стоящих у истоков индивидуальной жизни. Эти новые реальности биоэтики нашли отражение в документе ЮНЕСКО: Биоэтика: международные аспекты (октябрь, 2001). Подборку статей по юридическим и биоэтическим проблемам использования стволовых клеток можно найти в юбилейном выпуске журнала «Человек» (№ 1 за 2005 г.; см. также их он-лайновую версию).

В принятой ЮНЕСКО Универсальной декларации о геноме человека и правах человека (1999 г.) в статье 12 постулируется: «Свобода научных исследований, необходимых для прогресса знаний, есть часть свободы мысли. Практическое приложение знаний генома человека должно быть направлено на уменьшение человеческих страданий, улучшение здоровья как индивидов, так и всего человечества». Разработки с геномом ЭСК – часть этой общей стратегии.

В частности, по информации EUobserver. сот, министры по охране окружающей среды ЕС окончательно одобрили рекомендации ЕС по развитию генной терапии, оставляя за государствами-членами право наложения запрета на исследование эмбриональных клеток человека по этическим причинам. Законопроект, принятый 30 октября 2007 г., содержит технические детали о регулировании на уровне ЕС видов так называемой углубленной терапии – генной терапии, терапии стволовыми клетками взрослого организма и тканевой инженерии. Эта практика в настоящий момент законна только в нескольких странах – в том числе в Британии, а новые правила поддерживают право каждого государства – члена ЕС на запрет как исследований, так и продажи лекарств, которые получают из клеток человеческих эмбрионов.

Генная терапия с использованием стволовых клеток. Дискуссии об этичности клонирования человека и получении линий эмбриональных стволовых клеток (ЭСК) для «терапевтического» клонирования продолжаются и в настоящее время. Большинство специалистов отмечают, что возможность такого клонирования во многом лишь теория, пока только частично подтвержденная на практике. Наиболее перспективно использование комплексного подхода, сочетающего применение технологии ЭСК и генотерапии одновременно.

В подтверждение этого тезиса американские ученые из Института Уайтхед смогли продемонстрировать перспективность комплексного подхода, названного ими «клеточной терапией», при лечении наследственных заболеваний. Им удалось практически полностью восстановить иммунитет у мышей, страдающих серьезным наследственным иммунодефицитом, одновременно используя полученные в ходе «терапевтического клонирования» ЭСК и вводя животным их собственные гематопоэтические клетки с векторами, несущими определенные гены. В результате уже через две недели мыши из опытной группы были практически здоровы, в то время как существенных изменений у животных из контрольных групп не наблюдалось.

Еще более впечатляющее применение стволовых клеток человека – генная терапия *ex vivo*. В этом случае в организм больного можно ввести не обычные стволовые клетки, а генетически модифицированные, которые замешают дефектные клетки или восполняют недостаток продукта того гена, который включен в геном вводимых клеток. Стволовые клетки можно получать от самого пациента или от совместимых с ним доноров. Следует отметить, что генная терапия *ex vivo* с применением стволовых клеток человека делает лишь первые шаги.

Более реальным является использование модифицированных эмбриональных стволовых клеток для создания трансгенных животных. Соответствующие эксперименты уже широко проводятся на мышах (рис. 43).

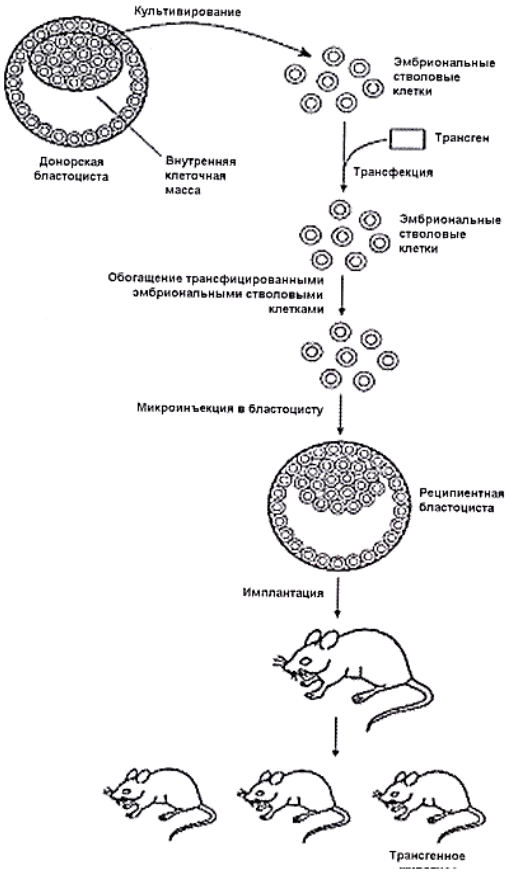


Рисунок 43 – Схема получения трансгенных мышей

Сначала получают эмбриональные стволовые клетки из внутренней клеточной массы бластоцисты мыши. Их генетически модифицируют с помощью вектора, несущего нужный ген (трансген), культивируют и отбирают тем или иным способом. Полученную популяцию клеток вновь культивируют и вводят в бластоцисты, которые затем имплантируют в матку «суррогатной» матери. Скрещивая животных, несущих трансген в клетках зародышевой линии мыши, получают линию трансгенных мышей.

В геном стволовой клетки можно не только встроить полезный ген, кодирующий какой-либо необходимый организму продукт, но и направленным образом вывести из строя («нокаутировать») ген, кодирующий, например, какой-нибудь токсин. Трансгенных мышей с нарушениями в определенном гене широко используют в качестве модели для изучения заболеваний человека на молекулярном уровне. Именно разработка и широкое применение метода «генного нокаута» стали одним из оснований присуждения Нобелевской премии 2007 г.

Сайты и базы данных по стволовым клеткам и клеточным технологиям. Банки эмбриональных стволовых клеток. Существует огромное количество ссылок на сайты и базы данных по стволовым клеткам и клеточным технологиям, включая международные журналы, находящиеся в свободном доступе. Только в Рунете на начало марта 2008 г. поиск по словам «стволовые клетки» и «базы данных» выдал 182 ООО ссылок, на слова «банки стволовых клеток» – 39 600 ссылок. В качестве примера можно привести сайт нашей лаборатории по старению, международный журнал «Cell», информационный сайт КлеткаРу, банки стволовых

клеток. Что касается конкретно банков стволовых клеток человека, то, по определению, – это система специализированных учреждений, основной задачей которых является подготовка образца к криоконсервации и длительное, практически не ограниченное во времени хранение ранее заготовленных стволовых клеток, вне зависимости от источника их получения. Существуют две системы банков стволовых клеток: банки персонального хранения и донорские банки. В банках стволовых клеток персонального хранения находятся клетки для конкретного реципиента. В донорских банках стволовые клетки хранятся для аллогенных трансплантаций.

Литература: Дромашко, С. Е. Очерки биоинформатики / С. Е. Дромашко ; НАН Беларуси, Ин-т генетики и цитологии ; науч. ред. Л. В. Хотылева. – Минск : Беларус. навука, 2009. – 400 с.

Занятие 6.1 Иерархические и диффузные системы мозга

Вопросы для изучения:

1. Кратковременные и долговременные изменения в синапсах.
2. Иерархические и диффузные системы мозга.
3. Стереотаксис.

Учебно-методические материалы.

Литература для изучения

1. Дромашко, С. Е. Мозг, интеллект, нейроинформатика : учеб.-метод. пособие / С. Е. Дромашко, Р. В. Телятников ; НАН Беларуси, Ин-т подгот. науч. кадров, Каф. естественнонауч. дисциплин. – Минск : ИПНК НАН Беларуси, 2010. – 54 с.
2. Дромашко, С. Е. Очерки биоинформатики / С. Е. Дромашко ; НАН Беларуси, Ин-т генетики и цитологии ; науч. ред. Л. В. Хотылева. – Минск : Беларус. навука, 2009. – 400 с.

Занятие 6.2 Нейроинформатика. Моделирование процессов мышления и человеко-машинного общения. Искусственный интеллект

Вопросы для изучения:

1. Математический нейрон Маккалока-Питтса.
2. Линейная передаточная функция. Сигмоидальная передаточная функция. Логистическая функция.
3. Гиперболический тангенс. Моделирование формальных логических функций.
4. Различия между биологическим и искусственным нейроном.
5. Нейронные сети. Приложение искусственных нейронных сетей к распознаванию образов. Будущее нейрокомпьютинга.

Учебно-методические материалы.

Литература для изучения

1. Дромашко, С. Е. Мозг, интеллект, нейроинформатика : учеб.-метод. пособие / С. Е. Дромашко, Р. В. Телятников ; НАН Беларуси, Ин-т подгот. науч. кадров,

Каф. естественнонауч. дисциплин. – Минск : ИПНК НАН Беларуси, 2010. – 54 с.
2. Дромашко, С. Е. Очерки биоинформатики / С. Е. Дромашко ; НАН Беларуси, Ин-т генетики и цитологии ; науч. ред. Л. В. Хотылева. – Минск : Беларус. навука, 2009. – 400 с.

Занятие 6.3 Биоинформатика и актуальные проблемы психофизиологии

Вопросы для изучения:

1. Современное состояние научной разработки проблемы коммуникации в физиологии и психологии.
2. Моделирование процессов мышления и искусственный интеллект. Искусственный интеллект и поведенческое мышление.
3. Современные представления о нейрофизиологических механизмах психических функций. РНК-интерференция и психические заболевания.

Учебно-методические материалы.

Современное состояние научной разработки проблемы коммуникации в физиологии и психологии. ***Психофизиология*** – пограничная область психологии, изучающая роль всей совокупности биологических свойств, и прежде всего свойств нервной системы, в детерминации психической деятельности и устойчивых индивидуально-психологических различий. Главная задача психофизиологии – причинное объяснение психических явлений путем раскрытия лежащих в их основе нейрофизиологических механизмов.

Психофизиология включает несколько областей исследования. ***Психофизиология ощущений и восприятий*** изучает нервные процессы в анализаторах, начиная с рецепторов и кончая корковыми отделами. Установлены специфические аппараты цветового зрения, специфические рецепторы и проводящие пути тактильной и болевой чувствительности, открыты нейроны, реагирующие на отдельные свойства зрительных и слуховых стимулов.

- ***Психофизиология речи и мышления*** изучает функциональную роль разных областей мозга и их взаимосвязей в осуществлении речевых процессов. Принципиально важным стало установление тесной связи мыслительных процессов с деятельностью речедвигательного анализатора.
- ***Психофизиология эмоций*** исследует нейрогуморальные механизмы возникновения эмоциональных состояний. Открыты нервные «центры» удовольствия и неудовольствия, расположенные в подкорковых областях мозга. Установлено, что важная роль в эмоциональном поведении принадлежит гормонам, выделяемым железами внутренней секреции (гипофизом, корой и мозговым слоем надпочечников и др.), а также различными биологически активными веществами.
- ***Психофизиология внимания*** исследует нейрофизиологические корреляты внимания (изменение электроэнцефалограммы и вызванных потенциалов, изменение кожно-гальванической и других реакций). Психофизиология внимания тесно связана с проблемами изучения ориентировочного рефлекса и второй сигнальной системы.
- ***Психофизиология произвольных действий*** вскрывает физиологическую структуру и механизмы их осуществления.
- ***Дифференциальная психофизиология*** изучает зависимость индивидуальных особенностей психики и поведения от индивидуальных различий в деятельности мозга.

К ***методам психофизиологии*** (см. статью «Методы психофизиологии» на сайте) относятся методы изучения работы головного мозга, измерение электрической активности кожи, снятие показателей работы сердечно-сосудистой системы, регистрация показателей активности мышечной системы, измерение показателей активности дыхательной системы, регистрация реакции глаз. Большинство этих методов объединяется в приборе, который по шпионским фильмам известен как ***детектор лжи.*** Остановимся кратко на этих методах.

Изучение работы головного мозга.

- ***Электроэнцефалография (ЭЭГ)*** – метод регистрации и анализа электроэнцефалограммы, т. е. суммарной биоэлектрической активности, отводимой как со скальпа, так и из глубоких структур мозга.
- ***Магнитоэнцефалографчя (МЭГ)*** – регистрация параметров магнитного поля, обусловленных биоэлектрической активностью головного мозга.
- ***Вызванные потенциалы*** (***ВП***) – биоэлектрические колебания, возникающие в нервных структурах в ответ на внешнее раздражение и находящиеся в строго определенной временной связи с началом его действия.
- ***Топографическое картирование электрической активности мозга (ТКЭАМ)*** – область электрофизиологии, оперирующая с множеством количественных методов анализа электроэнцефалограммы и вызванных потенциалов.
- ***Компьютерная томография (КТ)*** – новейший метод, дающий точные и детальные изображения малейших изменений плотности мозгового вещества. КТ соединила в себе последние достижения рентгеновской и вычислительной техники, отличаясь принципиальной новизной технических решений и математического обеспечения.
- ***Ядерно-магнитно-резонансная томография мозга*** – компьютерная томография стала родоначальницей ряда других, еще более совершенных методов исследования: томографии с использованием эффекта ядерного магнитного резонанса (ЯМР-томография), позитронной эмиссионной томографии (ПЭТ), функционального магнитного резонанса (ФМР).

— ***Измерение электрической активности кожи*** – кожно-гальванической реакции (КГР), обусловленной главным образом активностью потовых желез в коже человека, которые в свою очередь находятся под контролем симпатической нервной системы.

— Снятие показателей работы сердечно-сосудистой системы.

- ***Электрокардиограмма (ЭКГ)*** – запись электрических процессов, связанных с сокращением сердечной мышцы.
- ***Плетизмография*** – метод регистрации сосудистых реакций организма.

— ***Регистрация показателей активности мышечной системы.*** Это в первую очередь ***электромиография*** – метод исследования функционального состояния органов движения путем регистрации биопотенциалов мышц.

— ***Измерение показателей активности дыхательной системы.*** Для измерения интенсивности (амплитуды и частоты) дыхания используют специальный прибор – пневмограф. Этот метод обеспечивает хорошую запись изменений частоты и амплитуды дыхания.

— ***Регистрация реакции глаз.*** Для психофизиолога наибольший интерес представляют три категории глазных реакций: сужение и расширение зрачка, мигание и глазные движения, в соответствии с чем для регистрации используются:

- ***пупиллометрия*** – метод изучения зрачковых реакций;
- ***мигание*** (моргание) – периодическое смыкание век;
- ***движение глаз*** – разнообразные по функции, механизму и биомеханике вращения глаз в орбитах;
- ***электроокулография*** – метод регистрации движения глаз, основанный на графической регистрации изменения электрического потенциала сетчатки и глазных мышц.

Моделирование процессов мышления и человеко-машинного общения. Когнитивная психология и искусственный интеллект. ***Когнитивная психология*** (лат. ***cognitio*** – знание, познание) – направление в психологии, возникшее в США в конце 1950 – начале 1960-х годов XX в. в противовес бихевиоризму. Возникновению ***бихевиоризма*** способствовали разработки в области условных рефлексов и развитие зоопсихологии. Бихевиоризмом не принимались во внимание врожденные и приобретенные качества человека, его убеждения, основой всего считалось поведение. Развитие когнитивной психологии стало возможным благодаря предшествовавшим работам в области гештальтпсихологии, привлекшим внимание к вопросам восприятия и эффективному, творческому мышлению, исследование которого является одной из основных задач науки.

Когнитивная психология во многом основывается на аналогии между преобразованием информации в вычислительном устройстве и осуществлением познавательных процессов у человека, для пояснения этой аналогии было введено понятие компьютерной метафоры. ***Когнитивная система человека*** рассматривается как система, имеющая устройства ввода, хранения и вывода информации. Информация, достигшая когнитивной системы, преобразуется, обрабатывается, кодируется, хранится, запоминается и забывается, а затем преобразуется в знание. ***Методологическая платформа*** когнитивной психологии основана на компьютерных технологиях и новых теоретических исследованиях в области психологии.

Обобщая все сказанное выше, мы снова приходим к проблеме ***искусственного интеллекта.*** Эта отрасль компьютерных наук, моделирующая познавательные процессы человека, оказала огромное влияние на развитие когнитивной психологии, особенно когда для компьютерных программ искусственного интеллекта потребовались знания, как человек обрабатывает информацию.

Искусственный интеллект и поведенческое моделирование.

Поведенческие компьютерные модели или роботы, по мнению К. В. Анохина, в отличие от моделей классического искусственного интеллекта не имеют предварительной схемы мира, они должны строить ее под существующую среду подобно реальным животным. Такие системы адаптируются к среде с помощью набора элементарных поведенческих модулей, каждый из которых отвечает за какую-то независимую элементарную форму поведения, а затем методом проб и ошибок происходит отбор и фиксация таких модулей по результатам их действий. Такие системы получили название искусственных животных – ***«аниматов»*** (от англ. ***animal*** – животное и ***automat*** – автономно действующее устройство). Первыми существами, чье поведение пытались моделировать, были насекомые – пауки, пчелы, осы. Такие насекомоподобные роботы создаются в разных научных центрах в США (шестиногое 35-сантиметровое «насекомое- робот» Genghis), Франции (насекомоподобный ходящий робот Hexapod), Великобритании (целый ряд насекомоподобных «интеллектуальных» существ).

В настоящее время в лаборатории аниматов в Париже ***(Animat Lab)*** разрабатывается проект Psikharpax, где в роботе синтезируются некоторые из адаптивных механизмов и нервных структур, ответственных за пространственную навигацию у крыс. Способности такой крысы-робота будут расти за счет «обучения без учителя», т. е. анимат будет способен сам строить когнитивную карту среды и вырабатывать адаптивные стратегии поведения по механизмам, схожим с теми, что использует мозг крысы. В группе гуманоидной роботики в лаборатории искусственного интеллекта Массачусетского технологического института в США (***Humanoid Robotics Group)*** разрабатываются обезьяноподобные и мобильные роботы (Kismet, Coco) со сложным набором поведенческих реакций (способности к социальным взаимодействиям, эмоциональным реакциям и т. п.). На рис. 44 представлена главная страница этой группы.

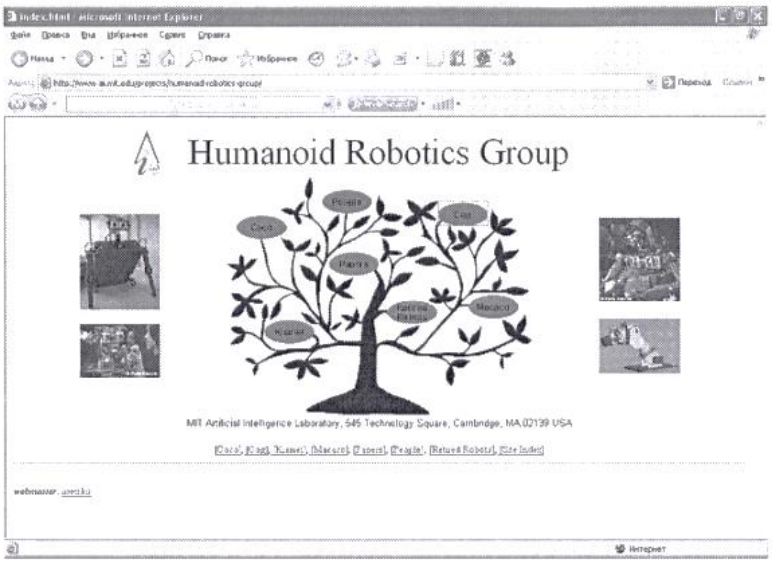


Рисунок 44 – Главная страница сайта группы гуманоидной роботики.

Коренные различия классического искусственного интеллекта и поведенческого моделирования в свое время остроумно отразил Бертран Рассел, объясняя национальную разницу в школах немецкой гештальтпсихологии и американского бихевиоризма: «Животные, которых наблюдали немцы, спокойно сидят, думают и, в конце концов, выдают решение из своего внутреннего сознания. А животные, которых изучали американцы, бешено прыгают с невероятным напором и темпераментом и, наконец, случайно получают желаемый результат».

На нынешней стадии работ по искусственному интеллекту, похоже, эти различия начинают сглаживаться. Многие нейробиологи настаивают на том, что теории работы мозга должны быть сформулированы алгоритмично, чтобы допускать моделирование. Нобелевский лауреат ***Джеральд Эделман (Gerald Edeiman),*** один из ведущих нейробиологов-теоретиков, в своем институте в Калифорнии (Neuroscience Institute) разрабатывает серию роботов ***NOMAD (Neurally Organized MobileAdaptive Devices).*** Эти роботы имеют родовое имя «Дарвин». Каждый новый «Дарвин» появляется на свет практически необученным, а когда сталкивается с объектами внешнего мира, начинает вырабатывать собственные абстрактные категории на основе имеющегося у него какого-нибудь врожденного предпочтения (учится, например, отличать друг от друга разные шары). Таким образом, у робота появляются знания, которые он может использовать и в других задачах – начинает работать один из принципов, по которым, видимо, шла в природе эволюция механизмов интеллекта.

Современные представления о нейрофизиологических механизмах психических функций.

Для физиологов все актуальнее становится проблема биологических предпосылок психических функций мозга человека.

Деятельность мозга животных проявляется в организации поведенческих актов, направленных на поддержание состояния устойчивого неравновесия с внешней средой. Активный характер поведения определяет стремление к удовлетворению потребностей. Очевидным, не требующим доказательств, представляется наличие у животных ***витальных и зоосоциалъных потребностей,*** направленных на выживание индивидуума и вида соответственно. К витальным потребностям относятся, например, пищевая и питьевая потребности, на базе которых может осуществляться выработка ***пищевых и питьевых условных рефлексов.*** На основе зоосоциальных потребностей формируются такие формы поведения, как ***половое, материнское, территориальное*** и некоторые другие. Не столь очевидно существование так называемых ***потребностей саморазвития,*** на базе которых

формируется игровое и исследовательское поведение. Тем не менее, многочисленные данные о формировании новых условнорефлекторных реакций в тех случаях, когда подкреплением является не некое лакомство, а возможность проводить исследовательскую деятельность, являются убедительным доказательством наличия таких потребностей. Например, обезьяны обучались нажимать на рычаг, если в качестве подкрепления они получали возможность увидеть через окно в соседнем помещении интересные вещи. Всем известно вошедшее в поговорки любопытство сорок и т. п.

На основе удовлетворения потребностей происходит условнорефлекторное обучение животных, проявляющееся в изменении их поведения, – один из вариантов обучения. Однако не для всех форм обучения доказано наличие фактора подкрепления, и поэтому имеются разные точки зрения на возможность сведения всех процессов обучения к выработке условных рефлексов. К таким формам обучения можно отнести ***импринтинг*** (запечатлевание), ***латентное обучение, имитацию*** (подражание).

Импринтинг в отличие от условнорефлекторного обучения не зависит от количества сочетаний и эффективно осуществляется лишь в определенный, критический, период жизни. Этот период, к примеру, у птенцов длится с 10 до 20 ч после рождения. Установление ранних контактов между матерью и детенышем, возможно, имеет особое значение и для материнского запечатлевания, обусловливающего материнскую привязанность.

Инсайт – это особый вид обучения, проявляющийся в осуществлении новой адаптивной реакции без предварительных проб и ошибок. Так, в свое время всемирной знаменитостью стала обезьяна по кличке Имо, которая в ходе эксперимента изящно решила задачу по очистке зерен злаков от песка. Она не стала выбирать зерна из песка, как это делали другие обезьяны, а бросила их в воду и потом собрала зерна с поверхности воды. А вот другие обезьяны, которые жили с ней в одном сообществе, переняли эти навыки от Имо на основе другого вида обучения – ***подражания.*** Подражание (имитация) проявляется на самых разных этапах развития животного мира. Однако особо важную роль эта форма обучения играет у приматов. С помощью подражания осуществляется передача навыков от поколения к поколению.

Исследования особенностей обучения позволили установить достаточно неожиданные способности животных, касающиеся процессов обработки информации. Так, например, у голубей выявлена способность дифференцирования фотографических снимков двух участков местностей, снятых в различных ракурсах, что позволяет предполагать наличие у птиц целостного характера зрительного восприятия. Люди моего поколения, наверное, при этих словах вспомнят веселую датскую кинопародию на шпионские фильмы «Бей первым, Фредди», где обыгрывается эта способность голубей.

Наконец, для животных установлена возможность оценки и использования информации о своем собственном состоянии. Подтверждением тому служат многочисленные документированные случаи обращения животных за помощью к человеку (для себя или своих сотоварищей).

Согласно широко распространенному до последнего времени заблуждению, бесспорным проявлением специфичности мозга человека считается структурная и функциональная асимметрия коры больших полушарий. Правое полушарие связывают с пространственно-синтетической деятельностью, левое – с речевой и аналитической. Однако наряду с этим накоплены убедительные доказательства функциональной асимметрии мозга животных. Обращаясь к проблеме специфичности речевой деятельности человека, следует отметить, что сегодня нет оснований для безоговорочного признания существования языка у животных. Видоспецифические акустические сигналы животных являются проявлением внутреннего состояния, а не отражают предметы и события внешнего мира. В то же время исследователями доказана потенциальная способность человекообразных обезьян овладевать знаковыми системами, с помощью которых они могли использовать названия предметов, в

частности их можно научить языку жестов американских глухонемых. Знаковые системы, которым обучают обезьян, состоят из нескольких сот знаков, причем животные могут употреблять знаки с переносом значения, создавать новые знаки и их комбинации. В целом считается, что знаковые системы, которыми в состоянии пользоваться обезьяны, сопоставимы с уровнем развития речи 2–3-летнего ребенка. Отметим, что этого именно уровня может достигнуть (в результате длительного обучения) ребенок типа Маугли – лишенный нормального развития в результате изоляции от людей.

Гены и поведение. Завершить рассказ о психофизиологических исследованиях интеллекта хотелось бы разговором о генетических основах поведения. В 1960-е годы шведский цитолог ***Хольгер Хиден*** разработал метод, позволяющий анализировать количество РНК и даже ее качественный состав в отдельных изолированных нервных клетках. В опытах по обучению крыс Х. Хиден и его сотрудники обнаружили, что количество РНК в тех нейронах крыс, которые предположительно имеют отношение к данной форме обучения, увеличивается, причем качественный состав, т. е. соотношение разных нуклеотидов, изменяется. Был сделан вывод, что в ходе обучения активируются новые, ранее «молчавшие» гены, которые синтезируют новую информационную РНК, являющуюся носителем памяти. Благодаря образованию этих молекул памяти, т. е. экспрессии в соответствующих нейронах неких новых генов, определенные поведенческие навыки запечатлеваются в мозгу животных.

За работами Х. Хидена последовало множество публикаций, в которых описывали опыты по влиянию торможения синтеза РНК и белков в мозгу с помощью антибиотиков актиномицина, пуромицина и циклогексимида на процессы обучения разных видов животных (крысы, мыши, рыбы, птицы), результаты которых, вроде бы, демонстрировали, что такое вмешательство препятствует овладению разными навыками, например, мешает нахождению правильного пути при обучении в лабиринте. Высказывалось даже утверждение, что кормление больных с потерей памяти препаратами РНК способствует частичному ее восстановлению.

Однако впоследствии, когда появилась более точная методика, оказалось, что качественных изменений РНК в нейронах при обучении не происходит, могут иметь место лишь изменения количественного соотношения разных фракций мРНК, уже существовавших в клетке, отражая изменение функционального состояния нейронов. Что же касается эффекта антибиотиков на процессы обучения, то член-корреспондент РАН Л. И. Корочкин еще во время работы в Новосибирске исследовал их влияние на морфологию клеток мозга в месте их введения в самых малых дозах, гораздо меньших, чем использовали в опытах по обучению, и показал, что даже эти малые дозы вызывали патологические изменения в нервых клетках, в том числе их гибель. Следовательно, введение антибиотиков в мозг препятствует обучению потому, что выводит из строя целые функциональные ансамбли нервных клеток в зоне введения препарата.

В целом, по мнению Л. И. Корочкина, наблюдаемые различия в активности генетического аппарата или отдельных генов между хорошо и плохо обучающимися линиями животных являются исходной характеристикой линий животных, сложившейся в процессе дифференцировки мозга в онтогенезе и обусловленной генетически детерминированными межлинейными различиями. Он называет и «те генетически детерминированные особенности мозга, от которых зависит эффективность его функционирования:

- масса ткани, количество составляющих ее клеток, что, в свою очередь, определяется соотношением размножения и гибели клеток в ходе онтогенеза;
- способность нейронов к образованию отростков и синаптических окончаний – контактов между клетками: чем больше отростков и синапсов образуется в ходе онтогенеза и клеточной дифференцировки и чем больше связей установится между клетками, тем лучше;
- способность образовавшихся синапсов к функционированию;
- региональные особенности распределения клеток внутри органа, соотношение нервных и вспомогательных, глиальных клеток;

- адаптивные и регулятивные способности клеток;
- способность определенных клеток погибнуть в определенный момент индивидуального развития, дабы обеспечить более эффективное функционирование оставшихся в живых нейронов».

Таким образом, генетическая детерминация поведенческих признаков осуществляется в процессе индивидуального развития, когда в результате экспрессии определенных генов формируются соответствующие нейральные сети. Различия в способностях к обучению и запоминанию связаны с различиями в эффективности функционирования этих сетей. Для желающих более подробно ознакомиться с проблемами генетической детерминации психофизиологических свойств человека и их эволюцией рекомендуем книгу замечательного советского генетика В. П. Эфроимсона «Генетика этики и эстетики». В Интернете имеются ее электронные версии в различных форматах, в частности в doc-формате на сайте.

РНК-интерференция и психические заболевания. Открытия последних лет, связанные с терапевтическим явления РНК-интерференции, заставляют, в какой-то мере, вспомнить о «молекулах памяти», рассмотренных выше. В частности, в Институте цитологии и генетики СО РАН в Новосибирске установлена возможность сиквенс-специфического подавления экспрессии гена-мишени короткой синтетической интерферирующей РНК в головном мозге млекопитающих ***in vivo***. Согласно результатам этих исследований, стресс, медикаментозное вмешательство или даже недостаточная материнская забота способны изменять экспрессию нейрогена в раннем онтогенезе, предрасполагая к развитию психопатологии. Методы siRNA открывают новые возможности изучения функции генов в центральной нервной системе и применения РНК-интерференции для воздействия на патологические процессы в головном мозге, а, следовательно, для нейробиологического моделирования умственных функций.

Литература: Дромашко, С. Е. Очерки биоинформатики / С. Е. Дромашко ; НАН Беларуси, Ин-т генетики и цитологии ; науч. ред. Л. В. Хотылева. – Минск : Беларус. навука, 2009. – 400 с.

Занятие 7.1 Сообщества организмов. Экологические системы

Вопросы для изучения:

1. Биоинформатика и оценка биологического разнообразия.
2. Экосистемы и теория информации. Сложность экосистем.
3. Теоретико-информационная мера оценки неопределённости.
4. Математический аппарат теоретико-информационного подхода. Простейшая модель Вольтерра-Лотки.

Учебно-методические материалы.

Биоинформатика и оценка биологического разнообразия. Биологическое разнообразие любой экосистемы может быть определено как полная генетическая совокупность определенной среды, состоящая из всех обитающих в ней видов, включая в целом биосферу нашей планеты. Здесь также нашла место для своего приложения биоинформатика. Основные направления исследования – изучение организации и функционирования биологических систем разного уровня методами и средствами информатики:

- разработка методологических подходов для изучения эволюции живой природы на основе количественных методов оценки генетической структуры популяции и филогенетических связей таксонов видового и надвидового уровней;
- исследование устойчивости экологических систем;

• компьютерное моделирование процессов получения, накопления, обработки и систематизации данных по биоразнообразию.

Так, для сбора видовых имен, описаний, ареала распространения, генетической информации используются разнообразные базы данных. Специализированное, объектно-ориентированное программное обеспечение применяется для поиска, визуализации и анализа информации и, что еще важнее, предоставления ее другим исследователям – пользователям. Компьютерные симуляторы моделируют популяционную динамику или вычисляют общее генетическое здоровье культуры в агрономии. Одним из важнейших направлений в этой области является анализ последовательностей ДНК или полных геномов целых вымирающих видов, что позволяет запомнить результаты генетического эксперимента природы в компьютере и дает возможность использовать их вновь в будущем, даже если эти виды полностью вымрут. На рис. 45 представлена принципиальная схема информационных потоков при модельных исследованиях структуры биосистем.

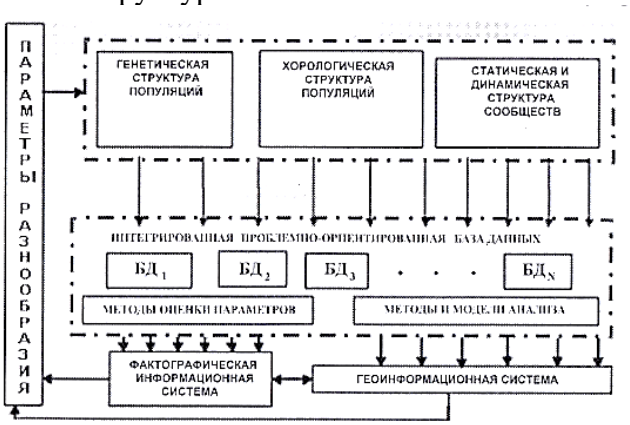


Рисунок 45 – Функциональная схема информационных потоков при модельных исследованиях структуры биосистем

Эволюция экосистем идет по пути эволюции сообществ – условно обратимого преобразования старых и образования новых сообществ в результате возникновения видов, формирования ими новых экологических ниш и отбора сочетаний видов. Как правило, при этом увеличивается экологическое разнообразие, т. е. соотношение между числом видов и численностью особей одного вида в экосистеме. Одна из отраслей биоинформатики – вычислительная эволюционная биология исследует происхождение и появление видов и их развитие с течением времени, используя для этого средства информатики. Применение информационных технологий позволяет изучать эволюцию большого числа организмов, измеряя изменения не только в их строении или физиологии, но и в ДНК, сравнивать целые геномы. Это позволяет количественно изучать более комплексные эволюционные события, такие, как дупликация генов или их латеральный перенос. На этой базе созданы компьютерные модели популяций – основа для предсказания поведения всей системы во времени. Наконец, биоинформатика дает средства для информационного обеспечения исследований, позволяя отслеживать появление публикаций, содержащих информацию о большом количестве видов.

Экосистемы и теория информации. Сложность экосистем. Разные экосистемы, в частности и биосфера в целом относятся к очень большим, сложно организованным комплексам. Одним из подходов к их описанию являются кибернетические методы, в частности теория информации. Любая сложная система состоит из частей, элементов, между которыми существуют более или менее сильные связи. Тогда наличие данных о состоянии одной части или элемента системы подразумевает наличие информации о состоянии другого, связанного с ним элемента системы. Эта информация тем больше, чем сильнее связь. С технической точки зрения любая такая взаимосвязь может быть представлена в виде канала, по которому от одного элемента к другому передается информация. Задача описания такой системы сводится, в

определенной степени, к оценке эффективности передачи информации между элементами системы.

Чем больше элементов в изучаемой системе, тем сложнее бывает проведение информационного анализа. Так, в двухкомпонентной системе необходимо учитывать связь между ее частями. Появление третьего элемента (или фактора) осложняет ситуацию так называемой связью между связями. В системе из четырех элементов приходится учитывать связь между одной из частей и комплексом связей и т. д. Кроме того, в системах, состоящих из большого числа элементов, возможен эффект объединения, когда некоторые компоненты системы сильно взаимодействуют между собой, выступая как единое целое по отношению к остальной части системы. Вряд ли можно рассчитать все информационные потоки в таких сложных системах вручную, без применения вычислительной техники. Одна из возможных схем, примененная к эколого-генетическим процессам, приводится ниже.

Теоретико-информационная мера оценки неопределенности. Теория информации, заложенная работами К. Шеннона и Н. Винера как сугубо техническая дисциплина, получила широкое распространение в биологии, психологии, лингвистике и других науках естественного и гуманитарного профиля. Теория информации представляет собой одну из возможных моделей измерения неопределенности. Основные особенности этого подхода изложены в классической работе Г. Кастлера «Азбука теории информации»:

1. Количество информации есть измеримая абстрактная величина, причем ее значение не зависит от объекта, точно так же, как длина, вес или температура имеют физический смысл независимо от природы нагретого объекта или объекта, имеющего длину или вес.
2. Информация связана с ансамблем возможных исходов некоторого события; ее величина зависит от вероятностей этих исходов, но не от их причин и не от их следствий».

Как теория информации, так и математическая статистика имеют дело с вероятностями тех или иных событий и занимаются разработкой методов извлечения максимально полной информации из ограниченного запаса наблюдений или экспериментов. Даже строгое определение информации было впервые введено знаменитым статистиком Р. Фишером в его работе, посвященной теории оценок. Это заставляет многих математиков считать теорию информации всего лишь одной из ветвей математической статистики, связанной с одной из возможных мер информации, а именно логарифмической мерой. Это приводит к замене обычных критериев проверки статистических гипотез на информационные (логарифмические).

Общность корней и имеющееся сходство приводят к тому, что в обеих этих дисциплинах часто используются одни и те же методические приемы, в частности многомерные таблицы сопряженности признаков. Однако цели, ради которых они применяются, в статистике и теории информации различны.

В математической статистике таблицы сопряженности признаков применяются в корреляционном и дисперсионном анализе для оценки достоверности различий действия разных факторов по отдельности и в совокупности. Та же самая цель оценки различий остается и при использовании в статистике информационной меры различий факторов. Иначе обстоит дело в теоретико-информационном подходе. Здесь таблицы сопряженности признаков являются инструментом для построения так называемых каналов связи и вычисления на этой основе коэффициентов передачи информации от одних факторов (совокупностей факторов) к другим.

Математический аппарат теоретико-информационного подхода. Пусть явление ***Y*** имеет m состояний $\boldsymbol{y_1}\ y_2,\ldots,\ y_m$, а каждый из ***N*** факторов X_i имеет $n(i)$ различных состояний $x_1\ x_2,\ldots,\ x_{n(j)}$. Когда состояния фактора и явления x_{ij}. и $\boldsymbol{y_к}$ независимы, вероятности их одновременного наблюдения связаны между собой хорошо известной формулой

$$p(y_к, x_{ij}) = p(y_к)p(x_{ij}). \qquad (17)$$

При этом для условной вероятности наблюдения состояния $\boldsymbol{y_к}$ при условии, что состояние x_{ij}. уже реализовалось, справедливо равенство

$\boldsymbol{P}(y_k/x_{ij}) = p(y_к, x_{ij})/p(x_{ij}) = p(y_к)p(x_{ij})/p(x_{ij}) = p(y_k).$ (18)

Любое взаимодействие между x_{ij} и y_k нарушает равенство (8.8). В частности, когда $p(y_к/x_{ij}) > p(y_к)$, говорят, что информация передается от x_{ij} к $y_к$. Задача исследователя на первом этапе сводится к отысканию таких взаимодействующих пар $y_к$, x_{ij} и к оценке степени этого взаимодействия.

Для этого в теории информации вводятся энтропийные функции

$$H(Y) = -\sum_k p(y_k)\log_2\left(p(y_k)\right), \quad (19)$$

$$H(X) = -\sum_j p(x_{ij})\log_2\left(p(x_{ij})\right), \quad (20)$$

$$H(Y/x_{ij}) = -\sum_k p(y_k/x_{ij})\log_2\left(p(y_k/x_{ij})\right), \quad (21)$$

$$I(Y, x_{ij}) = H(Y) - H(Y/x_{ij}), \quad (22)$$

$$T(Y, X_i) = \sum_j p(x_{ij})I(Y/x_{ij}). \quad (23)$$

Здесь $\boldsymbol{H(Y)}$ – максимальная энтропия явления Y, $\boldsymbol{H(X_i)}$ – максимальная энтропия действующего фактора i, $\boldsymbol{H(Y/x_{ij})}$ – условная энтропия явления Y для некоторого фиксированного состояния x_{ij} Физический смысл $\boldsymbol{I(Y, x_{ij})}$ – это условная информация, которую можно получить о любом состоянии явления Упри некотором фиксированном состоянии фактора x_{ij}. Средняя информация, содержащаяся в такой системе, дается выражением $\boldsymbol{T(Y, X_i)}$.

Из сопоставления формул (19 – 23) видно, что

$$T(Y, \boldsymbol{X}) = H(Y) + H(X_i) - H(Y, X_i) = \boldsymbol{T(X_i,}Y), \quad (24)$$

где $H(Y, X_i)$ определяются аналогично $H(Y)$ и $H(X_i)$. Формула (24) говорит, что с теоретико-информационной точки зрения совершенно безразлично, передается информация от $\boldsymbol{X_i}$ к Y или наоборот. Это может быть важно при решении обратной задачи – предсказании значения действующего фактора по известному состоянию явления. Эффективность передачи информации от X_i к Y и от Y к X_i. при этом определяется формулами

$$K(Y; X_i) = T(Y; X_i)/H(X_i), \quad K(X_i; Y) = T(X_i, Y)/H(Y). \quad (25)$$

Таким путем можно рассчитать все $\boldsymbol{2N}$ коэффициентов $\boldsymbol{K\{Y\backslash X)}$ (прямые информационные потоки) и $K(X_i; Y)$ (обратные информационные потоки) и ранжировать все действующие факторы по величине их влияния на результирующее явление и наоборот. Затем можно редуцировать количество действующих факторов, отбросив те из них, которые дают наименьшую информацию о системе. Дальнейший многофакторный анализ, в котором учитывается взаимодействие двух, трех, четырех и т. д. действующих факторов, ведется уже только по наиболее информативным из них. Формулы для оценки степени взаимодействия факторов аналогичны (8.9) – (8.13), где в качестве аргумента стоят сочетания $\boldsymbol{X_iX_j}$ $\boldsymbol{X_i,X_J,X_k}$ и т. д. Здесь, правда, встает проблема удобной для биологов интерпретации и верификации полученных оценок, что нуждается в специальном обсуждении.

Для анализа эколого-генетических взаимодействий в 2002 г. была разработана первая версия программного обеспечения, совместимая с форматом Excel, в настоящее время существенно модифицированная. В отличие от ранее созданных систем новая программа предоставляет пользователю все возможности, предусмотренные Excel. В программный блок входят ввод данных, скрининг факторов на наиболее информативные и собственно многофакторный анализ данных с использованием средств Excel.

Так, в частности, программа осуществляет расчет информативности тех или иных анализируемых факторов, комбинации которых, как и отнесение их к действующим или результирующим, можно задавать вручную, в зависимости от желания исследователя. Разбиение анализируемых данных на классы осуществляется в автоматическом режиме. В табличном виде приводятся расчеты всех основных энтропийных функций, необходимых для анализа. Окончательные результаты расчетов информативности исследуемых факторов и их

комбинаций (прямые и обратные коэффициенты передачи) автоматически выводятся на монитор как в табличном, так и в графическом виде. Возможно также их сохранение в соответствующих книгах Excel с последующей распечаткой на принтере. Кроме того, предусмотрен «послойный» вывод отдельных состояний факторов и визуализация их неслучайных комбинаций как элемент этапа прогнозирования.

Простейшая модель Вольтерра–Лотки. В экологии можно выделить три основные части.

• Эмпирическая часть содержит фактические сведения, полученные в экспериментах и наблюдениях, а также из первичной систематизации.

• Теоретическая часть развивает основные концепции, позволяющие объединить и объяснить с единых позиций эмпирические закономерности и явления.

• Математическая часть конструирует математические модели, служащие для проверки основных теоретических концепций, дает методы обработки экспериментальных данных и планирования экспериментов и наблюдений.

Математическое моделирование стимулирует накопление фактического материала, уточняет направление проводимых экспериментов.

Одним из классических объектов приложения математической экологии является система хищник–жертва. Рассмотрим простейшую математическую модель «хищник-жертва», описывающую совместное существование двух биологических видов (популяций) – хищников и их жертв, называемую ***моделью Вольтерра–Лотки.*** Впервые она была получена А. Лоткой (1925), американцем польского происхождения, который использовал ее для описания динамики взаимодействующих биологических популяций. Чуть позже и независимо от А. Лотки аналогичные (и более сложные) модели были разработаны итальянским математиком В. Вольтерра (1926), глубокие исследования которого в области экологических проблем заложили фундамент математической теории биологических сообществ или так называемой математической экологии.

Модель, которую мы рассмотрим, интересна как раз тем, что с нее фактически и началась математическая экология. Пусть есть два биологических вида, которые совместно обитают в изолированной среде. Среда стационарна и обеспечивает в неограниченном количестве всем необходимым для жизни один из видов, который будем называть ***жертвой.*** Другой вид – ***хищник*** также находится в стационарных условиях, но питается лишь особями первого вида. Это могут быть караси и щуки, акулы и тунцы, зайцы и волки, мыши и лисы, микробы и антитела и т. д. Пусть это будут, следуя: известной поговорке, караси и щуки, живущие в некоем изолированном пруду. Среда предоставляет карасям питание в неограниченном количестве, а щуки питаются лишь карасями.

Обозначим через y число щук, через x – число карасей. Со временем число карасей и щук меняется, но так как рыбы в пруду много, будем считать x и y непрерывными функциями времени t. Назовем пару чисел (x, y) состоянием модели. Как это состояние меняется с течением времени?

Рассмотрим dx/dt – скорость изменения численности карасей. Если щук нет, то число карасей увеличивается и тем быстрее, чем больше карасей. Для простоты будем считать эту зависимость линейной: $dx/dt \sim \varepsilon x$, причем коэффициент ε_1 зависит только от условий жизни карасей, их естественной смертности и рождаемости. Скорость изменения dy/dt числа щук (если нет карасей) зависит от числа щук y. Будем считать, что $dy/dt - \varepsilon_2 y$. Если карасей нет, то число щук уменьшается (у них нет пищи) и они вымирают. В экосистеме скорость изменения численности каждого вида также будем считать пропорциональной его численности, но только с коэффициентом y, который зависит от численности особей другого вида. Так, для карасей этот коэффициент y_j уменьшается с увеличением числа щук, а для щук y, увеличивается с увеличением числа карасей. Будем считать эту зависимость также линейной.

Тогда получим систему из двух дифференциальных уравнений:

$$\frac{dx}{dt} = \varepsilon_1 x - \gamma_1 yx,$$
$$\frac{dy}{dt} = -\varepsilon_2 y + \gamma_2 xy. \qquad (26)$$

Эта система уравнений и называется моделью Вольтерра–Лотки. Числовые коэффициенты ε_1,Yi, ε_2>Y2 называются ***параметрами модели.*** Очевидно, что характер изменения состояния (x, y) определяется значениями параметров. Изменяя их и решая систему уравнений модели, можно исследовать закономерности изменения состояния экологической системы.

В качестве примера на рис. 46 построены кривые изменения численности карасей x и щук y в зависимости от времени t для некоторых типичных значений параметров. Как видно, максимумы кривых чередуются, причем максимумы щук отстают от максимума карасей. Это отставание разное для разных экосистем типа «хищник-жертва», но, как правило, много меньше периода колебаний.

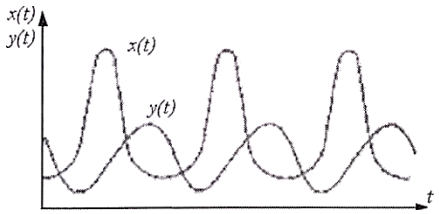


Рисунок 46– Колебания численности популяций в простейшей модели системы «хищник – жертва»

Литература: Дромашко, С. Е. Очерки биоинформатики / С. Е. Дромашко ; НАН Беларуси, Ин-т генетики и цитологии ; науч. ред. Л. В. Хотылева. – Минск : Беларус. навука, 2009. – 400 с.

Занятие 7.2 Эволюция живой природы как процесс передачи, накопления, хранения информации

Вопросы для изучения:

1. Микроэволюция на протоке (математическая модель).
2. Эволюция и ЭВМ. ЭВМ и макроэволюция.
3. Энергетические критерии эволюционного процесса.
4. Теоретико-информационные аспекты эволюции.

Учебно-методические материалы.

Эволюция живой природы как процесс передачи, накопления, хранения информации

Понятие об эволюции. Биологическая эволюция (biologic evolution, evolution; biological evolution – ***от лат.*** evolutio ***– развитие) – необратимое и в известной мере направленное историческое развитие живой природы, сопровождающееся изменением генетического состава популяций, формированием адаптации, образованием и вымиранием видов, преобразованием экосистем и биосферы в целом.***

Биологическая эволюция определяется изменчивостью, наследственностью, естественным отбором организмов, происходящими на фоне перемен в составе экосистем.

Различают макро- и микроэволюцию. ***Микроэволюция –*** эволюционные преобразования внутри вида на уровне популяций и демов, ведущие к внутривидовой дивергенции и видообразованию.

Надвидовая эволюция, или ***макроэволюция,*** – это эволюционный процесс образования из видов, возникших в результате микроэволюции, новых родов, из родов – новых семейств и т. д.

Классическим примером микроэволюционных изменений, не приводящих к образованию новых надвидовых структур, являются микроорганизмы. В самом деле, в благоприятных условиях всего за 1 год у любого микроорганизма сменяется столько поколений, что это соответствует сотням тысяч и даже миллионам лет эволюции для высших организмов. Основными факторами, лимитирующими рост всякой микробной популяции, являются нехватка питательных веществ и отравление продуктами собственной жизнедеятельности. Если постоянно обновлять в культиваторе (колбе или даже пробирке с отводами) среду, подводя свежий раствор питательных веществ и удаляя отработанную культуральную жидкость вместе с избытком биомассы, можно сделать популяцию практически «бессмертной», способной размножаться в течение любого необходимого исследователю времени.

Идея и разработка метода непрерывного культивирования микроорганизмов принадлежат французскому микробиологу Ж. Моно и американским ученым А. Новику и Л. Сциларду. В бывшем Советском Союзе пионером этого наиболее «технологичного» метода биологических исследований явился Н. Д. Иерусалимский, организатор Института биохимии и физиологии микроорганизмов АН СССР в Пущино. Математические аспекты автоселекции при непрерывном культивировании микробов интенсивно разрабатываются красноярскими учеными И. А. Терсковым и Н. С. Печуркиным в Институте биофизики СО РАН (Красноярск).

Управление ростом микробной популяции может осуществляться при ***непрерывном,*** или ***проточном, культивировании*** различными способами. В режиме ***хемостата*** управляющим параметром служит концентрация питательных веществ, лимитирующая неограниченное размножение микроорганизмов. В режиме без внешнего лимитирования работает, например, ***турбидостат,*** где постоянной задается плотность биомассы, или ***pH-стат,*** где постоянна концентрация водородных ионов.

Простейшая математическая модель, описывающая рост микроорганизмов на протоке, имеет вид:

$$dx/dt = \mu x - Dx,$$
$$dS/dt = -\alpha\mu x + D(S_0 - S), \qquad (27)$$

где x – плотность биомассы; $\boldsymbol{S}$ – концентрация питательного субстрата (пищи) внутри культиватора; $\boldsymbol{S_0}$ – концентрация «свежего» субстрата на входе; μ – удельная скорость размножения микробов; $\boldsymbol{D}$ – скорость протока культурной среды; a – экономический коэффициент, характеризующий степень использования данного субстрата для роста биомассы.

Скорость ц может зависеть от концентрации субстрата или же продуктов жизнедеятельности микроорганизмов. В частности, в первом случае эту зависимость можно описать формулой

$$\mu = \frac{\mu_{max} S}{Ks + S}, \qquad (28)$$

где μ_{max} – максимальная скорость роста; $\boldsymbol{K_s}$ – гак называемая константа Михаэлиса–Ментен. В условиях хемостата стационарный режим ($\boldsymbol{dx/dt} = 0$) устанавливается в силу субстратного лимитирования при скорости роста меньше μ_{max}. Турбидо- статное и pH-статное регулирование позволяет вести исследование стационарной популяции при максимальной скорости размножения μ_{max}. Изучение эволюции микробных популяций при проточном культивировании допускает все эти режимы.

Микроэволюция на протоке. Основным поставщиком материала для эволюции у микроорганизмов являются мутации. Происходят они с довольно низкими частотами 10^{-4}–10^{-}

^{10} на бактерию за одну генерацию. Однако огромное число микробов в популяции, доходящее до 10^{12}–$10^{м}$, приводит к тому, что в ней постоянно может находиться 10^3–10^8 мутантов. При этом может образоваться целый спектр мутантов, однако накопиться в популяции способна, по-видимому, только одна форма, обладающая максимальной удельной скоростью роста. Рассмотрим ее поведение.

В случае внешнего лимитирования (хемостатный режим) микроэволюция микробной популяции на протоке описывается следующей системой уравнений:

$$dx_m / dt = \frac{\mu_{m\max} S}{K_{Sm} + S} x_m - Dx_m + k_i^m x_i - k_m x_m,$$

$$dx_i / dt = \frac{\mu_{i\max} S}{K_{Si} + S} x_i - Dx_i + k_m^i x_m - k_i x_i,$$

$$dS / dt = -\alpha_i \frac{\mu_{i\max} S}{K_{Si} + S} x_i - \alpha_m \frac{\mu_{m\max} S}{K_{Sm} + S} x_m - (S_0 - S)D, \quad (29)$$

где k_i и k_m – скорости мутации *i*-й и *m*-й форм; k_i^m и k_m^i – скорости их мутаций в *m-ю* и *i*-ю формы.

В общем виде система этих нелинейных уравнений аналитически не решается. Однако можно установить, что между скоростью протока Д долей мутанта в популяции f_m и давлением отбора о существует простая зависимость:

$$D = \sigma \frac{x_m}{x_m + x_i} + \mu_i = \sigma f_m + \mu_i . \quad (30)$$

Это означает, что с ростом доли мутанта с $\mu_m > D$ в популяции при той же скорости протока *D* наблюдается усиленное вымывание «медленной» формы $\mu_i = D$. От величины f_mзависит и концентрация лимитирующего рост фактора. В целом отбор в хемостате, как показывает эта модель, идет в пользу штамма, способного с большей скоростью утилизировать лимитирующий субстрат по сравнению с исходной формой.

При описании микроэволюции без внешнего лимитирования достаточно двух первых уравнений системы (8.20). Рост культуры в турбидостате происходит при больших значениях ***S,*** поэтому можно считать, что $\mu_m = \mu_{m\ max}$, $\mu_i = \mu_{i\ max}$. Кроме того, режим турбидостата предполагает, что концентрация биомассы в культиваторе постоянна, т. е. $x_1 + x_m = x_r = \text{const}$. Вторым дополнительным условием турбидостата является равенство скорости протока *D* и удельной скорости μ_r роста популяции как единого целого: $D = \mu_r$. В силу этого соотношения скорость протока полностью определяется кинетикой роста турбидостатной популяции и меняется таким образом, чтобы соблюдалось основное требование турбидостата: постоянная плотность популяции.

В отличие от случая хемостатного культивирования популяция в турбидостате может быть описана аналитически. Однако получающиеся при этом выражения достаточно громоздки, поэтому мы ограничимся только некоторыми следствиями эволюционного характера. Оказывается, что соотношение исходной и мутантной форм целиком зависит от параметра δ, характеризующего селективные преимущества мутантного штамма:

$$\delta = \mu_m - \mu_i - k_m + k_m = \sigma_{\text{отб}} - \sigma_{m\ \text{мут}} + \sigma_{i\ \text{мут}}. \quad (31)$$

При $\delta > 0$ доля мутанта в популяции увеличивается, при $\delta < 0$ – уменьшается. Согласно этой формуле, эффективность отбора в турбидостате зависит как от чистого давления отбора, так и от давления мутаций.

Момент перестройки структуры турбидостатной популяции можно наблюдать по скачку в скорости протока *D*. В самом деле, при малых интенсивностях мутационного процесса

$$D = \mu_m = (\mu_i x_i + \mu_m x_m)/x_m = \mu_I + (\mu_m - \mu_i)x_m / x_{\text{т}}. \quad (32)$$

Отсюда следует, что, когда доля мутанта в популяции мала ($x_m/x_т$«1), скорость протока целиком определяется удельной скоростью роста исходного штамма: $D \sim \mu_i$. При замещении исходной формы мутантом x_m/x_r~1 и скорость протока возрастает до значения $D \sim \mu_m$

Сравнивая процессы микроэволюции при хемостатном и турбидостатном способах культивирования микробов, можно отметить следующее. При микроэволюционном переходе в лимитированных условиях происходит некоторое возрастание численности популяции за счет размножения формы, более полно использующей питательный субстрат. В результате популяция становится более «приспособленной» к условиям среды. Микроэволюция в нелимитированных условиях не приводит к уменьшению концентрации субстрата. В этом случае происходит увеличение максимальной удельной скорости роста популяции без увеличения ее размеров.

Это означает, что в зависимости от режима, в котором происходит проточное культивирование микроорганизмов, можно менять не только темпы микроэволюции, но и ее направление (рис. 47).

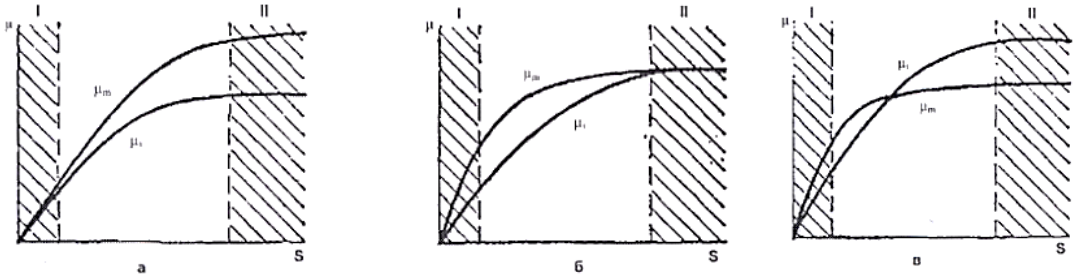


Рисунок 47 – Управление темпами и направлением микроэволюции при проточном культивировании микроорганизмов

На рис. 8.5 римскими цифрами I и II обозначены зоны устойчивой работы соответственно хемостата и турбидостата. Для случая, изображенного на рис. 8.5, а, в области работы хемостата автоселекции нет, поскольку исходная и мутантная формы не различаются по скорости роста. Турбидостатный режим дает заметное преимущество мутантной форме с большей максимальной скоростью размножения. На рис. 8.5, б представлена ситуация, в которой при хемостатном способе селекции преимущество получает мутантная форма, а в турбидостате ни одна из форм не будет различаться по скорости роста. Наконец, в ситуации, показанной на рис. 8.5, в, в хемостате селекция пойдет в пользу мутантного штамма, а в турбидостате – в пользу исходной формы.

Таким образом, изучение непрерывной культуры микроорганизмов показывает, что особенности процессов микроэволюции существенно зависят от условий развития популяции. Это сближает в какой-то мере математическую теорию автоселекции с теми вопросами, которые решают экологическая генетика и математическая экология. В целом же учет данных, доставляемых с помощью математических моделей микроэволюции непрерывной культуры микроорганизмов, необходим как при решении прикладных задач, связанных с получением новых промышленных штаммов – продуцентов полезных веществ, так и при формулировании фундаментальных вопросов теории эволюции.

Эволюция и ЭВМ. Положение с экспериментальным исследованием эволюционных процессов существенно изменилось с возникновением и совершенствованием современной вычислительной техники. Хотя первоначально развитие компьютерного моделирования ограничивалось быстродействием и объемом памяти ЭВМ, уже первый этап компьютерных экспериментов позволил по-новому взглянуть на влияние внешних (эколого-географических) и внутренних (генетических) факторов на процесс микроэволюции.

Еще в середине 1960-х годов им совместно с О. С. Кулагиной в Академгородке Сибирского отделения АН СССР (Новосибирск) был проведен ряд работ по моделированию

эволюции в гипотетической популяции гаплоидных организмов. Ограниченный объем памяти ЭВМ позволял следить одновременно за популяцией из 120–150 «особей» на протяжении 40–45 поколений. В основе модели лежала теория марковских процессов, что позволяло учесть положение об отсутствии у эволюционирующей биологической системы «памяти» о предшествующих состояниях.

Модель предполагала, что генотип каждой «особи» состоял из набора n генов, каждый из которых мог находиться в двух аллельных состояниях. Панмиксия, т. е. случайное скрещивание, в модели отсутствовало: в программу закладывалось, что при существовании у двух особей различий по достаточному числу этих генов *p* оставление потомства такой парой невоможно. При этом исходное строение популяции было таково, что допускало скрещивание между любыми парами особей.

Как показал компьютерный эксперимент, через несколько десятков поколений генотипическое строение популяции существенно менялось. Если *p* было достаточно малым, популяция неизбежно распадалась на несколько изолированных друг от друга генотипических групп. Оказалось также, что результаты всех экспериментов можно разбить на три класса.

1. Первоначально генотипически разнообразная популяция стягивается к некоему сгустку, в котором все особи генотипически достаточно близки.
2. Исходная популяция распадается на несколько биологически изолированных групп, скрещивание между которыми невозможно.
3. Неоднородность по генотипам в популяции возрастает, но не завершается распадом на изолированные группы.

Популяции двух первых типов соответствовали тупикам эволюции. В этих случаях потенциальная изменчивость популяции была крайне низка – популяция слишком приспосабливалась к существованию в данных внешних условиях и при любом изменении их была обречена на вымирание. В случае популяции третьего типа дивергенция видов, или расхождение формы, еще окончательно не завершилась. Такая «устойчивая неустойчивость» благоприятна для действия эволюционного процесса. По словам А. А. Ляпунова, здесь наблюдался процесс видообразования в стадии рождения.

Введение в математическую модель отбора привело к сильному дрейфу генотипов популяции. При этом основная масса популяции эволюционировала в сторону, диктуемую отбором. Однако машинный эксперимент показал, что в силу чисто статистических обстоятельств от популяции отщеплялись небольшие группы особей, сравнительно однородные генотипически и репродуктивно изолированные от остальной части. Эти реликтовые формы, как назвал их Ляпунов, эволюционировали гораздо медленнее и через 40 поколений заметно отставали от основной популяции в смысле генетического дрейфа.

В целом эти ранние эксперименты по «машинной» эволюции, в которых специально роль среды была сведена до минимума, показали решающую роль наследственности не только для сохранения уже существующих форм, но и для создания новых видов. Кроме того, оказалось, что помимо обычного дарвинского комплекса – наследственность, изменчивость, отбор – для устойчивого хода эволюции необходима еще и некоторая стохастическая неустойчивость. Чем больше число генов у особи и число особей в популяции, тем более вероятно возникновение такой неустойчивости, тем более благоприятны условия для дивергенции и видообразования.

ЭВМ и макроэволюция. Работы А. А. Ляпунова и О. С. Кулагиной вскрыли некоторые принципиальные механизмы микроэволюции. В то же время для оценки того, насколько соответствуют наши представления действительному положению дел, необходимы машинные эксперименты по моделированию эволюции реально существующих видов. Уже в 1970-х годах новое поколение быстродействующих ЭВМ сделало такую задачу возможной. Наиболее интересны в этом плане работы В. В. Меншуткина и Б. И. Медникова. Начало их работ по имитации процессов микро- и макроэволюции было положено компьютерным экспериментом

по эволюции популяции веслоногих рачков-циклопов под действием изменившихся условий среды. Построенная модель популяции впервые в практике «машинной» эволюции учитывала как генетические, так и экологические параметры. В частности, учитывались такие параметры, как температура, концентрация животного и растительного корма, избирательность и интенсивность воздействия хищников. Компьютерное моделирование показало, что уже через 200–700 поколений стали «рождаться» особи, приспособленные к новым условиям гораздо лучше, чем исходная популяция. Эта новая популяция по многим признакам значительно отличалась от материнской популяции и, главное, не могла скрещиваться с ней, т. е. налицо была репродуктивная изоляция – одно из условий видообразования.

Первая удача вдохновила исследователей на моделирование макроэволюции, и они решили проследить изменение признаков, по которым различаются крупные таксономические группы, такие, как род, семейство и т. д. Первая из разработанных для ЭВМ М4030 программ называлась «Из червя – в краба» и имитировала эволюцию членистоногих. В ее основу авторы положили три принципа: 1) равная вероятность прогрессивных и регрессивных измерений видов, 2) постепенность эволюции, т. е. отсутствие резких, скачкообразных изменений, 3) ограниченность машинной «биосферы», способной «прокормить» только около 100 видов. Поэтому виды- неудачники, наименее приспособленные к данной среде, вытирались из памяти ЭВМ («вымирали»), их место занимали («размножались») виды с наибольшей приспособленностью. Оценка приспособленности давалась интегральная – по всей совокупности из 24 жизненно необходимых признаков. Если, например, возникали конечности – ноги, пользоваться которыми мешал чрезмерно тяжелый панцирь, «цена» такого признака равнялась нулю.

Сначала «эволюция» шла чрезвычайно медленно: ни у одного из вновь возникавших видов не закрепился жесткий наружный скелет, составляющий характерную особенность членистоногих. И только когда в программу были введены ограничения на пищевые ресурсы и связанная с этим возможность хищного питания, темпы эволюции значительно ускорились. Довольно быстро возник жесткий скелет, типы которого вполне соответствовали тому, что наблюдается в реальной природе. Дважды осуществился выход зверей на сушу – сначала на трех парах ног (как у насекомых), затем – на четырех (как у паукообразных). Таким образом, за 1,5 часа машинного счета была имитирована эволюция членистоногих.

Самым интересным в этих опытах оказалось быстрое и одновременное возникновение жесткого скелета. До сих пор для палеонтологов было загадкой независимое появление примерно 600 миллионов лет назад скелета у самых разнообразных организмов: кораллов, губок, членистоногих. Машинный эксперимент позволил предположить, что в эту геологическую эпоху изменение условий существования живых организмов привело к открытию хищного способа питания. Появление хищников автоматически подняло эволюционную ценность прочного, предохраняющего от нападения скелета.

Не менее удачна оказалась вторая программа, имитировавшая эволюцию хордовых, к которым принадлежит и человек. По словам Б. М. Медникова, это было своего рода интеллектуальное хулиганство, и может быть поэтому результаты моделирования были изложены только в научно-популярном журнале «Знание – сила». Начав с первого предкового организма, так называемого праланцетника, В. В. Меншуткин и Б. И. Медников получили многообразный набор живых существ, известных по данным палеонтологии, – от ископаемых рыб и ящеров до птиц и млекопитающих (в том числе походивших на реально существовавших австралопитеков).

Энергетические критерии эволюционного процесса. В 1980-е годы появилась интересная концепция, позволяющая объяснить направленность эволюции без привлечения набившего оскомину «стремления к прогрессу». На основе математического анализа развития микроорганизмов и предбиологических структур (так называемых гиперциклов Эйгеиа) красноярский биофизик Н. С. Печуркин пришел к выводу, что падающий на Землю поток

свободной энергии и является тем организующим началом, которое формирует основной механизм эволюции. Биологическая эволюция определяется Н. С. Печуркиным как процесс вынужденного развития открытых автокаталитических систем, все более удаляющихся от исходного состояния равновесия под действием внешнего потока свободной энергии.

Такой энергетический взгляд на механизм эволюции позволяет заключить, что магистральное направление эволюции действительно существует. Суть его состоит в увеличении способности захватывать и использовать потоки свободной энергии. Сформулированный Н. С. Печуркиным энергетический принцип интенсивного развития гласит: «Любая живая система надорганизменного уровня развивается (эволюционирует) таким образом, что поток использованной энергии на единицу биологической структуры (за время существования этой структуры) возрастает». Количественную оценку этому принципу дает коэффициент интенсивного развития $К_м$, в простейшем случае деления клеток имеющий вид:

$$K_{\text{и.р}} = \frac{H_{\text{исп}} g}{\ln 2B}, \qquad (33)$$

где $H_{исп}$ – использованный поток энергии; g – длительность поколения; B – биомасса.

С точки зрения энергетического принципа живые организмы в ходе эволюции преобразуют все большее количество энергии на единицу своей структуры (веса, объема и т. д.). Когда иссякают источники энергии на каком-либо уровне организации живых систем, кризис преодолевается освоением нового источника. Так, возможно, возникли фотосинтезирующие растения, видимо, так был «открыт» хищный способ питания, о котором шла речь выше. Вершина эволюции – человек может не только наиболее эффективно использовать основной поток энергии, но и организовывать во все более возрастающих масштабах дополнительные потоки энергии.

Теоретико-информационные аспекты эволюции. Если концепция Печуркина лишь отталкивается от некоторых кибернетических принципов, лежащих в основе управляемого культивирования микроорганизмов, то в работах российского радиобиолога и генетика В. И. Корогодина (Дубна) теоретико-информационный подход является фундаментом для объяснения феномена жизни. Согласно взглядам В. И. Корогодина, информация, не будучи ни веществом, ни энергией, является необходимым компонентом информационных систем, определяющим их способность совершать целенаправленные действия.

Пусть S – «исходная ситуация», или пространство режимов, в котором могут происходить те или иные события, а Z – одно из таких событий («цель»), спонтанно осуществляющееся с некоторой вероятностью p >0. Отметим, что одновременно с Z могут происходить другие, сопутствующие ему события W. Тогда спонтанное осуществление Z можно описать с помощью преобразования

$$S \xrightarrow[p]{} (Z, W) \qquad (34)$$

«Целенаправленное действие», т. е. действие, направленное на реализацию цели Z, предполагает такое вмешательство в исходную ситуацию S, которое увеличивает вероятность осуществления Z с p до P. Соответствующее этой словесной формулировке преобразование имеет вид:

$$(R, S) \xrightarrow[p,P]{Q_j} (Z, W), \qquad (35)$$

где R – ресурсы, содержащиеся в S и идущие на осуществление цели Z; Q – «механизм», или оператор, применение которого в исходной ситуации S ведет к желаемому результату; j – информация, на основе которой строится оператор Q; W – побочные продукты достижения цели.

Из предыдущего рассмотрения ясно, что оператор не может возникнуть случайно – он должен быть построен в соответствии с заранее имеющейся программой. Эта «совокупность приемов, правил или сведений, необходимых для построения оператора», и называется информацией. Запись Q в формуле (35) тогда означает, что данный оператор $\boldsymbol{Q}$ построен в соответствии с данной информацией j.

Анализируя разные свойства информации и способы ее реализации, В. И. Корогодин показал, что все информационные системы можно разделить на два типа. Информационные системы 1-го рода характеризуются тем, что в них информация, считывающе-реализующие устройства и операторы связаны между собой столь тесно, что представляют единое целое. К ним относятся все живые организмы – от одноклеточных до человека. В информационных системах 2-го рода все эти три элемента могут существовать в пространственном разобщении, хотя функционировать и развиваться способны только совместно. Крайним примером таких систем является человеческое сообщество, где книги или другие системы записи несут в себе информацию, роль считывающих устройств играют нередко сами люди, а супероператором выступает современная технология.

Все информационные системы имеют два общих свойства – автокатализ и гетерокатализ. При этом автокатализ определяет воспроизводство самих информационных систем и кодирующей их информации, а гетерокатализ – те изменения, которые эти системы создают в окружающей их среде. Эволюция информационных систем происходит в результате изменчивости содержащейся в них информации, т. е. изменения ее количества или семантики, которые не лишают ее смысла. Развитие жизни на Земле является результатом этого самопреобразования информации (генетической, поведенческой, логической). При этом прогресс часто оказывается побочным продуктом решения других, более утилитарных задач. Так, например, усложнение структуры геномов живых систем (одноцепочечная ДНК или РНК $\rightarrow$ двунитевая ДНК $\rightarrow$ гаплоидный набор хромосом $\rightarrow$ диплоидный набор хромосом) первоначально, видимо, имело своей целью повышение защиты генетической информации от летальных мутаций. Все эволюционные преимущества такого решения задачи проявились только «в процессе жизни». В методологии такой феномен, когда метод решения задачи оказывается важнее для развития науки, чем сама задача, называется ***поризмом***.

К числу других эволюционно значимых проявлений принципа поризма, сказавшихся на развитии информационных систем, Корогодин относит также возникновение оогамии, т. е. способа полового размножения, когда женские особи формируют богатые цитоплазмой, крупные и неспособные самостоятельно перемещаться яйцеклетки, а мужские особи – мелкие, почти лишенные цитоплазмы и очень подвижные сперматозоиды. Хорошо известно, что одноклеточным эукариотам присущи самые разнообразные способы полового размножения – изогамия, различные варианты гетерогамии, истинная оогамия. Такое разнообразие форм полового размножения, видимо, связано с особенностями экологических ниш, заселенных этими организмами. Все эти способы хорошо решают задачу воспроизводства и размножения генетической информации, поэтому они сохранились до сих пор. Но только один из них (оогамия) в качестве «побочного продукта» содержал в себе возможность возникновения многоклеточных организмов. Для проявления ее достаточно было только возникнуть мутации, препятствующей расхождению продуктов первых дроблений оплодотворенной женской особи (яйцеклетки).

Таким образом, жизнь определяется как «форма существования информации и кодируемых ею операторов, обеспечивающих возможность воспроизведения этой информации в подходящих для этого условиях внешней среды». Цель жизнедеятельности всех без исключения живых организмов – это воспроизведение и видоизменение (совершенствование) кодирующей их информации. С ходом эволюции информационные системы совершенствуются, в дополнение к информации генетической возникают новые ее виды – поведенческая в популяциях животных и, с возникновением человека и общества, – логическая

информация, носителем которой являются речь, язык, слово. Феномен жизни и ее эволюцию тогда можно рассматривать как последовательный процесс возникновения и развития информации, который постепенно, по мере исчерпания емкости своих физических носителей, приобретает новые формы: от генетической информации к поведенческой, а от нее – к логической. Последняя форма информации на наших глазах начинает приобретать глобальный характер, объединяя в единую информационную систему все человеческое общество и всю биосферу.

Литература: Дромашко, С. Е. Очерки биоинформатики / С. Е. Дромашко ; НАН Беларуси, Ин-т генетики и цитологии ; науч. ред. Л. В. Хотылева. – Минск : Беларус. навука, 2009. – 400 с.

Занятие 7.3 Некоторые математические и кибернетические подходы в популяционной генетике человека

Вопросы для изучения:

1. Кибернетические особенности наследственной изменчивости. Исследование м-ДНК и популяционная геномика.
2. Генная реконструкция филогенеза популяций человека математическими методами.

Учебно-методические материалы.

Некоторые математические и кибернетические подходы в популяционной генетике человека

Кибернетические особенности наследственной изменчивости у эукариот. В середине 1960-х годов В. А. Геодакян, сотрудник Института биологии развития АН СССР, сформулировал принцип, что любая эволюционирующая система, которая находится во взаимодействии со средой, неизбежно вычленяет из себя подсистемы с постоянной и оперативной памятью. В плане наследственной изменчивости к ним относятся пары ДНК – белок, ядро – цитоплазма, половые клетки – соматические клетки, женский пол – мужской пол. В такой системе одна часть выполняет эволюционную задачу сохранения, а вторая – изменения. Обеспечивая информационный контакт со средой, элементы оперативной памяти должны обладать большей дисперсией признаков по сравнению с элементами постоянной памяти. Деление генотипа эукариот на обязательный и факультативный компоненты (ОК – гены и семейства генов и ФК – повторяющиеся последовательности ДНК, псевдогены, эндогенные вирусы и др.) соответствует данному принципу, обеспечивая разные по характеру типы наследственной изменчивости (табл. 6).

Таблица 6 – Три типа наследственной изменчивости у эукариот

Критерий оценки изменчивости	Мутационная	Вариационная	Эпигенетическая
Организация памяти: кодирование хранение передача	Нуклеотидный код, число и топография генетических единиц Структура ДНК Конвариантная редупликация	Коды сегрегации, комбинация ФК и ОК Структура хромосом Сегрегация	Онтогенетические цепочки генов Циклические и триггерные системы генов Распределение регуляторных молекул между дочерними клетками
Основные факторы, приводящие к появлению изменения	Ошибки систем редупликации, репарации, гомологичной рекомбинации	Рекомбинация, эписомы и другие эпигеномные элементы (мобильные элементы, эндогенные вирусы)	Характер соотношений регуляторных молекул
Характер появления новых изменений	Редкий, у отдельных особей; обычная частота 1 : 100 ООО, при действии ФК – до 20-40%	Массовый	Предопределенный
Характер наследования в ряду поколений	Менделевский	Менделевский и неменделевский	Системный (возможность «поглощения» признаков у гибридов, колебательный режим)

Взгляд на генотип эукариот как на систему взаимодействующих между собой информационных макромолекул, деление наследственной памяти на постоянную и оперативную, воплощенную в виде ОК и ФК, динамический способ хранения и передачи наследственной информации – все это позволяет построить стройную систему эволюции на разных уровнях организации.

В самом деле, на популяционном уровне получение информации зависит от интенсивности абиотического экологического фактора среды и сводится в конечном счете к гибели одних особей, отсутствию потомства у других и, наконец, к усиленному размножению третьих, наиболее устойчивых к влиянию данной среды (см. ставший классическим рис. 48). В результате такой перестройки популяция «выходит» из-под влияния вредного для нее фактора среды – ***эволюционирует.***

Однако для того чтобы популяция вовремя чувствовала приближение фронта вредного фактора, какая-то часть ее должна находиться в контакте с этим фактором, платя за информацию своей смертью. Эту роль, по В. А. Геодакяну, в процессе биологической эволюции принял на себя мужской пол, отличающийся повышенной чувствительностью ко всем факторам среды, т. е. обладающий большей дисперсией по всем признакам.

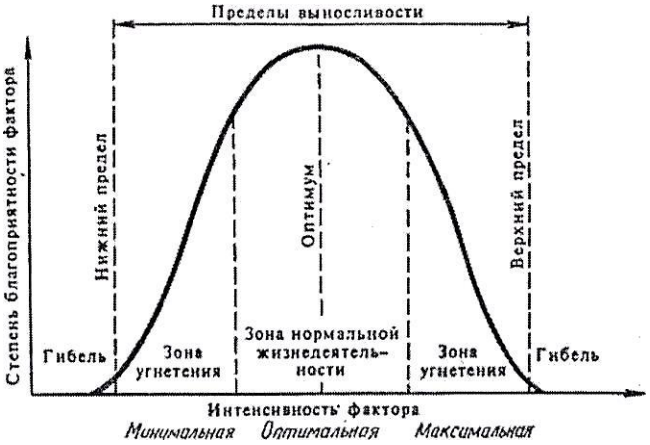


Рисунок 48 – Стандартная схема действия экологического фактора на популяцию

Таким образом, главной эволюционной задачей мужского пола является связь со средой, получение и передача от среды к популяции новой генетической информации, а женского пола – сохранение этой информации и ее закрепление в новых поколениях.

Поэтому соотношение полов оказывается одним из наиболее важных параметров раздельнополой популяции, определяющим эволюционную пластичность вида: чем больше в популяции самцов, тем больше шансов у всей популяции как единого целого приспособиться к действию вредных для нее факторов среды – и наоборот. Эта пластичность реализуется через дисперсию полов, означающую, что генотип у мужского пола связан с фенотипом гораздо «жестче», чем у женского. Это приводит к проявлению и «отбраковке» средой любого мужского генотипа, тогда как женские могут «ускользнуть» от действия среды. Иначе говоря, норма реакции женского пола шире, чем мужского, тогда как дисперсия признаков среди самцов больше, чем среди самок (рис. 49). По образному выражению В. А. Геодакяна, в лично-командных соревнованиях между сборными самцов и самок чемпионами в личном первенстве по всем программам будут самцы, а в командном зачете победят самки.

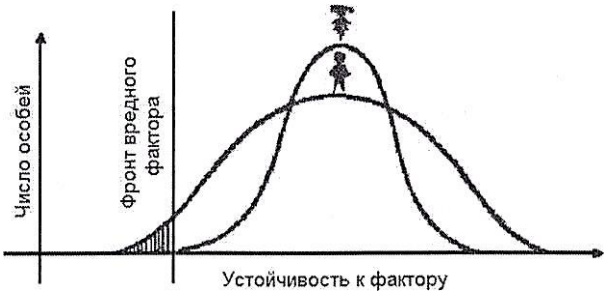


Рисунок 49 – Кривые разброса признаков мужского и женского полов по В. А. Геодакяну

Различие эволюционных ролей полов приводит к тому, что эволюционные изменения признаков, по которым наблюдается половой диморфизм, происходят обычно от нормы женского пола к норме мужского. Признаки, характерные для недалекого прошлого вида, будут чаще встречаться у женских особей, для недалекого возможного будущего – у мужских. Таким образом, по В. А. Геодакяну процесс эволюции рассматривается как метод проб и ошибок. Разделение единой популяции на две подсистемы (пола) позволяет в одной из них – мужском поле локализовать как пробы (поиск), так и неизбежные при этом ошибки. И только удачные находки, успешно «сдавшие» эволюционный «экзамен», получают возможность с помощью второй подсистемы (женского пола) закрепиться в популяции.

На молекулярном уровне такому механизму соответствует разделение на половые хромосомы и аутосомы. При этом эволюционно новая генетическая информация передается от Y- к X-хромосоме, а затем закрепляется в аутосомах. ***Эволюционно «старая» информация проходит обратный путь: аутосомы X-хромосома Y-хромосома и удаляется, укорачивая ее.***

Исследование мтДНК и популяционная геномика. Согласно стандартной модели эволюции митохондриального генома человека современного вида, подавляющее большинство мтДНК-мутаций в зародышевых линиях нейтральны. Они последовательно накапливались в пределах материнских филетических линий с момента происхождения человека современного вида на ограниченной территории в Восточной Африке около 130 тыс. лет назад и его последующего расселения по земному шару. Между тем все больше данных свидетельствует о том, что слабопатогенные мутации мтДНК в одних популяциях и материнских филетических линиях встречаются чаще, чем в других, предположительно благодаря адаптивному преимуществу на ранних этапах освоения человеком Северной Евразии.

Вполне вероятно, что некоторые варианты мтДНК, обладавшие адаптивным преимуществом в прошлом, ныне предрасполагают к развитию митохондриальной болезни. Так как система окислительного фосфорилирования обладает двоякой функциональной способностью – обеспечивать клетки энергией в форме АТФ и вырабатывать тепло для терморегуляции, дифференциальный отбор в предковых популяциях Северной Евразии мог сохранять те мутации, которые способствовали увеличению производства тепла и улучшению терморегуляции за счет снижения эффективности производства АТФ.

В Институте генетики и цитологии ПАН Беларуси под руководством члена-корреспондента О. Г. Давыденко в 2005– 2007 гг. выполнялся проект Инновационного фонда НАН Беларуси «Этногеномика белорусского народа». В ходе исследований было проведено секвенирование первого гипервариабельного сегмента митохондриальной ДНК в сочетании с последующим ПДРФ анализом кодирующего региона на 292 образцах тотальной ДНК белорусов из различных географических регионов. В результате установлено, что гаплогруппы H, J, K, N1, T, U4, U5, V, X, W, характеризующие европейский тип, составляют около 90% митохондриалыюго пула этнических белорусов. В то же время анализ распределения гаплогрупп среди различных географических регионов Беларуси позволил выявить внутреннюю гетерогенность популяции по распределению частот таких гаплогрупп, как J, U, V, T.

Кроме материнского компонента белорусской нации в рамках проекта исследовался также вклад мужской составляющей, для чего был проведен анализ двадцати пяти биаллельных маркеров Y-хромосомы у 574 этнических белорусов, проживающих в различных регионах страны. Выявленный генетический пул Y-хромосомы в основном соответствует данным, полученным на европейских популяциях. В белорусской популяции гаплогруппы Rla, lib и N3 являются доминантными и имеют частоты 50, 16 и 10% соответственно. Анализ внутрипопуляционной структуры Y-хромосомы современных белорусов позволил различить популяции южного (западное и восточное Полесье), а также северо-западного регионов на основании их композиций биаллельных маркеров, что отражает внутреннюю генетическую гетерогенность популяции. Результаты исследований демонстрируют высокое генетическое сродство белорусов с соседними славянскими народами (русскими, поляками, украинцами), а также с балтскими народами.

Генная реконструкция филогенеза популяций человека математическими методами. Надо отметить, что картирование отдельных генов человека в сочетании с математическими методами популяционной генетики, разработанными в первой трети XX в., также позволяет реконструировать филогенез отдельных групп и популяций человека. В прикладных задачах часто необходимо знать асимптотическое поведение процесса генного

дрейфа на небольших промежутках времени. Для дрейфа в случае двухаллельного гена такая аппроксимация была дана еще в 1920-е годы Р. Фишером. Предложенное им преобразование позволяло стабилизировать дисперсию генных концентраций. Для полиаллельного случая его подход дает возможность описать популяцию в любой момент времени в n-мерном декартовом пространстве. Если $p_1..., p_n$ и $x_1(t)$ – начальные и преобразованные концентрации аллелей, то положение популяции определяется вектором единичной длины $\{\sqrt{x_1(t)}, ..., \sqrt{x_i(t)}\}$, конец которого лежит на гиперсфере единичного радиуса в области положительных полуосей.

Отклонение от начального вектора $\{\sqrt{p_1}, ..., \sqrt{p_i}\}$ характеризуется углом $\Theta(t) = \arccos\left(\sum_{i=1}^{n} \sqrt{x_i(t)}\sqrt{p_i}\right)$. С помощью угла $\Theta(t)$ можно оценивать «угловое расстояние» между изолированными популяциями, имеющими общее происхождение, и реконструировать родословные древа родственных популяций. Особенно плодотворен такой подход при оценке появления новых племен изолированных популяций человека. Анализ обычно ведется на основе нейтральных по отношению к отбору генетических признаков, например отдельных групп крови. В качестве примеров такой оценки можно указать анализ населения Аландского архипелага в Балтийском море или религиозной секты дункеров в Пенсильвании. Еще в конце 1960-х годов советские ученые М. Б. Малютов, В. П. Пасеков и Ю. Г. Рычков провели интересное исследование реконструкции происхождения группы сибирских монголоидных племен, в образовании которых существенную роль сыграла изоляция (табл. 7).

Таблица 7 – Сопоставление генетической реконструкции с историческими сведениями

Генохронологическая реконструкция	Историческая хронология
Нет данных	VII–V вв. до н. э. Население европеоидного облика II в. до н. э. – V в. н. э. Распространение монголоидов в Туве 581–618 гг. Упоминание о древних тюрках теле в китайских хрониках
640 г. Разделение на горно-таежную (тофаларскую) и лесостепную (тувинскую) ветви	618–907 гг. Упоминание о тупо – предках современных тувинцев и тофаларов – как ветви теле в районе Восточного Саяна 627–648 гг. Существование каганата теле 750–840 гг. Уйгурское господство в Туве
1060 г. Разделение тоджинцев и тувинцев 1420 г. Отделение ветви гутарской популяции, основу которой составляет племя Хаш	XIII–XVI вв. Монгольская колонизация Тувы, основание золотых приисков в Тофаларии, положивших начало разделению тофаларов
1620 г. Отделение ветви алыгджерской популяции, основу которой составляет племя Чогда	XVII–XVIII вв. Маньчжурская колонизация Тувы, русская колонизация Тофаларии, введение административного деления для тофаларов в соответствии с родоплеменными группами

Литература: Дромашко, С. Е. Очерки биоинформатики / С. Е. Дромашко ; НАН Беларуси, Ин-т генетики и цитологии ; науч. ред. Л. В. Хотылева. – Минск : Беларус. навука, 2009. – 400 с.

Занятие 7.4 Биосфера и антропогенные воздействия

Вопросы для изучения:

1. Структура биосферы, её эволюция, условия устойчивости.

2. Антропогенные воздействия. Загрязнение окружающей среды.
3. Биомониторинг и экологический прогноз.
4. Видеоинформация по цитологии и цитогенетике.

Учебно-методические материалы.

Структура биосферы, ее эволюция, условия устойчивости. ***Биосфера*** охватывает часть атмосферы, гидросферу и верхнюю часть литосферы, которые взаимосвязаны сложными биогеохимическими циклами миграции веществ и энергии, начало которым дает трансформация солнечной энергии растениями и последующий синтез биогенных веществ. Термин «биосфера» ввел в 1875 г. австрийский геолог Э. Зюсс.

Общее учение о биосфере создано в 20–30-х годах XX в. В. И. Вернадским, развившим идеи В. В. Докучаева о комплексном естественно-историческом анализе взаимодействующих в природе разнокачественных объектов и явлений (факторов почвообразования) и выявлении самостоятельных природных объектов гетерогенной структуры и состава (почвы, природные зоны). Биосфера включает не только область жизни (биогеосферу, фитогеосферу, геомериду, витасферу), но и другие структуры

Земли, генетически связанные с живым веществом. По В. И. Вернадскому, вещество биосферы состоит из семи разнообразных, но геологически взаимосвязанных частей: живое вещество, биогенное вещество, косное вещество, биокосное вещество, радиоактивное вещество, рассеянные атомы, вещество космического происхождения. В пределах биосферы везде встречается либо живое вещество, либо следы его биогеохимической деятельности.

Биосфера – чрезвычайно сложная, саморегурирующаяся открытая система, в которой выполняется принцип JTe Шателье, обусловливающий устойчивость ее существования. Согласно этому принципу, внешнее воздействие на систему, находящуюся в состоянии равновесия, приводит к смещению этого равновесия в направлении, при котором эффект произведенного воздействия ослабляется. С кибернетической точки зрения здесь идет речь о сложной системе стабилизирующих обратных связей и зависимостей, пронизывающих всю биосферу. Стабильность биосферы основывается на высоком разнообразии живых организмов, отдельные группы которых выполняют различные функции в поддержании общего потока вещества и распределении энергии, на теснейшем переплетении и взаимосвязи биогенных и абиогенных процессов, на согласованности циклов отдельных элементов и уравновешивании емкости отдельных резервуаров. Однако эта стабильность имеет определенные пределы, и нарушение ее регуляторных возможностей чревато серьезными последствиями.

Антропогенные воздействия. Опасность дестабилизации биосферы связана с тем, что современное человечество использует не только огромные энергетические ресурсы биосферы, но и небиосферные источники энергии (например, атомной), ускоряя геохимические преобразования природы. Некоторые процессы, вызванные технической деятельностью человека, направлены противоположно их естественному ходу в биосфере (рассеивание металлов, руд, углерода и других биогенных элементов, торможение минерализации и гумификации, освобождение законсервированного углерода и его окисление, нарушение крупномасштабных процессов в атмосфере, влияющих на климат, и т. п.). В частности, за последние 100 лет численность человечества увеличилась в 4 раза, потребление энергии – в 10 раз, минерального сырья – в 29 раз, совокупный продукт – в 17,6 раза. 85% всех добытых за всю историю человечества полезных ископаемых приходится на XX в. Общее количество используемой энергии в конце XX в. было всего на 3–4 порядка величин меньше суммарной солнечной энергии, поступающей на верхнюю границу атмосферы Земли. К настоящему времени 1/4 суши занята агроценозами и пастбищами и 3/4 непокрытой вековыми льдами территории оказывается в зоне прямого хозяйственного воздействия. Мировой улов рыбы достиг своего теоретического предела. Происходит изменение глобального климата Земли, в

результате которого усиливаются стихийные бедствия, возрастают материальные и людские потери, начинается вымирание значительного количества видов.

Решением проблем взаимодействия и взаимовлияния биосферы и человечества заняты многие национальные и международные организации, научные центры и лаборатории. Большой вклад в разработку этой проблемы внесли доклады так называемому Римскому клубу, в которых кризис современной западной цивилизации рассматривался как комплекс взаимосвязанных кризисов: экологического, социального, демографического, ресурсного, экономического, политического и культурного. В серии докладов Римскому клубу «Пределы роста», «Динамика роста в конечном мире», «Человечество на перепутье», написанных в 1972–1974 гг., широкой мировой общественности было убедительно продемонстрировано, что экспоненциальный экономический рост имеет объективно обусловленные пределы, которые связаны с истощением невозобновляемых ресурсов и с приближением к потреблению всей продукции возобновимых ресурсов. Доклад Римскому клубу «Пределы роста» вызвал разнообразные отклики не только в научных кругах, но и в прессе; многими он был воспринят как предсказание скорого конца света, хотя речь шла скорее, по словам Д Медоуза, «не о будущем, а о выборе этого будущего».

Первой реакцией международного сообщества на эти вызовы стала ***Конференция по окружающей среде*** в Стокгольме в 1972 г. *(United Nations Conference on Human Environment in Stockholm, 5 –16 June 1972).* Конференция показала, что окружающую среду и развитие цивилизации нельзя рассматривать раздельно, второе неотделимо от первого. Окружающая среда – место жизни людей и источник всех благ, а развитие – процесс удовлетворения потребностей во благах, способы которого непрерывно меняются научно-техническим прогрессом. Главный итог Конференции 1972 г. – создание ***Программы ООН по окружающей среде (ЮНЕЛ),*** специальной структуры ООН для дальнейшей проработки обозначившихся проблем.

Вторая Конференция ООН по окружающей среде и развитию (***КОСР-2)*** состоялась 3–14 июня 1992 г. в Рио-де-Жанейро (Бразилия) (***United Nations Conference on Environment and Development in Rio de Janeiro, 3–14 June 1992).*** Это была действительно всемирная конференция, в которой участвовало 178 стран, 114 глав государств и правительств, представители 1600 неправительственных организаций, много журналистов. В это же время в Рио-де-Жанейро проходил «Глобальный форум» по экологическим проблемам, собравший около 9 тыс. организаций, 29 тыс. человек, 450 тыс. посетителей, а всего было проведено около тысячи заседаний и мероприятий разного типа. На Конференции был принят ряд документов, в частности «Декларация Рио-де-Жанейро по окружающей среде и развитию» (Декларация Рио), «Заявление о принципах глобального консенсуса по управлению, сохранению и устойчивому развитию всех видов лесов», «Повестка дня на XXI век» (Программа 21), «Рамочная конвенция ООН об изменении климата», «Конвенция ООН о биологическом разнообразии».

Можно также назвать ряд других мероприятий, в которых участвуют многие страны, например Международное гидрологическое десятилетие, Международная биологическая программа и т. п. Отметим также, что ЮНЕСКО учредило и координирует специальную ***Международную программу «Человек и биосфера»*** (***МАВ*** – от ее англ. названия ***«Man and the Biosphere»).*** Соответствующий Национальный комитет по программе МАВ около 40 лет назад создан в нашей стране.

Загрязнение окружающей среды. Результатом экстенсивного развития мировой экономики стало загрязнение окружающей среды. Человечество никогда ранее не задумывалось о судьбе отходов жизнедеятельности и не планировало замкнутых циклов производства. По сравнению с круговоротом веществ в биосфере отходы экономической деятельности человечества долгое время оставались незначительными. Однако многократное увеличение в течение XX в. промышленного и сельскохозяйственного производства привело к

столь же масштабному загрязнению воды, воздуха, почвы. При ограниченных размерах почти полностью заселенной планеты люди должны теперь сами обеспечивать переработку своих отходов так, чтобы не навредить биосфере.

Токсикантами среды обитания называются вредные вещества, распространяющиеся на обширные регионы биосферы далеко за пределы своего первоначального местонахождения, т. е. имеющие значительную миграционную подвижность и оказывающие скрытое вредное воздействие на животных, растения и впоследствии на человека.

К токсикантам среды обитания относят различные группы веществ, в частности многие органические соединения (например, бензол, асбест, бензапирен, пестициды), микроэлементы (марганец Мп, медь Си, кобальт Со), тяжелые металлы (ртуть Hg, свинец РЬ, кадмий Cd, цинк Zn, хром Сг и др.), радионуклиды ($^{134,137}Cs$, ^{90}Sr, трансурановые элементы), диоксид серы ($S0_2$), оксиды азота в воздухе (NO) и воде (нитраты, нитриты) и др.

По данным ООН, в результате деятельности человека в 1996 г. в атмосферу поступило 350 млн т оксида углерода СО, 145 млн т диоксида серы SO_2 20 млн т оксида азота N0, 2 млн т фреонов и полихлорвиниловых веществ, 200 тыс. т соединений свинца, 1 млн т углеводородов, в том числе канцерогенных, большое количество аэрозолей и твердых частиц (пыль, копоть, сажа) – по разным оценкам от 96 до 260 млрд т.

Высокие концентрации токсикантов могут вызвать специфическое химическое отравление различной степени. Низкие концентрации могут оказывать скрытое токсическое действие, которое способно проявиться в виде раковых заболеваний (канцерогенез), наследственных изменений (мутагенез), рождении уродов (тератогенез) или токсическом влиянии на плод (эмбриоцидное действие). Скрытое токсическое действие, в частности мутагенное, может сказываться на здоровье нескольких поколений людей. Токсиканты-мутагены вызывают изменения в молекулах ДНК, что ведет к развитию врожденных уродств и отклонений от нормального развития. Более 90% мутагенов одновременно являются канцерогенами, вызывая развитие злокачественных опухолей.

Биомониторинг и экологический прогноз. Согласно определению секретариата комиссии ЮНЕП, ***экологический мониторинг*** – это система повторных наблюдений за элементами окружающей среды в пространстве и во времени с определенными целями и в соответствии с определенными обстоятельствами. Экологический мониторинг должен включать звенья разного уровня.

1. Глобальный мониторинг, основанный на международном сотрудничестве.
2. Национальный мониторинг – общегосударственная система наблюдений.
3. Региональный мониторинг – система наблюдений в регионах с острой экологической ситуацией.
4. Локальный (импактный) мониторинг – наблюдения в зоне влияния предприятий.
5. Фоновый мониторинг – наблюдения в районах, где исключена всякая хозяйственная деятельность, в частности в биосферных заповедниках.

По результатам мониторинга составляются ***экологические прогнозы (ecological forecasts)*** – предсказания поведения экосистем, определяемого естественными процессами и воздействием на них человечества, в том числе с учетом данных соответствующих математических и компьютерных моделей. Экологические прогнозы служат основой для принятия соответствующих управленческих решений. По масштабам прогнозируемых явлений различают глобальный, региональный, национальный и локальный экологические прогнозы.

Важную роль в мониторинге и прогнозировании играют геоинформационные системы. ***Геоинформационная система*** (также ***ГИС – географическая информационная система)*** – система, предназначенная для сбора, хранения, анализа и графической визуализации пространственных данных и связанной с ними информации о представленных в ГИС объектах. По территориальному охвату различают глобальные ГИС (global GIS), субконтинентальные ГИС, национальные ГИС, часто имеющие статус государственных, региональные ГИС (regional

GIS), субрегиональные ГИС и локальные, или местные, ГИС (local GIS). В частности, можно назвать Геоинформационную систему «Республика Беларусь» (http://www.emaps- online.com/). ГИС различаются предметной областью информационного моделирования, например, городские ГИС, или муниципальные ГИС – МГИС (urban GIS), природоохранные ГИС (environmental GIS) и т. д. Проблемная ориентация ГИС определяется решаемыми в ней задачами (научными и прикладными), к ним относятся инвентаризация ресурсов (в том числе кадастр), анализ, оценка, мониторинг, управление и планирование, поддержка принятия решений. Интегрированные ГИС, или ИГИС (integrated GIS, IGIS), совмещают функциональные возможности ГИС и систем цифровой обработки изображений (данных дистанционного зондирования) в единой интегрированной среде.

Среди белорусских экологических геоинформационных ресурсов можно назвать ГИС «Беловежская пуша». Этот проект охватывает Государственный национальный парк «Беловежская пуща», расположенный в Беларуси на границе с Польшей и занимающий площадь около 90 тыс. га. ГИС «Беловежская пуша» включает следующие основные блоки: топографическая основа, лесные ресурсы, геологические ресурсы, экологическая информация, растительность, животный мир, климат, управление, изображения, энциклопедия. Дальнейшее развитие ГИС «Беловежская пуща» предполагает создание в республике единой геоинформационной системы охраняемых и заповедных территорий, включающей в первую очередь Березинский биосферный и Припятский ландшафтно-гидрологический заповедники. Национальный парк «Браславские озера».

Биомониторинг является составной частью экологического мониторинга и включает в себя такие подсистемы, как биотестирование, биоидикацию и биоаккумуляцию. Биотестирование и биоиндикация – это два схожих исследовательских приема, в которых о качестве среды, факторах, воздействующих на эту среду, судят по выживаемости, продуктивности, поведению, а также по различным физико-биологическим параметрам живых организмов. Различие состоит в том, что ***биотестирование*** подразумевает использование живых организмов, специально помещаемых в данную среду (тест-объекты), а ***биоиндикация*** исследует живые организмы, естественным образом обитающие в данной исследуемой среде. ***Биоаккумуляция*** – частный случай биотестирования или биоиндикации, когда о качестве среды и факторах, воздействующих на эту среду, судят по степени накопления вредных веществ в живых организмах.

Приемы и подходы при выборе тест-объектов и тест-реакций. Можно привести ряд примеров использования методов биотестирования. Так, в Голландии в качестве тест-объектов на больших площадях страны используются различные полезные для человека растения: гладиолусы и тюльпаны являются тест-объектами на накопление фторидов; итальянская ржаная трава – тест-объект на накопление ионов тяжелых металлов. В последнее время широкое применение во многих странах находит биотестирование с помощью таких высокочувствительных к загрязнению разного рода (ионы тяжелых металлов, нефтепродукты, УФ излучение и т. д.) тест-объектов, как дафнии (рачки – фильтраторы водоемов), пиявки, инфузории (простейшие), хлорелла (микроводоросль) и т. п.

В целях биоиндикации используется метод лихеноиндикации (от лат. ***lichenes*** – лишайник), основанный на учете количества лишайников в городских насаждениях, районах крупных предприятий и т. д. Установлена четкая связь между «полями загрязнения воздуха» в городах и встречаемостью лишайников на стволах деревьев. Весьма перспективным методом оценки качества поливной воды в садоводческих хозяйствах Подмосковья, Дальнего Востока и т. д. является так называемый ***coli-***тест на наличие в воде бактерии кишечной палочки (*Escherichia coli*)**.**

Биоаккумуляция примесей наблюдается у некоторых живых организмов. Например, накопление ртути в рыбе из реки Минамата в Японии привело к заболеванию населения, особенно детей, «болезнью Минамата». Ртуть и мышьяк, как известно, обладают способностью

накапливаться в волосах человека, что может быть использовано для индикации и предотвращения отравления человека этими ядами. В Японии весьма успешно разводят асцидии – морские организмы, которые концентрируют ванадий, затем их сжигают и получают из золы соединения ванадия. Исследования, проведенные в Беларуси после Чернобыльской катастрофы, показали, что существуют виды растений и грибов, накапливающие радионуклиды (виды-накопители), и виды-дискриминаторы, которые не поглощают радионуклиды из почвы. Эти данные чрезвычайно важны для мониторинга радиационно-загрязненных территорий, а также для безопасного ведения сельскохозяйственного производства.

Как мы видим, в биомониторинге в зависимости от поставленных целей и задач используются различные ***виды-индикаторы (indicators),*** т. е. виды, служащие показателем особенностей среды какого-либо биоценоза или экосистемы и относящиеся к разным уровням организации. При этом токсикологическая и экологическая характеристика токсикантов-ксенобиотиков дается по физиолого-биохимическому ответу живых систем.

Кроветворная ткань является одним из основных тест-объектов биоиндикации воздействия антропогенных факторов на организм животных. Повышение над контрольным уровнем числа клеток с хромосомными аберрациями является классической цитогенетической тест-реакцией повреждения ДНК. Однако в случае пойкилотермных животных с невысоким темпом деления клеток микроядерный тест имеет больше преимуществ над метафазным и анафазным методами. Другой тест-реакцией, указывающей на угнетающие действие на кроветворную ткань и возможное мутагенное воздействие, является клеточная гибель.

Филогенетически различные виды-биоиндикаторы могут отличаться по чувствительности на уровне генетической детерминации процессов формирования, сохранения и элиминации клеток с хромосомными аберрациями – маркерами мутагенного воздействия. Амфибии и млекопитающие (мышевидные грызуны) являются классическими видами-биоиндикаторами. Моллюски, как мы могли видеть выше, представляют собой перспективный вид-биоиндикатор. В лаборатории моделирования генетических процессов по инициативе к. б. н. В. Ю. Афонина проводились работы по введению новой тест-системы на основе легочного брюхоногого моллюска *большого прудовика (Lymnaea stagnalis)***.**

Lymnaea stagnalis среди моллюсков относится к наиболее развитым и по многим параметрам близок к млекопитающим. Продолжительность жизни и скорость смены поколений у этого вида моллюсков близка к таковым у мышевидных грызунов, используемых в биомониторинге. С другой стороны, на онтогенез моллюсков, как и амфибий, накладывают отпечаток сезонные факторы, что позволяет контролировать немутагенные параметры среды, влияющие на различные клеточные процессы. Связь частоты повреждений ДНК, индуцированных мутагенами, и генетической структуры популяции по различным ферментам указывает на возможное сочетанное применение этих критериев в комплексе с цитогенетическими повреждениями.

Возможность прижизненных наблюдений за различными биологическими характеристиками моллюсков делает данный вид удобным для биоиндикационных исследований в лаборатории. Способность к самооплодотворению и удобство содержания позволит создавать линии животных для изучения мутагенных воздействий как в лабораторных условиях, так и в природных. Различная степень дифференцировки гемоцитов и эмбриональных клеток моллюсков позволит разработать комплекс тестов ***in vitro*** и ***in vivo*** клеток. Так, используя одного моллюска, можно долгий период получать от него пробы иммунной и эмбриональной ткани. Разные темпы роста этих животных и выраженность нервных оборонительных реакций дадут дополнительную информацию о возможном модифицирующем действии физиологических факторов на результаты экспериментов по изучению действия ксенобиотиков.

Видеоинформация по цитологии и цитогенетике. В связи с проблемами биомониторинга остановимся на еще одном направлении исследований лаборатории

моделирования генетических процессов – создании электронного атласа по морфологическим формам клеток, отображающего процессы дифференцировки, апоптоза и цитогенетических нарушений под влиянием загрязнения окружающей среды. Изучение апоптоза, или программированной клеточной гибели, – одно из наиболее актуальных направлений биологических исследований, которому в последние годы посвящены сотни публикаций (см., например, обзор в журнале «Генетика»).

В этой области к. б. н. В. Ю. Афониным совместно с другими сотрудниками лаборатории на основе результатов собственных исследований создан широкий набор плакатов, отражающих различные аспекты указанных процессов у животных отдельных таксономических групп (рис. 50), в том числе следующие:

- Клетки белой крови амфибий.
- Цитогенетические нарушения в эритроцитах бурых лягушек.
- Эмбриональная гибель и цитогенетические нарушения у моллюсков (прудовики)
- Цитогенетические повреждения эритроидной ткани лабораторных мышей.
- Дифференцировка и гибель фибробластов культуры клеток мыши.
- Апоптоз в различных тканях европейской рыжей полевки.
- Сочетание окрасок акридиновым оранжевым и красителем Гимза.

Рисунок 50 – Плакаты из электронного атласа: апоптоз в разных тканях европейской рыжей полевки (а); цитогенетические нарушения в клетках мантийной жидкости моллюска прудовика (б)

Часть материалов видеоатласа можно найти в Интернете.

Поскольку ткани филогенетически разных животных обновляются с разной скоростью, данное пособие позволит выбирать комплекс критериев для оценки повреждений кроветворной ткани взаимозаменяемых модельных видов, а также отбора тест-объектов, наиболее подходящих для конкретной экологической ситуации, что важно для объективного планирования и проведения мониторинговых исследований.

Литература: Дромашко, С. Е. Очерки биоинформатики / С. Е. Дромашко ; НАН Беларуси, Ин-т генетики и цитологии ; науч. ред. Л. В. Хотылева. – Минск : Беларус. навука, 2009. – 400 с.

3. РАЗДЕЛ КОНТРОЛЯ ЗНАНИЙ

3.1. Вопросы к экзамену

1. Биоинформатика как наука. Понятие информации Свойства информации
2. Методологические аспекты моделирования. Понятие об аналогии. Ассоциативная аналогия. Определение модели. Аналогия и моделирование.
3. Типы моделирования. Материальные модели. Мысленные модели. Аналитические, конструктивные и имитационные модели. Материальное моделирование и модельный эксперимент.
4. Особенности моделирования в биологии. Модели данных. Модели систем. Применение некоторых информационных технологий в биологии.
5. Компьютерная алгебра и биоинформатика. Биологическое приложение теории нечётких множеств.
6. Сложные системы и оптимизация эксперимента. Полиномиальная модель и её адекватность.
7. Понятие линейного программирования. Особенности биометрической культуры. Статистика ДНК, как характеристика генома.
8. Программное обеспечение для статистической обработки биомедицинских исследований.
9. Базы и банки данных. Первичные, вторичные и курируемые базы данных.
10. Автоматическое аннотирование последовательности. Идентификаторы записей в базах данных.
11. Аналитические программы в биоинформатике. Энциклопедия KEGG и ее использование.
12. Выравнивание последовательностей. Алгоритм Нидлмана-Вунша.
13. Пути эволюции молекул и организмов. Молекулярная эволюция, как наука. Критерии сравнения нуклеотидных и белковых последовательностей.
14. Понятие гомологии. Работы Оуэна Р. В сфере создания естественной системы видов.
15. Ортологи, паралоги: аллопаралоги, симпаралоги. Ксенология.
16. Генетический перенос. Вертикальный и горизонтальный переносы генов.
17. Филогенетическое древо и методы его построения (UPGMA, NEIGHBOR-JOINING, MINIMAL EVOLUTION).
18. Биотехнология и генная инженерия. Генная инженерия микроорганизмов. Перенос генов в растения. Перенос генов животных. Биотехнология и биобезопасность.
19. Механизмы регуляции экспрессии генов прокариотов и эукариотов. Модель оперона.
20. Позитивная и негативная регуляции экспрессии генов. Этапы процесса экспрессии генов.

21. Значение анализа последовательностей ДНК. Секвенаторы. Автоматическое секвенирование ДНК. Секвенирование ДНК по Сэнгеру Ф. Секвенирование ДНК по Максаму и Гилберту: метод химической деградации. Проект «Геном человека».

22. Композиционная гетерогенность ДНК. Изохоры. Связь семейств изохор с кодирующими последовательностями ДНК и их экспрессией.

23. Значение праймеров для процесса амплификации и ПЦР. Принципы подбора праймеров. Требования, предъявляемые к праймерам.

24. Анализ частоты использования кодонов. Частота транзиций и трансверсий. Определение соотношения транзиций и трансверсий. ГН-насыщенность общая и отдельных положений кодона. Зависимость ГЦ-насыщенности отдельных положений кодона от общей ГЦ-насыщенности.

25. Анализ аминокислотного состава белков.

26. Системы классификации сайтов нуклеиновых кислот в связи с их положением в кодоне - система вырожденности и синонимичности / несинонимичности.

27. Методики изучения стратегии кодирования белков.

28. Фолдинг, белки шапероны Секреция белков у прокариот: Sec-аппарат и сигнальный пептид.

29. Распределение белков по компартментам клетки эукариот. Деградация белков.

30. Мотивы белков. Домены, их состав. Разница между мотивом и доменом.

31. Сворачивание белков. Цель и методы предсказания структуры белка. Алгоритмы предсказания вторичной структуры белков.

32. Предсказание функции и клеточной локализации белков. Методы прогнозирования структуры белков, основанные на гомологии, мотивах последовательностей, структуре белка, геномном контексте.

33. Типы клеточных популяций. Понятие о дифферoне. Апоптоз. Онтогенез и старение.

34. Новая компьютерная технология прижизненного наблюдения клеток. Бета-гал тест как маркер старения.

35. Злокачественные новообразования как следствие нарушения молекулярно-генетических и клеточных регуляторных механизмов.

36. Моделирование работы генов. Два типа моделей клеточной дифференцировки. К-генетические сети. Модель Сендова-Цанева.

37. История открытия стволовых клеток. Получение плюрипотентных клеток. Молекулярные основы тотипотентности эмбриональных стволовых клеток.

38. Стволовые клетки взрослого организма. Различия в потенции стволовых клеток разного происхождения.

39. Механизмы коммитирования стволовых клеток. Значение микроокружения для самоподдержания популяции стволовых клеток.

40. Молекулярные маркеры стволовых клеток. Источники стволовых клеток у взрослого организма.

41. Стволовые клетки в медицине. Сайты и базы данных по стволовым клеткам и клеточным технологиям.

42. Кратковременные и долговременные изменения в синапсах. Иерархические и диффузные системы мозга. Стереотаксис.

43. Математический нейрон Маккалока-Питтса. между биологическим и искусственным нейроном.

44. Нейронные сети. Приложение искусственных нейронных сетей к распознаванию образов. Будущее нейрокомпьютинга.

45. Современное состояние научной разработки проблемы коммуникации в физиологии и психологии.

46. Моделирование процессов мышления и искусственный интеллект. Искусственный интеллект и поведенческое мышление.

47. Современные представления о нейрофизиологических механизмах психических функций.

48. РНК-интерференция и психические заболевания.

49. Биоинформатика и оценка биологического разнообразия. Сообщество организмов. Экологические системы. Экосистемы и теория информации. Сложность экосистем.

50. Микроэволюция на протоке (математическая модель). Эволюция и ЭВМ. ЭВМ и макроэволюция.

51. Энергетические критерии эволюционного процесса. Теоретико-информационные аспекты эволюции.

52. Кибернетические особенности наследственной изменчивости.

53. Исследование м-ДНК и популяционная геномика.

54. Генная реконструкция филогенеза популяций человека математическими методами.

55. Биомониторинг и экологический прогноз. Видеоинформация по цитологии и цитогенетике.

3.2. Примерная тематика рефератов

(по Каленик Т. К., 2019)

1. Генерация и рецепция информации.
2. Особенности генетической информации.
3. Биоинформационные данные, сети и базы.
4. NCBI и сервисы
5. Форматы записи Fasta, Genbank, PDB
6. EMBL
7. Swiss-PDBviewer
8. Конструирование праймеров для ПЦР.
9. Геном человека и постгеномные проекты.
10. Геномная дактилоскопия.
11. Бластинг последовательностей нуклеиновых кислот и белков.
12. Использование доменов для предсказания структуры и функции белков.

13.Методы предсказания 3 D структуры белков.

14.Молекулярная филогения как средство изучения эволюционных связей между биологическими видами.

15.Микроэррей как средство массового анализа экспрессии генов.

16.Эволюция молекул и организмов.

17.Филогенетическое дерево как математический объект.

18.Модели эволюции.

19.Алгоритмы построения филогенетических деревьев.

20.Алгоритмические проблемы поиска оптимального дерева, bootstrapping, согласование деревьев.

21.Эволюция на уровне генома.

22.Анализ популяционных данных.

23.Статистика последовательностей ДНК

24.Статистика ДНК как характеристика генома

25.Вычислительная геномика

26.Метаболическая реконструкция

27.Позиционный анализ

28.Эволюция регуляторных взаимодействий

29.Эволюция белковых семейств, их доля в геноме

3.3 Тематика докладов для самостоятельной работы

(по Каленик Т. К., 2019)

1. Современные методы исследования первичной структуры белков (определение N-, C-, концевых аминокислот, секвенирование).
2. Протеомика: возможности и перспективы.
3. Процессинг и фолдинг синтезируемого белка, биологическое значение этих процессов.
4. Протеомика-лидер науки XXI века.
5. Филогенетические деревья.
6. Геномика и медицина.
7. Гипотеза «молекулярных часов»
8. Метаболомика и проблема антибиотикорезистентности.
9. Геномика и здоровье человека.
10. Метагеномика – обширная геномная информация из окружающей среды.

3.4 ИТОГОВЫЙ ТЕСТ

(по Каленик Т. К., 2019)

1. Мутация, при которой единичная замена основания оставляет аминокислотную последовательность неизменной, называется

а) нонсенс-мутация

б) обратная замена

в) «молчащая» мутация

г) мисенс-мутация

2. Основной постулат (центральная догма) молекулярной биологии

а) ДНК $\rightarrow$ РНК $\rightarrow$ белок

б) ДНК $\leftrightarrow$ РНК $\rightarrow$ белок

в) ДНК $\rightarrow$ РНК $\leftrightarrow$ белок

г) РНК $\rightarrow$ ДНК $\rightarrow$ белок

3. Вырожденность генетического кода – это
а) кодирование одним триплетом только одной аминокислоты
б) кодирование одним триплетом одной либо нескольких аминокислот
в) кодирование одной аминокислоты несколькими триплетами
г) кодирование одним триплетом разных аминокислот

4. Универсальность генетического кода – это
а) кодирование одним триплетом одной либо нескольких аминокислот
б) кодирование одной аминокислоты несколькими триплетами
в) кодирование одной аминокислоты одним триплетом
г) наличие единого кода для всех существ на Земле

5. Для нахождения консервативных регионов в наборе последовательностей применяется преимущественно
а) множественное выравнивание
б) локальное выравнивание
в) глобальное выравнивание
г) структурное выравнивание

6. Выравнивание - это:
а) сравнение последовательностей нуклеотидов с «липкими концами»
б) сравнение аминокислотных последовательностей белков по длине
в) сравнение нуклеотидных последовательностей по длине
г) сравнение последовательностей в поиске идентичных серий символов

7. Расстояние по Левенштайну или «редакционное расстояние» между двумя строками
а) минимальное число «операций редактирования» для того, чтобы превратить одну строку в другую
б) максимальное число «операций редактирования» для того, чтобы превратить одну строку в другую
в) минимальное число замен позиций в строке для того, чтобы превратить одну строку в другую
г) минимальное число вставок для того, чтобы превратить одну строку в другую

8. PSI-BLAST- это программа, которая
а) позволяет проводить анализ популяционно-генетических данных
б) осуществляет филогенетический анализ с использованием метода парсимонии
в) подбирает данные для последовательностей, аналогичных запрошенной
г) проводит множественное выравнивание нуклеотидных и аминокислотных последовательностей

9. В иерархии структур белка, домены располагаются …
а) после четвертичной структуры
б) между вторичными и третичными структурами мономера
в) подобный уровень отсутствует
г) между первичными и вторичными структурами мономера

10.Что из перечисленного не относится к основным типам генетических карт
а) генетические карты сцепленности генов

б) «бэндовые» схемы хромосом
в) последовательности ДНК
г) контиг

11.Контиг – это
а) набор перекрывающихся фрагментов ДНК, которые в совокупности представляют собой консенсусную область ДНК
б) локусы с варьирующим числом тандемных повторов
в) полиморфизм коротких тандемных повторов
г) короткий, секвенированный участок ДНК, локализованный в строго определенной области генома

12.Процесс спонтанного сворачивания полипептидной цепи в уникальную нативную пространственную структуру называется
а) сплайсинг
б) трансляция
в) фолдинг
г) процессинг

13.Основным инструментом биоинформатики является
а) выравнивание последовательностей
б) секвенирование
в) программирование
г) картирование генома

14.Точечная матрица – это
а) комплементарная, записанная в обратном порядке последовательность
б) палиндромность последовательности
в) матрицы расчета весов для замен в аминокислотных последовательностях
г) простейшее изображение, которое дает представление о сходстве между двумя последовательностями

15.Метод исследования вещества путём определения отношения массы к заряду и количества заряженных частиц, образующихся при том или ином процессе воздействия на вещество
а) ЯМР-спектроскопия
б) масс-спектрометрия
в) ИК-спекроскопия
г) электрофорез

16.Для аннотирования структуры белка не используется метод
а) метод выявления гомологии в последовательностях
б) метод распознавания фолда
в) экспериментальное определение структуры
г) метод Смита-Ватермана

17.При каком значении процента идентичных остатков в оптимальном выравнивании два белка будут, вероятно, иметь сходный паттерн фолдинга
а) более 25%
б) более 45 %

в) 18 %
г) 10 %

18.Паралогичные гены (paralogous genes) это:
а) гомологичные гены филогенетически родственных организмов
б) гены, которые произошли в результате внутригеномных дупликаций в геноме данного вида
в) гомологичные гены филогенетически родственных организмов, разошедшихся в процессе видообразования
г) гомологичные гены микроорганизмов, образовавшиеся в процессе горизонтального переноса

19.Трансмембранные сегменты состоят почти исключительно из гидрофобных аминокис-лотных остатков. Сколько остатков составляет длину транмембранных спиралей
а) 100 остатков
б) 150 остатков
в) 2-5 остатка
г) 15-30 остатков

20.Какая из перечисленных ниже программ используется для множественного выравнива-ния последовательностей ДНК и белков
а) ClustalW
б) BLAST
в) DALI
г) CASP

21.Модель для оценки эволюционного расстояния по нуклеотидным (либо аминокислотным) заменам в последовательности
а) модель Тамуры-Нея
б) модель Джукса-Кантора
в) марковская модель
г) скрытая марковская модель

22.Основная проблема постгеномной эры:
а) предсказание первичной структуры белка по последовательности ДНК
б) предсказание вторичной структуры белка по последовательности ДНК
в) предсказание третичной структуры белка по последовательности ДНК
г) предсказание четвертичной структуры белка по последовательности ДНК

23.Гомологичные нуклеотидные (или аминокислотные) последовательности называют паралогичными если…
а) они появились в результате видообразования
б) они появились в результате дупликации
в) они находятся в начале гена
г) они являются уникальными

24.Филогенетическое дерево (эволюционное дерево, дерево жизни) – дерево, отражающее эволюционные взаимосвязи между различными видами или другими сущностями, имеющими общего предка. Вершины филогенетического дерева делятся на три класса (отметить неверное):

а) листья
б) стволы
в) узлы
г) корень

25.Какое из перечисленных ниже выравниваний применяется к «похожим» последовательностям приблизительно одинаковой длины и наглядно показывает разницу между этими последовательностями
а) локальное
б) множественное
в) глобальное
г) структурное

26.Выравнивание нуклеотидных или аминокислотных последовательностей с самым высоким весом называют
а) оптимальным
б) множественным
в) глобальным
г) структурным

27.Методы предсказания структуры белков по аминокислотной последовательности включают в себя (отметить неверное)
а) моделирование по гомологии
б) распознавание способа укладки
в) предсказание новых фолдов
г) отсев вырожденных мишеней

Задания с выбором одного правильного ответа. Время выполнения задания 45 минут.
Количество заданий в каждом варианте – 4. Количество правильных ответов – 1.

4. ВСПОМОГАТЕЛЬНЫЙ РАЗДЕЛ

4.1. Учебно-программная документация

Учреждение образования
«Белорусский государственный педагогический университет
имени Максима Танка»

УТВЕРЖДАЮ

Проректор по учебной работе БГПУ

_____ А.В. Маковчик

«_____»_____2023 г.

Регистрационный № УД_____/уч.

БИОИНФОРМАТИКА

Учебная программа учреждения высшего образования
по учебной дисциплине для специальности
7-06-0113-03 Природоведческое образование
Профилизация: Биология, география, химия

2023 г.

Учебная программа составлена на основе образовательного стандарта углубленного высшего образования: ОСВО 7-06-1113-03-2023 по специальности 7-06-0113-03 Природоведческое образование (18.05.2023, № 160); учебного плана по специальности (23.02.2023, № 076-2023/у).

.

СОСТАВИТЕЛЬ:
Деревинский А. В., доцент кафедры биологии и методики преподавания биологии, кандидат сельскохозяйственных наук, доцент

РЕЦЕНЗЕНТЫ:
Звягинцев В. Б., доцент кафедры лесозащиты и древесиноведения учреждения образования «Белорусский государственный технологический университет», кандидат биологических наук, доцент
Ковалева О. А., доцент кафедры географии и экологии человека «Белорусский государственный педагогический университет имени Максима Танка», кандидат биологических наук, доцент

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:
Кафедрой биологии и методики преподавания биологии
(протокол № 4 от «22»ноября 2023 г.)

Заведующий кафедрой И.И.Жукова

Научно-методическим советом учреждения образования «Белорусский государственный педагогический университет имени Максима Танка»
(протокол № ___ от «_____»_____2023 г.)

Оформление учебной программы и сопровождающих ее материалов действующим требованиям Министерства образования Республики Беларусь соответствует

Методист отдела магистратуры БГПУ _____ А.И.Гарунчик

Директор библиотеки БГПУ _____ Н.П.Сятковская

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Учебная дисциплина «Биоинформатика» предусмотрена учебными планами подготовки обучающихся по специальности 7-06-0113-03 Природоведческое образование и входит в модуль по выбору «Модуль профилизации по биологии». Изучается во втором семестре.

Учебная программа по учебной дисциплине «Биоинформатика» направлена на изучение использования информационных технологий в биологии и медицине, информационные проблемы генетики и медицинские аспекты биоинформатики. Рассматриваются биоинформационные вопросы индивидуального развития организма, в том числе применительно к стволовым клеткам и их использованию в научных исследованиях и медицинской практике.

Цель преподавания и изучения учебной дисциплины «Биоинформатика» состоит в формировании у магистрантов и приобретении ими научных знаний об организации и функционировании биологических систем на основе теоретических представлений, методов и технических средств биоинформатики. Основными задачами изучения учебной дисциплины являются:

– изучение методов планирования биологического эксперимента;
– понимание принципов моделирования биологических систем и процессов;
– методы анализа и моделирование экологических процессов;
– исследование биологических систем на различных уровнях организации;
– приобретение навыка эффективного информационного поиска;
– понимание теоретических и методологических основ биоинформатики;
– использование информационных технологий в биологии и медицине;
– изучение информационных проблем генетики;
– понимание медицинских аспектов биоинформатики;
– анализ биоинформационных проблем индивидуального развития организма;
– понимание проблем коммуникации и регуляции на нейроэндокринном и иммунном уровне;
– развитие умений и навыков практического использования информационных технологий для математического моделирования биологических процессов.

Изучение учебной дисциплины «Биоинформатика» должно обеспечить формирование у магистрантов специализированных компетенций.

Требования к специальным компетенциям

Магистрант должен быть способен:

– СК-5. Применять методы биоинформатики, алгоритмы обработки молекулярно-биологических данных, использовать программные средства для решения прикладных научно-исследовательских задач.

В результате изучения учебной дисциплины «Биоинформатика» магистрант должен **знать:**

– принципы моделирования биологических систем и процессов;
– методы анализа и моделирования экологических процессов;
– теоретические и методологические основы биоинформатики;
– информационные проблемы генетики;
– медицинские аспекты биоинформатики;
– проблемы коммуникации и регуляции на нейроэндокринном и иммунном уровне;

В результате изучения учебной дисциплины «Биоинформатика» магистрант должен **уметь:**

– планировать биологический эксперимент;
– исследовать биологические системы на различных уровнях организации;
– эффективно осуществлять информационный поиск;
– использовать информационные технологии в биологии и медицине;
– практически использовать информационные технологии для математического моделирования биологических процессов.

В результате изучения учебной дисциплины «Биоинформатика» магистрант должен **владеть:**

– навыками статистической обработки данных биологических экспериментов.
– методами планирования биологического эксперимента;
– методами анализа и моделирование экологических процессов;
– навыком эффективного информационного поиска;

Основными формами организации учебного процесса по учебной дисциплине «Биоинформатика» являются лекции с применением мультимедийных средств обучения, практические занятия, самостоятельная работа.

В ходе изучения учебной дисциплины рекомендовано использовать следующие методы обучения: словесные, наглядные, практические, элементы проблемного обучения и научно-исследовательской деятельность.

В процессе самостоятельной работы магистранты работают с учебной и научной литературой, интернет-источниками, составляют аналитические таблицы, ведут терминологические словари.

Всего на изучение учебной дисциплины на дневной форме получения образования отводится 96 часов (3 з.е.), из них аудиторных 38 часов. Распределение аудиторных часов по видам занятий: 14 часов лекций, 24 часа практических занятий.

Промежуточная аттестация проводится в соответствии с учебным планом по специальности в форме экзамена во втором семестре.

На изучение учебной дисциплины на заочной форме получения образования отводится 96 часов (3 з. е.), из них аудиторных 38 часов. Учебная дисциплина изучается во 2, 3 семестрах. Распределение аудиторных часов по семестрам: во втором семестре аудиторная нагрузка составляет 4 часа, из них 4 часа лекций. В третьем семестре аудиторная нагрузка составляет 6 часов, из них 6 часов практических занятий. Промежуточная аттестация по учебной дисциплине проводится в соответствии с учебным планом специальности в форме экзамена в третьем семестре.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

РАЗДЕЛ 1. ТЕОРЕТИЧЕСКИЕ И МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ БИОИНФОРМАТИКИ

Тема 1.1. Предметная область биоинформатики

Биоинформатика как наука. Понятие информации Свойства информации Методологические аспекты моделирования. Понятие об аналогии. Ассоциативная аналогия. Определение модели. Аналогия и моделирование. Типы моделирования. Материальные модели. Мысленные модели. Аналитические, конструктивные и имитационные модели. Материальное моделирование и модельный эксперимент.

Тема 1.2. Принципы моделирования биологических систем, явлений и процессов

Особенности моделирования в биологии. Модели данных. Модели систем. Применение некоторых информационных технологий в биологии. Компьютерная алгебра и биоинформатика. Биологическое приложение теории нечётких множеств.

Тема 1.3. Статистические методы обработки медико-биологических данных

Биологическая статистика или биометрия. Сложные системы и оптимизация эксперимента. Полиномиальная модель и её адекватность. Понятие линейного программирования. Особенности биометрической культуры. Статистика ДНК, как характеристика генома.

Тема 1.4. Пакеты компьютерных программ

Программное обеспечение для статистической обработки биомедицинских исследований. Базы и банки данных. Первичные, вторичные и курируемые базы данных. Автоматическое аннотирование последовательности. Идентификаторы записей в базах данных. Аналитические программы в биоинформатике. Энциклопедия KEGG и ее использование.

РАЗДЕЛ 2. МОЛЕКУЛЯРНАЯ ЭВОЛЮЦИЯ

Тема 2.1. Парное и множественное выравнивание. Алгоритм и программы выравнивания

Выравнивание последовательностей. Алгоритм Нидлмана-Вунша.

Тема 2.2. Эволюция молекул и организмов

Пути эволюции молекул и организмов. Молекулярная эволюция, как наука. Критерии сравнения нуклеотидных и белковых последовательностей.

Тема 2.3. Гомология последовательностей. Ортологи и паралоги

Понятие гомологии. Работы Оуэна Р. В сфере создания естественной системы видов. Ортологи, паралоги: аллопаралоги, симпаралоги. Ксенология.

Тема 2.4. Биотехнология, генная инженерия и биобезопасность Горизонтальный перенос генов. Филогенетическое дерево и методы его построения.

Генетический перенос. Вертикальный и горизонтальный переносы генов. Филогенетическое древо и методы его построения (UPGMA, NEIGHBOR-JOINING, MINIMAL EVOLUTION). Биотехнология и генная инженерия. Генная инженерия микроорганизмов. Перенос генов в растения. Перенос генов животных. Биотехнология и биобезопасность.

РАЗДЕЛ 3. МЕТОДЫ АНАЛИЗА ПОСЛЕДОВАТЕЛЬНОСТЕЙ НУКЛЕИНОВЫХ КИСЛОТ

Тема 3.1. Гены и регуляция их экспрессии

Механизмы регуляции экспрессии генов прокариотов и эукариотов. Модель оперона. Позитивная и негативная регуляции экспрессии генов. Этапы процесса экспрессии генов.

Тема 3.2. Анализ последовательностей ДНК

Значение анализа последовательностей ДНК. Секвенаторы. Автоматическое секвенирование ДНК. Секвенирование ДНК по Сэнгеру Ф. Секвенирование ДНК по Максаму и Гилберту: метод химической деградации. Проект «Геном человека».

Тема 3.3. Нуклеиновый состав (изохоры, GC-острова) ДНК

Композиционная гетерогенность ДНК. Изохоры. Связь семейств изохор с кодирующими последовательностями ДНК и их экспрессией.

Тема 3.4. Подбор праймеров для ПЦР

Значение праймеров для процесса амплификации и ПЦР. Принципы подбора праймеров. Требования, предъявляемые к праймерам.

Тема 3.5. Анализ частоты использования кодонов

Частота транзиций и трансверсий. Определение соотношения транзиций и трансверсий. ГН-насыщенность общая и отдельных положений кодона. Зависимость ГЦ-насыщенности отдельных положений кодона от общей ГЦ-насыщенности. Анализ аминокислотного состава белков. Системы классификации сайтов нуклеиновых кислот в связи с их положением в кодоне - система вырожденности и синонимичности / несинонимичности. Методики изучения стратегии кодирования белков.

РАЗДЕЛ 4. МЕТОДЫ АНАЛИЗА ПОСЛЕДОВАТЕЛЬНОСТЕЙ БЕЛКОВ

Тема 4.1. Фолдинг и транспорт белков у про- и эукариот

Фолдинг, белки шапероны Секреция белков у прокариот: Sec-аппарат и сигнальный пептид. Распределение белков по компартментам клетки эукариот. Деградация белков

Тема 4.2. Мотивы и домены

Мотивы белков. Домены, их состав. Разница между мотивом и доменом.

Тема 4.3. Сворачивание белков, предсказание структуры белка, предсказание функции и клеточной локализации белков

Сворачивание белков. Цель и методы предсказания структуры белка. Алгоритмы предсказания вторичной структуры белков. Предсказание функции и клеточной локализации белков. Методы прогнозирования структуры белков, основанные на гомологии, мотивах последовательностей, структуре белка, геномном контексте.

РАЗДЕЛ 5. БИОИНФОРМАЦИОННЫЕ ПРОБЛЕМЫ ИНДИВИДУАЛЬНОГО РАЗВИТИЯ ОРГАНИЗМА. СТВОЛОВЫЕ КЛЕТКИ И ИХ ИСПОЛЬЗОВАНИЕ

Тема 5.1. Закономерности дифференцировки соматических клеток. Органогенез. Старение и раковая трансформация. Моделирование клеточной дифференцировки

Типы клеточных популяций. Понятие о диффероне. Апоптоз. Онтогенез и старение. Новая компьютерная технология прижизненного наблюдения клеток. Бета-гал тест как маркер старения. Злокачественные новообразования как следствие нарушения молекулярно-генетических и клеточных регуляторных механизмов. Моделирование работы генов. Два типа моделей клеточной дифференцировки. К-генетические сети. Модель Сендова-Цанева.

Тема 5.2. Стволовые клетки

История открытия стволовых клеток. Получение плюрипотентных клеток. Молекулярные основы тотипотентности эмбриональных стволовых клеток. Стволовые клетки взрослого организма. Различия в потенции стволовых клеток разного происхождения. Механизмы коммитирования стволовых клеток. Значение микроокружения для самоподдержания популяции стволовых клеток. Молекулярные маркеры стволовых клеток. Источники стволовых клеток у взрослого организма. Стволовые клетки в медицине. Сайты и базы данных по стволовым клеткам и клеточным технологиям.

РАЗДЕЛ 6. НЕЙРОИНФОРМАТИКА, ПСИХОФИЗИОЛОГИЯ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Тема 6.1. Иерархические и диффузные системы мозга

Кратковременные и долговременные изменения в синапсах. Иерархические и диффузные системы мозга. Стереотаксис.

Тема 6.2. Нейроинформатика. Моделирование процессов мышления и человеко-машинного общения. Искусственный интеллект

Математический нейрон Маккалока-Питтса. Линейная передаточная функция. Сигмоидальная передаточная функция. Логистическая функция.

Гиперболический тангенс. Моделирование формальных логических функций. Различия между биологическим и искусственным нейроном. Нейронные сети. Приложение искусственных нейронных сетей к распознаванию образов. Будущее нейрокомпьютинга.

Тема 6.3. Биоинформатика и актуальные проблемы психофизиологии

Современное состояние научной разработки проблемы коммуникации в физиологии и психологии. Моделирование процессов мышления и искусственный интеллект. Искусственный интеллект и поведенческое мышление. Современные представления о нейрофизиологических механизмах психических функций. РНК-интерференция и психические заболевания.

РАЗДЕЛ 7. БИОСФЕРА. ЭВОЛЮЦИЯ И БИОИНФОРМАТИКА

Тема 7.1. Сообщества организмов. Экологические системы

Биоинформатика и оценка биологического разнообразия. Сообщество организмов. Экологические системы. Экосистемы и теория информации. Сложность экосистем. Теоретико-информационная мера оценки неопределённости. Математический аппарат теоретико-информационного подхода. Простейшая модель Вольтерра-Лотки.

Тема 7.2. Эволюция живой природы как процесс передачи, накопления, хранения информации

Микроэволюция на протоке (математическая модель). Эволюция и ЭВМ. ЭВМ и макроэволюция. Энергетические критерии эволюционного процесса. Теоретико-информационные аспекты эволюции.

Тема 7.3. Некоторые математические и кибернетические подходы в популяционной генетике человека

Кибернетические особенности наследственной изменчивости. Исследование м-ДНК и популяционная геномика. Генная реконструкция филогенеза популяций человека математическими методами.

Тема 7.4. Биосфера и антропогенные воздействия

Структура биосферы, её эволюция, условия устойчивости. Антропогенные воздействия. Загрязнение окружающей среды. Биомониторинг и экологический прогноз. Видеоинформация по цитологии и цитогенетике.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА
УЧЕБНОЙ ДИСЦИПЛИНЫ «БИОИНФОРМАТИКА»

(дневная форма получения высшего образования)

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов		Самостоятельная (внеаудиторная) работа	Форма контроля знаний
		Лекции	Практические занятия		
1	2	3	4	5	6
2 семестр					
1	**ТЕОРЕТИЧЕСКИЕ И МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ БИОИНФОРМАТИКИ**	**2**	**4**	**10**	
1.1	**Предметная область биоинформатики** Биоинформатика как наука. Понятие информации Свойства информации Методологические аспекты моделирования. Понятие об аналогии. Ассоциативная аналогия. Определение модели. Аналогия и моделирование. Типы моделирования. Материальные модели. Мысленные модели. Аналитические, конструктивные и имитационные модели. Материальное моделирование и модельный эксперимент.	1		2	Устный опрос, защита практических заданий, тестирование
1.2	**Принципы моделирования биологических систем, явлений и процессов** Особенности моделирования в биологии. Модели данных. Модели систем. Применение некоторых информационных технологий в биологии. Компьютерная алгебра и биоинформатика. Биологическое приложение теории нечётких множеств.	1		2	Тематические доклады, устный опрос, тестирование
1.3	**Статистические методы обработки медико-биологических данных** Биологическая статистика или биометрия. Сложные системы и оптимизация эксперимента. Полиномиальная модель и её адекватность. Понятие линейного программирования. Особенности биометрической культуры. Статистика ДНК, как характеристика генома.		2	2	Устный опрос, защита практических заданий, тестирование

1.4	**Пакеты компьютерных программ** Программное обеспечение для статистической обработки биомедицинских исследований. Базы и банки данных. Первичные, вторичные и курируемые базы данных. Автоматическое аннотирование последовательности. Идентификаторы записей в базах данных. Аналитические программы в биоинформатике. Энциклопедия KEGG и ее использование.		2	4	Устный опрос, защита практических заданий, тестирование
2	**МОЛЕКУЛЯРНАЯ ЭВОЛЮЦИЯ**	**4**	**2**	**8**	
2.1 **2.2**	**Парное и множественное выравнивание. Алгоритм и программы выравнивания** Выравнивание последовательностей. Алгоритм Нидлмана-Вунша. **Эволюция молекул и организмов** Пути эволюции молекул и организмов. Молекулярная эволюция, как наука. Критерии сравнения нуклеотидных и белковых последовательностей.	1	2	2	Устный опрос, защита практических заданий, тестирование
2.3	**Гомология последовательностей. Ортологи и паралоги** Понятие гомологии. Работы Оуэна Р. в сфере создания естественной системы видов. Ортологи, паралоги: аллопаралоги, симпаралоги. Ксенология.	1		2	Устный опрос, защита практических заданий, тестирование
2.4	**Биотехнология, генная инженерия и биобезопасность Горизонтальный перенос генов. Филогенетическое дерево и методы его построения.** Генетический перенос. Вертикальный и горизонтальный переносы генов. Филогенетическое древо и методы его построения (UPGMA, NEIGHBOR-JOINING, MINIMAL EVOLUTION). Биотехнология и генная инженерия. Генная инженерия микроорганизмов. Перенос генов в растения. Перенос генов животных. Биотехнология и биобезопасность.	2		4	Устный опрос, защита практических заданий, тестирование
3	**МЕТОДЫ АНАЛИЗА ПОСЛЕДОВАТЕЛЬНОСТЕЙ НУКЛЕИНОВЫХ КИСЛОТ**	**4**	**4**	**12**	
3.1	**Гены и регуляция их экспрессии** Механизмы регуляции экспрессии генов прокариотов и эукариотов. Модель оперона. Позитивная и негативная регуляции экспрессии генов. Этапы процесса экспрессии генов.	1		2	Устный опрос, защита практических заданий, тестирование
3.2	**Анализ последовательностей ДНК** Значение анализа последовательностей ДНК. Секвенаторы. Автоматическое секвенирование ДНК. Секвенирование ДНК по Сэнгеру Ф. Секвенирование ДНК по Максаму и Гилберту: метод химической деградации. Проект «Геном человека».	1		4	Устный опрос, защита практических заданий, тестирование

3.3 **3.4**	**Нуклеиновый состав (изохоры, GC-острова) ДНК** Композиционная гетерогенность ДНК. Изохоры. Связь семейств изохор с кодирующими последовательностями ДНК и их экспрессией. **Подбор праймеров для ПЦР** Значение праймеров для процесса амплификации и ПЦР. Принципы подбора праймеров. Требования, предъявляемые к праймерам.	1	2	2	Устный опрос, защита практических заданий, тестирование
3.5	**Анализ частоты использования кодонов** Частота транзиций и трансверсий. Определение соотношения транзиций и трансверсий. ГН-насыщенность общая и отдельных положений кодона. Зависимость ГЦ-насыщенности отдельных положений кодона от общей ГЦ-насыщенности. Анализ аминокислотного состава белков. Системы классификации сайтов нуклеиновых кислот в связи с их положением в кодоне - система вырожденности и синонимичности / несинонимичности. Методики изучения стратегии кодирования белков.	1	2	4	Устный опрос, защита практических заданий, тестирование
4	**МЕТОДЫ АНАЛИЗА ПОСЛЕДОВАТЕЛЬНОСТЕЙ БЕЛКОВ**	**2**	**2**	**4**	
4.1 **4.2**	**Фолдинг и транспорт белков у про- и эукариот** Фолдинг, белки шапероны Секреция белков у прокариот: Sec-аппарат и сигнальный пептид. Распределение белков по компартментам клетки эукариот. Деградация белков **Мотивы и домены** Мотивы белков. Домены, их состав. Разница между мотивом и доменом.	1		2	Устный опрос, защита практических заданий, тестирование
4.3	**Сворачивание белков, предсказание структуры белка, предсказание функции и клеточной локализации белков** Сворачивание белков. Цель и методы предсказания структуры белка. Алгоритмы предсказания вторичной структуры белков. Предсказание функции и клеточной локализации белков. Методы прогнозирования структуры белков, основанные на гомологии, мотивах последовательностей, структуре белка, геномном контексте.	1	2	2	Устный опрос, защита практических заданий, тестирование
5	**БИОИНФОРМАЦИОННЫЕ ПРОБЛЕМЫ ИНДИВИДУАЛЬНОГО РАЗВИТИЯ ОРГАНИЗМА. СТВОЛОВЫЕ КЛЕТКИ И ИХ ИСПОЛЬЗОВАНИЕ**		**4**	**8**	
5.1	**Закономерности дифференцировки соматических клеток. Органогенез. Старение и раковая трансформация. Моделирование клеточной дифференцировки** Типы клеточных популяций. Схема развития эукариотического организма. Понятие о диффероне. Апоптоз. Онтогенез и старение. Новая компьютерная технология прижизненного наблюдения клеток. Бета-гал тест как маркер старения. Злокачественные новообразования как следствие нарушения молекулярно-генетических и клеточных регуляторных механизмов. Моделирование работы генов. Два типа моделей клеточной дифференцировки. К-генетические сети. Модель Сендова-Цанева.		2	4	Устный опрос, защита практических заданий, тестирование

5.2	**Стволовые клетки** История открытия стволовых клеток. Получение плюрипотентных клеток. Молекулярные основы тотипотентности эмбриональных стволовых клеток. Стволовые клетки взрослого организма. Различия в потенции стволовых клеток разного происхождения. Механизмы коммитирования стволовых клеток. Значение микроокружения для самоподдержания популяции стволовых клеток. Молекулярные маркеры стволовых клеток. Источники стволовых клеток у взрослого организма. Стволовые клетки в медицине. Сайты и базы данных по стволовым клеткам и клеточным технологиям.		2	4	Устный опрос, защита практических заданий, тестирование
6	**НЕЙРОИНФОРМАТИКА, ПСИХОФИЗИОЛОГИЯ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ**	**2**	**4**	**8**	
6.1	**Иерархические и диффузные системы мозга** Кратковременные и долговременные изменения в синапсах. Иерархические и диффузные системы мозга. Стереотаксис.	1	1	2	Устный опрос, защита практических заданий, тестирование
6.2	**Нейроинформатика. Моделирование процессов мышления и человеко-машинного общения. Искусственный интеллект** Математический нейрон Маккалока-Питтса. Линейная передаточная функция. Сигмоидальная передаточная функция. Логистическая функция. Гиперболический тангенс. Моделирование формальных логических функций. Различия между биологическим и искусственным нейроном. Нейронные сети. Приложение искусственных нейронных сетей к распознаванию образов. Будущее нейрокомпьютинга.	1	1	2	Устный опрос, защита практических заданий, тестирование
6.3	**Биоинформатика и актуальные проблемы психофизиологии** Современное состояние научной разработки проблемы коммуникации в физиологии и психологии. Моделирование процессов мышления и искусственный интеллект. Искусственный интеллект и поведенческое мышление. Современные представления о нейрофизиологических механизмах психических функций. РНК-интерференция и психические заболевания.		2	4	Устный опрос, защита практических заданий, тестирование
7	**БИОСФЕРА. ЭВОЛЮЦИЯ И БИОИНФОРМАТИКА**		**4**	**8**	
7.1	**Сообщества организмов. Экологические системы** Биоинформатика и оценка биологического разнообразия. Сообщество организмов. Экологические системы. Экосистемы и теория информации. Сложность экосистем. Теоретико-информационная мера оценки неопределённости. Математический аппарат теоретико-информационного подхода. Простейшая модель Вольтерра-Лотки.		1	2	Устный опрос, защита практических заданий, тестирование

7.2	**Эволюция живой природы как процесс передачи, накопления, хранения информации** Микроэволюция на протоке (математическая модель). Эволюция и ЭВМ. ЭВМ и макроэволюция. Энергетические критерии эволюционного процесса. Теоретико-информационные аспекты эволюции.		1	2	Устный опрос, защита практических заданий, тестирование
7.3	**Некоторые математические и кибернетические подходы в популяционной генетике человека** Кибернетические особенности наследственной изменчивости. Исследование м-ДНК и популяционная геномика. Генная реконструкция филогенеза популяций человека математическими методами.		1	2	Устный опрос, защита практических заданий, тестирование
7.4	**Биосфера и антропогенные воздействия** Структура биосферы, её эволюция, условия устойчивости. Антропогенные воздействия. Загрязнение окружающей среды. Биомониторинг и экологический прогноз. Видеоинформация по цитологии и цитогенетике.		1	2	Устный опрос, защита практических заданий, тестирование
	Итого:	**14**	**24**	**58**	**Экзамен**

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА
УЧЕБНОЙ ДИСЦИПЛИНЫ «БИОИНФОРМАТИКА»

(заочная формы получения высшего образования)

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов		Самостоятельная (внеаудиторная) работа	Форма контроля знаний
		Лекции	Практические занятия		
1	2	3	4	5	6
2 семестр					
1	**ТЕОРЕТИЧЕСКИЕ И МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ БИОИНФОРМАТИКИ**	**2**		**8**	
1.1	**Предметная область биоинформатики** Биоинформатика как наука. Понятие информации Свойства информации Методологические аспекты моделирования. Понятие об аналогии. Ассоциативная аналогия. Определение модели. Аналогия и моделирование. Типы моделирования. Материальные модели. Мысленные модели. Аналитические, конструктивные и имитационные модели. Материальное моделирование и модельный эксперимент.	1		2	Устный опрос, защита практических заданий, тестирование
1.2	**Принципы моделирования биологических систем, явлений и процессов** Особенности моделирования в биологии. Модели данных. Модели систем. Применение некоторых информационных технологий в биологии. Компьютерная алгебра и биоинформатика. Биологическое приложение теории нечётких множеств.	1		2	Тематические доклады, устный опрос, тестирование
1.3	**Статистические методы обработки медико-биологических данных** Биологическая статистика или биометрия. Сложные системы и оптимизация эксперимента. Полиномиальная модель и её адекватность. Понятие линейного программирования. Особенности биометрической культуры. Статистика ДНК, как характеристика генома.			2	Устный опрос, защита практических заданий, тестирование

1.4	**Пакеты компьютерных программ** Программное обеспечение для статистической обработки биомедицинских исследований. Базы и банки данных. Первичные, вторичные и курируемые базы данных. Автоматическое аннотирование последовательности. Идентификаторы записей в базах данных. Аналитические программы в биоинформатике. Энциклопедия KEGG и ее использование.			2	Устный опрос, защита практических заданий, тестирование
2	**МОЛЕКУЛЯРНАЯ ЭВОЛЮЦИЯ**	**2**		**6**	
2.1 **2.2**	**Парное и множественное выравнивание. Алгоритм и программы выравнивания** Выравнивание последовательностей. Алгоритм Нидлмана-Вунша. **Эволюция молекул и организмов** Пути эволюции молекул и организмов. Молекулярная эволюция, как наука. Критерии сравнения нуклеотидных и белковых последовательностей.	1		2	Устный опрос, защита практических заданий, тестирование
2.3	**Гомология последовательностей. Ортологи и паралоги** Понятие гомологии. Работы Оуэна Р. В сфере создания естественной системы видов. Ортологи, паралоги: аллопаралоги, симпаралоги. Ксенология.	1		2	Устный опрос, защита практических заданий, тестирование
2.4	**Биотехнология, генная инженерия и биобезопасность Горизонтальный перенос генов. Филогенетическое дерево и методы его построения.** Генетический перенос. Вертикальный и горизонтальный переносы генов. Филогенетическое древо и методы его построения (UPGMA, NEIGHBOR-JOINING, MINIMAL EVOLUTION). Биотехнология и генная инженерия. Генная инженерия микроорганизмов. Перенос генов в растения. Перенос генов животных. Биотехнология и биобезопасность.			2	Устный опрос, защита практических заданий, тестирование
3	**МЕТОДЫ АНАЛИЗА ПОСЛЕДОВАТЕЛЬНОСТЕЙ НУКЛЕИНОВЫХ КИСЛОТ**			**12**	
3.1	**Гены и регуляция их экспрессии** Механизмы регуляции экспрессии генов прокариотов и эукариотов. Модель оперона. Позитивная и негативная регуляции экспрессии генов. Этапы процесса экспрессии генов.			2	Устный опрос, защита практических заданий, тестирование
3.2	**Анализ последовательностей ДНК** Значение анализа последовательностей ДНК. Секвенаторы. Автоматическое секвенирование ДНК. Секвенирование ДНК по Сэнгеру Ф. Секвенирование ДНК по Максаму и Гилберту: метод химической деградации. Проект «Геном человека».			2	Устный опрос, защита практических заданий, тестирование

3.3 **3.4**	**Нуклеиновый состав (изохоры, GC-острова) ДНК** Композиционная гетерогенность ДНК. Изохоры. Связь семейств изохор с кодирующими последовательностями ДНК и их экспрессией. **Подбор праймеров для ПЦР** Значение праймеров для процесса амплификации и ПЦР. Принципы подбора праймеров. Требования, предъявляемые к праймерам.			4	Устный опрос, защита практических заданий, тестирование
3.5	**Анализ частоты использования кодонов** Частота транзиций и трансверсий. Определение соотношения транзиций и трансверсий. ГН-насыщенность общая и отдельных положений кодона. Зависимость ГЦ-насыщенности отдельных положений кодона от общей ГЦ-насыщенности. Анализ аминокислотного состава белков. Системы классификации сайтов нуклеиновых кислот в связи с их положением в кодоне - система вырожденности и синонимичности / несинонимичности. Методики изучения стратегии кодирования белков.			4	Устный опрос, защита практических заданий, тестирование
	Итого за 2 семестр	**4**		**26**	
	3 семестр				
4	**МЕТОДЫ АНАЛИЗА ПОСЛЕДОВАТЕЛЬНОСТЕЙ БЕЛКОВ**		**2**	**16**	
4.1 **4.2**	**Фолдинг и транспорт белков у про- и эукариот** Фолдинг, белки шапероны Секреция белков у прокариот: Sec-аппарат и сигнальный пептид. Распределение белков по компартментам клетки эукариот. Деградация белков **Мотивы и домены** Мотивы белков. Домены, их состав. Разница между мотивом и доменом.			8	Устный опрос, защита практических заданий, тестирование
4.3	**Сворачивание белков, предсказание структуры белка, предсказание функции и клеточной локализации белков** Сворачивание белков. Цель и методы предсказания структуры белка. Алгоритмы предсказания вторичной структуры белков. Предсказание функции и клеточной локализации белков. Методы прогнозирования структуры белков, основанные на гомологии, мотивах последовательностей, структуре белка, геномном контексте.		2	8	Устный опрос, защита практических заданий, тестирование

5	**БИОИНФОРМАЦИОННЫЕ ПРОБЛЕМЫ ИНДИВИДУАЛЬНОГО РАЗВИТИЯ ОРГАНИЗМА. СТВОЛОВЫЕ КЛЕТКИ И ИХ ИСПОЛЬЗОВАНИЕ**		**2**	**14**	
5.1	**Закономерности дифференцировки соматических клеток. Органогенез. Старение и раковая трансформация. Моделирование клеточной дифференцировки** Типы клеточных популяций. Схема развития эукариотического организма. Понятие о диффероне. Апоптоз. Онтогенез и старение. Новая компьютерная технология прижизненного наблюдения клеток. Бета-гал тест как маркер старения. Злокачественные новообразования как следствие нарушения молекулярно-генетических и клеточных регуляторных механизмов. Моделирование работы генов. Два типа моделей клеточной дифференцировки. К-генетические сети. Модель Сендова-Цанева.		1	6	Устный опрос, защита практических заданий, тестирование
5.2	**Стволовые клетки** История открытия стволовых клеток. Получение плюрипотентных клеток. Молекулярные основы тотипотентности эмбриональных стволовых клеток. Стволовые клетки взрослого организма. Различия в потенции стволовых клеток разного происхождения. Механизмы коммитирования стволовых клеток. Значение микроокружения для самоподдержания популяции стволовых клеток. Молекулярные маркеры стволовых клеток. Источники стволовых клеток у взрослого организма. Стволовые клетки в медицине. Сайты и базы данных по стволовым клеткам и клеточным технологиям.		1	8	Устный опрос, защита практических заданий, тестирование
6	**НЕЙРОИНФОРМАТИКА, ПСИХОФИЗИОЛОГИЯ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ**		**2**	**16**	
6.1	**Иерархические и диффузные системы мозга** Кратковременные и долговременные изменения в синапсах. Иерархические и диффузные системы мозга. Стереотаксис.		1	2	Устный опрос, защита практических заданий, тестирование
6.2	**Нейроинформатика. Моделирование процессов мышления и человеко-машинного общения. Искусственный интеллект** Математический нейрон Маккалока-Питтса. Линейная передаточная функция. Сигмоидальная передаточная функция. Логистическая функция. Гиперболический тангенс. Моделирование формальных логических функций. Различия между биологическим и искусственным нейроном. Нейронные сети. Приложение искусственных нейронных сетей к распознаванию образов. Будущее нейрокомпьютинга.		1	8	Устный опрос, защита практических заданий, тестирование

6.3	**Биоинформатика и актуальные проблемы психофизиологии** Современное состояние научной разработки проблемы коммуникации в физиологии и психологии. Моделирование процессов мышления и искусственный интеллект. Искусственный интеллект и поведенческое мышление. Современные представления о нейрофизиологических механизмах психических функций. РНК-интерференция и психические заболевания.			6	Устный опрос, защита практических заданий, тестирование
7	**БИОСФЕРА. ЭВОЛЮЦИЯ И БИОИНФОРМАТИКА**			**14**	
7.1	**Сообщества организмов. Экологические системы** Биоинформатика и оценка биологического разнообразия. Сообщество организмов. Экологические системы. Экосистемы и теория информации. Сложность экосистем. Теоретико-информационная мера оценки неопределённости. Математический аппарат теоретико-информационного подхода. Простейшая модель Вольтерра-Лотки.			4	Устный опрос, защита практических заданий, тестирование
7.2	**Эволюция живой природы как процесс передачи, накопления, хранения информации** Микроэволюция на протоке (математическая модель). Эволюция и ЭВМ. ЭВМ и макроэволюция. Энергетические критерии эволюционного процесса. Теоретико-информационные аспекты эволюции.			4	Устный опрос, защита практических заданий, тестирование
7.3	**Некоторые математические и кибернетические подходы в популяционной генетике человека** Кибернетические особенности наследственной изменчивости. Исследование м-ДНК и популяционная геномика. Генная реконструкция филогенеза популяций человека математическими методами.			4	Устный опрос, защита практических заданий, тестирование
7.4	**Биосфера и антропогенные воздействия** Структура биосферы, её эволюция, условия устойчивости. Антропогенные воздействия. Загрязнение окружающей среды. Биомониторинг и экологический прогноз. Видеоинформация по цитологии и цитогенетике.			2	Устный опрос, защита практических заданий, тестирование
	Итого за 3 семестр:		**6**	**60**	**Экзамен**
	Итого:	**4**	**6**	**86**	**Экзамен**

4.2. Список рекомендуемой литературы
ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

ЛИТЕРАТУРА

Основная литература

1. Дромашко, С. Е. Мозг, интеллект, нейроинформатика : учеб.-метод. пособие / С. Е. Дромашко, Р. В. Телятников ; НАН Беларуси, Ин-т подгот. науч. кадров, Каф. естественнонауч. дисциплин. – Минск : ИПНК НАН Беларуси, 2010. – 54 с.
2. Прищепа, И. М. Нейрофизиология : учеб. пособие для студентов учреждений высш. образования / И. М. Прищепа, И. И. Ефременко. – Минск : Выш. шк., 2013. – 285 с.

Дополнительная литература

3. Биоинформатика : учеб.-метод. комплекс / сост.: Д. Н. Дроздов, С. А. Зятьков, Г. Г. Гончаренко. – Гомель : Гомел. гос. ун-т, 2020. – 86 с.
4. Бородовский, М. Задачи и решения по анализу биологических последовательностей / М. Бородовский, С. Екишева ; под ред. А. А. Миронова ; пер. с англ. А. А. Чумичкина. – М. : Ин-т компьютер. исслед. ; Ижевск : R & C dinamics, 2008. – 438 с.
5. Вернадский, В. И. Биосфера и ноосфера / В. И. Вернадский. – М. : Айрис-пресс, 2004. – 576 с.
6. Дромашко, С. Е. Математические и компьютерные модели в биологии: взгляд генетика / С. Е. Дромашко ; НАН Беларуси, Ин-т генетики и цитологии. – Минск : Белорус. наука, 2006. – 167 с.
7. Дромашко, С. Е. Очерки биоинформатики / С. Е. Дромашко ; НАН Беларуси, Ин-т генетики и цитологии ; науч. ред. Л. В. Хотылева. – Минск : Беларус. навука, 2009. – 400 с.
8. Каменская, М. А. Информационная биология : учеб. пособие для студентов вузов / М. А. Каменская. – М. : Академия, 2006. – 361 с.
9. Огурцов, А. Н. Методы биоинформационного анализа : учеб. пособие для студентов / А. Н. Огурцов. – Харьков : Нац. техн. ун-т «Харьковский политехнический институт», 2011. – 114 с.
10. Огурцов, А. Н. Основы биоинформатики : учеб. пособие для студентов / А. Н. Огурцов. – Харьков : Нац. техн. ун-т «Харьковский политехнический институт», 2013. – 400 с.
11. Реброва, О. Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA / О. Ю. Реброва. – М. : Медиа Сфера, 2002. – 305 с.
12. Чернавский, Д. С. Синергетика и информация : динамич. теория информ. / Д. С. Чернавский ; Рос. акад. наук, Физ. ин-т. – М. : Наука, 2001. – 245 с.
13. Шульговский, В. В. Основы нейрофизиологии : учеб. пособие для студентов вузов / В. В. Шульговский. – М. : Аспект-пресс, 2000. – 277 с.
14. Яновская, В. В. Биоинформатика : курс лекций / В. В. Яновская. – Витебск : Витеб. гос. ун-т, 2022. – 83 с.

ПЕРЕЧЕНЬ РЕСУРСОВ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ «ИНТЕРНЕТ»

1. http://elibrary.ru Научная электронная библиотека eLIBRARY.RU
2. Электронно-библиотечная система «Лань» http://e.lanbook.com/
3. Электронно-библиотечная система «IPRBOOK» http://www.iprbookshop.ru
4. База данных Scopushttp://www.scopus.com/home.url
5. База данных Web of Science http://apps.webofknowledge.com/
6. База данных полнотекстовых академических журналов Китая http://oversea.cnki.net/
7. Электронная библиотека диссертаций Российской государственной библиотеки http://diss.rsl.ru/
8. Электронные базы данных EBSCO http://search.ebscohost.com/

ПЕРЕЧЕНЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

1. http://www.ncbi.nlm.nih.gov/BLAST
2. http://www.genome.jp/tools/clustalw/
3. http://www.ncbi.nlm.nih.gov/Entrez
4. http://www.expasy.org/ Expasy
5. http://www.drive5.com/muscle/
6. http://evolution.genetics.washington.edu/phylip.html
7. Entrez cross-database search page – http://www.ncbi.nlm.nih.gov
8. Биоинформационные ресурсы для геномики и протеомики – http://www.expasy.org
9. Биологические банки и базы данных –
10. http://www.nsu.ru/education/i4biol/noframes/reviewdb.html
11. Программы анализа полинуклеотидных и полипептидных последовательностей - http://blast.ncbi.nlm.nih.gov
12. Программы множественного выравнивания - www.genome.jp/tools/clustalw/
13. Форум по молекулярной биологии - http://molecularstation.com/
14. Unipro UGENE: [Электронный документ] (http://ugene.unipro.ru/ru).
15. Институт биоинформатики: [Электронный документ] (http://bioinformaticsinstitute.ru/).
16. Справочник по биоинформатике http://www.cellbiol.ru/book/bioinformatika

Протокол согласования учебной программы учебной дисциплины «Биоинформатика» с другими учебными дисциплинами специальности

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)
Цитология Эволюционная биология Микробиология Генетика Эпигенетика Биологическая химия	Кафедра биологии и методики преподавания биологии Кафедра химии и методики преподавания химии	При составлении программ по учебным дисциплинам учитывать содержание учебной программы по учебной дисциплине «Биоинформатика»	Протокол № 4 от 22.11.2023 г.

ПЕРЕЧЕНЬ ИСПОЛЬЗУЕМЫХ СРЕДСТВ ДИАГНОСТИКИ РЕЗУЛЬТАТОВ УЧЕБНОЙ ДЕЯТЕЛЬНОСТИ

Для текущего контроля и самоконтроля знаний и умений магистрантов по учебной дисциплине «Биоинформатика» можно использовать следующий диагностический инструментарий:

- устный опрос;
- тестовый контроль;
- подготовка рефератов и презентаций;
- подготовка тематических докладов;
- подготовка обзоров литературы по отдельным темам;
- выполнение поисковых заданий;
- экзамен.

Текущий контроль успеваемости проводится в форме устного или тестового опроса на практических занятиях с выставлением текущих оценок по десятибалльной шкале.

Учебным планом в качестве формы текущей аттестации по учебной дисциплине «Биоинформатика» предусмотрен экзамен.

ТРЕБОВАНИЯ К ВЫПОЛНЕНИЮ САМОСТОЯТЕЛЬНОЙ (ВНЕАУДИТОРНОЙ) РАБОТЫ МАГИСТРАНТОВ (СРМ) ПО УЧЕБНОЙ ДИСЦИПЛИНЕ «БИОИНФОРМАТИКА»

№ п/п	Название раздела	Количество часов на СРМ	Задание	Форма выполнения
1	Теоретические и методологические основы биоинформатики	10	Изучить рекомендованную литературу; составить краткий конспект по вопросам темы (согласно программе). Подготовить реферат или презентацию (на выбор) по темам раздела.	Конспект, реферат, презентация
2	Молекулярная эволюция	8	Изучить рекомендованную литературу; составить краткий конспект по вопросам темы (согласно программе). Подготовить доклад и презентацию (на выбор) по темам раздела.	Конспект, доклад, презентация
3	Методы анализа последовательностей нуклеиновых кислот	12	Изучить наследственность и изменчивость на разных уровнях организации живого (молекулярном, клеточном, организменном и популяционном); мутагенез, рекомбинацию и картирование генома.	Конспект, реферат
4	Методы анализа последовательностей белков	4	Составить опорный конспект по вопросам темы	Конспект
5	Биоинформационные проблемы индивидуального развития организма. Стволовые клетки и их использование	8	Закономерности дифференцировки соматических клеток. Органогенез. Старение и раковая трансформация. Моделирование клеточной дифференцировки.	Отчет, реферат
6	Нейроинформатика, психофизиология, искусственный интеллект	8	Изучить коммуникационные и информационные проблемы биоинформатики. Составить опорный конспект. Подготовить презентацию по выбранным темам.	Конспект, реферат
7	Биосфера. Эволюция и биоинформатика	8	Проанализировать основную и дополнительную литературу по вопросам темы и составить опорный конспект	Реферат

КРИТЕРИИ ОЦЕНКИ ЗНАНИЙ И КОМПЕТЕНЦИЙ МАГИСТРАНТОВ

Форма итогового контроля по учебной дисциплине**– экзамен**

10 баллов – десять:

систематизированные, глубокие и полные знания по программе учебной дисциплины, а также по основным вопросам, выходящим за ее пределы; точное использование научной терминологии, грамотное, логичное выполнение заданий и умение делать обоснованные выводы; безупречное владение инструментарием учебной дисциплины, умение его эффективно использовать в постановке и решении профессиональных задач; выраженная способность самостоятельно и творчески решать проблемы в нестандартной ситуации; полное и глубокое усвоение основной и дополнительной литературы, рекомендованной учебной программой дисциплины; умение ориентироваться в теориях, концепциях и направлениях по методикам биологических исследований, использовать научные достижения других дисциплин; творческая самостоятельная работа при выполнении общих и индивидуальных заданий семинарских занятий, активное участие в групповых обсуждениях, высокий уровень культуры исполнения заданий.

9 баллов – девять:

систематизированные, глубокие и полные знания по всем разделам программы учебной дисциплины; точное использование научной терминологии, грамотное, логичное выполнение заданий и умение делать обоснованные выводы; хорошее владение инструментарием, умение его эффективно использовать в постановке и решении научных и профессиональных задач; способность самостоятельно и творчески решать сложные проблемы в нестандартной ситуации в рамках учебной программы учебной дисциплины; полное усвоение основной и дополнительной литературы, рекомендованной учебной программой; умение ориентироваться в основных теориях, концепциях и направлениях по методикам биологических исследований; творческая самостоятельная работа при выполнении общих и индивидуальных заданий семинарских занятий, высокий уровень культуры исполнения заданий.

8 баллов – восемь:

систематизированные, глубокие и полные знания по всем разделам программы учебной дисциплины; точное использование научной терминологии, грамотное, логичное выполнение заданий и умение делать обоснованные выводы; владение инструментарием, техникой информационных технологий; умение его использовать в постановке и решении научных и профессиональных задач; способность самостоятельно решать сложные проблемы в рамках учебной программы; освоение основной и дополнительной литературы, рекомендованной учебной программой учебной дисциплины; умение ориентироваться в основных теориях, концепциях и направлениях по ботанике; активная самостоятельная работа

при выполнении общих и индивидуальных заданий семинарских занятий, высокий уровень культуры исполнения заданий.

7 баллов – семь:

систематизированные, глубокие и полные знания по всем разделам учебной программы учебной дисциплины; использование научной терминологии, грамотное, логичное выполнение заданий и умение делать обоснованные выводы; владение инструментарием, умение его использовать в постановке и решении научных и профессиональных задач; усвоение основной и части дополнительной литературы, рекомендованной учебной программой дисциплины; умение ориентироваться в основных теориях, концепциях и направлениях по методикам биологических исследований; самостоятельная работа при выполнении общих и индивидуальных заданий семинарских занятий, хороший уровень культуры исполнения заданий.

6 баллов – шесть:

достаточно полные и систематизированные знания в объеме учебной программы учебной дисциплины; использование необходимой научной терминологии, грамотное, логичное выполнение заданий и умение делать в основном обоснованные выводы; владение инструментарием, умение его использовать в решении учебных и профессиональных задач; способность самостоятельно применять типовые решения в рамках учебной программы учебной дисциплины; усвоение основной литературы, рекомендованной учебной программой учебной дисциплины; умение ориентироваться в базовых теориях по изучаемой учебной дисциплине; консультативная помощь преподавателя для организации самостоятельной работы при выполнении общих и индивидуальных заданий семинарских занятий, достаточный уровень культуры исполнения заданий.

5 баллов – пять:

достаточные знания в объеме учебной программы; использование научной терминологии, грамотное, логичное выполнение заданий и умение делать в основном обоснованные выводы; владение инструментарием учебной дисциплины, умение его использовать в решении типовых учебных задач; усвоение основной литературы, рекомендованной учебной программой; консультативная помощь преподавателя для организации самостоятельной работы при выполнении общих и индивидуальных заданий семинарских занятий, достаточный уровень культуры исполнения заданий.

4 балла – четыре:

достаточный объем знаний в рамках программы; усвоение основной литературы, рекомендованной учебной программой; использование основной научной терминологии, выполнение заданий и умение делать наблюдения и выводы без существенных ошибок; владение инструментарием методик биологических исследований, умение его использовать в решении стандартных (типовых) задач; умение под руководством преподавателя решать стандартные (типовые) задачи; выполнение общих и индивидуальных

заданий семинарских занятий с большой долей помощи преподавателя или товарища-консультанта, допустимый уровень культуры исполнения заданий.

3 балла – три:

недостаточно полный объем знаний в рамках программы; знание части основной литературы, рекомендованной учебной программой; использование научной терминологии, выполнение заданий лабораторных занятий и изложение ответа на вопросы с существенными ошибками; слабое владение инструментарием методик биологических исследований, некомпетентность в решении стандартных (типовых) задач; пассивность при выполнении общих и индивидуальных заданий семинарских занятий.

2 балла – два:

фрагментарные знания в рамках программы; знания отдельных литературных источников, рекомендованных учебной программой; неумение использовать научную терминологию методик биологических исследований; не выполнение общих и индивидуальных заданий семинарских занятий.

1 балл – один:

отсутствие знаний и компетенций в рамках учебной программы практики по методикам биологических исследований.

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ОРГАНИЗАЦИИ И ВЫПОЛНЕНИЮ САМОСТОЯТЕЛЬНОЙ (ВНЕАУДИТОРНОЙ) РАБОТЫ МАГИСТРАНТОВ

Одной из основных задач высшего образования, является формирование творческой личности специалиста, способного к саморазвитию, самообразованию, инновационной деятельности. Символами нового взгляда на образование становятся компетентность, индивидуальное творчество, самостоятельный поиск знаний, потребность их совершенствования.

Только знания, добытые самостоятельным трудом, делают выпускника продуктивно мыслящим специалистом, способным творчески решать профессиональные задачи, уверенно отстаивать свои позиции.

При изучении учебной дисциплины «Биоинформатика» могут использоваться различные подходы в организации самостоятельной работы магистрантов.

Деятельность магистрантов состоит в изучении обзорного лекционного материала, содержания литературных источников, включающих учебники и учебные пособия, интернет источники, составлении аналитических таблиц, схем, терминологических словарей.

Работа преподавателя состоит в обучении магистрантов способам самостоятельной учебной работы и формировании у них соответствующих компетенций; в выделении отдельных тем или их частей для самостоятельного изучения по учебникам и учебным пособиям, а также в разработке программы контроля самостоятельной работы магистранта.

Самостоятельная работа магистрантов протекает в форме делового взаимодействия: студент получает непосредственные указания, рекомендации преподавателя об организации и содержании самостоятельной деятельности, а преподаватель выполняет функцию управления через учет, контроль и коррекцию ошибочных действий.

С первой недели семестра магистранты получают от преподавателя задания для самостоятельной работы с требованиями к качеству ее выполнения.

К основным формам межсессионного контроля работы магистрантов по изучению дисциплины «Биоинформатика» можно отнести:

- устный и письменный опрос, выполнение тестовых заданий;
- краткие контрольные задания и проверка конспектов;
- подготовка сообщений, тематических докладов, рефератов, презентаций;
- составление терминологических словарей, аналитических таблиц.

При изучении дисциплины рекомендуется использовать следующие формы самостоятельной работы в оптимальном сочетании:

- составление аналитических обзоров учебной и научной литературы;
- выполнение практических типовых и нестандартных заданий.

МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ПОДГОТОВКЕ ДОКЛАДА

Данный вид образовательной деятельности со студентами предполагает решение следующих задач: повысить уровень самостоятельности студентов, активизировать познавательную деятельность, стимулировать творческий подход при решении профессиональных задач. Подготовка докладов в рамках данного модуля является обязательным и предполагает индивидуальную или групповую работу.

Этапы работы над докладом:

1. Определение темы доклада.
2. Формулировка проблемы, постановка цели и задач.
3. Активная и самостоятельная работа над докладом;
4. Консультации преподавателя; оформление материалов.
5. Подготовка к представлению доклада.

На выполнение доклада отводится одна неделя (время обучения в рамках модуля). Объем выполненной работы должен быть не менее 5 страниц.

Доклад считается выполненным полностью в случае

1. Предоставления полного объема учебных материалов по заранее утвержденной теме, полностью раскрывающих заявленную тему;

2. Предоставления материалов на электронном носителе и в печатном виде.

Выполненный доклад должен быть представлен в электронном и печатном виде. Работа должна быть оформлена соответствующим образом:

- введение;
- основная часть;
- заключение;
- библиографический список, ссылки на Интернет-ресурсы;

Текст печатается на одной стороне стандартного формата А4 через один интервал, Times New Roman, 14 pt, красная строка – 1,25 см, выравнивание по ширине. Размер левого поля – 30 мм, правого – 10 мм, верхнего и нижнего – по 20 мм. Нумерация страниц начинается с титульного листа, но номер его страницы не указывается. Все остальные страницы нумеруются по порядку, размещая номер в середине верхнего или нижнего поля.

Подготовка к защите заключается в оформлении электронного и печатного варианта доклада, а также подготовке выступления, отражающего цели и задачи работы, основное содержание темы выступления, наиболее сильные стороны выполненной работы. Продолжительность выступления – не более 10 минут.

Выступление с докладом предполагает выступление студента группы, перед студентами и преподавателем. После каждого выступления присутствующие на защите участники задают вопросы, чтобы прояснить некоторые моменты, выяснить насколько глубоко проработана тема исследования и насколько эффективно. Каждая работа оценивается: при этом оценку своей работы получает каждый участник группы, учитывается выступление на защите, наконец, оценивается вся работа в целом.